

33rd Workshop of the Swedish
**Artificial Intelligence
Society (SAIS 2021)**



Welcome Message

Table of Contents

Technical Papers

www.sais.se

Visit the website for more information!

2021 Swedish Artificial Intelligence Society Workshop (SAIS)

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For reprint or republication permission, email to IEEE Copyrights Manager at pubs-permissions@ieee.org. All rights reserved. Copyright © by IEEE.

2021 Swedish Artificial Intelligence Society Workshop (SAIS)

WELCOME MESSAGE FROM THE CHAIRS

On behalf of the Organizing Committee, it is our greatest pleasure to welcome you to the 33rd annual workshop of the Swedish Artificial Intelligence Society, SAIS 2021.

Since its first edition, the SAIS workshop has promoted research and applications of Artificial Intelligence (AI) and nurtured networks across academia and industry in national and international contexts. The SAIS workshop provides a forum for researchers and professionals in AI and related fields, and a valuable opportunity to get feedback on work in progress and discuss the latest advances.

This edition of the workshop is hosted by the Machine Learning group at the Luleå University of Technology and was planned as a hybrid event with both physical and online participation. Unfortunately, due to the recent increased concern for the situation regarding the Coronavirus (COVID-19) also this year's conference will be held online.

Out of the 31 submissions, SAIS 2021 accepted 12 papers for inclusion in the IEEE proceedings and 16 non-archived contributions, including 6 Ph.D. project descriptions presented in mentoring sessions. We are happy to see a good number of contributions from collaborations including most Swedish universities active in the field and also some research organizations, and companies. Most contributions are non-archived presentations of work in progress, which are further improved using the review comments and experience from discussions at the workshop and possibly lead to archived publications in other venues. This spirit, i.e., giving feedback to peer researchers and mentoring PhD students from other groups, is one of the major missions of the SAIS workshop – with the goal to increase the scientific quality of AI research in Sweden.

The program of SAIS 2021 includes a variety of topics, ranging from AI applications in fields such as computer vision, epidemic modeling, fluid dynamics, healthcare, human-machine interaction, mining, and retail, to investigations of methods for problems such as gesture recognition, multi-modal domain translation, neuromorphic sensing, privacy-preserving learning using homomorphic encryption, self-supervision, and visual instance segmentation. We would like to express our gratitude to all the authors who submitted their work, to the program committee and reviewers, and to the organizers, student mentors, and session chairs.

Finally, we thank the highly reputed plenary speakers – Seiichi Uchida, Elena Fersman, and Elisabetta Chicca, and the contributors to the special session focusing on Swedish-Brazilian research collaboration.

Fredrik Sandin and Marcus Liwicki (Program chair and General Chair of SAIS 2021)

2021 Swedish Artificial Intelligence Society Workshop (SAIS)

TABLE OF CONTENTS

Paper #	Title and Author(s)	Page #
1	<i>Identifying regions most likely to contribute to an epidemic outbreak in a human mobility network</i> Alexander Bridgwater, Andras Bota	1-4
2	<i>Predicting Signed Distance Functions for Visual Instance Segmentation</i> Emil Brissman, Joakim Johnander, Michael Felsberg	5-10
3	<i>Robot First Aid: Autonomous Vehicles Could Help in Emergencies</i> Martin Cooney, Felipe Valle and Alexey Vinel	11-14
4	<i>Towards motivation-driven intelligent interfaces: formal argumentation meets activity theory</i> Esteban Guerrero, Helena Lindgren	15-18
5	<i>Towards a Machine Learning Framework for Drill Core Analysis</i> Christian Günther, Nils Jansson, Marcus Liwicki, Foteini Simistira-Liwicki	19-24
6	<i>Class-Incremental Learning for Semantic Segmentation - A study</i> Karl Holmquist, Lena Klasén, Michael Felsberg	25-28
7	<i>Identifying cheating behaviour with machine learning</i> Elina Kock, Yamma Sarwari, Nancy Russo, Magnus Johnsson	29-32
8	<i>AI Transformation in the Public Sector: Ongoing Research</i> Einav Peretz-Andersson, Niklas Lavesson, Albert Bifet, Patrick Mikalef	33-36
9	<i>Rock Classification with Machine Learning: a Case Study from the Zinkgruvan Zn-Pb-Ag Deposit, Bergslagen, Sweden</i> Filip Simán, Nils Jansson, Tobias Kampmann, Foteini Simistira Liwicki	37-41
10	<i>Hit Detection in Sports Pistol Shooting</i> Elinore Stenhager, Niklas Lavesson	42-45
11	<i>Machine Learning Computational Fluid Dynamics</i> Ali Usman, Muhammad Rafiq, Muhammad Saeed, Ali Nauman, Andreas Almqvist, Marcus Liwicki	46-49
12	<i>Smart Sewage Water Management and Data Forecast</i> Qinghua Wang, Viktor Westlund, Jonas Johansson, Magnus Lindgren	50-53

Identifying regions most likely to contribute to an epidemic outbreak in a human mobility network

Alexander Bridgwater¹ and András Bóta²

Abstract—The importance of modelling the spreading of infectious diseases as part of a public health strategy has been highlighted by the ongoing coronavirus pandemic. This includes identifying the geographical areas or travel routes most likely to contribute to the spreading of an outbreak. These areas and routes can then be monitored as part of an early warning system, be part of intervention strategies, e.g. lockdowns, aiming to mitigate the spreading of the disease or be a focus of vaccination campaigns.

In this paper we present our work in developing a network-based infection model between the municipalities of Sweden in order to identify the areas most likely to contribute to an epidemic. We first construct a human mobility model based on the well-known radiation model, then we employ a network-based compartmental model to simulate epidemic outbreaks with various parameters. Finally, we adopt the influence maximization problem known in network science to identify the municipalities having the largest impact on the spreading of infectious diseases.

We only present the first part of our work in this paper. In the future, we plan to investigate the robustness of our model in identifying high-risk areas by simulating outbreaks with various parameters. We also plan to extend our work to selecting the most likely infection paths contributing to the spreading of infectious diseases.

I. INTRODUCTION

The Covid-19 pandemic has shown that countries all over the world have to be more prepared to tackle highly contagious diseases. Proper preparation involves developing early detection systems, making sure the health infrastructure is ready to handle the peak hospitalization of patients and in general, the ability of society to get information and react to the outbreak. A critical component of each of these tasks is risk detection, that is identifying the geographical areas or routes that are likely to contribute the most to the spreading of diseases. Much progress has been made in this field in recent years [1, 5, 14, 17, 20, 22], highlighting the importance of large population centers [6, 24], travel hubs [1] and the effect of efficient global transportation [5, 10, 17, 22].

Unfortunately, our global transportation network provides an efficient platform for contagious diseases to propagate by [1, 5, 10]. However, modern technology also makes it easy to trace human mobility, allowing researchers to study its role in epidemic spreading. Apart from simply estimating the importance of short distance commuting [4, 24] or long

distance travel patterns [5, 17], individual route-level risk can also be identified [2, 10]. Techniques also exist to infer the most likely infection paths contributing to a contagion [9, 16].

Epidemic modelling is responsible for creating mathematical tools to aid the study of disease spreading. The most widely used model family in this field is the family of compartmental models [7, 18]. Compartmental models have been adapted to complex networks [2, 10, 12], and were shown to be an excellent tool for modelling the geospatial spreading of epidemics [2, 10].

Here we aim to show how a well-known problem of network science can be applied to the field of epidemic modelling. In a general complex network setting, the influence maximization problem [8, 12] aims to identify the set of nodes capable of influencing (or infecting, depending on interpretation) the largest fraction of susceptible nodes in the network. In [13], a greedy optimization algorithm was proposed to solve the problem, which is still widely used due to its simplicity and efficiency. In an epidemiological context, the most “influential” nodes in a network are the ones that are most effective at spreading the outbreak to other nodes, thus having the largest contribution to the outbreak. In our interpretation, these nodes are super-spreaders.

In this paper we aim to identify the super-spreading nodes of a human mobility network defined between the municipalities of Sweden. We construct the mobility network from publicly available data using the radiation model [19]. Then, we implement a network-based generic SEIR compartmental epidemic model [7, 18]. Finally, we implement the greedy influence maximization algorithm in [13] to identify the super-spreading nodes. As our secondary goal, we also aim to investigate the effect the SEIR model parameters have on the super-spreading nodes, as they can be used to fine-tune the model to represent diseases with different characteristics, in this case latency and the length of the infectious period. This will allow our model to be adapted to real-life outbreaks, such as influenza or even coronavirus.

II. BACKGROUND

All of the methods and algorithms employed in this paper are defined on networks or graphs. We define a graph G as a set of nodes $v \in V(G)$ and a set of edges $e_{u,v} \in E(G)$, where edge $e_{u,v}$ connects or links nodes $u \in V(G)$ and $v \in V(G)$. A graph is undirected, if the connections between its nodes are symmetrical, that is if there is an edge $e_{u,v}$, then both u is connected to v and v to u . In directed graphs however, an edge $e_{u,v}$ only defines a link from u to v , while there is

¹A. Bridgwater is with Luleå University of Technology, Department of Computer Science, Electrical and Space Engineering, 97187 Lulea, Sweden. alexander_bridgwater at hotmail.com

²A. Bóta is with Luleå University of Technology, Department of Computer Science, Electrical and Space Engineering, Embedded Intelligent Systems Lab, 97187 Lulea, Sweden. andras.bota at ltu.se

no link between v and u unless another edge, $e_{v,u}$ is present. It is possible to assign weights to both the nodes and edges of the graphs, denoted as w_v or $w_{e_{u,v}}$. The infection model introduced in section II.B requires a specific value, an edge infection probability to be present on all edges of its input graph, we denote this value as p_e , where $0 \leq p_e \leq 1$.

A. The radiation model for human mobility

Representing human mobility patterns is a critical part of our work. Gathering accurate data on the required spatial resolution is difficult however. Fortunately, if accurate, measured data is unavailable there are methods available to estimate these missing values. The radiation model for human mobility estimates travel frequencies or fluxes between geographical areas, and has been successfully used as a proxy for mobility patterns in an epidemiological setting before [23]. The model is parameter free, and depends only on the population size of the areas and the distance between them [19], The model can be defined in network terms, where nodes represent the areas, and the model estimates the fluxes between two nodes $i \rightarrow j$ using the formula

$$T_{ij} = T_i \frac{n_i n_j}{(n_i + s_{ij})(n_i + n_j + s_{ij})} \quad (1)$$

where n_i and n_j represents the populations of areas i and j , while s_{ij} denotes the population within a circular area around i with the radius being the distance between i and j , and subtracting the population of i and j . The proportional variable $T_i = n_i (\frac{N_c}{N})$, where N is the total population in the model and N_c denotes the number of outgoing commuters of a region.

B. Network-based SEIR infection model

We define the network-based SEIR infection model in the following way, similar to [2, 10]. Like their traditional counterparts, network-based infection models work with states. Network-based models assign these states to the nodes of the network. In the case of the SEIR model, these states are: Susceptible (S) - signaling the node is vulnerable to infection, Exposed (E) - the node is infected but not yet contagious, Infectious (I) - infected and infectious, Removed (R) - the node is not infectious and cannot be infected anymore. In the beginning of the infection process only a subset of the nodes are in the infectious I state, we denote this set as A_0 . The rest of the nodes are in the S state. The infection process itself takes place in discrete time steps, where the nodes change states according the infection mechanics described in the next paragraphs.

The model has two parameters

- τ_e denoting the number of iterations a node stays in the E state, indicating the latency period of the disease
- τ_i denoting the number of iterations a node stays in the I state, indicating the infectious period of the disease

Nodes in the S state stay in the state until infected. Nodes in the E state stay in this state for τ_e iterations, then they transition into the I state. Nodes in the I state stay in the state for τ_i iterations, after which they move into the R state

and no longer take part in the infection process. The transition between the S and E states drive the infection process forward. An infectious (I) node u may turn a susceptible (S) node v into the E state, if there is a link $e_{u,v}$ between them, according to the probability p_e on the link. If multiple nodes are trying to infect node v in the same iterations, the attempts are made in an arbitrary order independently of each other until one of them succeeds. It is easy to see, that if τ_e and τ_i are finite, then after a finite number of iterations, no infection events will take place anymore, and all nodes of the network will be in the S or R states, terminating the spreading process.

C. Influence maximization

One of the problems associated with information diffusion or infection models is influence maximization. This task was first described by Domingos and Richardson in [8], seeking to enhance the efficiency and effectiveness of viral marketing. The problem was later reformulated to network terms by Kempe, Kleinberg and Tardos in [12] and became very popular. Using the concepts introduced in the previous subsection, the problem of influence maximization aims to select the set A_0 that results in the largest fraction of infected nodes, where the size of A_0 is limited to $|A_0| = k$, and the expected fraction of infected nodes is denoted as $\sigma(A)$.

In a subsequent publication [13], the same authors proposed a greedy optimization heuristic, providing a guaranteed precision of at least 63% of the optimum. The greedy algorithm is still very popular due to its simplicity, ease of implementation and efficiency on general maximization tasks. Algorithm 1 shows the pseudocode of the greedy algorithm, as appears in [11].

Algorithm 1 Greedy method

- 1: **Input:** Graph $G(V, E)$, k : desired size of the A_0
 - 2: **Output:** A_0
 - 3: $A_0 \leftarrow \emptyset$
 - 4: **While** $|A_0| \leq k$
 - 5: $A_0 = A_0 \cup \arg \max_{v \in V(G) \setminus A_0} \sigma(A_0 \cup \{v\})$
-

Starting from an initially empty set for A_0 , the algorithm iteratively selects the node v maximizing $\sigma(A_0)$ in $V(G) \setminus (A_0 \cup v)$, considering all available nodes $V(G) \setminus A_0$ in a brute-force manner. The selected node v is then added to A_0 , and if $|A_0| < k$, the process is repeated. There are multiple methods for computing $\sigma(A)$ in the literature. In this paper we used the Complete Simulation method [3] adapted to a SEIR infection model.

III. INPUTS

The human mobility network was constructed between the municipalities of Sweden using publicly available data. The nodes of the network are the municipalities of Sweden, while the edges between them indicate the rate of human mobility according to the radiation model in section II.A.

In order to compute the mobility values on the edges according to the radiation model, population sizes and out-going commuters for each municipality were gathered from Statistiska CentralByrån [21]. We used a simple method to approximate the population size of a circular area with a given diameter d around a given municipality i . We used the QGIS software [15] with shapefiles for the municipalities of Sweden downloaded from [21] to compute the centroids of each municipality. Then, we simply summarized the population sizes of all municipalities with centroids closer to the centroid of i than d . We acknowledge, that this is an oversimplification of the problem, and introduces inaccuracies to our results. However, in this early phase of our work, we deem it an acceptable solution for showing the usefulness of our approach.

In order to convert the mobility rates to edge infection probabilities we use the following scaling formula:

$$p_{e_{i,j}} = s \frac{w_{e_{i,j}}}{\max_{e \in E(G)} w_e} \quad (2)$$

for all municipalities i and j , where w_e denotes the mobility rate on edge e . A scaling value $s < 1$ is introduced to ensure that no edge has an infection probability of 1, which would be unrealistic. We used $s = 0.9$ in our initial experiments.

IV. METHODS

After constructing the human mobility network, we implemented the SEIR infection model and the greedy influence maximization heuristic to identify the super-spreading municipalities. While we aim to thoroughly investigate the effect the SEIR model parameters have on these super-spreaders, in this early phase of our work we only consider three parameter settings, with the latency period τ_e set to 5 and the infectious period τ_i set to 5, 10 and 15. In the context of network-based compartmental models, these parameters denote the number of iterations a node stays in a specific state, where each iteration corresponds to a specific time period, usually a day, thereby representing diseases with different characteristics.

The influence maximization algorithm requires $\sigma(A)$, the expected number of infected nodes for a given A_0 to be known. We estimate this number according to the algorithm in [3] by running the infection model r times and counting and averaging how many nodes were infected in each run. Following the recommendations in [3], we set $r = 10000$.

V. RESULTS AND DISCUSSION

We can see the list of super-spreading nodes selected in each iteration of the greedy algorithm for $k = 10$ in Table 1 for $\tau_i = 5, 10, 15$, while Figure 1 shows them on the map of Sweden¹. Since the nodes are selected in a greedy fashion, it is possible to rank them in the order of their selection and compare the resulting rankings.

¹Figure 1 was drawn with QGIS

Order/ τ_i	5	10	15
1	Stockholm	Stockholm	Stockholm
2	Göteborg	Malmö	Malmö
3	Malmö	Göteborg	Kungsbacka
4	Uppsala	Örebro	Karlstad
5	Haninge	Helsingborg	Kristianstad
6	Helsingborg	Vänersborg	Timrå
7	Örebro	Kristianstad	Vänersborg
8	Österåker	Hammarö	Falun
9	Norrköping	Gävle	Mörbylånga
10	Vänersborg	Värmdö	Boden

TABLE I
SUPER-SPREADING MUNICIPALITIES WITH $\tau_i = 5, 10, 15, \tau_e = 5$
ITERATIONS

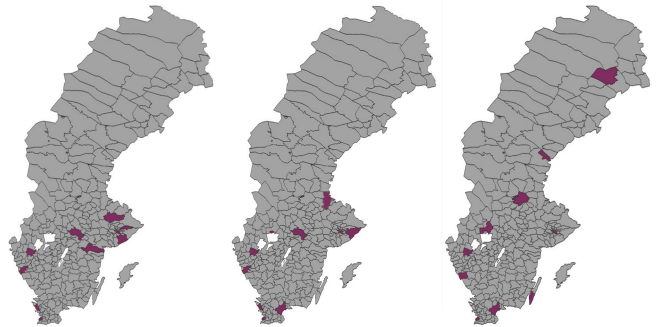


Fig. 1. Super-spreading municipalities with $\tau_i = 5, 10, 15$ iterations from left to right.

As can be seen in Table 1, Stockholm is selected as a first choice by the algorithm for all parameter settings we tried, which is unsurprising considering the capital city's role in the travel network of the country and its population size. Malmö takes the second or third place due to similar reasons. The rest of the list shows remarkable variations depending on model parameters. Figure 1 indicates, that the greater the τ_i value is, the more geographically diverse the initial spreaders are. When $\tau_i = 5$ the selected nodes are close to Stockholm, Malmö and Göteborg. With $\tau_i = 10$ the super-spreaders are more spread out, still in Southern Sweden. With $\tau = 15$ however, the selected municipalities include Boden, a travel hub in Norrbotten, Timrå, a satellite town close to Sundsvall housing the latter's airport, and Falun, which together with Borlänge is a major industrial center in Central Sweden.

Our definition of super-spreaders in this paper is the set of nodes or municipalities able to start an epidemic outbreak in the most effective way. In this early phase of our work, we did not model how the infection arrived in the country itself. However, it should be noted that several of the selected municipalities are transport hubs having airports themselves or having airports close by, allowing the introduction of an infectious traveller. Considering the variation between the super-spreading nodes according to the model parameters we can hypothesise, that the longer a node stays infectious, that is the longer the outbreak lasts in a municipality, the more effective it is in infecting its neighbors in the mobility network. This means, that an outbreak can be the most damaging if

it begins in multiple distant municipalities at the same time, allowing it to reach the whole country in a short time period. In contrast, an infection that passes quickly is most dangerous if it begins in a few highly populated areas near travel hubs.

The super-spreading nodes selected by our model can be part of an early detection strategy, with increasing infection rates in these municipalities signalling the beginning of a larger outbreak. Non-pharmaceutical interventions, such as lockdowns or travel restrictions may also be employed here to reduce the ability of the outbreak to spread to neighboring regions. Vaccination strategies can also be focused here to build up resistance to the disease and decrease the chances of it gaining a foothold.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper we have shown how the influence maximization problem can be applied to a human mobility network in an epidemiological setting. This provides researchers with a modelling tool that can be applied to model the spreading of diseases and identify super-spreading municipalities. Due to the flexibility of the SEIR model, this framework can be tailored to the specifics of different diseases. Our initial results confirm existing results, that travel hubs and large population centers are ideal for disease spreading [6, 14]. These areas can be part of early detection or mitigation strategies.

According to our initial results, the geographical position of the super-spreading nodes depends highly on the infection model parameters. However, we only investigated a few parameter configurations in this phase of our work. Clearly investigating the relationship between the selected nodes and the model parameters will be the first task in the continuation of our work. It is also clear, that the mobility rate estimations calculated by the radiation model are not perfect: they produce inaccuracies in both short and long distances. Correcting these inaccuracies of the model, or replacing them at least partially with real-life data will have a priority in our future work. It is also possible to extend the geographical resolution of our work to the DeSo (demographic statistical areas) level. Since DeSo areas differentiate between urban and rural regions, we would be able to examine spreading characteristics in different environments.

A natural extension of our model would be to identify the most likely infection routes starting in the selected municipalities. To do this we intend to employ techniques similar to those in [9, 16]. Identifying these routes would make early detection and monitoring systems even more accurate.

ACKNOWLEDGMENT

Parts of the work have been funded by the Applied AI Digital Innovation Hub North project, funded by the European Regional Development Fund.

REFERENCES

- [1] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases", *Proceedings of the National Academy of Sciences* 106(51), pp. 21484-21489, 2009.
- [2] A. Bóta, M. Holmberg, L. Gardner, M. Rosvall, "Socio-economic and environmental patterns behind H1N1 spreading in Sweden", *medRxiv preprint: medRxiv 2020.03.18.20038349*, 2020.
- [3] A. Bóta, M. Krész, A. Pluhár, "Approximations of the generalized cascade model", *Acta Cybernetica* 21(1), pp. 37–51, 2013
- [4] V. Charu, S. Zeger, J. Gog, O. N. Bjørnstad, S. Kissler, L. Simonsen, B. T. Grenfell, C. Viboud, "Human mobility and the spatial transmission of influenza in the United States", *PLoS Computational Biology* 13(2), e1005382, 2017.
- [5] V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, "The role of the airline transportation network in the prediction and predictability of global epidemics", *Proceedings of the National Academy of Sciences* 103(7), pp. 2015–2020, 2006.
- [6] B. D. Dalziel, S. Kissler, J. R. Gog, C. Viboud, O. N. Bjørnstad, C. J. E. Metcalf, et al., "Urbanization and humidity shape the intensity of influenza epidemics in US cities", *Science*, 362(6410), pp. 75–79, 2018.
- [7] O. Diekmann, J. A. P. Heesterbeek, *Mathematical epidemiology of infectious diseases. Model Building, Analysis and Interpretation*. John Wiley Sons, 2000.
- [8] P. Domingos, M. Richardson, "Mining the network value of customers", In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 57–66, 2001.
- [9] D. Fajardo, L. M. Gardner, "Inferring Contagion Patterns in Social Contact Networks with Limited Infection Data", *Netw. Spat. Econ.* 13, pp. 399–426, (2013). <https://doi.org/10.1007/s11067-013-9186-6>
- [10] L. M. Gardner, A. Bóta, K. Gangavarapu, M. U. Kraemer, N. D. Grubaugh, "Inferring the risk factors behind the geographical spread and transmission of Zika in the Americas", *PLoS neglected tropical diseases* 12(1):e0006194, 2018.
- [11] L. Hajdu, M. Krész, A. Bóta. "Community based influence maximization in the Independent Cascade Model." In *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*, IEEE, pp. 237–243, 2018.
- [12] D. Kempe, J. Kleinberg, É Tardos, "Maximizing the spread of influence through a social network", In: *Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 137–146, 2003.
- [13] D. Kempe, J. Kleinberg, É Tardos, *Influential nodes in a diffusion model for social networks*, In: *International Colloquium on Automata, Languages, and Programming*. Springer, pp. 1127–1138, 2005
- [14] S. E. Morris, B. Freiesleben de Blasio, C. Viboud, A. Wesolowski, O. N. Bjørnstad, B. T. Grenfell, "Analysis of multi-level spatial data reveals strong synchrony in seasonal influenza epidemics across Norway, Sweden, and Denmark", *PLoS one* 13(5), e0197519, 2018.
- [15] QGIS version 3.16.3. A Free and Open Source Geographic Information System. Available from: <https://qgis.org/en/site/>.
- [16] D. Rey, L. Gardner, S. T. Waller, "Finding Outbreak Trees in Networks with Limited Information", *Netw. Spat. Econ.* 16, pp. 687–721, 2016. <https://doi.org/10.1007/s11067-015-9294-6>
- [17] C. A. Russell, T. C. Jones, I. G. Barr, et al., "The global circulation of seasonal influenza A (H3N2) viruses", *Science* 320(5874), pp. 340–346, 2008.
- [18] L. Russo, C. Anastassopoulou, A. Tsakris, et al., "Tracing day-zero and forecasting the COVID-19 outbreak in Lombardy, Italy: A compartmental modelling and numerical optimization approach", *Plos one* 15(10), e0240649, 2020.
- [19] F. Simini, M. C. González, A. Maritan, A. L. Barabási, "A universal model for mobility and migration patterns", *Nature* 484(7392), pp. 96–100, 2012.
- [20] L. Skog, A. Linde, H. Palmgren, H. Hauska, F. Elgh, "Spatiotemporal characteristics of pandemic influenza", *BMC Infectious Diseases* 14(1), p. 378, 2014
- [21] Statistics Sweden <https://www.scb.se/en/> (Accessed at 2021/1/16).
- [22] M. Tizzoni, P. Bajardi, C. Poletto et al., "Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm", *BMC Medicine* 10, p. 165, 2012.
- [23] M. Tizzoni, P. Bajardi, A. Decuyper et al., "On the use of human mobility proxies for modeling epidemics.", *PLoS Comput Biol* 10(7), e1003716, 2014.
- [24] C. Viboud, O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, B. T. Grenfell, "Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza", *Science* 312(5772), pp. 447–451, 2006.

Predicting Signed Distance Functions for Visual Instance Segmentation

Emil Brissman^{1,2}, Joakim Johnander^{1,3}, Michael Felsberg¹

¹Computer Vision Laboratory, Dept. of Electrical Engineering, Linköping University

²Saab, Sweden

³Zenseact, Sweden

{emil.brissman, joakim.johnander, michael.felsberg}@liu.se

Abstract—Visual instance segmentation is a challenging problem and becomes even more difficult if objects of interest varies unconstrained in shape. Some objects are well described by a rectangle, however, this is hardly always the case. Consider for instance long, slender objects such as ropes. Anchor-based approaches classify predefined bounding boxes as either negative or positive and thus provide a limited set of shapes that can be handled. Defining anchor-boxes that fit well to all possible shapes leads to an infeasible number of prior boxes. We explore a different approach and propose to train a neural network to compute distance maps along different directions. The network is trained at each pixel to predict the distance to the closest object contour in a given direction. By pooling the distance maps we obtain an approximation to the signed distance function (SDF). The SDF may then be thresholded in order to obtain a foreground-background segmentation. We compare this segmentation to foreground segmentations obtained from the state-of-the-art instance segmentation method YOLACT. On the COCO dataset, our segmentation yields a higher performance in terms of foreground intersection over union (IoU). However, while the distance maps contain information on the individual instances, it is not straightforward to map them to the full instance segmentation. We still believe that this idea is a promising research direction for instance segmentation, as it better captures the different shapes found in the real world.

I. INTRODUCTION

Instance segmentation is a core computer vision problem. Given an image, the aim is to detect and segment all objects in an image. A class label is associated to each pixel in the image, similar to semantic segmentation, except that multiple objects of the same class shall be treated as separate entities. This problem has been extensively studied in recent years but remains challenging. The dominant approaches rely on *anchorboxes* [1], [6]. Anchorboxes is a predefined set of axis-aligned bounding boxes with one or a few different aspect ratios. The idea is that each object shape and object location in the scene should be approximately described by one or more anchorboxes. These approaches then train a neural network to classify each anchorbox as an object or as background. Usually, most objects are fairly well described by one or more of these anchorboxes. However, several common objects such as ropes or road lanes, are not well described by this approach.

In this work, we explore a new direction with the aim to tackle instance segmentation. The idea is to at each pixel predict the distance to the closest object contour, in a



Fig. 1. We show two example segmentations predicted by our method. In both cases, we have long, slender objects that are not well described by anchor-boxes. Our network instead predicts, for each pixel, the signed distance to the closest contour along a set of directions. By taking the sign of these functions, we directly obtain a segmentation of the input image.

given direction. We use Signed Distance Functions (SDFs), giving us a positive distance if we are within an object and a negative distance if we are outside an object. By thresholding, we immediately know whether we are within an object or not. Furthermore, the object contours are found at the zero-crossings of the SDF. In practice, the SDF will deviate slightly; negative if the contour is too far inside and positive if it is too far outside. This is like a spring model and the global equilibrium determines the location of the contour. By representing detected objects with the SDF, avoiding the predefined set of anchorboxes, we are able to capture objects of arbitrary shapes.

Our contributions are

- We propose to represent segmentations with signed distance functions
- We show that standard encoder-decoder neural networks are able to learn to predict these functions
- We quantitatively demonstrate that using the signed distance function output, we are able to produce a foreground-background image segmentations of quality comparable to what is obtained with state-of-the-art instance segmentation algorithms

We are able to efficiently get both the foreground and the contours of objects. An efficient way to separate the different instances from the signed distance function is subject to future work.

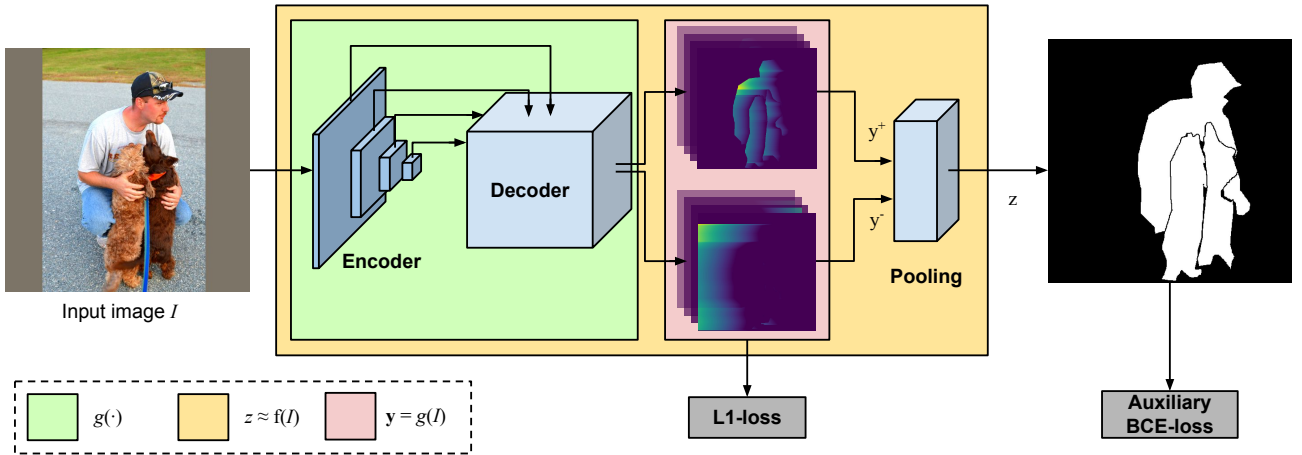


Fig. 2. Overview of our approach. The encoder-decoder network outputs distance maps along the directions of θ . The output y is divided into a positive part $y^+ \in \mathbb{R}^{4 \times H \times W}$ that defines the foreground and a corresponding negative part, that defines the background. We apply min-pooling to the parts separately and subtract the positive part with the negative. After subtraction, $z \approx f(I)$ and the foreground segmentation is directly obtained by taking the sign of $z > 0$.

II. RELATED WORK

A lot of research work has been put down into increasing the accuracy of instance segmentation and a lot of work is still ongoing. In three sections we review related methods (Section II-A) and highlight alternative approaches to the instance segmentation problem (Sections II-B to II-C).

A. Anchor-boxes

Instance segmentation approaches that rely on anchor-boxes still dominate, to a large extent, the baseline of state-of-the-art methods. Mask R-CNN [6] is anchor-based and splits the task into two subtasks, first object detection then object segmentation. Subsequent computations restrict this approach for real-time speed. Single-stage instance segmentation, like YOLACT [1], improve this aspect but yield lower accuracy. This work directly predicts pixel position representations that are assembled into final results. For segmentation, Bolya et al. [1] suggest to compute object foreground as a logistic regression problem where regression hyper parameters are inferred at each pixel position. In contrast to our approach we do not have any prior assumption on position or shape for the instance segmentation task.

B. Spatial embeddings

Subsequent methods attempts to perform instance segmentation without anchor-boxes and a popular branch is proposal-free methods. CenterNet [15] suggest to model each object as its centerpoint. Similarly, but for the instance segmentation task, CenterMask [14] propose to adopt local shape information from the representation of object centerpoints to compute the object segmentations. In Neven et al. [12], segmentations are acquired by clustering pixel embeddings that assigns a pixel to an object. A major drawback is that objects could acquire the same centerpoint, which makes it hard to discriminate between the instances. A recent method [2] considers all the pixels as object queries from

which self-attention retain those with coherent activations. The top remaining queries are subsequently assigned as background or foreground, and in the latter case, class and bounding box are predicted.

C. Signed distance function

TensorMask [3] created a foundation for exploring new directions for research within instance segmentation. Chen et al. [3] motivated to building a strong image understanding from structured 4D tensors to represent complex object shapes. This inspired our work to consider signed distance functions. Such functions are simplistic but able to model complex object shapes. In 3D-graphics the signed distance function show more frequent usage, like in previous work [10], [13], but that only considers a single object.

III. METHOD

We aim to develop a neural network that learns to approximate a Signed Distance Function (SDF) and then use the SDF for visual instance segmentation. The trained network should at each pixel predict the distance to the closest object contour in a set of different directions. These distances provide the information necessary to segment object instances, and the image is easily segmented into foreground and background by thresholding the SDF. An overview of the approach is shown in Fig. 2.

A. Distance maps

A signed distance function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a continuous function that, for a given pixel position $\mathbf{p} \in \mathbb{R}^2$ in an image \mathcal{I} , outputs the distance to the closest object contour. The function sign encodes whether \mathbf{p} is on the object (positive) or outside of the object (negative). Furthermore, all underlying object contours are implicitly represented by the iso-curves $f(\cdot) = 0$. The SDF, f , is defined as

$$f(\mathbf{p}) = s \cdot \min_{d \in \mathcal{D}_\theta^p} d, \quad (1)$$

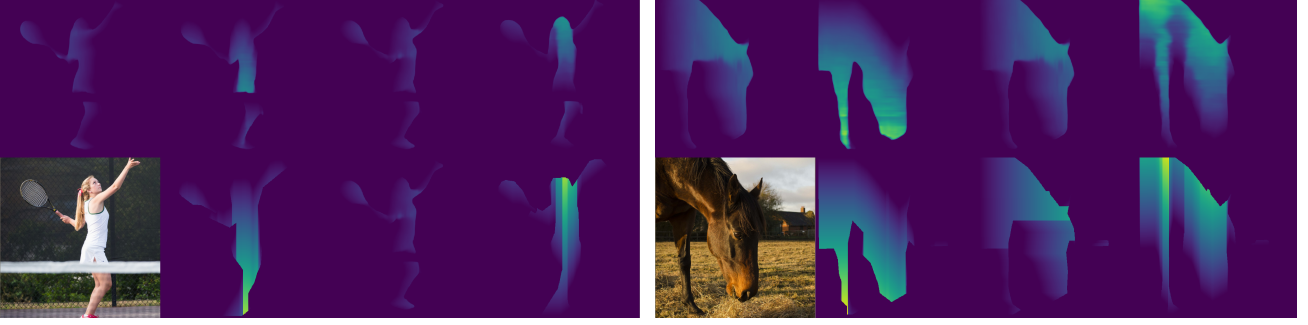


Fig. 3. We show two examples by the positive part, $y^+ \in \mathbb{R}^{4 \times H \times W}$, along the directions $\{0, 90, 180, 270\}$. The leftmost image depicts a woman playing tennis and the rightmost image depicts a horse. The bottom row show the original image and the three last distance annotations, $\{90, 180, 270\}$. Further details about the generation of distance annotations can be found in Section III-C. Although the ground truth neglects the tennis net that occludes the woman who plays tennis, the distance predictions do not.

where

$$\mathcal{D}_\theta^{\mathbf{p}} = \{\infty\} \cup \{d \in \mathbb{R}^+ : \mathbf{p} + d \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \in \Omega\} . \quad (2)$$

Here, d represents distances to contours in the direction θ and $s \in \{-1, 1\}$ is the sign. If \mathbf{p} is within an object we let s be positive, and otherwise negative. The set Ω contains a subset of image pixel positions, each of which belongs to object contours. That is, $\mathcal{D}_\theta^{\mathbf{p}}$, in (2), contains the distances to object contours that are intersected by the lines $\mathbf{p} + d \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}^T$. The signed distance value is the minimum of this set, see (1). The set will only contain the value of infinity if there is no intersection for each of the directions. In other words, the network encodes the distance to the closest object contour in a discrete number of directions. Since the number of directions is discrete, we consider the SDF in (1) as an approximation. See Fig. 3 for an example on the output maps.

B. Our approach

Similar to semantic segmentation, we want to predict a vector that encodes distances to the closest object contours for each pixel. That is, given an input image, the output $\mathbf{y} = g(\mathcal{I})$ encodes the geometric structures established by the training dataset. For this purpose we use an encoder-decoder network $g : \mathbb{R}^{3 \times H \times W} \rightarrow \mathbb{R}^{2 \times 4 \times H \times W}$ that predicts distances in four directions of $\theta_k = 90k$ degrees, where $k \in \{0, \dots, 3\}$. The output \mathbf{y} is divided into two parts. A negative part $y^- \in \mathbb{R}^{4 \times H \times W}$ that defines the background, and a corresponding positive part, y^+ , that defines the foreground. To compute an approximate SDF $z_{uv} \approx f(\mathbf{p})$, we use

$$z_{uv} = \min_k y_{kuv}^+ - \min_k y_{kuv}^- , \quad (3)$$

where $\mathbf{p} = (u \ v)^T$. In other words, we take the minimum value along the first dimension of both parts and subtract these for all pixels.

We apply a L_1 -loss to the distance output of g , such that

$$\mathcal{L}_{\text{dist}} = \sum_{(u,v) \in \mathcal{I}} \sum_{\forall k} |\hat{y}_{kuv}^+ - y_{kuv}^+| + |\hat{y}_{kuv}^- - y_{kuv}^-| \quad (4)$$

is minimised. In (4) we denote y^+ as the annotated foreground distance and \hat{y}^+ as the predicted foreground distance. Similarly for the background part of (4).

In $\mathcal{L}_{\text{dist}}$ we utilise the value zero for two purposes, 1) to minimise the distance between the predicted distance and the true distance to an object contour and 2) for invalid distances, that is, when $\mathcal{D}_\theta^{\mathbf{p}}$ contains the value of infinity. From a pixel position that in a given direction does not intersect any object contours is in fact invalid. To use zero in the latter case maintains a simple loss. For the L_1 -loss this means that the expected value for each invalid location with a predicted positive distance should be suppressed to zero.

We initially observed that the predicted distances were noisy. We therefore regularise the network by adding a binary cross-entropy loss \mathcal{L}_{aux} to z . This loss drives the network to predict correct foreground-background segmentations, consistent with the distance output \mathbf{y} . We sum the two sub-losses into the total loss $\mathcal{L}_{\text{tot}} = \mathcal{L}_{\text{dist}} + \mathcal{L}_{\text{aux}}$ and backpropagate.

C. Implementations details

We train the model, including the encoder, with batch size 8 on one GPU using the COCO 2017 benchmark [9]. This dataset is divided into 118K training samples and 5K validation samples. The encoder used ImageNet [4] pretrained weights and the decoder was randomly initialised sampling from a uniform distribution. We train with SGD for 54 epochs starting at an initial learning rate of $2 * 10^{-3}$ and divide with 10 reaching epochs 19, 40, 48 and 51. Weight decay and momentum was 10^{-4} and 0.9 respectively. For data augmentation we used the same as in SSD [11], also used by YOLACT [1]. Training takes 4 days on a single Nvidia V100 GPU.

1) *Encoder-decoder*: In our model we use a ResNet50 [7] feature extractor as the encoder and a DFN [9] as the decoder. DFN [9] predicts a high-resolution output by fusing the deep feature maps with successively shallower features. Our distance model is motivated by this aspect. Hence, we hypothesise that consistency are kept for distance predictions to the contours of multi-scale objects. In the pooling layer the predicted distance maps are heuristically merged using two min-pooling operations that are concatenated with the

correct sign s , for background and foreground. Finally, this is summed along the first dimension, see (3).

2) *Distance annotations*: COCO [9] comprises 80 classes of objects. These classes were selected to well represent frequently used words to describe visually identifiable objects. The selection is based on the PASCAL VOC dataset [5], a list of the 1200 most common words to describe objects, where several children between 4 and 8 were asked to name every object encountered in indoor and outdoor environments [9]. The result is a set of 80 distinct categories, most fairly high-level, e.g. car, human, or dog. In each image, each object is annotated with a segmentation mask. Based on these masks, we compute the index map. In some cases, multiple object masks overlap, leading to some pixels being claimed by multiple objects. Our distance annotations rely on each pixel mapping the distance to the single closest object contour. We therefore, for those pixels, select the single object mask corresponding to the smallest mask of all the overlapping object masks. The rationale is that we then give all objects a chance to appear. This can of course turn out wrong in certain cases. As in the upper left image of Fig. 4, where people in the background, which are occluded by people at the front, will be prioritised incorrectly.

Distance maps, used in the loss which is described in Section III-B, are computed for the directions $\{0, 90, 180, 270\}$. We first rotate the index map according to the directions. This enables row-wise computation of the distance to the closest object contour. Finally the distance maps are rotated back to their initial state. The procedure is vectorised to increase overall GPU utilisation, which means faster training.

IV. RESULTS

We report a result on COCO instance segmentation [9]. Our model as, well as YOLACT [1], is trained on the 118K `train2017` images. We test out approach on the 5K `val2017` images and compare our foreground segmentation with the foreground segmentation obtained from YOLACT [1]. We use the intersection-over-union averaged over all 5K samples in `val2017`.

Table I compares the IoU of the foreground segmentation by *topk* detections and changing thresholds on z . Our approach yields a 4% increase over the state-of-the-art method YOLACT [1]. This result demonstrates that managing of complex shapes can be beneficial. We also observe a negative effect when increasing the number of *topk* detections as well as for larger and larger thresholds on z . However, the latter is expected, since small objects should disappear as result of the distance modelling. The former is contradictory towards this anchor-based method, since it should be expected that performance would go up when enabling more objects to be detected.

Using our method to predict foreground segmentation is faster compared to YOLACT [1]. We use the same feature extractor as in Bolya et al. [1], but without a computationally expensive Feature Pyramid Network [8]. However, this speed is not fair to compare, since more computations in our approach is necessary to achieve full instance segmentation.

TABLE I
COMPARING FOREGROUND SEGMENTATION BETWEEN YOLACT [1]
AND OUR APPROACH ON COCO `val2017` [9]. BOTH WITH A
RESNET50 [7] BACKBONE.

Method	z	topk	mIoU
YOLACT [1]	-	5	0.66
	-	50	0.65
	-	100	0.64
Ours	0.0	-	0.70
	3.0	-	0.67
	5.0	-	0.63
	10.0	-	0.53

In Fig. 3 we show two qualitative examples of our predicted distances. The rightmost show a simple example of a hoarse where prediction and ground truth are comparable. In the leftmost example a woman is playing tennis. The distance predictions are harder to compare to the ground truth since our model finds that the woman is occluded by a tennis net.

Furthermore, in Fig. 4 we show that our method predicts accurate distance maps. These are used to approximate a signed distance function where the positive set of function values describe the foreground segmentation. We also show cases that our method has difficulty with, such as object reflections, object shadows, low-light scenes, and very thin object parts.

V. CONCLUSION AND FUTURE WORK

We explored a new direction for the instance segmentation problem, where the predictions made by a neural network has a direct geometric interpretation. That is, we train the network to predict distances to object contours that are used to compute a signed distance function. We analyse this approach and show that it outperforms the state-of-the-art instance segmentation YOLACT at foreground-background segmentation. A possible explanation is that some shapes are not well described by anchor-boxes.

Although the predicted distance maps contain information on the individual instances it remains challenging to map this representation to a full instance segmentation. For this representation, an object can consist of different segmentation blobs that are separated due to e.g. occlusion. To label these object related parts with the same label identification is one of the challenges with instance segmentation, even though the blobs fit well to the object scope. We therefore believe this representation to be a promising direction for instance segmentation methods.

VI. ACKNOWLEDGEMENT

This work was partially supported by the Wallenberg AI, Autonomous Systems, and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

REFERENCES

- [1] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee. Yolact: Real-time instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.



Fig. 4. Our method predicts accurate distance maps that are used to separate the foreground from the background (first row). However, it has difficulty with very thin parts that belong to objects, such as the bicycle frame or the legs of the giraffes (first column). Our method also has difficulty with object reflections and object shadows (second row), as well as with poorly lit scenes (third row).

- [2] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko. End-to-end object detection with transformers. In *ECCV*, 2020.
- [3] X. Chen, R. Girshick, K. He, and P. Dollar. Tensormask: A foundation for dense object segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.
- [4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2), July 2010.
- [6] K. He, G. Gkioxari, P. Dollar, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.
- [7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [8] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and L. Zitnick. Microsoft coco: Common objects in context. In *ECCV*, September 2014.
- [10] S. Liu, Y. Zhang, S. Peng, B. Shi, M. Pollefeys, and Z. Cui. Dist:

- Rendering deep implicit signed distance function with differentiable sphere tracing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C.-Y. Fu, and A. Berg. Ssd: Single shot multibox detector. In *ECCV*, 2016.
- [12] D. Neven, B. D. Brabandere, M. Proesmans, and L. V. Gool. Instance segmentation by jointly optimizing spatial embeddings and clustering bandwidth. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [13] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove. DeepSDF: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [14] Y. Wang, Z. Xu, H. Shen, B. Cheng, and L. Yang. Centermask: single shot instance segmentation with point representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9313–9321, 2020.
- [15] X. Zhou, D. Wang, and P. Krähenbühl. Objects as points. *arXiv preprint arXiv:1904.07850*, 2019.

Robot First Aid: Autonomous Vehicles Could Help in Emergencies*

Martin Cooney, Felipe Valle and Alexey Vinel

Abstract— Safety is of critical importance in designing autonomous vehicles (AVs) that will be able to perform effectively in complex, mixed-traffic, real-world urban environments. Some prior research has looked at how to proactively avoid accidents with safe distancing and driver monitoring, but currently little research has explored strategies to recover *afterwards* from emergencies, from crime to natural disasters. The current short paper reports on our ongoing work using a speculative prototyping approach to explore this expansive design space, in the context of how a robot inside an AV could be deployed to support first aid. As a result, we present some proposals for how to detect emergencies, and examine and help victims, as well as lessons learned in prototyping. Thereby, our aim is to stimulate discussion and ideation that—by considering the prevalence of Murphy’s law in our complex world, and the various technical, ethical, and practical concerns raised—could potentially lead to useful safety innovations.

I. INTRODUCTION

A few decades into the future, if our cities are full of autonomous vehicles (AVs)—what could an AV do to help in an emergency? The current paper explores this question, also bringing together various ongoing work that we have been conducting since 2014 [1], [2], [3], [4].

Research on AVs, in fields such as Connected Cooperative and Automated Mobility, autonomous platooning, and mobile robotics, is progressing rapidly, toward improving people’s quality of life and saving money and time. One focus is on what can be done *before* an emergency occurs to prevent damage, via driver monitoring [5] and prediction of safe boundaries and paths [6]. Less attention has been given to what an AV could do *after* an emergency has occurred. Information about injuries could help expedite treatment, but contact-based sensing on a car wheel or seats can be limited to drivers or challenging due to clothing, and stationary cameras can be occluded. Thus, we see a potential benefit of having an on-board robot that can actively sense, and even physically seek to help, as depicted in Fig. 1.

Along similar conceptual lines, some teleoperated robots have been designed to facilitate first aid. For example, a military system was proposed to enable remote observation and surgery on a soldier on a stretcher [7]. Another design proposed that a teleoperator can pilot a drone with a defibrillator to go to a location when called [8]. Samani and Zhu

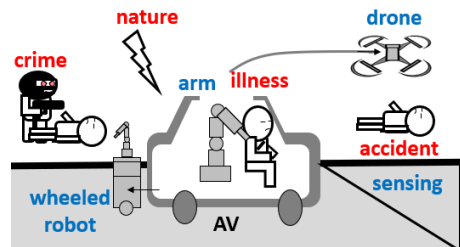


Fig. 1. Basic concept: first aid robots in an AV (blue) could save lives in emergencies (red) by detecting and treating health problems

also built a wheeled robot with a defibrillator that could be called via a smart phone app to go to a map location by calculating its own path [9].

Designs that can autonomously carry out medical procedures have also been proposed. For example, Wong et al. proposed a high-level process flow for a tracked robot with an arm to assess casualties on a battlefield; aside from following soldiers’ reports, it is suggested that the robot could automatically find casualties and conduct triage with Glasgow Coma Scale (GCS) evaluations [10]. Furthermore, similar to the current paper, Kurebwa and Mushiri proposed an on-board robot in a passenger car that could conduct triage and first aid without human intervention [11]. A follow-up study also proposed that the best placement for a first aid robot might be in the middle of a car near the floor, and that it should have its own power source [12]. These studies highlight interesting and useful possibilities, but what is missing is a proposal for how a robot could carry out a first response, which, in line with Moravec’s paradox, we assume will involve overcoming highly challenging and non-trivial sensorimotor challenges.

Thus, there seems to be a high potential value in this area, but so far little research has been done, possibly due to its high-risk and emerging nature.¹ Some questions that should be explored are as follows:²

- Q1 What emergency scenarios could exist?
- Q2 How can such emergencies be detected?
- Q3 How can victims’ health states be assessed?
- Q4 What can be done to improve victims’ health states?

II. METHODS

To explore the four questions we identified above, we use a *speculative prototyping* approach. “Speculative design”

¹Legal culpability is unclear, and AVs and assistive robots are not yet common in our cities.

²These are not all possible questions but merely some that we identified.

*We gratefully acknowledge support from the Swedish Knowledge Foundation (KKS) for the “Safety of Connected Intelligent Vehicles in Smart Cities – SafeSmart” project (2019–2023), the Swedish Innovation Agency (VINNOVA) for the “Emergency Vehicle Traffic Light Pre-emption in Cities – EPIC” project (2020–2022), and the ELLIIT Strategic Research Network.

¹M. Cooney, F. Valle and A. Vinel are with the School of Information Technology, Halmstad University 301 18 Halmstad, Sweden martin.daniel.cooney at gmail.com

involves conducting simplified problem-finding and scenario-building to extract thought-provoking ideas from expansive, ambiguous design spaces. Its pragmatic sister, "prototyping", offers a sanity check that balances speed of exploration with the accuracy of insights obtained.

A. Emergency Scenarios

Emergencies can involve accidents, medical problems (e.g. heart attacks, strokes, or epilepsy), crime (e.g. shootings or stabbings from robberies, carjackings, violence toward taxi drivers, or terrorist bombings) and nature (e.g. fires, earthquakes, floods, tornadoes, thunderstorms, tsunamis, or volcanic eruptions). Some considerations include if the AV should help (relating to what has sustained damage, people or the AV?), and if so what its robotic first aid embodiment might look like (relating to where the injured people are, inside or outside of the AV).

1) *Responsibility*: Generally, emergencies should be reported. Pathways being developed to report collisions such as Advanced Automatic Crash Notification (AACN) or Decentralized Environmental Notification Messages (DENM) could be adapted to remit telemetry data in various scenarios [13]. If humans are injured, AVs with useful capabilities could be legally required to stay and help, even if the AV has not caused the incident. Conversely, if injured humans are not detected, e.g. when vandalism and rioting are occurring, an AV might not need to become involved; in addition to sending a warning, an AV with some public duty could even play a siren sound through its speakers as it passes. Or, if the AV itself has been damaged, it can report to its owner or platoon that it will be delayed and request assistance.

2) *Embodiment*: If a human is injured, a robot can seek to help. Various kinds of robot exist, including humanoid, zoological, soft, wheeled and flying robots. To sense and touch injured passengers, a stationary robot arm affixed to the center floor could be sufficient. To deal with external victims, mobility would likely be useful, as the AV should probably not drive onto sidewalks and go where vehicles are prohibited, and an arm might not be practical at larger distances. Wheeled robots and drones could be used: A wheeled robot capable of clearing obstacles and helping humans could also be heavy and large; a detachable robot might be better placed at the rear of a car, where nowadays a first aid kit might be kept. Conversely, a drone placed at the center of a car could rise through a convertible top. Although highly mobile and capable of scanning a wide area from a high height, drones can suffer from lower payloads, short operation times due to light batteries, problems dealing with wind, and restrictions in where they can operate, etc.

B. Emergency Detection

Detection could occur either directly (e.g. the AV observing a fall or crime) or indirectly (e.g. the AV receiving a call for help from passengers, observers, remote emergency services, smart infrastructure, or other AVs). Interoceptive detection of medical problems or crimes can occur, as noted, with contact-based or remote sensors; exteroceptive detection

could involve audiovisual sensors such as microphones, cameras, thermal cameras, lidars, and radars, to detect distress (shouting, crying) and emergencies that are medical (falls, e.g. from heart attacks or strokes), accidental (crashes, explosions), natural (e.g. haze reducing camera visibility, bright flashes from lightning), or criminal. Thermal cameras could be useful to detect humans by body temperature, although various confounds arise from occlusions, other heat sources, and thermal reflections.

C. Health Inference

If no medical personnel is present at the scene, the robot should first call emergency services and allow teleoperation. If no one is available to control the robot, it should first move to a safe distance near injured humans (this could be trivial inside the AV but not externally). With multiple conscious victims, the robot can use a Patient Assist Method (PAM) to bootstrap triage, calling for victims to go to a casualty collection point if they need help, or leave if they do not, then identifying the immobilized remainder. The problem of identifying an efficient path (minimizing the competitive ratio) between partially observable victims could potentially be formulated as a variant of the Canadian Traveller Problem, where not only edge but also vertex membership might be unknown [14]. Then START (Simple Triage and Rapid Treatment) can be used to identify those likely to be affected by treatment (those who will either live or die regardless, or who can be saved with either immediate or delayed care) [15]. The robot can ask conscious victims to explain their injuries and needs, but must find another way to form an opinion about the incapacitated. For this, scoring systems like the Injury Severity Score (ISS) can be used, which use three categories—A head, B midsection, C outer parts—to assign a score from 0 to 75 based on severity of injury.

Based on the triage results, the robot can follow the CABD procedure to assess circulation, airway, breathing, and any deadly bleeding [16]. Furthermore, a person's state of awareness can be evaluated with various similar scales—e.g. GCS, Full Outline of UnResponsiveness (FOUR), Simplified motor scale, and AVPU ("alert, verbal, pain, unresponsive"); GCS is common but suffers from poor inter-rater reliability, whereas FOUR and SMC haven't gained consensus [17]. Here we focus on AVPU, whose simplicity suggests its usefulness as a starting point for prototyping; furthermore, AVPU can be conducted before a more complex method.

The AVPU scale, however, is formulated for humans rather than robots. Fig. 2 shows how AVPU could be encoded. Here, dlib can be used to threshold eye aspect ratios to detect open eyes.³ The sounddevice library and CMU PocketSphinx or Google Speech can be used to detect utterances.⁴ OpenPose, Kinect skeletons, foreground extraction, or optic flow could be used to detect motion.^{5,6} Touching a motionless person's hand or chest can involve inverse kinematics to position the

³<http://dlib.net/>

⁴<https://cmusphinx.github.io/>

⁵<https://github.com/CMU-Perceptual-Computing-Lab/openpose>

⁶<https://opencv.org>

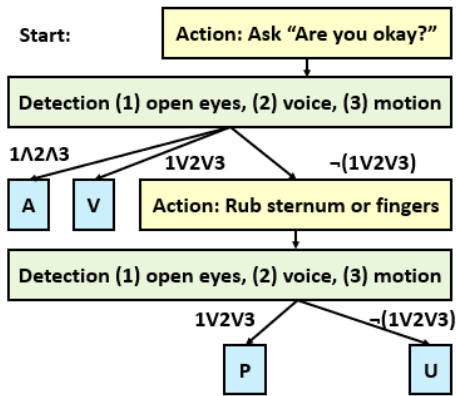


Fig. 2. Process flow for how a robot can check a person’s state of consciousness via the AVPU scoring system

robot’s end effector at inferred joint locations. Contingency can be inferred with a simple threshold pair (0.5s to 3s), or in a richer manner using Poisson threshold learning [18]. Thus, victims will be assessed at various times, resulting in a time series of assessments.

D. Health Improvement

Based on recognition results, a lone robot might have to act to save lives. Aside from tasks such as attaching a triage tag, one decision is if the patient should be moved to the AV to be transported to a medical facility, or treated onsite.⁷ If CABD is required, a first response could include chest compressions, adjusting posture to lift the chin or place in a recovery pose, rescue breathing, and bleeding control, potentially with tourniquets. Specific health scenarios could also require other actions, like the use of a defibrillator for cardiac arrest. Before such ideas can be realized, however, a robot should first be able to conduct health inference.

E. Prototyping and Implementation

To obtain some practical insight, several short prototyping design studies were conducted, focusing on the example of a fallen person, as shown in Fig. 3. We focused on detection and inference tasks that seemed feasible and to have not received much attention yet, for two kinds of robots, wheeled and flying: The wheeled robot was a Turtlebot 2, a small differential drive mobile base with a Microsoft Kinect sensor, and a homemade arm with sensors on top (0.35 x 0.35 x 0.42m, 6.3 kg). The flying robot was an off-the-shelf Parrot AR Drone. Robot Operating System (ROS) was used for navigation and visualization, and Festival and CMU PocketSphinx for speech-based interactive capabilities.

1) *Emergency detection: Direct detection:* A simplified approach for fall detection was implemented by thresholding vertical displacement of the shoulder joint of a frontally located person (mannequin) detected via Kinect, while also noting fall direction (forward, backward, or sideways), which

⁷For example, something like Panasonic’s Transfer Assist Robot could help people to change pose to be loaded into the AV.

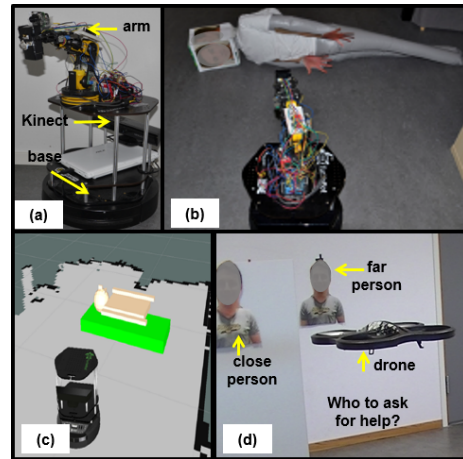


Fig. 3. Scenes from prototyping: (a-c) wheeled robot, (d) drone

could be valuable for estimating injury locations. The robot was furthermore set to detect fallen persons while patrolling through a lab environment, by comparing the size and temperatures of clustered laser scans within a known map with thresholds corresponding to human size and temperature. The prototype estimated the location of body parts of interest for first aid (chest, hands, chin, mouth, and nose) based on face and skin detection, and a simplified prior model. The estimated pose was visualized over a map of the environment [4].⁸

Indirect detection: We also explored the detection of anomalies, such as a human not moving from a location for a long time, via smart infrastructure. Data from eleven sensors, comprising four pressure, three contact and four passive infrared sensors, were processed via a Random Forest, triggering a database "brain" to wirelessly command the mobile robot to head to a map location [1].

2) *Health Inference:* After detecting an emergency and sending a warning, a robot should try to find help. We know that medical interventions through teleoperation can be successful. For the case when no remote expert is available, we explored having a drone prototype rotate at head height to detect nearby adults, and then infer their ages and proximities via the height and size of detected faces [2]. For the case when there are no other people, we designed a mobile robot to verbally ask a fallen person if they require emergency medical services, and decide not to call only if they respond that everything is okay [1]. For the case when a person is unresponsive, we developed algorithms to check for relative blueness in the distal portion of the hands to assess circulatory state (*peripheral cyanosis*), chin pose for airway, speed and normalcy of sound for breathing, and location and rate of expansion of red color for bleeding, via support vector machines and audiovisual features [3].

⁸To detect averted faces, the prototype navigated around the anomaly while scanning with its arm camera and running face detection on data rotated through various angles; additionally a visual servoing algorithm was built for the robot to indicate points of interest with a laser pointer for debugging and as a potential aid for medical personnel

F. Evaluation results

A simplified evaluation was conducted for the developed capabilities: *Direct emergency detection*. Accuracy was 85% for emergency detection: 80% for detecting fall direction, and 90% for detecting fallen persons (85% for anomalies, and 95% for detecting human temperature). Faces were detected in 70% of cases, for which average error was 0.015m.

Indirect emergency detection. The prototype required an average of 13.8s (SD: 7.9) to navigate to four locations within the space we created. Verbal responses from the experimenter at the anomaly location were correctly recognized 76.9% of the time, with problems arising due to the timing of when the robot should recognize.

Finding help. Two printouts of human upper bodies were attached to a wall at varying heights and distances; the drone approached the correct target 90% of the time (18/20 trials) but crashed into the wall and a target one time each due to drifting, erratic air currents, and low ground contrast, leading to a reasonable initial rate.

Detecting vital signs. Average accuracy for vital signs was 79%: 65% for cyanosis, 75% for chin pose, 85% for breathing, and 91% for bleeding (97% location/85% speed).

Accuracies were imperfect due to various factors: sensor noise and inaccuracies, some challenging conditions (proximity to walls, warm objects, extreme angles, and low resolution data), and simplifications in our models. We believe these results were reasonable given the exploratory study design, but suggest the high challenge that designers will face in bringing such systems into the real world.

III. DISCUSSION

The contribution of the current paper is in proposing ideas for how an AV can deal with accidents after they occur, with a focus on perception for robot first aid. We proposed four emergency categories, with a strategy for when an AV should become involved, how its onboard robot could be embodied, and how emergencies could be detected. The adaptation of medical approaches for triage and health assessment was discussed, as well as what robot actions will be required to conduct first aid. Finally, we reported on some experiences prototyping initial solutions.

A. Limitations and Future Work

Our work is limited by its exploratory nature; future work will look at various technical, ethical, and practical challenges: Interventions such as chest compressions should be explored, along with how to conduct treatment in a moving AV (e.g. taking into account road conditions during surgery?), implement other assessment approaches such as GCS, and detect potential confounds (e.g. people lying in a park might not require emergency assistance). As well, ethical and legal concerns must be considered. For example, a difficulty with triage is that not all factors can be considered and boundaries can be vague. Also, could criminals hack AVs to corner, rob, or hurt people, or seize control of an AV by faking an injury? (And if so, could reinforcement learning be used to provide an AV with some basic concept

of self-preservation and defence?) Practically, robots are also expensive, not all AVs might require a first aid robot, and a robot could get in the way in life or death situations. By encouraging such discussion, our aim is to facilitate the realization of AVs and robots that can provide enhanced safety in our daily lives.

ACKNOWLEDGMENT

We thank in particular our students who helped with some of the prototyping!

REFERENCES

- [1] J. Lundström, W. O. De Moraes, and M. Cooney, "A holistic smart home demonstrator for anomaly detection and response," in *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 2015, pp. 330–335.
- [2] J. Heyne, "Assistance-seeking strategy for a flying robot during a healthcare emergency response (internship report, halmstad university)," 2015.
- [3] T. Zhang and Y. Zhao, "Recognition for robot first aid: Recognizing a person's health state after a fall in a smart environment with a robot," 2016.
- [4] W. Hotze, "Robotic first aid: Using a mobile robot to localise and visualise points of interest for first aid," 2016.
- [5] M. Q. Khan and S. Lee, "A comprehensive survey of driving monitoring and assistance systems," *Sensors*, vol. 19, no. 11, p. 2574, 2019.
- [6] J. Thunberg, G. Sidorenko, K. Sjöberg, and A. Vinel, "Efficiently bounding the probabilities of vehicle collision at intelligent intersections," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 2, pp. 47–59, 2021.
- [7] G. Martinic *et al.*, "Glimpses of future battlefield medicine—the proliferation of robotic surgeons and unmanned vehicles and technologies," *Journal of Military and Veterans Health*, vol. 22, no. 3, p. 4, 2014.
- [8] A. Katz, "News: Drones: The (possible) future of medicine," 2015.
- [9] H. Samani and R. Zhu, "Robotic automated external defibrillator ambulance for emergency medical service in smart cities," *IEEE Access*, vol. 4, pp. 268–283, 2016.
- [10] K. H. Wong, S.-C. B. Lo, C.-F. Lin, B. Lasser, and S. K. Mun, "Imaging components for a robotic casualty evaluation system," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 467–470.
- [11] J. G. Kurebwa and T. Mushiri, "Internet of things architecture for a smart passenger-car robotic first aid system," *Procedia Manufacturing*, vol. 35, pp. 27–34, 2019.
- [12] J. Kurebwa and T. Mushiri, "A study of damage patterns on passenger cars involved in road traffic accidents," *Journal of Robotics*, vol. 2019, 2019.
- [13] G. Bahouth, J. Graygo, K. Digges, C. Schulman, and P. Baur, "The benefits and tradeoffs for varied high-severity injury risk thresholds for advanced automatic crash notification systems," *Traffic injury prevention*, vol. 15, no. sup1, pp. S134–S140, 2014.
- [14] A. Bar-Noy and B. Schieber, "The canadian traveller problem," in *Proceedings of the second annual ACM-SIAM symposium on Discrete algorithms*. Citeseer, 1991, pp. 261–270.
- [15] H. Nakao, I. Ukai, and J. Kotani, "A review of the history of the origin of triage from a disaster medicine perspective," *Acute medicine & surgery*, vol. 4, no. 4, pp. 379–384, 2017.
- [16] A. H. Travers, T. D. Rea, B. J. Bobrow, D. P. Edelson, R. A. Berg, M. R. Sayre, M. D. Berg, L. Chameides, R. E. O'Connor, and R. A. Swor, "Part 4: Cpr overview: 2010 american heart association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care," *Circulation*, vol. 122, no. 18_suppl_3, pp. S676–S684, 2010.
- [17] M. Fischer, S. Rüegg, A. Czaplinski, M. Strohmeier, A. Lehmann, F. Tschan, P. R. Hunziker, and S. C. Marsch, "Inter-rater reliability of the full outline of unresponsiveness score and the glasgow coma scale in critically ill patients: a prospective observational study," *Critical care*, vol. 14, no. 2, pp. 1–9, 2010.
- [18] K. Gold and B. Scassellati, "Learning acceptable windows of contingency," *Connection Science*, vol. 18, no. 2, pp. 217–228, 2006.

Towards motivation-driven intelligent interfaces: formal argumentation meets activity theory

Esteban Guerrero

Department of computing science
Umeå university
Umeå, Sweden
esteban@cs.umu.se

Helena Lindgren

Department of computing science
Umeå university
Umeå, Sweden
helena@cs.umu.se

Abstract—Theories about human activity and motivation point out that motives are driving forces behind human activities and development of healthy and unhealthy habits. Activity theory is one of these that has been applied to develop activity-centered user interfaces. Activity theory differentiates between sense-making and stimuli-oriented types of motives that have a strong influence on our daily behavior. Two main challenges are explored in this paper: 1) the personalisation of graphical user interfaces to mediate representations of motivation-based activities to support behaviour change processes; and 2) the proactiveness of such visual representations.

As methods, we use *activity theory* as a framework for defining the motivations' dynamics, and *formal argumentation theory* as the underlying mechanism for interactive reasoning and decision-making in the process of generating the user interface.

Our contributions are two-folded: 1) a dynamic graphical user interface where the background responds to behaviors linked to sense-making motives, and the foreground to stimuli motivation; and 2) a non-monotonic reasoning mechanism endowing the user interface with proactiveness (not only react to the user interactions but trigger and direct attention to potential conflicts), and a motive-based behavior conflict resolution process. Future work includes user studies to explore how triggering of focus may create increased awareness in an individual of conflicting motives in daily activities and how this may support changes of unhealthy habits.

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

This paper presents a novel computational mechanism to visually represent activity and their conflicting motives for a person based on the underlying motives for activity execution.

Two main challenges relating to humans interacting and collaborating with intelligent coaching systems are explored in this paper: 1) the personalization of graphical user interfaces to represent motivation-based activities as part of the individual's attempt to change behaviour for improving health; and 2) proactiveness of visual representations, i.e. interfaces that not only react to the user's interactions but have a certain level of autonomy to trigger the user's attention and direct a change of focus.

Research was supported by Forte, the Swedish Research Council for Health, Working Life and Welfare, which supports the STAR-C project during 2019–2024 (Dnr. 2018-01461) 978-1-6654-4236-7/21/\$31.00 ©2021 IEEE

Studies have explored the connection between motivation and general strategies for designing intelligent interactive coaching systems [1], [2]. To increase the social intelligence of intelligent software agents, social science theories have recently been adopted (e.g., [3]). Also social-cultural theories such as Activity Theory (AT) [4] earlier adopted by researchers in human-computer interaction (HCI), pedagogy and work development studies, have been incorporated for personalisation purposes of intelligent coaching systems (e.g. [5]). Activity theory presents two general types of motives: *sense-forming motives*, which give the activity its meaning, and *motive-stimuli* motives that provoke various emotional reactions in a person [6]. This general description of human motivation is similar to how motivation is described in some behavior change models [7] where stimuli and sense-forming motives can be seen as opposite poles of a motivation line. Moreover, Activity theory provides a framework for understanding conflicts, e.g., between motives.

When intelligent software agents are endowed with common-sense reasoning abilities and the ability to take the initiative to deduce a sufficiently wide class of immediate consequences, “change its mind” when more information is available, and in addition act based on its decisions, we may call it *proactive* behaviour [8].

In this paper, we address these problems **hypothesizing** that: 1) Visual representations of an individual's sense-forming motives and motive stimuli are necessary to be displayed in a user interface to support behavior change; 2) Personalization of a user interface can be leveraged by the integration of manual (by a user) and automatic (by a software agent) inputs to tailor visual representations, i.e. allow the human and agent to collaborate in the process; and 3) the agent's proactiveness and common-sense reasoning can be implemented through a computational mechanism that reasons about motives and activities (habits) of a user, and be mediated by a user interface that is organised based on the motivational structures.

In this setting, this paper presents two **main contributions**: 1) a mapping between interface design assumptions, theoretical background, and formal implications of those assumptions, and 2) the *MOArg* framework, that is a computational mechanism that reasons about motives and activities of a person to provide proactiveness of a user interface.

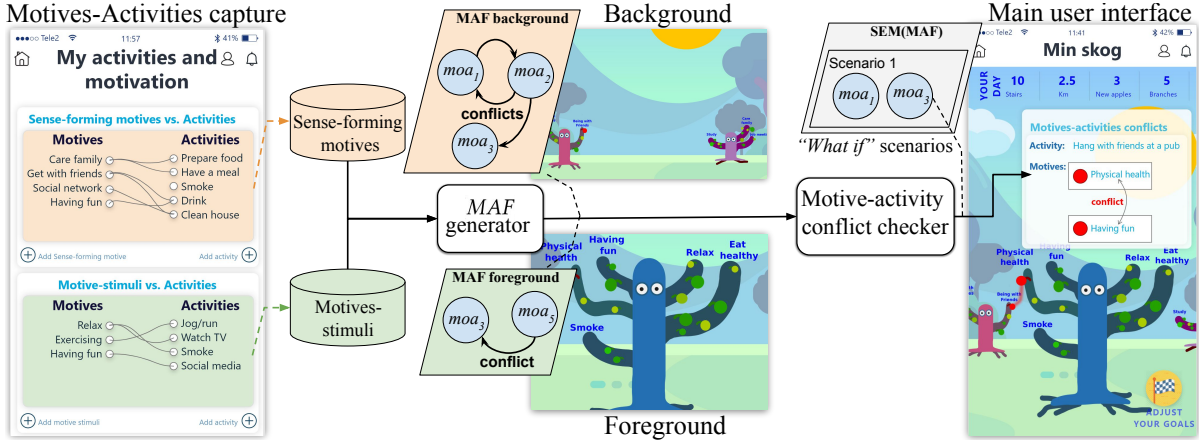


Fig. 1. Sense-forming motives and motive-stimuli used to control background and foreground of an interactive proactive user interface

This paper is structured as follows: Section II introduces key concepts about motivation from an AT perspective, and definitions from *formal argumentation theory* that is a formalism for capturing various approaches to common-sense reasoning, in particular non-monotonic reasoning and cooperative games [9]. In Section III, our main design and formal results are presented. Section IV highlights our novel contributions and draws our main line of future work.

II. METHODS

A. Motives behind an activity

Leontiev proposed two types of motives of an activity [4]: *sense-forming* motives, which give the activity its meaning, and *motive-stimuli*. Motive-stimuli can stimulate a person and elicit various emotional reactions (sometimes coming into conflict with the general purpose of an activity) but they are of secondary importance compared to the sense-forming motives. Therefore, in the case of a conflict, sense-making motives prevail over motive-stimuli [6]. Table I exemplifies this categorization.

TABLE I
TWO TYPES OF GENERAL MOTIVES IN ACTIVITY THEORY

Sense-forming motives	Motives-stimuli
Object-oriented	Emotional-oriented
Stable	Temporary
Internal	External
<i>Examples:</i>	
<ul style="list-style-type: none"> Maintain physical health Nurture social relationships 	<ul style="list-style-type: none"> Feeling calm and rested Feeling part of a social context

B. Formal syntax

In this paper, we use *propositional logic* with a syntax language that uses the connectives \wedge, \leftarrow, \neg . A *program* P is a set of propositional atoms encoding information, such as motives and activities of a user, but also decisions of a software agent. We will use the following notation regarding motives and activities: $M_{sense} = \{se_1, \dots, se_i\}$ and

$M_{stimuli} = \{st_1, \dots, st_j\}$ representing sets of sense-forming and motives-stimuli; $Ac = \{ac_1, \dots, ac_k\}$ are sets of activities such as smoke, a physical activity, alcohol drinking, etc.; two conflicting motives se_i, se_j are denoted as $\text{Conf}(se_i, se_j)$. We call $\mathcal{M} = M_{stimuli} \cup M_{sense}$ the entire set of motives. We note $\text{Directs}(se, ac)$ a motive se as the force behind an activity ac .

C. Non-Monotonic Reasoning about Motives and Activity

We will use argumentation theory as a formal mechanism for capturing conflicts between activities and motives, and for providing suggestions to a user about activity. In the following, we will introduce a basic notation and definitions connected with argument-based structures.

Definition 1 (Motivation-based activity: moa): Let $s \in \mathcal{M} = M_{stimuli} \cup M_{sense}$ be a motive, and $ac \in Ac$ be an activity, where $\mathcal{M}, Ac \subseteq P$. The tuple $moa = \langle s, ac \rangle$ is an argument-based structure where $s \vdash ac$, and s is minimal for the sets s satisfying the previous condition. s is called the *support* and ac is the *conclusion* of a *moa*.

Intuitively, a *moa* following the previous definition, says that an activity ac has the potential to be achieved if a motive s is present. We denote MOA the set of all moas that can be obtained from a program P . A *sub-moa* is a moa that is part of the support of another moa, which introduces a notion of *multi-motivated activity*. We use handy functions $\text{Activ}(moa)$ and $\text{Motiv}(moa)$ to retrieve the activity and the motive from a given moa. We can formally define a moa *conflict*:

Definition 2 (Conflicting moas): Let $\langle s_1, ac_1 \rangle, \langle s_2, ac_2 \rangle \in MOA$. $\langle s_1, ac_1 \rangle$ is in *motivation conflict* with $\langle s_2, ac_2 \rangle$ iff $s_1 \cap s_2$, and is in *activity conflict* iff $ac_1 \equiv \neg ac_2$.

In this paper, a *graph* structure considering moas as nodes and conflicts (Definition 2) will be called a motivation-activity framework (MAF). Then, *argumentation semantics* [10], which are abstract patterns of selection, can be used on a MAF

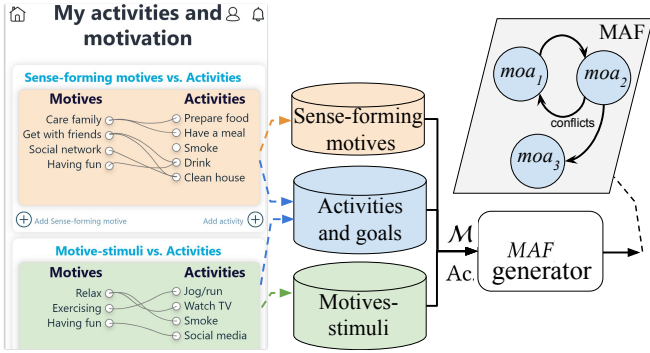


Fig. 2. Capturing motives and activities to build MOAs to obtain *extensions*, which are non-conflicting sets of moas, more formally: $SEM(MAF) = \{Ext_1, \dots, Ext_m\}$ where every extension is a set of moas $Ext_i = \{moa_1, \dots, moa_n\}$.

III. RESULTS

A. Motivation-Driven User Interface Generation

Based on the abstract division of motives in [4], our interface design was divided in two main visual representations where sense-forming motives (M_{sense}) as a stable, internal driving forces, define an *interactive background*; and motive-stimuli motives ($M_{stimuli}$), temporary, emotion-driven, determine an *interactive foreground* (see Figure 1 and Figure 3). Our main graphical character is a tree-like representation, with motives as branches, and *moas* as small fruits. Given this general visual approach, different assumptions regarding the design and the interaction with visual elements (*e.g.* trees, branches and fruits) are necessary to make in order to have a sound design. A novel result of this paper is a *mapping* between design assumptions, underlying theories and formal implications, which is summarized in Table II. In this mapping, we use $tree_{back}$, $tree_{front}$, $branch_{back}$, $branch_{front}$, $fruit_{back}$ and $fruit_{front}$ to designate those visual representations in the background and the foreground. Additionally, we use $context_{user}$ to refer to user's context, *e.g.* time and location.

B. MOArg framework for proactive interfaces

Our framework to build proactive interfaces has three steps: 1) *MOA* generation; 2) *MOA* interface update/display; and 3) “What if” scenario creation based on moas.

1) *MOA generation*: Sense-forming and stimuli-motives are formulated by the user to distinguish between their motives and activities, and their relationships. This step can be seen as a *baseline assessment* of an individual's situation, *e.g.*, as a starting point to improve health. A user may list activities such as *exercising by going for a run*, *get together with friends at the pub*. These activities serve sense-forming motives such as *maintain physical health*, *having fun*, *nurture social relationships*. The connections between these are defined by the user and visualised in the background. Motive-stimuli are the additional motives for an activity in terms of triggered positive and negative emotions, *e.g.*, *calm*, *excited*, *happy*, *low*, *worried*, *anxious*, *stressed*. These connections are visualised in

the foreground (Figure 2). At this point, we have as *input* from the MOA generator a full set of motives \mathcal{M} , a set of activities \mathcal{Ac} and their relations.

2) *Interface Update and Visualisation of MOAs*: Personalization of the user interface is performed by updating and linking motives and activities, using a simple user interface (see left in Figure 1). Every branch of a tree has a random shape, which is generated by modifying the structural forming vectors in JavaScript. *moas* as fruits, have parameterized random sizes and colors, except for red indicating conflict (see Figure 3). Proactiveness of our interface is achieved by the automatic creation of moas using an AI-approach for detecting four types of activities, walking, running, cycling and bus/car transportation (see Section III-C listing tools used in our implementation).

Then, those moas are displayed in the main tree. Other type of moas are added explicitly by a user when they achieve them, for example, partying with friends, drinking, etc. Moas' conflicts, are detected using our argumentation framework and highlighted when they are identified (see Figure 3).

The interface *background* contains two types of visual elements:

- Tree-like visual representations of sense-forming-based moas. These trees have the same configuration as the main foreground tree image, but the relative position and size, makes the background resemble a collection of other trees, *i.e.* a forest.
- Sun/moon and sky changing with time and context. We obtain from the user's device (*e.g.* mobile phone or browser) time and approximate location to update visual background behind the background trees.

3) *Argument-Based “what if” Scenarios*: A key feature of our argument-based approach is the generation of “what if” scenarios, which technically are non-conflicting moas highlighting those *non-admissible* or *non-supported* by the argumentation semantics. We use a *non-skeptic* argumentation semantics in $SEM(MAF)$ to generate as many scenarios as possible, which are highlighted in the user interface.

Proposition 1 (Potential scenarios of motives-activities): Sets of non-conflicting moas generated from $SEM(MAF)$ are scenarios where motivation-based activities generating conflicts are hypothetically discarded, therefore user interfaces using non-conflicting moas sets can display hypothetical scenarios where a user can speculate about their motives and activities framed in a particular behaviorchange context.

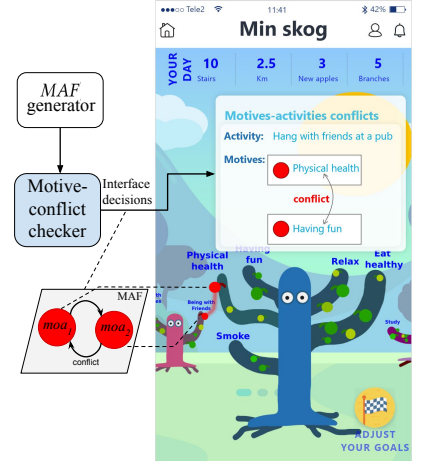


Fig. 3. Displaying conflicts among moas.

TABLE II
DESIGN ASSUMPTIONS, THEIR FORMAL INTERPRETATIONS AND UNDERLYING THEORY

Design assumption	Ref.	Characterization
1. Multi-motivated activities <ul style="list-style-type: none"> Background branches as sense-forming motives can be linked to foreground fruits in foreground branches. Foreground visual entities (fruits) can be associated with multiple stimuli (branches). 	[4], [11]	$\exists moa \in \mathcal{M} \in tree_{back} \text{Activ}(moa) \in tree_{front}$ with $ac_d \in Ac$
2. Motivation conflicts <ul style="list-style-type: none"> Conflicting moas exist, and they are displayed in trees as red fruits. 	[6]	$\exists moa \in MOA \in tree_{front} \text{Motiv}(moa) \in tree_{front}$
3. Hypothetical activity-motives scenarios <ul style="list-style-type: none"> The user selects a red fruit (conflicting moa) to display hypothetical alternatives. 	[12]	$\exists moa_a, moa_b \in MOA \text{Conf}(moa_a, moa_b)$ where $moa_a, moa_b \in tree_{front} \vee tree_{back}$
	[5]	$\nexists moa_a, moa_b \in MOA$ and $\text{Conf}(moa_a, moa_b) $ moa_a, moa_b are in the same extension of $SEM(MAF)$

Proposition 1 is a key contribution since it provides a formal relationship between visual alternatives for representing activity change, and a theoretical characterization of non-monotonic reasoning, i.e. a link between the system’s reasoning and the human’s, thus enabling a transparent collaborative reasoning allowing the human to engage in the process. This contribution is in line with the ambition to develop human-centric AI [13].

C. Implementation

The prototype was built using the following technologies; the argument construction was accomplished using TweetyProject and Google API; the web implementation was built using Vue.js, GreenSock and Ionic Framework; and the graphical user interface was implemented using Inkscape and Adobe XD. Part of the software is released as open-source.

IV. CONCLUSIONS

By combining an individual user’s activities, their motives, and the relationships between these based on Activity theory in a common formal framework that can be used by a system for reasoning about conflicts, a *narrative* of user experiences can be visualised, trigger attention and direct focus, and be used as tool for reason about how to solve conflicts. By distinguishing between the two types sense-forming motives and motive-stimuli following Activity Theory, a user may become aware of how he/she make decisions, and may chose to alter habitual behaviours that may be less healthy.

The key contribution is the MOArg framework, which is an abstract mechanism for reasoning about activities and motives. It has a three-fold purpose: 1) capture user’s information as activity-motives tuples, called *moas*, 2) provide a proactive behavior to the interface through non-monotonicity of the argumentation-based process, e.g., when a moa is created/added, other moas may be in conflict, which is detected automatically and mediated to the user; and 3) it provides a mechanism for showing non-conflicting scenarios, which we denote “what if” perspectives that are hypothetical views when conflicting activities and motives are eliminated (see Proposition 1).

The graphical interface generated by the MOArg framework was implemented using state-of-the-art Web technology, making our prototype extensible to be evaluated on different

platforms. Part of our designs and tools are released as open-source. As future work, the current proposal needs to be evaluated in users studies to explore how triggering of focus may create increased awareness in an individual of conflicting motives in daily activities and how this may support changes of unhealthy habits. We anticipate that our framework with collaborative features would generate a *gaming nature* to the framework.

REFERENCES

- [1] H. Lindgren, E. Guerrero, M. Jingar, K. Lindvall, N. Ng, L. Richter Sundberg, A. Santosa, and L. Weinehall, “The star-c intelligent coach: a cross-disciplinary design process of a behaviour change intervention in primary care,” in *pHealth 2020*, vol. 273. IOS Press, 2020, pp. 203–208.
- [2] S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen *et al.*, “Guidelines for human-ai interaction,” in *Proceedings of the 2019 chi conference on human factors in computing systems*, 2019, pp. 1–13.
- [3] F. Dignum, “Autonomous agents with norms,” *Artificial Intelligence and Law*, vol. 7, no. 1, pp. 69–79, Mar 1999.
- [4] A. N. Leontyev, “Activity and consciousness,” *Moscow: Personality*, 1974.
- [5] E. Guerrero, J. C. Nieves, M. Sandlund, and H. Lindgren, “Activity qualifiers using an argument-based construction,” *Knowledge and Information Systems*, vol. 54, no. 3, pp. 633–658, 2018.
- [6] V. Kaptelinin, “The object of activity: Making sense of the sense-maker,” *Mind, culture, and activity*, vol. 12, no. 1, pp. 4–18, 2005.
- [7] J. O. Prochaska and C. C. DiClemente, “Stages and processes of self-change of smoking: toward an integrative model of change,” *Journal of consulting and clinical psychology*, vol. 51, no. 3, p. 390, 1983.
- [8] J. McCarthy *et al.*, *Programs with common sense*. RLE and MIT computation center, 1960.
- [9] P. Baroni, F. Toni, and B. Verheij, “On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games: 25 years later,” *Argument & Computation*, no. Preprint, pp. 1–14, 2020.
- [10] P. M. Dung, “On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games,” *Artificial Intelligence*, vol. 77, no. 2, pp. 321–357, 1995.
- [11] V. Kaptelinin and B. A. Nardi, *Acting with Technology: Activity Theory and Interaction Design*, ser. Acting with Technology. MIT Press, 2006.
- [12] H. Lindgren, E. Guerrero, and R. Janols, “Personalised Persuasive Coaching to Increase Older Adults’ Physical and Social Activities: A Motivational Model,” in *PAAMS conference*. Cham, Switzerland: Springer, Jun 2017, pp. 170–182.
- [13] Luc Steels, “Personal dynamic memories are necessary to deal with meaning and understanding in human-centric ai,” in *NeHuAI@ECAI*. CEUR-WS.org, 2020.

Towards a Machine Learning Framework for Drill Core Analysis

Christian Günther¹, Nils Jansson², Marcus Liwicki³, and Foteini Simistira-Liwicki⁴

Abstract— This paper discusses existing methods for geological analysis of drill cores and describes the research and development directions of a machine learning framework for such a task. Drill core analysis is one of the first steps of the mining value chain. Such analysis incorporates a high complexity of input features (visual and compositional) derived from multiple sources and commonly by multiple observers. Especially the huge amount of visual information available from the drill core can provide valuable insights, but due to the complexity of many geological materials, automated data acquisition is difficult. This paper (i) describes the difficulty of drill core analysis, (ii) discusses common approaches and recent machine learning-based approaches to address the issues towards automation, and finally, (iii) proposes a machine learning-based framework for drill core analysis which is currently in development. The first major component, the registration of the drill core image for further processing, is presented in detail and evaluated on a dataset of 180 drill core images. We furthermore investigate the amount of labelled data required to automate the drill core analysis. As an interesting outcome, already a few labelled images led to an average precision (AP) of around 80 %, which indicates that the manual drill core analysis can be made more efficient with the support of a Machine Learning/labeling workflow.

I. INTRODUCTION

In mining, quantitative and qualitative geological models of mineral deposits form the basis of the value chain. The output of the models' effects all subsequent decisions on valuation, mining method, processing method, and measures to alleviate the environmental impact of mining. The primary input data in such models are derived from the analysis of drill cores; continuous cylindrical rock samples of the sub-surface, extracted from drill holes produced by exploration drilling. By interpolating data (e.g., metal grade) and features (e.g., geological contacts) between several different drill holes, geologists can construct 3D models of geological bodies, including mineral deposits.

A common problem in drill core logging lies in the complexity of geological materials, requiring the logging

geologist to integrate numerous visual features (e.g., texture, structure) with compositional features (e.g., mineralogy, geochemistry). The complexity of in particular the visual features makes it hard to easily log them systematically, which in turn complicates interpolation between different drill holes and thus impacts negatively on the geological models. This is even more problematic in modern situations, where mines commonly drill tens of thousands of meters a year, and several different observers need to be engaged in drill core logging to keep pace with production planning. Since visual characterization can be subjective, this can insert uncertainties and ambiguities into the geological interpretations. Furthermore, the resulting time-pressure results in that generally, a pragmatic approach is adopted in which only a fraction of the visual information stored in the core is systematically logged. Hence, the full information stored in the rocks is seldom tapped, which may impact negatively on the mines in the long-run, for example; by producing geological models that may be sufficient for production-planning, but which are too rudimentary to be useful for mineral exploration around the mine. This in turn can eventually lead to depletion of resources and mine closure. Besides that, the number of drill core samples and images created from it is in itself a challenge since archives with several thousand images are created, and the complete manual processing is very time-consuming and cost-intensive.

One way to mitigate these problems is by implementing new innovative Computer Vision (CV) and Machine Learning (ML) tools for the analysis of visual and compositional data. The developed models and algorithms can assist in making logging both more time-efficient and consistent. They can also be applied to e.g., data and images from legacy drill cores, as need arises (e.g., extracting geotechnical data from early exploration drill cores). A pre-requisite for this to be successful lies in a successful synergy between domain-expertise in hard rock geology and technological-expertise in machine learning, and integration with human experts and users. Of critical importance is the need to create quantitative and qualitative reliable training data for supervised machine learning, such as the ground truth for the drill core imagery. The labels that the machine learning models can be trained on needs to be in harmony with the images. Geological data is commonly stored as depth intervals (e.g., lithologies), where depth has been manually measured by the logging geologist in relation to depth marks in the core boxes. However, drill core images are seldom depth-registered, whereby pixels in the image

¹C. Günther is with the Machine Learning Research Group, Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, 971 87 Luleå, Sweden christian.gunther at ltu.se at ieee.org

²N. Jansson is with the Ore Geology research group, Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, 971 87 Luleå, Sweden nils.jansson at ltu.se

³M. Liwicki is with the Machine Learning Research Group, Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, 971 87 Luleå, Sweden marcus.liwicki at ltu.se

⁴F. Simistira-Liwicki is with the Machine Learning Research Group, Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, 971 87 Luleå, Sweden foteini.liwicki at ltu.se

cannot be directly related to such tabulated intervals without pre-processing. Furthermore, geological boundaries can be distinct, gradational, interfingering etc., and there is always a degree of simplification and subjectivity based on experience when geologists assign hard interval boundaries to gradational transitions, presenting a challenge for automation. The drill core images furthermore need additional pre-processing because of exposure to surroundings. Without such pre-processing, further analysis is deemed to generate suboptimal classification models that may be too over-fitted to highly specific conditions (e.g., things around the core) which are not the geological features of interest (things in the core).

The focus in the current paper is the preprocessing of drill core data with advanced ML methods, which will be the initial step towards a semi-automated ML framework for drill core analysis. The remainder of this paper is organized as follows. Section II presents the state of the art in drill core analysis. Section III presents the dataset and Section IV illustrates the various methods employed in this work and discuss about the results and findings of this work. Finally, the conclusion is in Section V.

II. STATE OF THE ART

Machine Learning is still on hype, and standard ML and Data Mining frameworks are already used in various industrial products and services, e.g., IBM Watson¹, Microsoft Azure², or the freely available Weka³. Also, RapidMiner⁴, Talend⁵, BayesianLab⁶ are frequently used AI toolkits.

The great potential of ML methods is reaching the field of mining geology and mineral exploration which is shown by current research in [1], [2], or in [3]. These papers have in common that they are using a novel approach to identify rock types with the help of deep learning, specifically with the use of Convolutional Neural Network (CNN). In [1], an overall classification accuracy of 97.96% in the context of rock type identification in the field is reported. A 100% accuracy of automatically classifying four groups of Martian rock is reported in [2], and a single-type rock image recognition out of nine different rock types in [3] reports more than 96% accuracy. Such approaches and results indicate that deep learning methods can bring significant improvements in mining geology. Also, there are industrial example cases related to mining by TOMRA⁷, and ML applications in commercially available software such as ioGAS⁸.

However, ML is based on data, and digitalization is still in an early stage when it comes to systematically implementing ML workflows in mining geology and exploration applications. A comprehensive study on digitalization trends in the mining industry revealed that ML is one of the technologies

of most interest -especially concerning automation- but the industrial processes are very complex, and it is challenging to integrate the technologies[4]. Many companies are re-designing their workflows and business models to allow for more rapid and continuous compositional and textural data collection using modern XRF, XRT, or hyperspectral sensor technologies. [5] shows an innovative procedure in creating a training dataset for supervised machine learning by fusing high-resolution mineralogical analysis and HS data, embedded in a Machine Learning Framework for Drill-Core Mineral Mapping. New sensor technologies are creating valuable data and opportunities for advanced machine learning approaches, yet are impeded by factors such as 1) the general complexity of geological materials, 2) a lower analytical precision and accuracy in many online methods compared to traditional methods for compositional analysis and 3) immaturity when it comes to selection of machine learning models that are most fit for purpose.

Additionally, there exist a vast amount of legacy data and drill core imagery collected for decades of mineral exploration, for which huge potential exist, but where advanced analysis is not possible without first solving problems related to preprocessing. With the help of ML methods, the industry brought solutions for this preprocessing step to the market in the last years^{9 10 11}. Only a few freely available preprocessing approaches are available like the CoreBreakout Python package[6]. With this package, images of core sample boxes can be converted into depth-registered datasets. This workflow includes a labeling step for the dataset to be processed so that the use of own datasets is possible to a certain extent. We follow a similar approach as CoreBreakout with a stronger focus on a human expert-ML workflow and extended to our dataset, despite the adaptation function to other datasets from CoreBreakout. Reasons for the own development are the need to cut out the boxes from their environment, recognize handwritten markings, or integrate automatic annotations.

Especially for the preparation of the legacy core imagery, there is a lack of free applications and accessible training data for supervised machine learning methods.

III. DATASET

The dataset used in this work consists of legacy drill-core imagery, produced from an active exploration drilling program in northern Sweden. These photos of drill-core have been made available for this project by the Swedish mining company Boliden. In the context of the Future Logging Assistant (FLA) project, the dataset will be extended to include many thousands of images from various exploration projects from Boliden, presenting different geological settings, and thus widening the scope of the training data. The first step of this study is presented here, and involves a small-scale, proof-of-concept study centered on a dataset from one single drill-core.

¹<https://www.ibm.com/watson/products-services>

²<https://azure.microsoft.com>

³<https://www.cs.waikato.ac.nz/ml/weka/>

⁴<https://rapidminer.com>

⁵<https://www.talend.com>

⁶<https://www.bayesia.com>

⁷<https://www.tomra.com/en/sorting/mining>

⁸<https://reflexnow.com/product/iogas/>

⁹<https://www.datarock.com.au/>

¹⁰<https://goldspot.ca/our-expertise/image-analytics/>

¹¹<https://www.imago.live/>

The selected drill core is systematically imaged in 180 photographs, from a starting at a depth of 12.35 meters below surface (the contact between overburden and bedrock) down to a depth of 852.70 meters (end of hole). The drill-core samples are stored in wooden boxes with a length of one meter per row, distributed over five rows in each box. The boxes were photographed in a wet state, whereby light reflections from lamps hanging above the box can be seen on the rock. Selected descriptions and depth information are handwritten on the wooden boxes and on wooden blocks. The wooden blocks were inserted between each drill run, when core was unloaded into the box, and the depth of hole registered by the drillers.

The dataset consists of high-resolution RGB JPG images, with 3648x2736 pixel dimensions, 500MB each. This size is a good fit for image resolution and the possibility of fast data processing, which is essential for integration into a human workflow. We could reach integration acceptable training times with NVIDIA's GeForce GTX 1080 of around 20 minutes.

The process of taking photos did not follow strict guidelines, and they were taken manually by different people and at different periods, which means that the photos all look different in detail.

The setup for taking photos consists of a storage area where the boxes can be lined up with an adjustable metal arm as a camera holder placed over the boxes. In the photos, the box is always aligned from top/left to bottom/right.

The region of interest (ROI) in the photographs is the drill-core and the handwritten depth meter indications. This region makes up 3400x1200 pixel dimensions on average and still is a high resolution for further image analysis. We assume that anything but the ROI in the photos will have a negative impact on further analysis of the samples, like showing in Figure 1 and listed in the following:

- surroundings
- other boxes in the surroundings
- slightly different perspectives of the boxes
- rotation of the boxes
- light reflections on the samples
- slightly different color and light conditions

The difficulty of reading out the depth meter indications are given by:

- handwritten, from different persons
- appears in different positions
- appears in different orientations

Besides the drill-core imagery, Boliden provides related rock-quality designation (RQD) and lithological analyses in the form of datasheets. The data were recorded manually in reference to the depth meter indications on the boxes, which is common practice. Nevertheless, a digital reference to the exact position/pixel in the photo is not given. Without a digital depth reference, the dataset is not suitable for supervised machine learning training since the dataset does not contain the content to be learned (GroundTruth). Without preprocessing of the dataset or different approaches, further

analysis is not possible. The dataset used is exemplary for most of the legacy drill-core imagery globally; therefore, preprocessing influences the context of the FLA also.

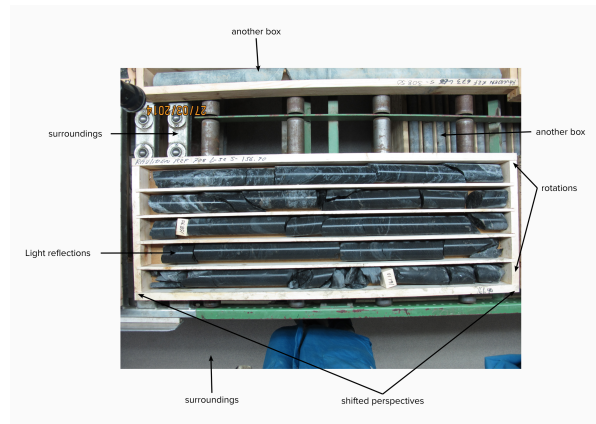


Fig. 1. Image quality issues

IV. METHODOLOGY AND RESULTS

Our approach in developing an ML framework for drill core analysis consists of integrating ML methods into the human workflow of drill core logging. Thereby we pursue the following goals:

- 1) *Digitization of the logging process based on the photos.* Data in the logging process, like the lithology or RQD, are recorded directly in the photograph by labeling. By placing labels in the photo, a direct digital reference between the data and the photo is created.
- 2) *Encapsulation of the depth reference problem.* Because logging process data is labeled in the photos, a depth reference is no longer necessary at this step. The non-trivial depth reference is then used to establish relations to other data sources or sets, such as data from a core-scanner. However, the problem can be dealt with independently in the respective context.
- 3) *Creation of training data.* Core logging experts are creating valuable training data through the labeling process, which is mandatory for supervised machine learning methods. The process can take place on-site at the samples or remotely.
- 4) *ML support for core logging and expert correction for ML models.* ML models can support experts by automated analysis by integrating ML models in the core logging processes. Incorrect automated analysis can be corrected by the experts and processed as training input in the model.

The most important property for implementing this approach is modularity; to reach the necessary flexibility and adaptability for various datasets and analyses. Each task is viewed independently and integrated as a module. At the beginning of the process, we are defining the following tasks:

- 1) Crop the ROI -drill core boxes- in the images.
- 2) Crop the core in the images (without box) and drill hole line up based on the output of 1.

- 3) Determination of the lithology based on the output of 2.
- 4) Determination of the RQD based on the output of 2.
- 5) Set the image depth reference based on the output of 1.

Figure 2 visualizes these tasks, including an ML-Expert workflow. Each task consists of a circle, where an expert labels an initial dataset entry used for initial training. After the initialization, experts can use and correct automated analysis. Corrections are used as new training input. The output of a circle can be used as input for another task to chain tasks.

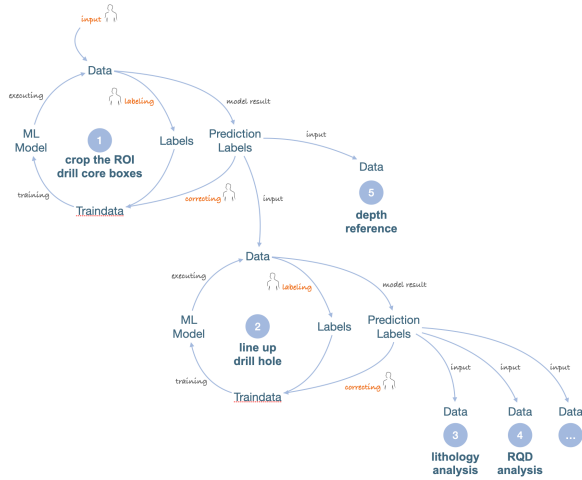


Fig. 2. Human Expert - ML Workflow

The weak link in this concept is the amount of work that the experts need for labeling. If the workload for the labeling is too high, this concept becomes impractical and thus infeasible. The amount of workload can be represented by the number of images that need to be labeled to reach acceptable precision in the predictions of the ML models. To evaluate the number of labeled images in relation to the ML model precision, we implemented and tested the first task of the framework; i.e., crop the ROI -drill core boxes- in the images. Cropping is a simple task that can also be carried out manually in less than a minute, but generally, considering the amount of several thousand images in drill core archives, even this simple task is worth digitizing.

We use various open source solutions for the technical implementation of our framework. Computer Vision Annotation Tool (CVAT)¹² is used for labeling, which provides a feature for automatic annotation/labeling. This feature is based on Nuclio¹³, a serverless platform through which we can deploy our trained models to CVAT. To train our models, we are using Detectron2¹⁴, which provides options to train on predefined models and adapt them to our core-logging use cases.

¹²<https://github.com/openvinotoolkit/cvat>

¹³<https://nuclio.io/>

¹⁴<https://github.com/facebookresearch/detectron2>

The experiment setup contains 180 images (see Section III Dataset), in which we manually labeled the area of the wooden drill core box. Labeling one box took around 30 seconds with CVAT. By that, the classes to be learned in our proof-of-concept are limited to two classes: drill core box and not drill core box. Each photo contains exactly one box to be found and are therefore evenly distributed over the photos. We trained different models to predict the drill core box based on these labels, in which different numbers of labeled images were available for training; defined as training configurations. In each training configuration, we run five iterations in which the data between training, validation, and test were randomly shuffled before every iteration and are mutually exclusive. Table I shows the training configurations:

TABLE I
TRAINING CONFIGURATIONS: TRAIN, VAL AND TEST SPLIT

Conf.	Train		Validation		Test	
	%	images	%	images	%	images
t80:	80%	144	10%	18	10%	18
t60:	60%	108	20%	36	20%	36
t40:	40%	72	30%	54	30%	54
t20:	20%	36	40%	72	40%	72
t10:	10%	18	45%	81	45%	81
t05:	05%	9	45%	81	50%	90
t00:	0%	0	0%	0	100%	180

We use the *mask R-CNN R 50 FPN 3x* from the Detectron2[7] Model Zoo and retrained it for 3000 epochs. For the evaluation -how precise the model can recognize the drill core box- we use the Detectron2 implementation of the COCOevaluator¹⁵. We are taking the AP values from the COCOevaluator as the evaluation measure, which are specified for the bounding box and the segmentation: AP at IoU=.50:.05:.95 (primary challenge metric). Figure 3 and 4 are showing exemplary a labeled image and a prediction of the bounding box and segmentation.



Fig. 3. Manually labeled drill core box used for training data

The tables II and III showing the AP results of the bounding box (bbox) and segmentation (segm) prediction

¹⁵<https://cocodataset.org/#detection-eval>



Fig. 4. Prediction of the bounding box (AP: 89.99%) and segmentation (AP: 89.99%)

on the test dataset. The iterations (iter.) for each training configuration are ordered from best to lowest AP value to increase readability. A box plot visualization of the tables is added in Figure 5 and 6.

TABLE II
AP BOUNDING BOX (BBOX) RESULTS FOR THE TRAINING CONFIGURATIONS ON THE TEST DATASETS

	t00 bbox AP	t05 bbox AP	t10 bbox AP	t20 bbox AP	t40 bbox AP	t60 bbox AP	t80 bbox AP
iter. 1	0.00	79.46	79.36	81.20	80.77	84.91	85.39
iter. 2		79.20	75.53	81.05	80.10	81.42	78.81
iter. 3		77.55	75.13	77.99	79.35	76.62	71.86
iter. 4		76.48	71.75	75.99	73.72	76.53	71.07
iter. 5		74.62	67.45	66.96	67.54	73.20	71.05
std		2.00	4.48	5.83	5.64	4.61	6.35
mean		77.46	73.85	76.64	76.30	78.54	75.64

In general, all values show a high level of precision, ranging from 70% to 87% for segmentation and 71% to 85% for the bounding box on the test set. These results are sufficient to crop out the core box, as most of the surrounding area is cut away, and the drill core remains completely visible. Interestingly, the number of labeled images is not significantly decisive for higher precision. For example, the average of all iterations for the training configuration *t05 segm AP* (80.4%) is higher than the average for *t80 segm AP* (79.7%). Recognition of the box without training on the base model (*t00* on *mask R-CNN R 50 FPN 3x*) did not work and returned an AP of zero. All training configurations have a relatively high standard deviation (std). It is lowest on the *t05* training configuration, but there is no clear trend that std increases with the number of labeled images in training. The fact that the test precision in the randomly shuffled iterations fluctuates highly - as with the configuration *80 segm AP* from 85.7% - 75.7%, it is reasonable to assume that a unique selection of training data is the decisive factor for high precision. This would mean that a small amount of

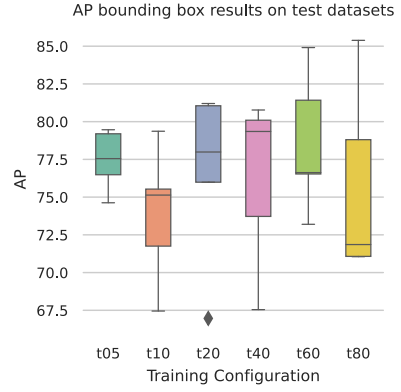


Fig. 5. AP bounding box (bbox) results for the training configurations on the test datasets

TABLE III
AP SEGMENTATION (SEGM) RESULTS FOR THE TRAINING CONFIGURATIONS ON THE TEST DATASETS

	t00 segm AP	t05 segm AP	t10 segm AP	t20 segm AP	t40 segm AP	t60 segm AP	t80 segm AP
iter. 1	0.00	83.70	83.29	84.30	84.74	87.90	85.65
iter. 2		81.83	81.16	83.58	83.57	85.62	83.41
iter. 3		79.56	79.33	82.10	80.88	80.83	76.96
iter. 4		78.80	75.73	78.99	80.15	80.35	76.91
iter. 5		78.41	71.90	68.68	70.94	79.73	75.72
std		2.24	4.52	6.40	5.43	3.65	4.48
mean		80.46	78.28	79.53	80.05	82.89	79.73

well-selected labeled data which covers the variations and unique features in the images is sufficient to achieve a high level of precision. However, this assumption needs further investigation.

V. CONCLUSION AND FUTURE WORK

A framework for the drill core analysis based on a labeling/training workflow between human experts and ML models can introduce the required digitalization in the still manual dominant core logging process. This digitalization enables the creation of high-quality training datasets for ML models and various other analyzes in the direction of lithology or RQD determination. Whether such a framework can be beneficial depends on the required amount of labeled images related to the precision of ML model predictions. The evaluation of the relation between the number of labeled images and the models' AP showed that a minimal number of labeled data is sufficient to achieve a high level performance on the test data. The assumption that, instead of the number of labeled images, the selection of the labeled variants in the data is decisive for a high level of precision has to be further evaluated. However, since only a few images need to be labeled for relatively high performance and the time required for labeling is manageable (30 seconds per box) as well as short retraining times around 20 minutes for a drill

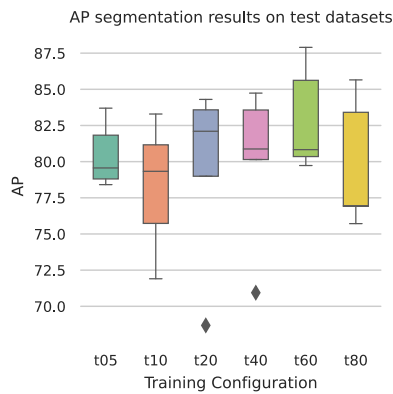


Fig. 6. AP segmentation (segm) results for the training configurations on the test datasets

core are given, such a framework can be a feasible solution to digitize the core logging process.

Since we implemented a relatively simple task for the evaluation, the framework must also be evaluated on more complex tasks, such as the lithological analysis. Furthermore, a test is needed to determine whether such a framework can support human experts and which level of the ML model's performance is needed for it. For this, the framework has to be integrated and tested in a natural working environment. This is planned as a next step in the context of the ML4DrillCore and FLA project and can thus form the required basis for advanced analyses on drill core. Finally, we intend to make our framework open source and available for the research community in the future.

ACKNOWLEDGMENT

The current work is funded by the LTU-SUN seed project ML4DrillCore¹⁶ and by the LTU-Boliden collaborative project FLA. We would like to thank the Swedish mining company Boliden for the valuable discussions and for providing the dataset used in this work. Especially we would like to thank Paul McDonnell and Tobias Hermansson from Boliden for their magnificent support and collaboration to bring drill core logging with ML into a vision of a Future Logging Assistant.

REFERENCES

- [1] X. Ran, L. Xue, Y. Zhang, Z. Liu, X. Sang, and J. He, "Rock classification from field image patches analyzed using a deep convolutional neural network," *Mathematics*, vol. 7, no. 8, p. 755, 2019.
- [2] J. Li, L. Zhang, Z. Wu, Z. Ling, X. Cao, K. Guo, and F. Yan, "Autonomous martian rock image classification based on transfer deep learning methods," *Earth Science Informatics*, pp. 1–13, 2020.
- [3] X. Liu, H. Wang, H. Jing, A. Shao, and L. Wang, "Research on intelligent identification of rock types based on faster r-cnn method," *IEEE Access*, vol. 8, pp. 21 804–21 812, 2020.
- [4] L. Barnewold and B. G. Lottermoser, "Identification of digital technologies and digitalisation trends in the mining industry," *International Journal of Mining Science and Technology*, vol. 30, no. 6, pp. 747–757, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2095268620300744>
- [5] I. C. C. Acosta, M. Khodadadzadeh, L. Tusa, P. Ghamisi, and R. Gloaguen, "A machine learning framework for drill-core mineral mapping using hyperspectral and high-resolution mineralogical data fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 4829–4842, 2019.
- [6] R. G. Meyer, T. P. Martin, and Z. R. Jobe, "Corebreakout: Subsurface core images to depth-registered datasets," *Journal of Open Source Software*, vol. 5, no. 50, p. 1969, 2020.
- [7] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, "Detectron2," <https://github.com/facebookresearch/detectron2>, 2019.

¹⁶<https://www.ltu.se/research/Framtidsomraden/SUN/Projekt/Seed-projekt/ML4DrillCore-1.205897?l=en>

Class-Incremental Learning for Semantic Segmentation - A study

Karl Holmquist¹, Lena Klasén^{1,2} and Michael Felsberg^{1,3}

Abstract—One of the main challenges of applying deep learning for robotics is the difficulty of efficiently adapting to new tasks while still maintaining the same performance on previous tasks. The problem of incrementally learning new tasks commonly struggles with catastrophic forgetting in which the previous knowledge is lost.

Class-incremental learning for semantic segmentation, addresses this problem in which we want to learn new semantic classes without having access to labeled data for previously learned classes. This is a problem in industry, where few pre-trained models and open datasets matches exactly the requisites. In these cases it is both expensive and labour intensive to collect an entirely new fully-labeled dataset. Instead, collecting a smaller dataset and only labeling the new classes is much more efficient in terms of data collection.

In this paper we present the class-incremental learning problem for semantic segmentation, we discuss related work in terms of the more thoroughly studied classification task and experimentally validate the current state-of-the-art for semantic segmentation. This lays the foundation as we discuss some of the problems that still needs to be investigated and improved upon in order to reach a new state-of-the-art for class-incremental semantic segmentation.

I. INTRODUCTION

Incremental Learning, sometimes referred to as continual learning or life-long learning, is the task of expanding the knowledge of a system to encompass new concepts. It has both been formulated as a continuous and a discrete task in which new experiences are collected and learned from continuously with little to no memory [1]. Class-Incremental Learning (CIL) is defined as a sequential task where the learning is done incrementally in discrete batches representing separate tasks. The task being disjoint set of semantic classes that all should be learned at the end of the last batch of classes. The most common formulation uses very limited amounts to no data from previous classes and primarily addresses the problem of catastrophic forgetting [2], [3], [4], [5]. One alternative formulation of CIL is the Generalized CIL (GCIL) [6] framework that attempts to model a more realistic scenario in which there is little control of which data is collected and annotated. They do this by allowing classes

*This work was supported by Vinnova project 2020-02838 Model Agnostic Meta Learning (MAML) for 3D Forestry Artificial Intelligence

¹All authors are with the Computer Vision Group at the Department of Electrical Engineering, Linköping University, Sweden

²L. Klasén is also with the Office of the National Police Commissioner, The Swedish Police Authority

³M. Felsberg is also with the University of KwaZulu-Natal, School of Engineering, Durban 4000, South Africa

to appear in multiple increments and allowing for a larger class imbalance.

These frameworks are focusing primarily on the classification task in which the entire image is categorized as a single class. Instead, we are investigating semantic segmentation which introduces a couple of different challenges and possibilities. The pixel-wise classification done in semantic segmentation is a much more information intensive problem as multiple classes exists in each image and as such, it normally has a much larger class imbalance than in classification.

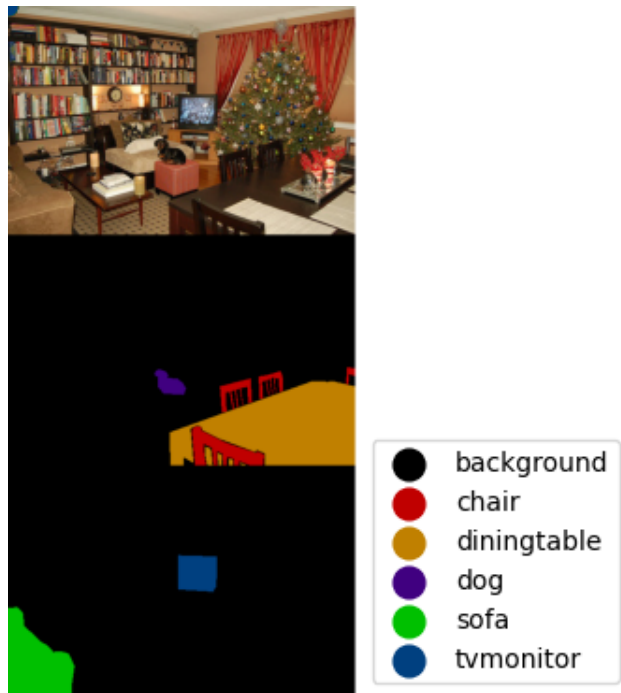


Fig. 1: Illustration of how the tasks could be separated into disjoint sets. Top: RGB image, Middle: First increment showing chair, diningtable and dog, Bottom: Second increment showing the sofa and tvmonitor classes. Background is always present and note that classes not belonging to the current set is labeled as background.

A. Problem formulation

Given a set of classes $\mathcal{C} := \{c_0, \dots, c_k\}$ separated into M disjoint sets D^t for $t \in \{1, \dots, M\}$, the task is to learn the full set of classes \mathcal{C} . The classes are learned incrementally by only providing one of the disjoint class sets at a time and

only once. However, each set of classes could potentially contain unlabeled pixels from one of the other sets labeled as background, see figure 1.

In the rest of the paper we will refer to the classes up to (and including) time step t by $\mathcal{D}^{0:t} := \bigcup_{k=0:t} \mathcal{D}^k$

II. RELATED WORK

In this section we will present some of the main approaches used for class incremental learning, primarily in terms of classification since this has been the more well-studied problem so far.

A. Weight regularization methods

The first type of methods for incremental learning attempts to minimize the deterioration by calculating the importance of the individual weight parameters for the old classes and primarily updating the less important ones. Elastic Weight Consolidation (EWC) [7] calculates the importance offline using a Fisher information matrix, while Memory Aware Synapses (MAS) [8] observes the gradients with respect to the output given a small noise in the parameter space.

Another related set of approaches concerns itself with preserving the topology of the feature space [9], [10] or correcting for the semantic drift [11].

These methods aims at avoiding catastrophic forgetting by preserving the weights and/or topology for the old tasks. However, this might limit the capabilities of the network to learn new classes.

B. Memory based methods

Orthogonal to the weight regularization methods the memory based methods utilizes exemplars of previous classes in order to avoid forgetting. One of the first methods to use exemplars for training a deep neural network for the class incremental classification task was Incremental Classifier and Representation Learning (iCaRL) [2] which used a distillation loss together with a limited memory bank and uses the *nearest-mean-of-exemplars* rule for later classifying new samples. ReMIX [6] combines the idea of exemplar replay (ER) with Mixup [12] to learn more robust representations by augmentation of the training data.

The drawback with an explicit memory is that either the memory size needs to increase or the number of samples per class will decrease as more classes are added to the model. This restriction also commonly leads to a class imbalance between the exemplars of old classes and the samples from the new classes. These problems has been addressed by using class-conditioned generative models [13] at the expense of increasing the network size.

C. Distillation based methods

Both of the above methods are also commonly combined with distillation, using the model from the previous step to teach the new model. These methods are based on the Learning-without-forgetting idea (LwF) [4] in which a cross-entropy term is used to learn the new classes while a distillation loss is used to maintain the knowledge from the

previous increment. Knowledge distillation has previously been used for training a smaller more efficient model from large ensembles by calculating the cross-entropy between the predictive distributions of the teacher model(s) and the student model [14]. Distillation has also been used to merge the information of two models trained on different increments [15].

There have also been studies regarding the role that the distillation loss has for incremental learning and how it maintains the discriminativity between old classes [16].

D. Class Incremental Semantic Segmentation

While the classification problem has been studied in depth in terms of CIL, semantic segmentation has been mostly left unattended. The challenges present for semantic segmentation differs largely from the classification problem since there is often a much larger class imbalance in the dataset. There is also a special class, *background*, which needs to be taken into consideration since it encompasses all unlabeled classes, some of which will be learned in later increments.

In this work we train and evaluate the method proposed by Cermelli et.al [17] and highlight some of the problems that still exists that need to be addressed to further improve on the state-of-the-art for class-incremental semantic segmentation.

III. METHOD

This section will describe the current state of the art method, Modeling the Background (MiB), proposed in [17] starting with the loss functions used and continuing with the initialization of new classification layers.

A. Loss functions

The standard cross-entropy loss is one of the standard losses used for classification tasks and is commonly used for learning new classes in the class incremental setting as well. It is formulated as follows

$$\mathcal{L}_{CE}(\mathbf{x}, y) = - \sum_{k \in \mathcal{D}^t} \delta_{k=y} \log(p_k(\mathbf{x})), \quad (1)$$

in which $\delta_{k=y}$ is a binary indicator function, taking the value 1 when the predicted label k is equal to the ground-truth label y .

Similarly, the knowledge of previous classes is maintained by a knowledge distillation loss between the teacher trained in the previous increment and the student that is currently trained. It is formulated as

$$\mathcal{L}_{KD}(\mathbf{x}) = - \sum_{k \in \mathcal{D}^{0:t-1}} q_k(\mathbf{x}) \log(p_k(\mathbf{x})), \quad (2)$$

where $p_k(\mathbf{x})$ and $q_k(\mathbf{x})$ is the probabilities for the currently training model and the model trained in the last increment respectively.

Both of these losses compares two probability distributions and as such, we normalize the predicted logits from the models using softmax.

$$p(\mathbf{x}) = \text{Softmax}(z(\mathbf{x})) \quad (3)$$

However, in contrast to classification, semantic segmentation also has a background class which contains unseen classes. In order to utilize this information, MiB was proposed [17] in which the assumption that new classes would have been classified as background previously and previous classes should be part of the background in later increments.

This assumption is made based on that the environment is relatively static while primarily the classes that was labeled is changing. However, they also show that even without having samples with both new and old labels together it improves the performance over not modelling the background.

The proposed method constitutes two modifications. First, instead of simply masking the new classes from the KD-loss and the old classes from the CE-loss they accumulate the logits for the unused classes as background.

$$z_{bkg}^{CE}(\mathbf{x}) = \sum_{k \in \mathcal{D}^{0:t-1}} z_k(\mathbf{x}) \quad (4)$$

$$z_{bkg}^{KD}(\mathbf{x}) = \sum_{k \in \mathcal{D}^t} z_k(\mathbf{x}) \quad (5)$$

Secondly, the last classification layer for the new classes at increment t is initialized based on the background class from increment $t - 1$.

$$w_c^t = \begin{cases} w_{bkg}^{t-1}, & \text{if } c \in \mathcal{D}^t \\ w_c^{t-1}, & \text{otherwise} \end{cases} \quad (6)$$

$$\beta_c^t = \begin{cases} \beta_{bkg}^{t-1} - \log(|\mathcal{D}^t|), & \text{if } c \in \mathcal{D}^t \\ \beta_c^{t-1}, & \text{otherwise} \end{cases} \quad (7)$$

B. Experimental setup

Our evaluation methodology follows the one proposed in [17]. We evaluate on the Pascal VOC dataset [18] which contains 20 classes plus background and focus on the overlapped setting in which all samples containing one or more of the new classes is used. The classes that does not belong to the current increment \mathcal{D}^t is relabeled as background for training and are ignored during validation on the new classes.

The upper performance limit (UPL) is represented by the same underlying model structure (DeepLab-v3 [19]) trained simultaneously on all classes.

Cermelli et.al[17] proposed multiple tasks varying by number of classes per increment and number of increments. Some of the tasks are described in table I.

Task	Classes	#Increments
15-5	{1-15}, {16-20}	2
15-5s	{1-15}, {16}, {17}, {18}, {19}, {20}	6
10-5-5	{1-10}, {11-15}, {16-20}	3

TABLE I: Description of tasks

The models was trained without early stopping for 30 epochs using SGD with nesterov momentum. The learning rate was set to $1e - 2$ for the first step and $1e - 3$ for subsequent increments. The momentum was set to 0.9, a weight decay of $1e - 4$ was applied to all parameters and a

batch-size of 24 was used for the training. The backbone is a resnet101 trained with In-Place Activated BatchNorm [20]. All done the same way as [17].

IV. RESULTS

In this section we present some of our reproduced results from MiB [17]. We also show some of the failure cases that exists that requires further attention in future research.

A. Quantitative results

By training the model on the 15-5 task we evaluated the model at three steps. First, the model trained on the first increment (classes 1-15) and evaluate this model on the subset of samples which contain at least one of the active classes. Second, we continue to train the model on the second increment and report the validation results on the second subset which contain one or more classes from the second increment. Finally, we evaluate the model which has been trained on both increments on all classes using the entire validation dataset.

The results of the incremental learning was compared to the UPL, trained on all classes simultaneously and some other common methods used for classification, see table II.

Model	1-15 (mIoU)	16-20 (mIoU)	1-20
MiB (1-15)	0.78	-	-
MiB (16-20)	-	0.87	-
MiB	0.71	0.44	0.64
Finetuning ¹	-	-	0.10
EWC ¹	-	-	0.27
LwF ¹	-	-	0.53
LwF-MC ¹	-	-	0.52
UPL	0.78	0.72	0.77

TABLE II: Result table for the 15-5 task (¹ Result as reported in [17])

Figure 2 shows the confusion matrices of the offline trained model and the sequentially trained model. As can be seen there is a trend that some of the old classes are confused with one of the new ones and to a lesser extent the opposite. However, the largest confusion seems to be between background and foreground.

B. Problems

Based on the results presented in table II it seems that directly applying some of the methods commonly used for class-incremental classification problems is not sufficient, instead we need to handle the background as a special entity.

While the background needs to be handled differently from the other classes, we also realize that the assumption that all new classes should be similar to the background might be a too strong assumption if the new classes haven't been commonly present in the previous increment. This could be a contributing factor as to why similar classes such as 'sheep' and 'cow' is commonly confused as in figure 3. Instead of mainly adjusting the last classification layer the error is also back-propagated to earlier layers causing unnecessary large changes in the network that possibly decreases the performance of the old classes.

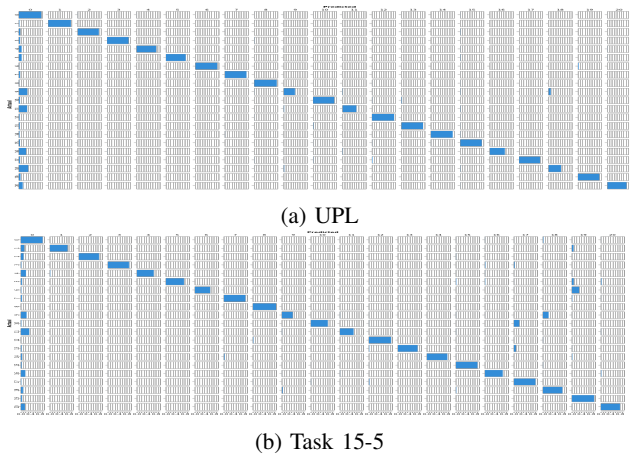


Fig. 2: Class confusion matrices for the UPL and the model trained on task 15-5

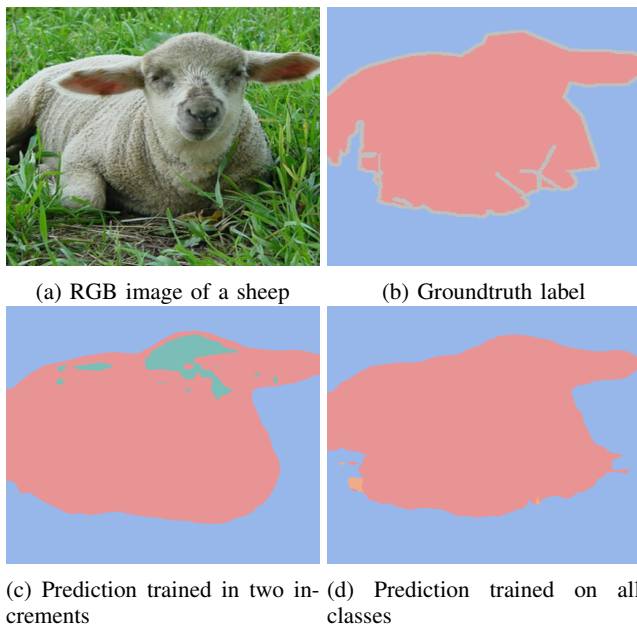


Fig. 3: Example images illustrating confusion caused between old and new classes. The model trained in two increments miss-classifies part of the new class (sheep) as an old class (cow).

V. CONCLUSIONS

In this paper, we have provided a problem description of class incremental learning for semantic segmentation, experimentally validated the current state-of-the-art and illustrated both the need for a different approach than previously used for classification as well as discussed some of the problems that still needs addressing in current state-of-the-art. Future work will try to address these issues to further improve the performance for class-incremental semantic segmentation.

REFERENCES

[1] R. Aljundi, K. Kelchtermans, and T. Tuytelaars, “Task-free continual learning,” in *Proceedings of the IEEE/CVF Conference on Computer*

Vision and Pattern Recognition, pp. 11254–11263, 2019.

[2] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, “icarl: Incremental classifier and representation learning,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 2001–2010, 2017.

[3] G. M. Van de Ven and A. S. Tolias, “Three scenarios for continual learning,” *arXiv preprint arXiv:1904.07734*, 2019.

[4] Z. Li and D. Hoiem, “Learning without forgetting,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2935–2947, 2017.

[5] Y. Liu, Y. Su, A.-A. Liu, B. Schiele, and Q. Sun, “Mnemonics training: Multi-class incremental learning without forgetting,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12245–12254, 2020.

[6] F. Mi, L. Kong, T. Lin, K. Yu, and B. Faltings, “Generalized class incremental learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[7] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, et al., “Overcoming catastrophic forgetting in neural networks,” *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[8] R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, and T. Tuytelaars, “Memory aware synapses: Learning what (not) to forget,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 139–154, 2018.

[9] X. Tao, X. Chang, X. Hong, X. Wei, and Y. Gong, “Topology-preserving class-incremental learning,” in *European Conference on Computer Vision*, pp. 254–270, Springer, 2020.

[10] X. Tao, X. Hong, X. Chang, S. Dong, X. Wei, and Y. Gong, “Few-shot class-incremental learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12183–12192, 2020.

[11] L. Yu, B. Twardowski, X. Liu, L. Herranz, K. Wang, Y. Cheng, S. Jui, and J. v. d. Weijer, “Semantic drift compensation for class-incremental learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[12] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “mixup: Beyond empirical risk minimization,” *arXiv preprint arXiv:1710.09412*, 2017.

[13] X. Liu, C. Wu, M. Menta, L. Herranz, B. Raducanu, A. D. Bagdanov, S. Jui, and J. v. de Weijer, “Generative feature replay for class-incremental learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 226–227, 2020.

[14] G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” *arXiv preprint arXiv:1503.02531*, 2015.

[15] J. Zhang, J. Zhang, S. Ghosh, D. Li, S. Tasci, L. Heck, H. Zhang, and C.-C. J. Kuo, “Class-incremental learning via deep model consolidation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, March 2020.

[16] B. Zhao, X. Xiao, G. Gan, B. Zhang, and S.-T. Xia, “Maintaining discrimination and fairness in class incremental learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[17] F. Cermelli, M. Mancini, S. R. Bulò, E. Ricci, and B. Caputo, “Modeling the background for incremental learning in semantic segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[18] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results.” <http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html>.

[19] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017.

[20] S. Rota Bulò, L. Porzi, and P. Kotschieder, “In-place activated batchnorm for memory-optimized training of dnns,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

Identifying cheating behaviour with machine learning

Elina Kock¹, Yamma Sarwari¹, Nancy Russo¹ and Magnus Johnsson²

Abstract—We have investigated machine learning based cheating behaviour detection in physical activity-based smartphone games. Sensor data were acquired from the accelerometer/gyroscope of an iPhone 7 during activities such as jumping, squatting, stomping, and their cheating counterparts. Selected attributes providing the most information gain were used together with a sequential model yielding promising results in detecting fake activities. Even better results were achieved by employing a random forest classifier. The results suggest that machine learning is a strong candidate for detecting cheating behaviours in physical activity-based smartphone games.

I. INTRODUCTION

Smartphones are convenient tools that are extensively used for everyday activities, entertainment, and more. The sophisticated sensors found in these devices have been found to be extremely valuable for several applications, human activity recognition (HAR) being one of them.

HAR using machine learning has been widely studied and several combinations of sensors have been used. Examples of sensor combinations are: accelerometer only [1, 2]; accelerometer and gyroscope [3, 4]; accelerometer, gyroscope, and magnetometer [5, 6]; tri-axial accelerometer, linear accelerometer, gyroscope, and orientation sensors [7]; accelerometer, gyroscope, visible light, barometer, magnetometer [8]. Cameras can also be used and one of the authors has previously led the development of a hierarchical neural network architecture that uses a hierarchy of self-organizing maps to recognize actions [9, 10, 11]. However, a majority of previously conducted HAR research has utilized the accelerometer in combination with the gyroscope.

Something that goes hand-in-hand with human activity recognition is physical exercise. Humans are less physically active than before. It has been shown that gamification can help children and young adults to be more active by providing motivation for physical activity [12], and several physical activity-based games have been developed. Pokémon Go is a well-known example of how gamification motivated millions of people to be physically active.

However, the possibility to cheat in physical activity-based games demotivates play as intended [13] and it is fairly common both in games and non-gamified physical activity-based applications [13, 14, 15]. Cheating in a physical activity-based game can come in a variety of forms. A

common way to cheat is to shake the phone in order to fool the game that the player is moving [13].

There are previous works using various kinds of machine learning methods for HAR, e.g.: k-nearest-neighbours (KNN) [16], support vector machines (SVM) [3, 17], and naïve Bayes [3]; and neural networks such as multilayer perceptron (MLP) networks [1, 18, 19, 3, 8], recurrent neural networks (RNN) [17, 20, 21, 22], deep neural networks (DNN) [16], convolutional neural networks (CNN) [5], and hierarchical self-organizing maps (SOM) [9, 10, 11].

Most existing works on HAR focus on everyday activities such as walking, standing, sitting, running, going upstairs or downstairs, etc. [1, 2, 4, 23]. As many studies investigate these same activities, most of them are able to focus on improving the accuracy and perfecting the models that already exist. Very few works, e.g. [8, 24], aim to investigate how more vigorous activities such as jumping or squatting can be recognized, and how machine learning can be used for cheat detection in physical activity-based games for smartphones.

This paper investigates how machine learning can be utilized to detect whether a player is actually performing the intended physical activity of a physical activity-based smartphone application, or if the player is simply pretending to perform the activity. The activities investigated are jumping, squatting, and stomping along with their fake counterparts fake jumping, fake squatting, and fake stomping.

II. EXPERIMENTS

We have used two different models two categorize the activities and fake activities: an MLP network, and a random forest (RF) classifier.

Data was collected from the accelerometer and the gyroscope of an iPhone 7 by letting 12 test subjects (heights 160-197 cm, weights 60-116 kg) perform the activities: jumping, squatting, stomping, fake jumping, fake squatting, fake stomping while holding the iPhone. The fake activities are activities where the test subject perform physical motions or manipulating the phone in their own way to attempt to simulate the activity in an attempt to cheat. 15 data entries of each of the six activities were obtained, which means that some test subjects performed the activities repeatedly.

The features provided by the accelerometer and gyroscope were userAcceleration (accelerometer), gravity (accelerometer), rotationRate (gyroscope), and attitude (gyroscope). Each of these objects contain three values. Three of them, userAcceleration, gravity, and rotationRate, contain X, Y and Z values, whereas attitude contains pitch, roll, and yaw. This means that there were 12 sensor features in total, 6 from each sensor, see Table 1.

¹E. Kock, Y. Sarwari and N. Russo is with the Faculty of Computer Science and Media Technology, Malmö University, Sweden mau.se

²M. Johnsson is with the Faculty of Computer Science and Media Technology, Malmö University, Sweden and with Magnus Johnsson AI Research AB, Sweden magnusjohnsson.se

TABLE I

OVERVIEW OF THE DIFFERENT SENSORS AND THE ABBREVIATION FOR EACH VALUE RETRIEVED.

Sensor Data	Sensor	Abbreviation
userAcceleration x	Accelerometer	accx
userAcceleration y	Accelerometer	accy
userAcceleration z	Accelerometer	accz
gravity x	Accelerometer	gravx
gravity y	Accelerometer	gravy
gravity z	Accelerometer	gravz
rotationRate x	Gyroscope	rotationx
rotationRate y	Gyroscope	rotationy
rotationRate z	Gyroscope	rotationz
attitude pitch	Gyroscope	pitch
attitude roll	Gyroscope	roll
attitude yaw	Gyroscope	yaw

Feature selection was performed by determining which attributes provided the most information gain. These were: yaw, accx, rotationz, rotationy, gravz, roll, rotationx, accy, accz, pitch, gravi, gravx. Attribute selection was subsequently used to determine the subset that would yield the best results, namely: yaw, accx, rotationz, rotationy, gravz, rotationx, accy, accz, pitch. The selected features were normalized to a range between 0 and 1.

Multilevel Perceptron Network. We used an MLP network with nine hidden layers with 100 neurons each. The input layer takes an 200x9 element vector as input, corresponding to a sequence window of 200 consecutive inputs, which in turn corresponds to around four seconds. The output layer is consists of six neurons (one for each possible output, jumping, fake jumping, stomping, fake stomping, squatting, fake squatting).

The model outputs its predictions as a vector with six floating point numbers that represent the percentage of certainty the model has for each of the six activities.

The model was trained by feeding it with input representing 200 consecutive sensor readings at a time with a 100 step distance, meaning that the first training sample was readings 1-200, the second training sample was readings 100-300, the third was 200-400 etc. That means there is a 50% overlap in the data between the samples. This was chosen to make as much use of the data as possible without making training samples too similar to each other, and to make sure that all data had a chance to occur at various positions in a training sample.

The recorded data were split into a training and a test set where 80% of the data were used for training and 20% for testing. The model was trained using 12 epochs.

The accuracy of the trained MLP network was 83.3% on the validation set and 66% on the test set. Out of 84 instances, 56 were classified correctly and 28 were classified incorrectly. This means that a total of 34% of the activities from the test set were classified incorrectly. Table 2 shows the confusion matrix for the MLP network model.

Some activities are classified with high accuracy, while others are classified poorly. As can be seen in Table 3,

TABLE II

CONFUSION MATRIX FOR THE MLP NETWORK MODEL.

	F. Ju	F. Sq	F. St	Ju	Sq	St
F. Ju	9	0	4	0	0	0
F. Sq	0	13	1	0	0	0
F. St	0	0	14	0	0	0
Ju	0	1	3	3	1	6
Sq	0	9	0	0	4	1
St	2	0	0	0	0	11

fake activities and stomping were classified well, whereas jumping and squatting were classified poorly. Fake stomping was classified with 100% accuracy and fake squatting was classified with 92.9% accuracy. These two activities had the highest recognition rate. On the contrary, the two activities with the lowest accuracies were jumping and squatting. Jumping and squatting were classified with an accuracy of 21.4% and 28.6% respectively.

TABLE III

TABLE SHOWING THE CLASSIFICATION ACCURACY OF EACH ACTIVITY WITH THE MLP NETWORK MODEL.

Activity	Classification Accuracy
Fake jumping	69.2%
Fake squatting	92.9%
Fake stomping	100%
Jumping	21.4%
Squatting	28.6%
Stomping	78.6%

Random Forrest Classifier. In addition to the MLP network, we tested an RF classifier on the recorded dataset using 10-fold cross validation. The RF classifier classified 90.5% of the instances correctly and 9.5% incorrectly. As can be seen in Table 4, the activities with the highest classification accuracy with the RF classifier were squatting with 95.8% and fake stomping with 94.4%.

TABLE IV

TABLE SHOWING THE CLASSIFICATION ACCURACY OF EACH ACTIVITY WITH THE RANDOM FORREST CLASSIFIER.

Activity	Classification Accuracy
Fake jumping	92.0%
Fake squatting	89.1%
Fake stomping	94.4%
Jumping	83.8%
Squatting	95.8%
Stomping	88.1%

III. DISCUSSION

The results show that the MLP network model accurately classify fake squatting, fake stomping, and stomping. Fake jumping is classified reasonably well, but jumping and squatting were classified with low accuracy.

The underlying reason for the better performance on fake activities is unknown. It could be that the test subjects tended

to perform such activities in a more stereotypical manner. That would lead to a smaller within category variance.

The MLP network model performed particularly bad at jumping and squatting, with classification accuracies of only 21.4% and 28.6% respectively. This brought the overall classification accuracy by almost 20 percentage points. 36.6% of the authentic activities were misclassified as fake activities.

Around 36% of the predicted fake stomplings, 43.5% of the predicted fake squattings and 38.9% of the predicted stomplings were false positives.

The most confused activity is squatting. This activity was misclassified as fake squatting in 64.3% of the cases. A possible reason could be that this activity has been particularly easy to fake. 30.8% of the fake jumpings were misclassified as fake stomping. Jumping was misclassified as stomping in 42.9% of the cases, whereas stomping was misclassified as fake jumping in 18.2% of the cases. Fake stomping was never misclassified.

When it comes to the RF model it is important to stress that because it was never tested on a separate test set the accuracy numbers must be taken lightly. However, it seems that the RF model had a much higher accuracy than the MLP network model.

The activities which were best predicted with the RF model were squatting, fake stomping, and fake jumping. This was not the case for the MLP network model, which predicted fake stomping, fake squatting, and stomping best.

One similarity between the results of the two models is that jumping was most often confused with stomping in both cases. However, the RF model most often confused stomping firstly with squatting, and secondly with jumping, whereas the MLP network model most often confused stomping with fake jumping. The RF model, confused jumping and stomping the most. This coincide with the results from the MLP network model.

Interestingly, squatting was often misclassified as fake squatting by the MLP network model, whereas the RF model almost never did that (only in 0.6% of the cases).

IV. CONCLUSIONS

This paper has presented and discussed the results of two machine learning models employed for human activity recognition (HAR). The machine learning models classified sensor data from a smartphone obtained while a number of test subject carried out the exercise activities stomping, jumping, and squatting, and their fake counterparts fake stomping, fake jumping, and fake squatting. The sensor data was obtained through readings of the accelerometer and the gyroscope of an iPhone 7.

We used a multilevel perceptron (MLP) network and obtained a classification accuracy of 66%. We also used a random forest (RF) model and obtained an accuracy of 90.5% in validation.

The result from the MLP network model could be further improved in many ways, whereof one would be by increasing the amount of data in our dataset to more thoroughly train and fine tune the model. The results of our work contribute

to the study of HAR by introducing three activities that have previously not been thoroughly investigated, namely stomping, jumping, squatting, and in particular their faked counterparts.

Our work also introduces a position previously unexplored within the field of HAR, in which the user holds the smartphone during an activity. Our work can be used as a foundation for further studies in this field.

Potential future work could be to obtain more data with larger variation. For example, data obtained from a wider a more varied sample of the human population. The dataset should capture a wide range of ages, physiques, including e.g. tall and thin, short and thin, tall and average, short and heavy, etc.

Several previous studies attempting to classify human activities, such as sitting, walking, going upstairs, jumping, have had success utilizing statistical features in their datasets [7, 8, 24, 25]. Therefore, these kinds of features should also be utilized with our dataset.

Future work could also be done to improve the understanding of which classifier best classifies various kinds of activity. This would mean that various classifiers had to be tested and evaluated for each activity to be able to choose the optimal one. The RF classifier must also be more thoroughly evaluated. Classifiers can be tested either on the entire dataset or on individual activities, but the best understanding would likely come from testing them on individual activities.

ACKNOWLEDGMENT

We acknowledge the research centre Internet of Things and People (IoTaP) at Malmö university that supported this work through the KK Foundation founded project Motion Action Recognition through Sensors (MARS). We also acknowledge the Cybercom Group for providing us with the physical activity based smartphone game Bamblup to work with, and in particular their former employee Dennis Zikovic, who provided us with feedback and ideas. This paper is based on the bachelor thesis of Elina Kock and Yamma Sarwari [26].

REFERENCES

- [1] J. Wannenburg, R. Malekian, "Physical Activity Recognition From Smartphone Accelerometer Data for User Context Awareness Sensing", *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (Volume: 47 , Issue: 12 , Dec. 2017), 2016, pp. 3142 - 3149.
- [2] W. Hung, F. Shen, Y. Wu, M. Hor and C. Tang, "Activity Recognition with sensors on mobile devices," 2014 International Conference on Machine Learning and Cybernetics, Lanzhou , 2014, pp. 449-454.
- [3] X. Yin, W. Shen, J. Samarabandu, X. Wang, "Activity Detection and Analysis Using Smartphone Sensors", 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD) , 2015.
- [4] E. Bulbul, A. Cetin and I. A. Dogru, "Human Activity Recognition Using Smartphones", 2018 2nd International Symposium on Multi-disciplinary Studies and Innovative Technologies (ISMSIT), 2018, pp. 1-6.
- [5] R. Zhu, Z. Xiao, Y. Li, M. Yang, Y. Tan, L. Zhou, S. Lin, H. Wen, "Efficient Human Activity Recognition Solving the Confusing Activities Via Deep Ensemble Learning", *IEEE Access* (Volume: 7), 2019, pp. 75490 - 75499.

- [6] M. S. Astriani, G. P. Kusuma, Y. Heryadi and E. Abdurachman, "Smartphone sensors selection using decision tree and KNN to detect head movements in Virtual Reality Application", 2017 International Conference on Applied Computer and Communication Technologies (ComCom), 2017, pp. 1-5.
- [7] X. Yin, W. Shen, J. Samarabandu, X. Wang, "Human activity detection based on multiple smart phone sensors and machine learning algorithms", 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design (CSCWD), 2015.
- [8] N. Jablonsky, S. McKenzie, S. Bangay, T. Wilkin, "Evaluating sensor placement and modality for activity recognition in active games", ACSW '17: Proceedings of the Australasian Computer Science Week Multiconference , 2017, pp. 1 - 8.
- [9] M. Buonamente, H. Dindo, M. Johnsson, "Hierarchies of Self-Organizing Maps for Action Recognition". Cognitive Systems Research, 2016.
- [10] Z. Gharaee, P. Gärdenfors, M. Johnsson, "First and Second Order Dynamics in a Hierarchical SOM system for Action Recognition". Applied Soft Computing, 59, 2017, pp. 574-585.
- [11] Z. Gharaee, P. Gärdenfors, M. Johnsson, "Online Recognition of Actions Involving Objects". Journal of Biologically Inspired Cognitive Architectures, 2017.
- [12] J. Sween, S. F. Wallington, V. Sheppard, T. Taylor, A. A. Llanos, L. L. Adams-Campbell, "The Role of Exergaming in Improving Physical Activity: A Review", Journal of Physical Activity & Health , 2014, pp. 864-870.
- [13] Y. Lee, Y. Lim, "How and Why I Cheated On My App: User Experience of Cheating Physical Activity Exergame Applications", DIS '17 Companion: Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems, 2017, pp. 138 - 143.
- [14] J. Paay, J. Kjeldskov, D. Internicola, M. Thomsen, "Motivations and practices for cheating in Pokémon Go", MobileHCI '18: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, 2018, pp. 1-13.
- [15] A. Gal-Oz, O. Zuckerman, "Embracing Cheating in Gamified Fitness Applications", In Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15) , 2015, pp. 535-540.
- [16] P. Mishra, S. S. Ghosh, D. B. Seal, S. Goswami, "Human Activity Recognition using Deep Neural Network", 2019 International Conference on Data Science and Engineering (ICDSE) , 2019.
- [17] N. Damodaran, J. Schäfer, "Device Free Human Activity Recognition using WiFi Channel State Information", 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Leicester, United Kingdom, 2019, pp. 1069-1074.
- [18] O. M. Prabowo, K. Mutijarsa and S. H. Supangkat, "Missing data handling using machine learning for human activity recognition on mobile device", 2016 International Conference on ICT For Smart Society (ICISS), 2016, pp. 59-62.
- [19] M. N. S. Zainudin, Md N. Sulaiman, N. Mostapha, T. Perumal, "Activity recognition based on accelerometer sensor using combinational classifiers", 2015 IEEE Conference on Open Systems (ICOS), 2015.
- [20] F. Hernández, L. F. Suárez, J. Villamizar and M. Altuve, "Human Activity Recognition on Smartphones Using a Bidirectional LSTM Network", 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), Bucaramanga, Colombia, 2019, pp. 1-5.
- [21] S. W. Pienaar, R. Malekian, "Human Activity Recognition using LSTM-RNN Deep Neural Network Architecture", 2019 IEEE 2nd Wireless Africa Conference (WAC) , Pretoria, South Africa, 2019, pp. 1-5.
- [22] X. Wang, W. Liao, Y. Guo, L. Yu, Q. Wang, M. Pan, P. Li, "PerRNN: Personalized Recurrent Neural Networks for Acceleration-Based Human Activity Recognition", ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, 2019, pp. 1-6.
- [23] Z. Li, X. Xie, X. Zhou, J. Guo, R. Bie, "A Generic Framework for Human Motion Recognition Based on Smartphones", 2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI), 2015.
- [24] A. Almeida, A. Alves, "Activity recognition for movement-based interaction in mobile games", In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17), 2017, pp. 1-8.
- [25] M. Ahmed, A. D. Antar, Md A. R. Ahad, "An Approach to Classify Human Activities in Real-time from Smartphone Sensor Data", International Conference on Informatics, Electronics & Vision & International Conference on Imaging, Vision & Pattern Recogn., 2019.
- [26] E. Kock, Y. Sarwari, "How can machine learning help identify cheating behaviours in physical activity-based mobile applications?". Bachelor thesis, department of computer science and media technology, malmö university, Sweden, 2020.

AI Transformation in the Public Sector: Ongoing Research

Einav Peretz-Andersson

Department of Computing, School of Engineering
Jönköping University, Gjuterigatan 5
SE-553 18 Jönköping, Sweden
Email: Einav.Peretz.Andersson@ju.se

Niklas Lavesson

Department of Computing, School of Engineering
Jönköping University, Gjuterigatan 5
SE-553 18 Jönköping, Sweden
Email: Niklas.Lavesson@ju.se

Albert Bifet

Te Ipu o te Mahara AI Institute
University of Waikato, Gate 1, Knighton Road
Hamilton 3240, Waikato, New Zealand
Email: abifet@waikato.ac.nz

Patrick Mikalef

Department of Technology Management
SINTEF Digital, Strindvegen 4
7034 Trondheim, Norway
Email: patrick.mikalef@sintef.no

Abstract—Real-world application of data-driven and intelligent systems (AI) is increasing in the private and public sector as well as in society at large. Many organizations transform as a consequence of increased AI implementation. The consequences of such transformations may include new recruitment plans, procurement of additional IT, changes in existing positions and roles, new business models, as well as new policies and regulations. However, it is unclear how this transformation varies across different types of organizations. We study the effects of bottom-up approaches, such as pilot projects and mentoring to specific groups within organizations, and aim to explore how such approaches can complement the top-down approach of strategic AI implementation. Our context is the public sector. Our goal is to acquire an improved understanding of how and when AI transformation occurs in the public sector, which are the consequences, and which strategies are fruitful or detrimental to the organization. We aim to study public sector organizations in Sweden, Norway, New Zealand, Germany, and The Netherlands to learn about potential similarities and differences with regard to AI transformation.

I. INTRODUCTION

The digital transformation of society has been steadily increasing during the last five decades. The Internet era of the last two decades has boosted this development and it has made possible the information sharing between individuals and organizations, and across nations, that we experience today. The early industrialization, later industrialization, and the digital transformation along with the introduction of renewable energy sources, are referred to as the first, second, and third industrial revolution. We are now entering the era of the fourth industrial revolution.

The advent of social media and electronic commerce platforms, such as Amazon, Facebook, and Twitter – along with the development of cheaper and more powerful hardware – has significantly increased real-world application of connected,

data-driven and intelligent systems (AI) in the private and public sectors as well as in society at large. As was the case in earlier societal revolutions, old systems, schools of thought, organizations, jobs, and cultures will change or disappear. However, we argue that the AI revolution – the shifting of the cognitive workload from humans to computers – may have characteristics and consequences quite unlike the earlier revolutions.

We are interested in understanding how and why organizations are transformed as a consequence of AI introduction, implementation, and increased readiness. We believe the strategic planning, organizational culture, and decision making processes differ significantly between typical public and private sector organizations. The most prominent explorations of AI transformation have arguably been carried out by and for the private sector. In particular, the focus has been on the information technology industry. Two notable developments in this respect represent fitting illustrations:

- **Landing.AI:** Andrew Ng – founder of Coursera – released the AI Transformation Playbook¹ on December 14, 2018. The playbook draws from experience of AI transformation from the Google Brain team and the Baidu AI Group. The purpose of the playbook is to deliver insights to organizations that can help them to transform into strong AI actors. It specifically addresses large private enterprises.
- **ML-Ops:** In 2015, a paper about the technical debt of ML systems was published at the influential Neural Information Processing Systems (NIPS) conference [1]. The paper, which was written by a number of Google employees, reported that it was common to incur massive ongoing maintenance costs in real-world ML systems. As a consequence, initiatives such as ML-Ops have

been started to establish effective practices and processes around designing, building, and deploying ML models into production².

It is conceivable that the rules of the AI playbook and the guidelines and principles of ML-Ops can be adapted successfully to suit small and medium sized private enterprises. Similarly, the rules and guidelines may provide insights that could benefit non-information technology actors or public sector stakeholders. The question is which rules and guidelines will apply in such contexts, and to what extent. Another question is whether insights gained from AI transformation in the public sector carries across nations the same way such insights would from AI transformation in the private sector.

II. OUTLINE

This abstract seeks to sketch and motivate our long-term AI transformation research agenda and international collaboration, and to describe an ongoing case study of bottom-up AI transformation approaches in the Swedish public sector.

We first provide a brief background on organizational change, digital transformation, and the relationship to AI implementation.

III. BACKGROUND

Organizations change and develop their business models in relation to digitalization. Organizations are operating in a complex environment where new technologies require continuous change. The introduction of AI represents one important type of change. The new age of AI will change the composition, business models, and tasks required in an organization. New business models can be a result of strategy or a strategizing action [2]. AI significantly changes the composition of the resources, operations, and structure that an organization can employ in order to become more efficient and to create value. AI is viewed by many as a revolutionary and societal change agent, which will affect organizational strategy. It will involve economical, psychological, technological, political, and ethical aspects. This change will directly and indirectly affect the business models, including the purpose, process, strategies, infrastructure, organizational structures, and operational processes and policies.

The theory of dynamic capabilities implies that organizations are experiencing constant change, which force them to adapt, integrate, and reconfigure their internal and external competences in order to maintain their competitive advantage [3].

AI technologies affect regulations and policies. From an institutional point of view, some aspects need to receive special consideration:

- 1) **Governance.** Governance concerns the responsibility for the risks involved in AI technologies. Governance should be the bridge of information and knowledge to society in regard to AI. It should ensure that AI technologies that

have an impact on human lives should communicate in a transparent, fair, and accountable way [4].

- 2) **Accountability and Responsibility.** The legal issues concerning AI requires more discussion to determine who should be responsible for the consequences of AI actions and behavior, and to define rules and guidelines for privacy, safety, and integrity.
- 3) **Economy of Scale:** AI may create imbalances in the economy of scale or disrupt small companies [5]. This is a situation that can lead to the increased monopoly of large organizations.

The introduction of smart technologies creates ambiguity and uncertainty. AI may cause *destructive creation* in the future organization. However, from the nature of capitalism it may also create paths to growth [6].

According to the Cambridge Dictionary, *transformation* refers to a complete change in the appearance or character of something or someone, especially so that that thing or person is improved. However, in the area of business administration, transformation usually refers either to radical change or to incremental change in organizations. Organizational Transformation (OT) has been discussed in the organization and management literature from various perspectives. It has been considered as a form of radical change [7]–[9], strategic change [10], revolutionary change [11], continuous and confluent organizational change [12], and organizational discontinuous change [13], [14]. The discussion in the literature about OT dispels ambiguity. However, we consider OT as a consequence of changes happening due to AI technology use.

In our view, AI transformation is an interdisciplinary research topic, and we believe that researchers should opt for new, innovative approaches when exploring this phenomenon. The diverse aspects of the topic can be psychological, socio-cognitive, socio-technical, economic, and political. Research on the topic therefore needs to consider the relevant aspects for the context [15].

In order to strategize AI, it is important to quantify and understand the organization's AI maturity, since maturity is a measure that relates to the organizational readiness and AI capability. The organizational inertia is a vital aspect that makes OT an important theoretical and practical problem [16]. In terms of AI-enabled OT research, the existence of organizational inertia raises the question where the focus should be put: Which are the relevant aspects and which unit of analysis should be considered? (individual vs. departmental, organization vs. sector, and so forth) [16].

We are unable to find much empirical research concerning AI-related OT in the public sector. Meanwhile, public sector companies and authorities face a multitude of challenges related to digital and AI transformation of society. We therefore argue that AI-related OT deserves more attention from the research community. This attention needs to be interdisciplinary in nature.

²ML-Ops, <https://ml-ops.org>

A. Related Work

We are unable to find a concise and useful definition of AI transformation [17]. The available research that brings up this phenomenon is often focusing on digital transformation. There is a substantial scientific discussion around digital transformation but very few works focus on AI. The digital transformation phenomenon refers to the relationship between three domains: data, digital technology, and people [18]. Transformation of an organization or a network of organizations occurs in a variety of contexts, such as: cultural, technological, and governance strategy [19]. Various definitions related to digital transformation are presented [20], however, we observe the need to provide a definition of AI transformation specifically. The reason for this is that, unlike other forms of digital transformation, AI will shift cognitive work from human actors to computers. The consequences for many organizations will be significant. Organizations often recognize the need to implement AI as part of their vision and strategy but training and tutoring regarding AI and its capabilities must be a first anchor in all levels in the organization. AI is affecting organizations in at least two directions. It results in increased digitization of the economy and it enables the automation of existing processes [21]. In addition, AI has a disparate effect on organizations: on the one hand it creates promising opportunities for the future, but on the other hand it involves challenges, uncertainties, and risks. Defining and explaining AI transformation is therefore crucial, since a useful definition will support organizations to adapt to AI.

IV. ONGOING PROJECTS

This section describes ongoing projects and collaborations regarding various aspects of AI transformation research and real-world state-of-practice.

A. AI Strategy Development

AI is an important tool to handle strategic issues which also affects the ability of the organization to achieve strategic change [2]. AI can be both a narrow, specific strategic issue or broad and general. For example, if AI is viewed as a technology, a strategic issue might be to procure an AI system. However, if AI is viewed as a revolutionary and societal change agent, the strategic issue will be broad and involve economical, psychological, technological, political, and ethical aspects. We believe it is important to learn about various organizations' strategic planning processes and their view on how to properly adopt AI so that it fits the strategic planning process. We assume that adaptation of AI is inevitable. However, we observe that few sectors and industries understand AI and its potential impact on organizations.

We have supported Jönköping Municipality (Sweden) in developing an AI strategy. The city director and official management are now implementing the AI strategy. We refer to this implementation as a *top-down approach* for AI transformation.

B. Systematic Literature Review

We have conducted a systematic literature review [17]. The aim of the SLR was to aggregate the body of knowledge on the relationship between AI and organizational transformation, to map the field, and to identify the gaps in research that represent an opportunity for future study. We have used three databases (Science Direct, Scopus, and IEEE) and found 966 papers related to the topic. We have followed the procedure and guidelines of Kitchenham [22] and narrowed the article selection to a final number of 52 relevant articles, which help us to explore AI transformation.

C. Interviews with Jönköping Municipality

We have performed interviews with 23 officials from the official management of Jönköping Municipality. The purpose is to understand how AI implementation transforms public sector organizations. The municipal organization represents an interesting context. The aim is to contribute to the knowledge of how the municipal board and civil servant management and other management groups create, formulate, and implement digitization and AI strategies. A special focus will be put on how AI can contribute to the municipality's work to improve the organization's efficiency.

D. Bottom-up AI Transformation

We have asked Jönköping Municipality to define use cases or challenges within its respective branches and municipal companies. We have then supported the municipality in developing the most mature use cases into research and development projects. We are now assigning students and researchers to develop prototypes and experiments related to four of the use cases:

- UC1 **Contact Center.** The goal is to learn from a data set of historical incoming citizen requests and responses to generate a decision support system that can recommend categorizations of requests and formulation of responses human operators.
- UC2 **Technical Office.** The goal is to learn from a data set of historical citizen incident reports to generate a model that can recommend actions to human operators (for example, to file a police report or provide a specific response).
- UC3 **Technical Office.** The goal is to learn from a data set of historical citizen incident reports about winter weather consequences to generate a model that can formulate automated answers to citizens and prioritize planned actions.
- UC4 **City Planning Office.** The goal is to find discrepancies between a high-precision digital base map of the geography and built environment of the municipality and actual aerial footage.

We aim to perform an intervention study in which we explore the effects of conducting these types of AI-student projects as well as other bottom-up approaches, such as mentoring and competence development courses, in the public sector.

E. Ongoing Work in New Zealand

We asked Waikato Regional Council to find environmental use cases where AI can help to improve decision making. They provided four use cases, and now we have researchers from the TAI AO project working on these cases. The *Time-Evolving Data Science / Artificial Intelligence for Advanced Open Environmental Science (TAIAO)* is an AI program over seven years, funded by the New Zealand Ministry of Business, Innovation, and Employment (MBIE). It is a collaboration between the Universities of Waikato, Auckland and Canterbury, Beca and MetService.

F. Ongoing Work in Norway

We have been collaborating with several municipalities within Norway on developing their AI capabilities, and accumulating the necessary resources to commence their AI transformation. An early study revealed some of the main challenges municipalities face in their quest of deploying AI applications [23]. The current work examines the role of technological, organizational, and environmental enablers and inhibitors of AI deployment in municipalities, as well as the realized organizational value from such investments. This work, as well as ongoing research examining a process view of AI deployment, aims to understand how the value of novel technologies such as AI can be maximized.

V. JOINT RESEARCH AGENDA

We are aiming for an innovative research approach. We view AI transformation as an interdisciplinary phenomenon, and we aim to investigate how organizations change due to AI technology maturation and implementation. In this particular research agenda, our main interest is in organizations which operate in public sector. We aim to collaborate in an international setting, and to perform joint studies and comparative studies across various contexts, as well as different public sector organizations and countries, including: Sweden, Norway, New Zealand, Germany, and The Netherlands. We believe that the interdisciplinary and international collaboration will yield a comprehensive understanding of AI transformation and its impact on the public sector. Also, we intend to establish a foundation for decision and policy makers, so that they are able to engage fruitfully in AI transformation.

ACKNOWLEDGMENT

We are thankful to Jönköping Municipality and city director Johan Fritz for opening up the organization to us and for enabling us to interview decision makers, perform collaborative pilot projects, and follow the strategic implementation of digitalization and AI carried out by the municipality.

REFERENCES

- [1] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, "Hidden technical debt in machine learning systems," in *Neural Information Processing Systems*, 2015.
- [2] J. E. Dutton and S. E. Jackson, "Categorizing strategic issues: Links to organizational action," *Academy of Management Review*, vol. 12, no. 1, pp. 76–90, 1987.
- [3] D. J. Teece, G. Pisano, and A. Shuen, "Dynamic capabilities and strategic management," *Strategic Management Journal*, vol. 18, no. 7, pp. 509–533, 1997.
- [4] U. Gasser and V. A. Almeida, "A layered model for ai governance," *IEEE Internet Computing*, vol. 21, no. 6, pp. 58–62, 2017.
- [5] B. W. Wirtz, J. C. Weyerer, and C. Geyer, "Artificial intelligence and the public sector – applications and challenges," *International Journal of Public Administration*, vol. 42, no. 7, pp. 596–615, 2019.
- [6] J. A. Schumpeter, *The Theory of Economic Development*. Harvard, 1934.
- [7] R. Greenwood and C. R. Hinings, "Understanding radical organizational change: Bringing together the old and the new institutionalism," *Academy of Management Review*, vol. 21, no. 4, pp. 1022–1054, 1996.
- [8] D. Anderson and L. A. Anderson, *Beyond change management: Advanced strategies for today's transformational leaders*. John Wiley & Sons, 2002.
- [9] M. L. Tushman, W. H. Newman, and E. Romanelli, "Convergence and upheaval: Managing the unsteady pace of organizational evolution," *California Management Review*, vol. 29, no. 1, pp. 29–44, 1986.
- [10] A. Pettigrew, *The Awakening Giant*. Oxford: Blackwell, 1985.
- [11] C. J. Gersick, "Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm," *Academy of Management Review*, vol. 16, no. 1, pp. 10–36, 1991.
- [12] H. Arazmjoo and H. Rahmanseresht, "A multi-dimensional meta-heuristic model for managing organizational change," *Management Decision*, vol. 58, no. 3, pp. 526–543, 2019.
- [13] H. Mintzberg and F. Westley, "Cycles of organizational change," *Strategic Management Journal*, vol. 13, no. S2, pp. 39–59, 1992.
- [14] H. Tsoukas and R. Chia, "On organizational becoming: Rethinking organizational change," *Organization Science*, vol. 13, no. 5, pp. 567–582, 2002.
- [15] P. Mikalef, R. van de Wetering, and J. Krogstie, "Building dynamic capabilities by leveraging big data analytics: The role of organizational inertia," *Information & Management*, 2020.
- [16] P. Besson and F. Rowe, "Strategizing information systems-enabled organizational transformation: A transdisciplinary review and new directions," *Journal of Strategic Information Systems*, vol. 21, no. 2, pp. 103–124, 2012.
- [17] E. Peretz-Andersson, N. Lavesson, and R. Torkar, "Ai transformation: A systematic literature review," 2021, submitted for peer-review.
- [18] M. L. Ashwell, "The digital transformation of intelligence analysis," *Journal of Financial Crime*, vol. 24, no. 3, pp. 393–411, 2017.
- [19] L. Heilig, E. Lalla-Ruiz, and S. Vob, "Digital transformation in maritime ports: Analysis and a game theoretic framework," *Netnomics: Economic Research and Electronic Networking*, vol. 18, no. 2–3, pp. 227–254, 2017.
- [20] S. Akter, K. Michael, M. R. Uddin, G. McCarthy, and M. Rahman, "Transforming business using digital innovations: The application of ai, blockchain, cloud and data analytics," *Annals of Operations Research*, pp. 1–33, 2020.
- [21] S. L. Wamba-Taguimdje, S. F. Wamba, J. R. K. Kamdjoug, and C. E. T. Wanko, "Influence of artificial intelligence (ai) on firm performance: The business value of ai-based transformation projects," *Business Process Management Journal*, 2020.
- [22] B. Kitchenham, "Procedures for performing systematic reviews," Keele University. Technical Report TR/SE-0401, Department of Computer Science, Keele University, UK, Tech. Rep., 2004.
- [23] P. Mikalef, S. O. Fjortoft, and H. Y. Torvatn, "Artificial intelligence in the public sector: A study of challenges and opportunities for norwegian municipalities," in *Conference on e-Business, e-Services and e-Society*, 2019, pp. 267–277.

Rock Classification with Machine Learning: a Case Study from the Zinkgruvan Zn-Pb-Ag Deposit, Bergslagen, Sweden

1st Filip Simán

*Civil, Environmental and Natural Resources Engineering
Luleå University of Technology
Luleå, Sweden
filip.siman@ltu.se*

2nd Nils Jansson

*Civil, Environmental and Natural Resources Engineering
Luleå University of Technology
Luleå, Sweden
nils.jansson@ltu.se*

3rd Tobias Kampmann

*Civil, Environmental and Natural Resources Engineering
Luleå University of Technology
Luleå, Sweden
tobias.kampmann@ltu.se*

4th Foteini Liwicki

*Computer Science, Electrical and Space Engineering)
Luleå University of Technology
Luleå, Sweden
foteini.liwicki@ltu.se*

Abstract—In this paper we assess two traditional machine learning (ML) methods which can be used for automatic rock type classification: (1) the Self-Organising Map (SOM) with k-means clustering, and (2) Classification and Regression Trees (CART). The dataset used for this paper were chemical compositional data of rocks acquired through X-Ray Fluorescence (XRF) analysis. The ground truth of the dataset was generated by human experts in the field of geology. The complexity of the chosen dataset influenced the evaluation performance of the two ML models. We achieve an overall accuracy of 68.02 % and 62.79 % respectively when using SOM with k-means and CART.

Index Terms—Rock classification, Self Organising Map, Classification and Regression Trees

I. INTRODUCTION

Metals used in the development of green technologies require new mineral deposits to be discovered. However, after centuries of mining in Sweden, new economic deposits are becoming increasingly difficult to discover. To find and delineate new deposits, mineral exploration relies on using exploration drilling for collecting cylindrical rock samples of the sub-surface, i.e. drill cores. An example of drill core is shown in Figure 1. The drill cores are investigated and sampled by geologists to produce core logs, and the collected data and information are interpolated between drill holes to produce 3D geological models. A core log is a graphical representation of data collected from a drill core, and can include continuous measurements (e.g. core scan data), spot observations with depth mark (e.g. a fracture) or sequences, such as the sequence of rock units intersected in a drill hole, provided as depth intervals. These depth intervals are manually labelled according to identified rock type by the logging geologist, using textural, physical and/or compositional criteria. This study compares two different Machine Learning (ML) algorithms for

classifying rock types using continuous X-Ray Fluorescence (XRF) core scan data of the chemical compositions of rocks in the drill core.

The benefits of a ML approach are objectivity and rapidity, as opposed to the slower and commonly more subjective method of manual core logging done by a geologist. A key finding of this study is that different ML algorithms and input parameters will yield different classification accuracies, naturally leading to the question of which is the optimal ML algorithm. Crucial for the choice of ML algorithm is the intended purpose of a hypothetical project, i.e. how the classification outputs will be used during subsequent geological modelling, and the degree of prior knowledge. Templ et al. [1] conclude that clustering algorithms are suitable for exploratory data analysis, i.e. if the data are unlabelled. Supervised methods used e.g. by Hood et al. [2] could, on the other hand, replace the manual task of chemically grouping rocks to decrease work load for the geologist. Once systematic and unambiguous rock classification has been performed, the results from several drill cores can be interpolated to construct geological models of the sub-surface allowing for prediction of the location, volume and tonnage of mineral deposits. Although fully automated mineral exploration is unlikely in the near future, the industry may soon benefit from partial automation using ML algorithms, granted that high enough classification accuracies can be achieved.



Fig. 1. Example optical photograph of a core box with rock samples. Notice the heterogeneity of rock types.

II. METHODOLOGY AND DATADSET

This study uses rocks from the Zinkgruvan Zn-Pb-Ag and Cu deposit, Bergslagen, Sweden because of its relatively predictable and discrete succession of rocks in the stratigraphy [3], making it a suitable test site for applying ML algorithms for rock clustering and classification. The Zinkgruvan mine is one of the largest and most high-grade zinc deposits in Europe. The mined zinc ore body is sheet-like and constitutes a minor part of the stratigraphic succession. However, it occupies a highly predictable stratigraphic position in the sense that the sequence of rock types on one side of the zinc ore are different from the ones on the other side. Hence, by being able to correctly identify the sequence of rock types in drill core, correct interpolation between them can be done and the mine can effectively vector in on the ore-bearing unit in 3D space and trace it horizontally over distances of several kilometres, and to a depth of at least 1.6 km [3].

Figure 2 shows a simplified workflow used in the study resulting in three different representations of the same drill cores. Four drill cores, consisting of 1204 metres of rock, were studied and a detailed manual core log was made for each of them based on conventional rock classification methods and examination of XRF core scan data for mapping the chemistry of the stratigraphy. Each manual core log consisted of depth references, the geologist assigned rock types and the XRF chemistry data. Two ML algorithms were studied, an unsupervised and a supervised ML algorithm: (1) k-means clustering applied to the Self-Organising Map (SOM) and (2) Classification and Regression Trees (CART). CART and SOM, presented here, both used 1-metre data resolution and includes the elements Al, Si, K, Ca, Ti and Fe. These elements were chosen since they best represented the rock types in the stratigraphy.

In the case of CART, the manual core logs were divided so that three were used as training and validation data and the fourth was used as test data. To classify drill core 4442 CART was trained on 602 metres of XRF sample intervals and 258 metres were used as validation data. The distribution of classes within the 602 metres of training data was unbalanced. Table I shows the distribution of classes, i.e. rock types. Notice

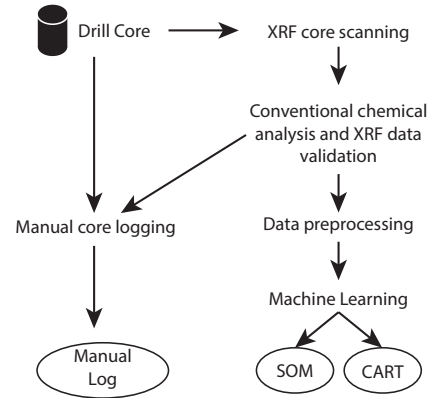


Fig. 2. Schematic of the workflow used for data acquisition, pre-processing and core logging. The elliptical boxes represent the core logs shown in this paper.

that diabase does not occur in any other drill cores than 4442, hence it was not represented in the training data.

TABLE I
CLASS DISTRIBUTION ON THE TRAINING DATA

Rock Type	Class (%)
Pelite	3.84
Garnet+biotite+quartz rock (GBK)	1.05
Metatuffite	33.02
Diopside Skarn	11.63
Interbedded calcite skarn and metatuffite (KSL)	7.33
Stratiform sphalerite mineralisation	0.81
Dolomitic marble	17.65
Metamafite	2.09
Microcline-quartz rock	22.67
Granite	0.46
Pegmatite	1.71

The SOM used all four core logs, i.e. 1204 metres of XRF sample intervals, as training data. After training the SOM it was handed to k-means clustering to produce meaningful clusters. The k-value for k-means clustering was chosen by plotting the variation of values within clusters, i.e. delta value or sum of squares (SS), against the number of clusters, as shown in Figure 3. This way an optimal k-value of 7 was found. The output from k-means was classified by assessing to which extent the clusters corresponded to identifiable rock types in the manual core log by means of both visual and compositional characteristics.

For determining the test accuracies of the two ML methods a simple calculation was applied. The test accuracy was calculated as the percentage of intervals between the manual core log and an ML core log that were assigned with the same rock type. The validation accuracy for CART was calculated in a similar way.

A potential need for pre-processing prior to running ML models stems from a special feature of compositional data, namely that it inhabits the simplex space. This means that each data point represents a part of a constant sum, and

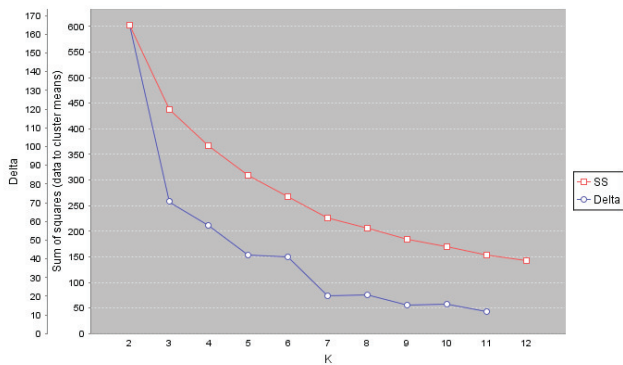


Fig. 3. Plot showing the decrease of variation within clusters as the k-value increases. This is the plot used assisting in the choice of k=7.

variables are not free to vary independently; increasing one variable leads to decrease in all others to maintain constant sum. This can lead to spurious correlations. It is therefore recommended by some authors to apply a data transformation to move the data from the simplex to Euclidean space [4], [5]. Chemical data from rocks also carry artifacts from the analytical procedure. XRF detectors have detection limits. That is, when a specific element occurs in low concentrations in a rock, it is not possible for the detector to properly quantify that element. In such a case, no unique numeric value is returned for that element in that sample, and instead, the non-numeric 'below limit of detection (LOD)' becomes the entry for that element, commonly tabulated as μX , where X is the lowermost value that can be quantified. This study uses a common replacement rule used in geochemistry, namely replacing μX with the numeric value $0.5X$. It is emphasized that different LOD apply to different elements during XRF analysis; the lower the atomic number an element has, the less detectable it is [6]. For the specific XRF scanner used in this study, Al was the element with the highest LOD.

III. RESULTS AND DISCUSSION

Figure 4 presents rock units in a selected drill core, represented as coloured bars assigned to specific depths in the drill hole. The zinc ore is seen at a depth of 180 metres (termed Stratiform sphalerite mineralization) in the manual log, but is missing from the CART and SOM logs. Where the logs visually agree better are for the dolomitic marble, that is seen at 220 to 240 metres depth. The overall test accuracy for the four drill cores was 68.02 % and 62.79 % for SOM and CART respectively. CART achieved a validation accuracy of 77.4 %.

Figure 5 shows a correlation matrix for SOM clusters to different rock types. An important insight from the correlation matrix is the confusion between clusters, or in other words the overlap of a rock type over several clusters. Metatuffite for example is not uniquely correlated with a specific cluster. This is because there is no distinct chemical composition for a metatuffite. Metatuffite is a particular type of rock of clastic sedimentary origin, for which the composition and

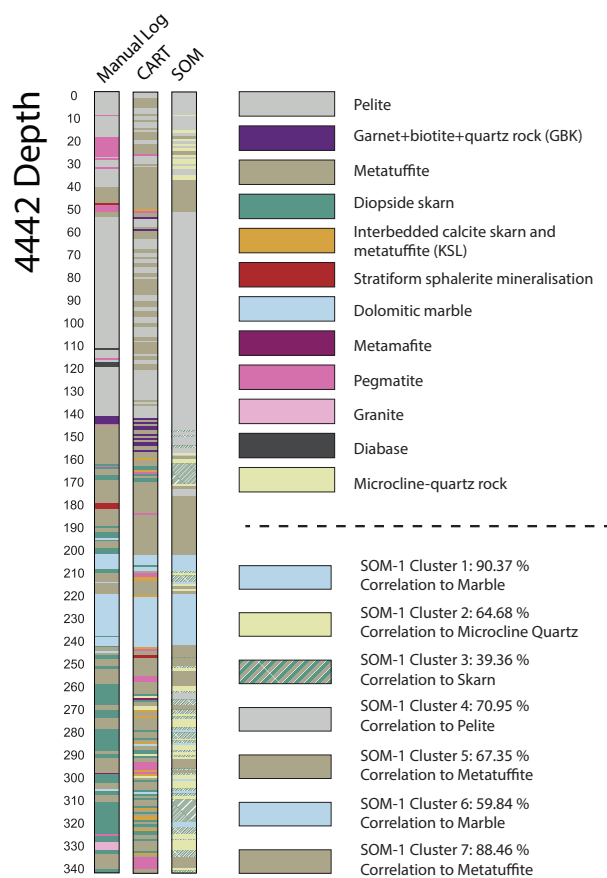


Fig. 4. Comparison of three methods of core logging on one of the drill cores, hole ID 4442.

	Pelite	Garnet+Bi- tite+quartz rock (GBK)	Metatuffite	Diopside skarn	Interbedded calcite skarn and metatuffite (KSL)	Stratiform sphalerite mineralisation	Dolomitic marble	Metamafite	Pegmatite	Granite	Diabase	Microcline-quartz rock
Cluster 1	0.00	0.00	0.00	9.63	0.00	0.00	90.37	0.00	0.00	0.00	0.00	0.00
Cluster 2	2.23	0.37	21.19	4.83	0.74	0.00	0.00	0.74	2.97	2.23	0.00	64.68
Cluster 3	0.00	0.00	30.85	39.36	9.57	0.00	0.00	11.17	0.53	0.53	0.00	5.85
Cluster 4	70.95	6.67	18.10	1.43	0.00	0.00	0.00	0.95	1.43	0.00	0.48	0.00
Cluster 5	0.00	0.00	67.35	11.22	10.20	9.18	0.00	1.02	1.02	0.00	0.00	0.00
Cluster 6	0.00	0.00	0.82	29.51	6.56	0.00	59.84	0.82	0.00	0.00	0.00	0.00
Cluster 7	1.65	0.00	88.46	0.00	0.55	0.00	0.00	0.00	8.79	0.55	0.00	0.00

Fig. 5. Correlation matrix for SOM between clusters and rock types. Green represents stronger correlation and red represents weaker correlation.

texture is commonly heterogeneous, reflecting variable and non-systematic admixture of clays and volcanic detritus in the rock precursor. In fact, from a geological point of view, this heterogeneity is a defining criteria of metatuffite, whereby it reflects a challenge for any classification aiming to sort data into discrete compositional groups. In contrast, the rock type dolomitic marble is compositionally homogeneous and chemically distinctive from other rocks, and hence correlates strongly to a single cluster. Further insight may be found by comparing Figure 5 to Table I. Notice that rock types that occur sparsely, i.e. small class percentage in training data, generally do not correlate with any clusters. Examples of these rocks are diabase, granite and pegmatite. Thus, the SOM has difficulty finding rock types that are uncommon in the stratigraphy. This is also apparent by the fact that the k -value was found to be 7 when there are in fact 12 different rock types. Most importantly, none of the methods successfully identifies the Stratiform sphalerite mineralization as a distinct group. Nevertheless, if the entire sequence of CART and SOM outputs are considered, then both classifications are useful for predicting the position of the Stratiform sphalerite mineralization. This position is apparent in all methods as the transition from Metatuffite/Cluster 5&7 alternating with dolomite marble/Cluster 1&6, to Pelite/Cluster 4. This underlines the importance of studying the spatial context and sequence of the ML outputs.

Shortcomings in test accuracy for both ML algorithms may in part be explained by the insights from Figure 5. However, there is more to be discussed in terms of error analysis. The reason why studying ML applications in geosciences is interesting include challenges highlighted by this contribution. According to Dramsch [7], ML is successful in applications where there is a lot of data and the environment is relatively simple, which is generally not the case for geological data. As this case study has shown, the environment is difficult with data being heterogeneous, unbalanced and noisy. A further challenge is the choice of environment, i.e. chemical data, which does not fully describe all properties of rocks. For example, two different rock types may have similar chemical compositions with differences only in visual features such as grain size. To mitigate this issue different data sources would have to be fused, such as visual and physical characteristics. ML classification is also a challenge since rock materials are often heterogeneous and inaccessible (i.e. the very reason why drill cores are produced) making it difficult to assert a ground truth for assessing test accuracy and training supervised learning algorithms. This challenge is specifically relevant to supervised learning since the ground truth is used for labelling. Ground truth data for labelling is acquired by geological experts. However, a specific issue with defining a ground truth lies in that geological experts commonly cannot objectively agree upon or define all aspects of the ground truth data, such as e.g. terminology schemes and genetic interpretations. This is in part due to general subject-specific state-of-the-art questions when it comes to evolving geological understanding and different schools of thought. This is partially related to

individual geologists being biased by their own experience from a specific mineral deposit or deposit types, and how e.g. routines for rock subdivision changes with exposure to more and more core [8]. Inherent to the word 'exploration' is the fact that it explores the unknown rather than categorises the already known. Besides searching for mineral deposits, exploration geologists commonly find themselves in a situation where they gradually have to build up geological understanding of a new area, including devising a scheme for being able to classify rocks in a way that is not too complicated, and not too simplified. Establishing a ground truth under these conditions can be very difficult.

Other shortcomings may be due to the aforementioned unique features of chemical data from rocks. Rock chemistry data rarely conform to a gaussian distribution [5], instead they inhabit the compositional data space (the simplex). The issue when performing correlation analysis on compositional data is that they are not independent, leading to spurious correlations. However, there is no consensus on whether data transformation to Euclidean space is necessary. Trépanier et al. [9] argue that such a data transformation makes no difference whereas Reimann et al. [4] argue that is crucial for successful statistical analysis of compositional data. Also, rock chemistry datasets often have gaps due to detection limits of the method of analysis, in the case of this study detection limits of the XRF sensor. Missing values are an issue since they can lead to spurious correlations, which in turn can impact negatively on any method which relies on correlations for data dimension reduction or clustering.

This study concludes that: 1) k -means clustering applied to SOM is the most successful approach in classifying rock types of the two tested ML algorithms, and 2) the complexity of the rocks is a challenge for ML leading to low test accuracies. However, both the supervised method of CART and the unsupervised method of k -means clustering applied to SOM show potential to assist core logging procedures, and successfully outline the prospective part of the studied drill core when it comes to the zinc ore mined at Zinkgruvan. Hence, despite the many challenges ML faces in geosciences, further work is strongly recommended.

ACKNOWLEDGMENT

This contribution is based on a MSc thesis by the first author, presented at LTU. We thank Marcus Liwicki (LTU) for answering ML questions during the work. We also thank the staff at Zinkgruvan Mining AB. We especially thank Anja Hagerud (Zinkgruvan Mining AB) who helped organise visits to Zinkgruvan and Filip Ivarsson (Zinkgruvan Mining AB) who conducted the rock sampling. Thanks also go to Minalyze AB employees Andreas Inerfeldt and Torbjörn Svensson for answering questions concerning XRF core scanning.

REFERENCES

- [1] Templ, M., Filzmoser, P., Reimann, C., 2008: Cluster analysis applied to regional geochemical data: Problems and possibilities. *Applied Geochemistry* 23, pp. 2198–2213.

- [2] Hood, S.B., Cracknell, M.J., Gazley, M.F., 2018: Linking protolith rocks to altered equivalents by combining unsupervised and supervised machine learning. *Journal of Geochemical Exploration* 186, pp. 270–280.
- [3] Jansson, N.F., Zetterqvist, A., Allen, R.L., Billström, K., Malmström, L., 2017: Genesis of the Zinkgruvan Zn-Pb-Ag deposit and associated dolomite-hosted Cu ore, Bergslagen, Sweden. *Ore Geology Reviews* 82, pp. 285–308.
- [4] Reimann C., Filzmoser P., 1999. Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. *Environmental Geology*, Vol. 39, pp. 1001–1014.
- [5] Aitchison, J., 1982: The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society* 44, pp. 139–177.
- [6] Beckhoff, B., Kanngießer, B., Langhoff, N., Wedell, R., Wolff, H., 2006. *Handbook of Practical X-Ray Fluorescence Analysis*. Springer.
- [7] Dramsch, J.S., 2020. 70 Years of Machine Learning in Geoscience in Review. *Advances in Geophysics*, Vol. 61, pp. 1–55.
- [8] Baddeley, M.C., Wood, R., Curtis, A., 2004. An introduction to prior information derived from probabilistic judgements: Elicitation of knowledge, cognitive bias and herding. Geological Society London Special Publications. DOI: 10.1144/GSL.SP.2004.239.01.02
- [9] Trépanier, S., Mathieu, L., Daigneault, R., Faure, S., 2016. Precursors predicted by artificial neural networks for mass balance calculations: Quantifying hydrothermal alteration in volcanic rocks. *Computers & Geosciences*, Vol. 89, pp. 32–43.

Hit Detection in Sports Pistol Shooting

Elinore Stenhager

Department of Computing, School of Engineering
Jönköping University, Gjuterigatan 5
SE-553 18 Jönköping, Sweden
Email: elinorestenhager@hotmail.com

Niklas Lavesson

Department of Computing, School of Engineering
Jönköping University, Gjuterigatan 5
SE-553 18 Jönköping, Sweden
Email: Niklas.Lavesson@ju.se

Abstract—Score calculation and performance analysis of shooting targets is an important aspect in the development of sports shooting ability. An image-based automatic scoring algorithm would provide automation of this procedure and digital visualization of the result. Existing solutions are able to detect hits with high precision. However, these methods are either too expensive or adapted to unrealistic use cases where high quality paper targets are photographed in very favorable environments. Usually, precision pistol shooting is performed outdoors and bullet holes are covered with stickers between shooting rounds. The targets are reused until they are destroyed. This paper introduces the first generation of an image-based method for automatic hit detection adapted to realistic shooting conditions. It relies solely on available image processing techniques. The proposed algorithm detects hits with 40 percent detection rate in low-quality targets, reaching 88 percent detection rate in targets of higher quality.

I. INTRODUCTION

In sports analytics, the motion patterns of athletes and their handling of related equipment are analyzed in order to produce predictions and improve coaching techniques and instruction. The collection of relevant data has in recent years been increasingly automated and is currently gathered using advanced sensors and cameras in many sports [1].

Recreational and competitive shooting sports as well as hunting are popular activities in many countries. In Sweden, an estimated 600,000 members of the population hold firearms licenses for purposes of hunting or shooting sports. Recreational and competitive shooting is the second most popular sport in Sweden, based on the number of athletes.

In several shooting sport disciplines, training and competition are carried out outdoors. The targets in these disciplines are usually made out of paper. It is common for competitive shooters to fire thousands of rounds per year for training purposes. The detection and marking of hits and the score calculation during training and competition is almost exclusively carried out manually. Professional sports shooters may use expensive acoustic target technology for automatic target scoring.

In this paper, we propose a method for automatically detecting hits and calculating the score. We perform experiments on a small data set of digitalized paper targets. Our aim is to develop the method further to improve the hit detection rate

and to introduce automatic shot grouping analysis. This way, our method would be able to provide support to the athletes to make training more effective.

The paper is structured as follows. First, we provide a brief background on the shooting sport discipline of study, precision pistol shooting. Then, we review related work. We present the first generation of the proposed method and then describe a pilot experiment design. This is followed by a summary and discussion of the results. Finally, we draw conclusions and provide pointers to future work.

II. BACKGROUND

A. Precision Pistol Shooting

Sports shooting depends on accuracy, speed, and control in using firearms. The sport consists of a multitude of diverse disciplines, including precision pistol, which is performed by standing in the off-hand position, aiming at a target at distance 25 meters, whereupon a series of shots are fired. According to Swedish national contest rules, a contest commonly includes six rounds or more, where one round consists of a series of five shots fired within the time limit of five minutes. Between each round, the paper targets are investigated to identify and mark the hits. Then the score is calculated and the bullet holes are covered using stickers¹.

B. The International Pistol Target

In precision pistol contests, the official 25-meter precision target is used. It is referred to as the International Pistol Target (see Figure 1). This target is sectioned into eleven circular score zones, referred to as score rings. The diameter of the outer border of the outermost ring is 500 millimeters. Each score ring represents scores from 1 (the outer most score ring) to 10 (center). The five inner score rings representing scores 7 to 10 are black. The two most inner score rings both represent 10 points and are collectively referred to as the bullseye. The center score ring, referred to as the inertia, is used to distinguish top scoring competitors.

Each round produces a *shot grouping*, or cluster of hits. The size of the cluster represents spreading of the hits (high spread equals low precision). The difference between the cluster's center of mass and the center of the bullseye represents

¹<https://www.pistolskytteforbundet.se/om-pistolskytte/banskyttegrenar/precisionskjutning/>

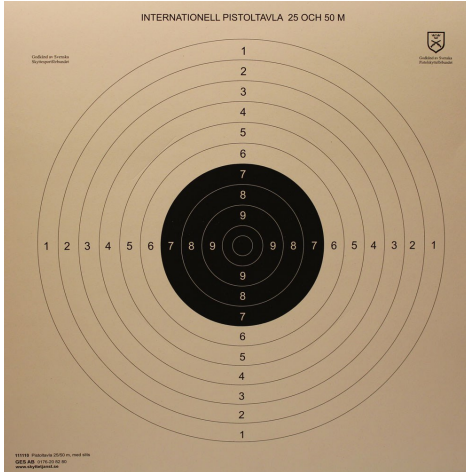


Figure 1. The official international pistol target.

accuracy (large difference equals low accuracy). Analysis of this shot grouping may provide information on how movement, position, posture, breathing, and trigger pull can be adjusted to improve the performance. Currently, the hit identification, scoring, and shot grouping analysis are manually performed.

One approach to automate these steps for training and education purposes is to develop a software application that can be installed on inexpensive smartphones. The application would allow the user to capture a digital image of the target. The image is then processed by the application, producing a score calculation and a shot grouping analysis. A challenge regarding this approach is to develop a robust method to detect hits from a digital image of the paper target. This would require a calibrated chain of suitable image processing and image analysis techniques.

C. Shot Grouping Analysis

Shot group analysis often leads to an increased understanding of the performance of a shooter and the strengths and weaknesses. If the shooter fires a series of, say, five shots, the grouping can be analyzed by determining its size or dispersion. It is possible to determine whether the shooter lacks precision (large spread of shots) or accuracy (the center of mass of the grouping is not located in the center of the target). For precision pistol purposes, the precision of groupings is a key success factor. With multiple, subsequent series and shot grouping analyses, it is possible to determine the shooter's tendency to achieve *fliers* (tight groupings with the exception of one or two poor shots).

III. RELATED WORK

For military applications, traditional Location of Miss and Hit (LOMAH) radar-based technology provides automatic location of misses and hits by detecting the presence of rounds passing over or about detecting sensors². LOMAH technology is too expensive for most shooting sports organizations to

invest in. Moreover, up until recently, LOMAH was only available for supersonic projectile scenarios, which excludes most pistol calibers and projectiles. For civilian purposes, there exist electro-mechanical acoustical projectile detection systems for the same purpose (LOMAH). However, such systems are typically too expensive for most sports organizations [2]. Image recognition-based methods have been developed and evaluated as an alternative to acoustical systems. One study performed an experimental evaluation of a proposed image recognition-based hit recognition method [3]. The method, which was primarily based on segmentation and erosion, yielded a recognition accuracy of 98.3%. Segmentation was used to identify the inner score rings (the black area, sometimes referred to as the aiming point). Erosion was then used to highlight and pinpoint the hits. The drawback of the proposed method is that it was developed for high quality paper targets photographed in indoor environments with favorable lighting conditions. Another similar method was proposed and it yielded a recognition accuracy of 92% [4]. The method was broken down into three steps:

- 1) to identify score rings,
- 2) to recognize hits,
- 3) to determine the score of each hit.

The proposed approaches for identification and recognition were based on standard image recognition and processing algorithms, such as circle Hough transform (CHT), Prewitt edge filter, and flood fill. This method and study share the drawbacks of the previously reviewed method and study, namely that only high quality paper targets were used and photographs were taken under favorable conditions. Under realistic conditions, athletes practice and compete outdoors in varying weather and lighting conditions. Depending on the lighting and shading conditions, the characteristics of the target and the bullet holes change drastically. Additionally, the paper targets are typically reused by using stickers to cover bullet holes. After a number of reuses, especially when using large calibers, the paper target quality deteriorates. We were unable to find any existing image recognition-based methods or commercial systems for such realistic conditions.

IV. AIM AND SCOPE

The aim of this paper is to present the first generation of an automatic method for hit detection, under realistic conditions, in precision sports pistol shooting. Refer to Figure 2 for two illustrations of real-world paper targets. The end goal is to achieve sufficiently high accuracy to minimize manual human labor for post-processing (deletion, addition, or movement of hits). If this goal is reached, the recognition system could be implemented as a mobile application to support and assist athletes in the shooting sports community.

A. Delimitation

This work is limited to the precision pistol shooting discipline, subject to the official Swedish national contest rules. The data collection is limited during this iteration of the research

²<https://www.polytronic.ch/en/military-and-police/lomah>

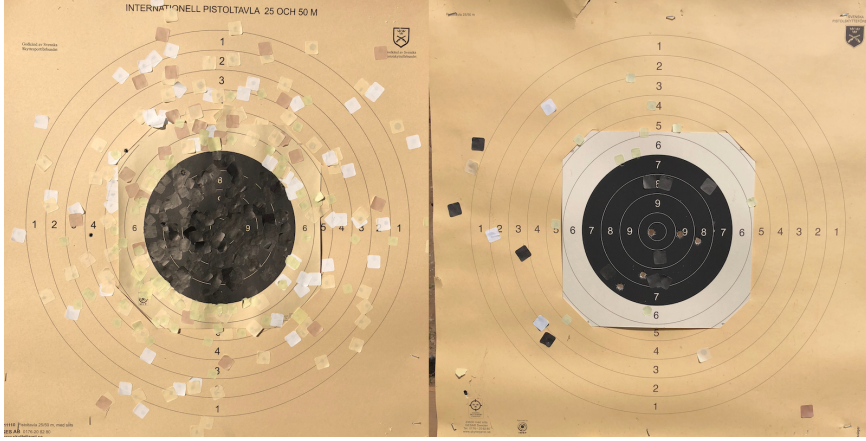


Figure 2. Two examples of real-world paper targets, photographed in realistic conditions.

project. Data have been collected from multiple local shooting ranges, belonging to Swedish sport shooting organizations.

V. METHOD

The proposed approach is divided into three steps. In Step A, the image is pre-processed to identify the outer (light brown) and inner (black) scoring ring of the target. In Step B, each scoring zone is identified, including the bullseye (the two innermost rings). In Step C, hits are detected and the score is estimated.

A. Target Segmentation

Each image is cropped to improve the target segmentation. The process of cropping starts by performing a binarization of the image using Otsu's method [5].

From the resulting binarized image, the contour of the target is extracted. Using the contour line, the image can then be cropped and rotated.

Detection of Inner Score Rings: Refer to Figure 1 for a visualization of the paper target. The black circle represents the inner score rings. This region is sometimes referred to as the aiming point. It is central to the target and it also serves as a suitable reference point for target segmentation.

Since the paper target projection is usually slightly rotated in more than one dimension, the black area appears as a rotated ellipse in the image, rather than a circle. After the binarization, the inner score rings are simple to detect and their contour can be extracted and defined as an ellipse $E_0 = (O_0, a_0, b_0, \theta_0)$, where O_0 represents the center coordinate, a_0 and b_0 represent the main axes, and θ_0 the orientation.

1) *Cropping to Region of Interest:* The region of interest in hit recognition is the contour of the outer most score ring of the paper target and the complete area inside of this contour. We refer to this region as the complete *scoring region* of the target. We isolate the scoring region through image masking, effectively removing all information in the image outside of the scoring region contour.

The dimensions of the scoring region are proportionate to the inner score rings. The scoring region has an actual diameter

of 500 millimeter and the outer bound of the inner score rings has an actual diameter of 200 millimeter. Consequently, the location and shape of the scoring region can be estimated with an ellipse $E_{roi} = (O_0, ka_0, kb_0, \theta_0)$. Here, $k = 0.3$ is a multiplying factor that ensures that the complete scoring region is captured even in situations where proportions are incorrect due to skew in projection. By performing another image masking operation, we are able to remove all image information surrounding the scoring region.

B. Score Ring Estimation

Based on the perspective of the image, the score rings are projected as ellipses with varying center coordinates and unproportionally increasing dimensions. The estimation of these score rings based on the position and dimensions of the inner score rings (the black area) would yield a potential score estimation accuracy which is correlated with the position and angle of the camera. We therefore opt for an approach that is based on recognition of visible (drawn) circle contours in the target. However, the inner circles are usually more fragmented and obfuscated by stickers. The score rings are therefore divided into two groups and treated separately: inner score rings (the black area) and outer score rings (the light brown area). A detailed description of the score ring estimation is out of scope for this summary. An intermediate result can be viewed in Figure 3.

C. Hit Recognition and Score Estimation

Potential hits are detected using the circle Hough transform algorithm. Each potential hit is classified by a Support Vector Machine to reduce the number of false positives. A hit yields the score of the corresponding score ring if it is contained inside the ring or breaks the ring (outer border).

VI. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method was evaluated experimentally on 15 digitalized paper targets with 65 actual hits. The photographs of the paper targets were taken in varying lighting conditions and with slight variation in the angle. The paper targets also

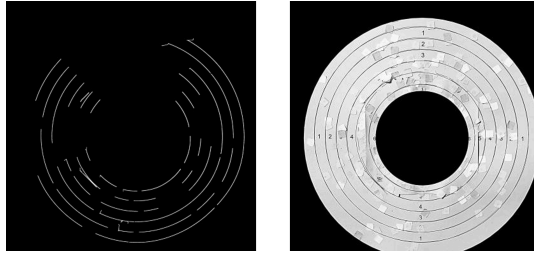


Figure 3. Intermediate result of score ring estimation.

Table I
PAPER TARGET IMAGE QUALITY CATEGORIES.

Category	Description
1	High brightness, high quality paper target
2	High brightness, low quality paper target
3	Low brightness, low quality paper target
4	Low brightness, high quality paper target

varied in quality. These variations were introduced to create more realistic conditions. The 15 targets were manually inspected and classified into four categories of quality according to Table I. The method was able to detect 210 objects out of which 41 represented correctly classified hits and 169 represented false positives. Image quality category 4 yielded the highest number of detected actual hits (88%). However, 71% of the detected objects represented false positives. Image quality category 1 yielded, in comparison, yielded less detected actual hits (72%) but also lower number of false positives (68%). Image quality category 3 yielded the poorest results with a low number of detected actual hits (40%) and a high number of false positives (85%). The score estimation was evaluated by inspecting the locations of detected hits in relation to the actual score rings. The actual scores were then compared to the scores estimated from judging the detected hits in relation to the estimated score rings. The score estimation method managed to assign the correct score for 90% of the detected hits. Out of 15 targets, 10 targets yielded high precision in score estimation. 2 targets yielded low precision and 3 targets yielded no score estimation since the score rings could not be estimated. The precision of the score estimation was to some degree affected by false positive score ring contours. The summarized results for each image category are presented in Table II. The proposed method is currently only suitable for training purposes for amateur shooters interested in digital storage of results and performance³. Ongoing and future work will employ supervised learning techniques to increase accuracy.

VII. CONCLUSIONS

We propose a method for automatic hit detection. The method is based on basic image processing approaches, including circle Hough transform and image classification. The purpose of this work is to contribute to solutions for automatic

Table II
SUMMARIZED RESULTS PER IMAGE CATEGORY.

a	b	c	d	e	f	g	h
1	2	11	8	0.72	17	0.32	100
2	6	24	16	0.66	74	0.21	100
3	5	21	9	0.40	53	0.15	0.55
4	2	9	8	0.88	25	0.29	100

^aImage category, ^bNumber of targets.

^cNumber of actual hits (PI).

^dNumber of correct detections (TP).

^eRecall: $\frac{TP}{PI}$, ^fFalse positives, ^gPrecision: $\frac{TP}{TP+FP}$.

^hScore estimation. This represents the number of correctly estimated scores in relation to the number of correctly identified hits. A result of 100% hence does not imply that all score circle zones were detected with 100% precision.

detection and score calculating in the domain of precision pistol shooting. However, the general approach proposed can be transferred to other shooting disciplines with minor modifications. The proposed method is able to detect up to 88% of shots in high-quality paper targets and up to 40% of shots in low-quality paper targets. In the latter category of target, the false positive rate is high. Currently, the method is suitable for generating a draft digital version of the target and hits. This draft can be modified by the user (by moving, deleting, or adding detected hits). Future work involves improving the detection method to strive for a fully automated solution and to include shot grouping analysis to support learning during practice.

REFERENCES

- [1] M. Gowda, A. Dhekne, S. Shen, R. R. Choudhury, L. Yang, S. Golwalkar, and A. Essanian, "Bringing iot to sports analytics," in *14th USENIX Symposium on Networked Systems Design and Implementation*, 2017.
- [2] A. Bergman, "Elektroniskt skyttesystem," Master's thesis, Karlstad University, 2016.
- [3] F. Ali and A. B. Mansoor, "Computer vision based automatic scoring of shooting targets," in *IEEE International Multitopic Conference*, 2008, pp. 515–519.
- [4] J. Rudzinski and M. Luckner, "Low-cost computer vision based automatic scoring of shooting targets," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, 2012, pp. 185–195.
- [5] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

³Code and data: <https://github.com/stenhager/HitPointDetection>

Machine Learning Computational Fluid Dynamics

Ali Usman
EISLAB Machine Learning
Luleå University of Technology
Luleå, Sweden
ali.usman@ltu.se

Muhammad Rafiq
Data Science Lab
Yeungnam University
Gyeongsan-si, South Korea
rafiq@ynu.ac.kr

Muhammad Saeed
Mechanical Engineering Department
Khalifa University of Science and Tech
Abu Dhabi, United Arab Emirates
muhammad.saeed1@ku.ac.ae

Ali Nauman
WINLab
Yeungnam University
Gyeongsan-si, South Korea
anauman@ynu.ac.kr

Andreas Almqvist
Division of Machine Element
Luleå Tekniska Universitet
Luleå, Sweden
andreas.almqvist@ltu.se

Marcus Liwicki
EISLAB Machine Learning
Luleå University of Technology
Luleå, Sweden
marcus.liwicki@ltu.se

Abstract— Numerical simulation of fluid flow is a significant research concern during the design process of a machine component that experiences fluid-structure interaction (FSI). State-of-the-art in traditional computational fluid dynamics (CFD) has made CFD reach a relative perfection level during the last couple of decades. However, the accuracy of CFD is highly dependent on mesh size; therefore, the computational cost depends on resolving the minor feature. The computational complexity grows even further when there are multiple physics and scales involved making the approach time-consuming. In contrast, machine learning (ML) has shown a highly encouraging capacity to forecast solutions for partial differential equations. A trained neural network has offered to make accurate approximations instantaneously compared with conventional simulation procedures. This study presents transient fluid flow prediction past a fully immersed body as an integral part of the ML-CFD project. MLCFD is a hybrid approach that involves initialising the CFD simulation domain with a solution forecasted by an ML model to achieve fast convergence in traditional CFD. Initial results are highly encouraging, and the entire time-based series of fluid patterns past the immersed structure is forecasted using a deep learning algorithm. Prepared results show a strong agreement compared with fluid flow simulation performed utilising CFD.

Keywords—Machine learning, fluid-structure interaction, computational fluid dynamics, numerical analyses, flow past a cylinder

I. INTRODUCTION

Multiphysics involved in fluid-structure analyses creates time and computational efficiency issues related to design approaches based on conventional numerical simulations. Such approaches typically discretise the governing equations in time and space, solve them using iterative numerical schemes, and computational time increases exponentially with an increase in mesh density. Similar to CFD and FSI, discretised numerical computations are involved in numerous engineering fields, e.g., electric- and electromagnetic- fields, acoustics, chemical species transport, fluid flow, heat transfer, optics, semiconductors, tribology and structural mechanics.

Recently, Berg and Nystrom [1] used deep feedforward artificial neural network (ANN) and unconstrained gradient-based optimisation to approximate partial differential equation (PDE). They showed an example where classical mesh-based methods could not be used, and neural network is an attractive alternative. Similarly, Long et al. used feedforward ANN to approximate a PDE solution and learned the PDE under consideration in an inverse approach [2]. They showed that the ability of ANN to learn the entire family of PDE based on available data is remarkable. Nabian and Meidani enforced initial and boundary conditions during the feedforward ANN training, and the loss function were optimised for randomly selected spatial points during each iteration [3]. This makes the outcome of ANN so-called mesh-less. Sun et al. proposed an approach for extracting governing PDEs for dynamic datasets fed to an ANN-based model [4].

Most recently, Li et al. [5] introduced a Fourier Neural Operator (FNO) to learn the entire family of parametric PDEs, including the Navier–Stokes equations. This work also presents ingeniously interpreting results between discrete mesh points. The method was shown to outperform traditional methods to solve PDEs in terms of error. Moreover, CPU-time required to get an output from a neural network for varying initial conditions is far lesser than that required by traditional numerical approaches, for instance, finite element method, to solve PDEs. Similarly, a study has reported successful ML utilisation to classify 2D surface pressure data [6]. The 2D surface data can be viewed as spatial and temporal variations in the output of a multivariable function similar to PDE output. This study shows the capabilities of ML to study a domain that is not directly considered eligible for transfer learning.

The Navier–Stokes equations are utilised extensively to estimate fluid-structure interaction in machine elements, for instance, air and blade interaction in a wind turbine. Therefore, the successful utilisation of ML to approximate and forecast fluid vortices and eddies, e.g., [7], is a proof of concept for the present study. Shortly, we foresee that computational mechanics involved in engineering design will rely heavily on ML because of the widespread benefit offered by ML, including time and efficiency issues related to the traditional design approach. Moreover, ML algorithms also suggest efficient strategies to reduce three-dimensional data to the planar transport network, such as turbulent CFD superstructures [8]. Knowledge gap: even though ANN-based physics-informed and data-driven models have been shown to accurately approximate and instantaneously forecast multivariable functions, the application of such models is

rarely explored toward developing an alternative framework to perform a computational hybrid ML-CFD approach in particular and computational engineering physics in general.

This study evaluates the utilisation of FNO [5] for FSI between incompressible fluid and a circular immersed body. A 2D FSI is computed utilising a commercially available CFD package. A comprehensive dataset is generated to train the FNO for varying boundary conditions at the inlet of the fluid domain and corresponding fluid dynamics developing in the computational domain the passage of time.

II. MODELLING

A. Computational fluid dynamic model

The physical description and the applied boundary conditions for the current work are shown in Fig. 1. The computational domain size is $30D \times 30D$, where D is the diameter of the cylinder. The Inlet boundary is placed $10D$ upstream of the cylinder while the outlet is positioned at $20D$ downstream. The distance of the side walls is maintained at $15D$ on each side of the cylinder. The structured mesh was generated using hexahedral elements, and an O-grid was introduced around the cylinder walls to ensure high-quality mesh in the vicinity of the boundary layer thickness providing 50 nodes within the boundary layer region. Mesh topology and node distribution near the cylinder wall are displayed in Fig. 1.

$$\frac{\partial}{\partial x}(\rho \mathbf{v}) = 0 \quad (1)$$

$$\boldsymbol{\sigma} = -p\mathbf{I} + \mu(T)(\nabla \mathbf{v} + \nabla \mathbf{v}^T) \quad (2)$$

$$\rho(D\mathbf{v}/Dt) = \text{div}\boldsymbol{\sigma} \quad (3)$$

Transient and incompressible forms of the continuity and Navier-Stokes equations (Eq. 1 – 3) [9] are considered where $\boldsymbol{\sigma}$ is Cauchy stress, and are solved using a commercial software ANSYS-CFX. The normal velocity condition is imposed at the inlet boundary, and average pressure is specified at the outlet boundary. At the same time, the side walls are assigned with no-slip conditions. The time step and

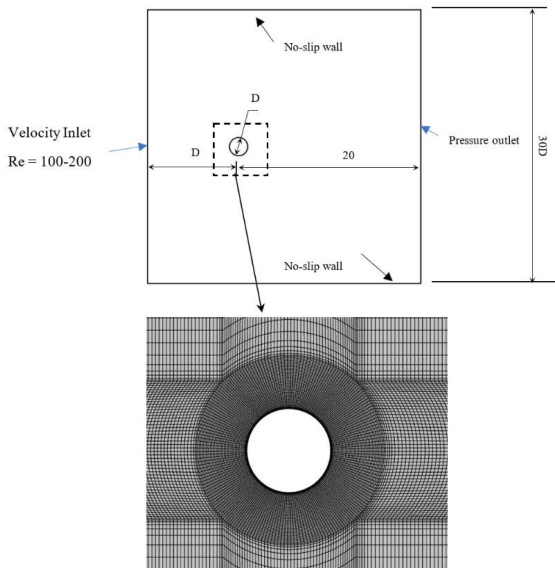


Fig. 1 Computational domain and boundary conditions (top), and mesh around the cylinder(bottom)

the total simulation time of 0.001s and 1s, respectively, were used for all calculations performed within the present study.

TABLE I. THE DATASET: FIRST FEW COEFFICIENT OF PRESSURES (CoP) FOR A TRANSIENT CFD AT VARYING REYNOLD NUMBERS. THE BLUE COLOUR REPRESENTS THE LOWEST CoP, AND RED REPRESENTS THE HIGHEST CoP FOR THE ENTIRE DATASET.

Re	t	t + 5T	t + 10T
100			
110			
120			
130			
140			
150			
160			
170			
180			
190			
200			

B. The dataset

A complete dataset is developed by numerical simulations of the aforementioned governing equations on the mesh shown in Fig.1b. The developing fluid flow pattern in the computational domain is divided into 1000 timesteps to capture detailed dynamics. The Reynolds number (Re) at the inlet of the boundary is varied from 100 to 200 with an increment of 10. Hence, a comprehensive dataset comprising 11000 distinct pressure field corresponding to each time step was created. The Coefficient of pressure at each timestep is non-dimensionalised by utilising all-time maximum CoP at Re = 200. The data for varying Reynolds number is shown in Table I.

C. The deep learning model

An FNO for the Navier–Stokes Equations [5] was adopted when designing the neural network architecture. This network consists of four spectral convolution blocks and two fully connected layers with 128 neurons each, and the architecture is shown in Fig. 2.

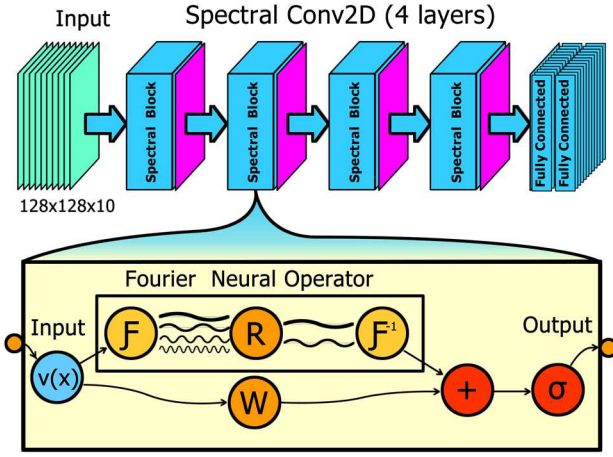


Fig. 2. Deep learning model based on Fourier Neural Operator (a part of this figure is adopted from [5] to illustrate the data flow).

a) *Input*: The model accepts a sequence of 10 images as a single input in the form of a 3D array; every single input is with a shape of 128x128x10. We employ an image sequence generated from the CDF system, as explained in the dataset section. The original sequence is processed to create input blocks with ten images in each sequence at a particular time t .

$$\text{Input_array}[t] = \text{ImageSequence}[t:t + S] \quad (3)$$

where t represent a particular input time, and S represents an input sequence size for one block, which is in our case, is set to 10.

b) *Spectral Convolutional Block*: A spectral convolution block consists of two streams. The FNO converts the input into the frequency domain and applies a low-pass filter to it, meaning that only the low-frequency content is considered in the multiplication between the Fourier transformed input and the systems Greens' function. Finally, the processed data is converted back to the existing domain. The weights used in the network to represent the spectral blocks are then combined with the weights that directly scale the input, as

illustrated in Fig. 2. An activation function is finally applied to obtain the block-wise output.

III. NUMERICAL EXPERIMENTATION

The flow of data and training method utilised for the present study is shown in Fig. 3. Data preprocessing is employed before further the data to FNO model. This includes the resizing, centering, clipping, and aligning of the input image sequence. The preprocessed images are further processed to generate input sequenced blocks to match the model input shape of 128x128x10. The prepared data is split into a training- and a test set sequences. The network is trained using a training loop, which employs loading the next batch, compute the loss function, updates optimiser and backpropagation derivatives for parameter updates. The model is saved at every best validation loss, and in case loss is not further improving, an early stop mechanism ends the training to avoid overfitting with the best weights in hand. Loss function for the training and test splits are shown in Fig. 4.

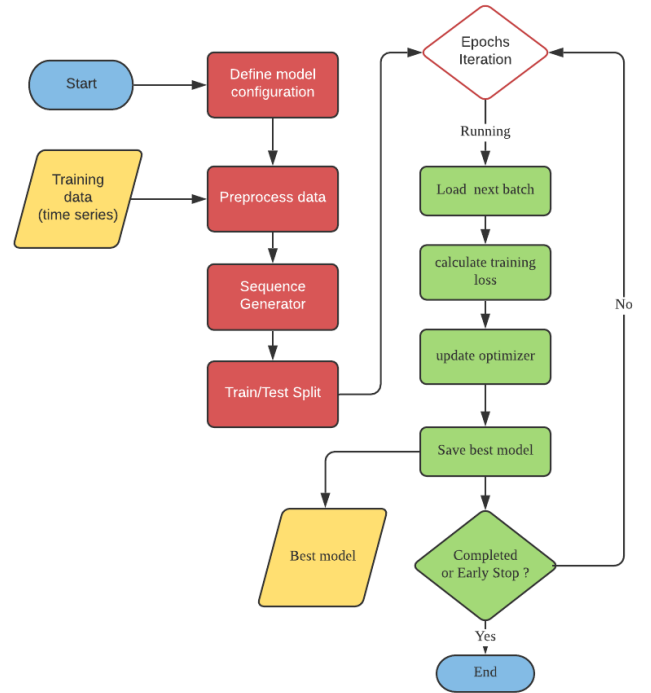


Figure 3. Illustration of model implementation and training procedure utilised in this study.

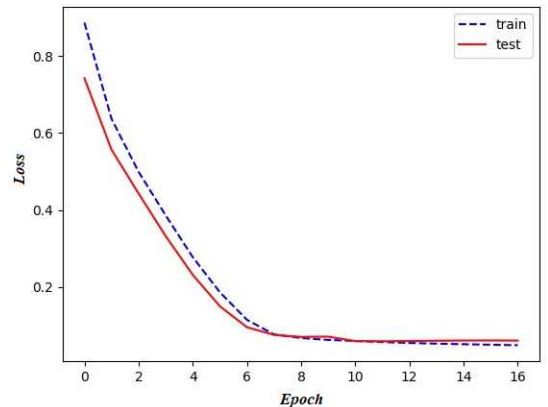
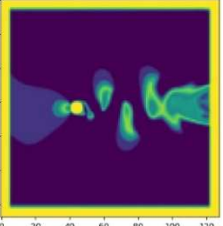
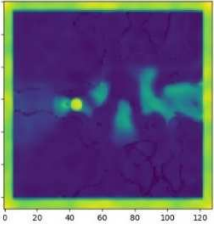
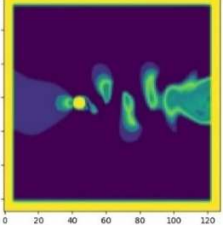
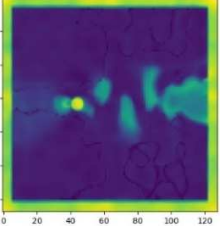
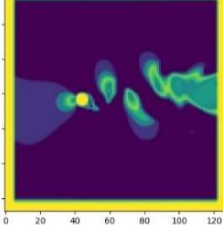
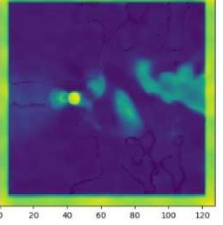
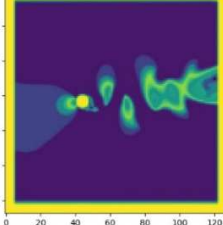
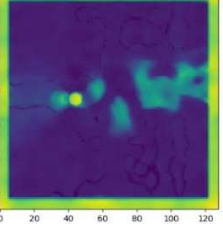
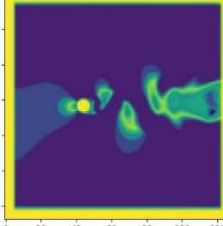
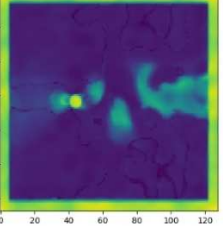


Fig. 4. Loss function value for training and test splits of the dataset.

TABLE II. PREDICTION OF DEEP LEARNING MODEL (DLP) COMPARED WITH THE GROUND TRUTH (GT) BASED ON THE TRADITIONAL NUMERICAL APPROACH OF COMPUTATIONAL FL.

Time	Ground Truth	Prepared
t (s)		
Pixel to Pixel Comparison = 1.303711%		
t + 20 (ms)		
Pixel to Pixel Comparison = 1.227539%		
t + 40 (ms)		
Pixel to Pixel Comparison = 1.218750%		
t + 60 (ms)		
Pixel to Pixel Comparison = 1.290039%		
t + 80 (ms)		
Pixel to Pixel Comparison = 1.361328%		

IV. RESULTS AND DISCUSSION

The output sequence is generated given the input development of the solution. CoP solutions at selected instances are shown in Table II. Although there are apparent visual discrepancies between the DLM prediction and the GT, they agree qualitatively. The pixel-to-pixel binary image comparison with a threshold pixel value of 125, the solution prepared while utilising the aforementioned deep learning model, are in high agreement with the solution generated through traditional CFD computations. The model, remarkably, captured vortices generated in fluid past the

cylindrical structure. Here, it is noteworthy that the model forecasts the complete transient solution. Given the initial ten solutions at the beginning of the fluid-structure interaction, the following ten solutions are purely predicted. The solution at the earliest instant, i.e., 1st in the generated sequence, replaces the latest solution in the input sequence. This predict-and-replace process is continued, and after 11 iterations, the complete input sequence has been replaced with the predicted series of solutions. All the iterations from the 12th and until the 1000th are purely based on the predicted sequence of PDE solutions. Since it is an ongoing project and the results presented here are initial developments, the prepared results show diverged solution at some specific grid points. We believe this problem could be addressed by optimising the network architecture, which is our future goal in order to achieve state-of-the-art accuracy. Presently, the learned model is useful to the configuration of the flow field as presented in Table I with the circular body fully immerses in the fluid field; however, we are generating datasets with varying shapes of bodies with varying orientations to make predictions widely generalised.

ACKNOWLEDGMENT

The Swedish Kempe Foundation has funded parts of this work. The authors would also like to acknowledge the support from the Swedish Research Council: DNR 2019-04293.

REFERENCES

- [1] J. Berg and K. Nyström, "A unified deep artificial neural network approach to partial differential equations in complex geometries," *Neurocomputing*, vol. 317, pp. 28-41, 2018/11/23/ 2018, doi: <https://doi.org/10.1016/j.neucom.2018.06.056>.
- [2] Z. Long, Y. Lu, X. Ma, and B. Dong, "PDE-Net: Learning PDEs from Data," presented at the Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research, 2018. [Online]. Available: <http://proceedings.mlr.press>.
- [3] M. A. Nabian and H. Meidani, "A deep learning solution approach for high-dimensional random differential equations," *Probabilistic Engineering Mechanics*, vol. 57, pp. 14-25, 2019/07/01/ 2019, doi: <https://doi.org/10.1016/j.probenmech.2019.05.001>.
- [4] Y. Sun, L. Zhang, and H. Schaeffer, "NeUPDE: Neural Network Based Ordinary and Partial Differential Equations for Modeling Time-Dependent Data," *arXiv pre-print server*, 2019-08-08 2019, arxiv:1908.03190.
- [5] Z. Li *et al.*, "Fourier Neural Operator for Parametric Partial Differential Equations," *arXiv pre-print server*, 2020-10-18 2020, arxiv:2010.08895.
- [6] Monit, V. Pondenkandath, B. Zhou, P. Lukowicz, and M. Liwicki, "Transforming Sensor Data to the Image Domain for Deep Learning - an Application to Footstep Detection," *arXiv pre-print server*, 2017-07-14 2017, arxiv:1701.01077.
- [7] D. Kochkov, Jamie, A. Alieva, Q. Wang, Michael, and S. Hoyer, "Machine learning accelerated computational fluid dynamics," *arXiv pre-print server*, 2021-01-28 2021, arxiv:2102.01010.
- [8] S. Pandey, J. Schumacher, and K. R. Sreenivasan, "A perspective on machine learning in turbulent flows," *Journal of Turbulence*, vol. 21, no. 9-10, pp. 567-584, 2020, doi: 10.1080/14685248.2020.1757685.
- [9] A. Almqvist, E. Burtseva, K. Rajagopal, and P. Wall, "On lower-dimensional models in lubrication, Part A: Common misinterpretations and incorrect usage of the Reynolds equation," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, p. 135065012097379, 2020, doi: 10.1177/1350650120973792.

Smart Sewage Water Management and Data Forecast

Qinghua Wang¹, Viktor Westlund¹, Jonas Johansson¹, Magnus Lindgren²

Abstract—There is currently an ongoing digital transformation for sewage and wastewater management. By automating data collection and enabling remote monitoring, we will not only be able to save abundant human resources but also enabling predictive maintenance which is based on big data analytics. This paper presents a smart sewage water management system which is currently under development in southern Sweden. Real-time data can be collected from over 500 sensors which have already been partially deployed. Preliminary data analysis shows that we can build statistical data models for ground water, rainfall, and sewage water flows, and use those models for data forecast and anomaly detection.

Index Terms—Anomaly detection; Big data; Internet of things; Sewage water management

I. INTRODUCTION

Sewage and wastewater systems are critical to a functioning society, and they are facing an ever increasing challenge from the climate change and fast urbanization. In the recent years, we have seen an increasing number of floods around the world. We have also experienced a fast population growth and urbanization mainly in developing countries, but also in developed countries due to mass immigration and the process that is called reurbanisation [1]. In southern Sweden, the threats from climate change have been critical due to the rising sea levels and the increasing risk of flooding [2]. Meanwhile, we have cities that have been built for several hundred years and their sewage and wastewater systems need to be updated to accommodate nowadays population and climate challenges. To ensure the access and sustainable management of water and sanitation for all is one of the key goals for sustainable development that is coined in the United Nations' 2030 Agenda [3].

Kristianstad is a city that is located in the Scania County, which is the southernmost county in Sweden. The city is located below sea level and has high groundwater, which easily cause floods that often result in major damages to both property owners and communities. The floods can overwhelm the capacity of sewage and wastewater systems, which may eventually lead to the disposal of waste water in environmentally sensitive areas, public bathing areas or canals that are close to the city. These problems are not unique to Kristianstad but occur throughout Sweden.

*This work is supported by the strategic innovation programme IoT Sweden, which is a joint effort by Vinnova, the Swedish research council Formas and the Swedish Energy Agency.

¹Q. Wang, V. Westlund and J. Johansson is with the Department of Computer Science, Faculty of Natural Sciences, Kristianstad University, SE-29188 Kristianstad, Sweden. qinghua.wang at hkr.se

²M. Lindgren is with Kristianstad Municipality, Sweden. magnus.lindgren at kristianstad.se

Digitalization has the potential to revolutionize the management of sewage and wastewater systems, and provide us tools to address the challenges raised by population growth and climate change. Digital solutions can make data acquisition and data analysis more efficient, and lead to cost savings and optimization of management processes. Kristianstad Municipality is a pioneer in applying the digital transformation in its sewage and wastewater systems. Quite recently, Kristianstad Municipality, together with Kristianstad University and C4 Elnät AB were granted a research project on smart real time monitoring of sewage water systems with data analysis. This project is implemented within the strategic innovation program IoT Sweden [4].

According to the project plan, Kristianstad Municipality will place 550 sensors around the entire municipality to build a network that can report what is happening in the municipality's water supply and sewerage network in real time. All these data will be centrally managed in a data portal and will then be made visible to residents. Data acquisition forms the basis in the municipality's action plans to reduce flooding and to reduce the volume of added water in wastewater drainage systems. Besides data acquisition, machine learning-based algorithms will also be developed to process data from all the sensors. The algorithms will be able to detect water leakage and other anomalies, make predictions for future trends on the sensor readings, and trigger alarms when there are storms and other environmental hazards. The details of the project implementation and some preliminary results are presented in the following sections in this paper.

II. RELATED WORKS

A study released from Uppsala University was conducted on municipalities in Sweden and explored if leaking drink water pipes could be a significant source for additional water in sewage pipes. The machine learning data analysis was made using simple linear regression and multiple linear regression analysis on data collected from the municipalities. The conclusion of the study found that there was a correlation between leakages in drink water pipes and additional water entering sewage treatment plants shown by the linear regression [5].

In the sub-tropical climate of Hong Kong that usually experiences heavy rainfalls and risk of the continuous rising of the ocean, a study was conducted on potentially implementing a real-time monitoring system in drainage pipes using IoT-devices. The goal of the study was to create a smart drainage system using sensors and machine learning to be able to prevent potential flooding. Sensors collected water-level and water-flow data which was then processed using artificial neural network and cross-validation. The conclusion of the

study showed that the use of Artificial Neural Network could both benefit new city planning as well as predicting any potential flood in real-time based on the collected data [6].

A research comparing different AI-methods for analyzing Inflow and Infiltration in Sewer Sub-catchments was released in August 2020. The aim of the research was to develop an AI-based system to analyze real-time Inflow and Infiltration data. The two AI-methods that was developed and tested in the study involved an Adaptive Neuro-Fuzzy Inference System (ANFIS) and a Multi-layer Perceptron Neural Network (MLPNN). The conclusion of the study was that the ANFIS-method showed an overall better performance [7].

A literature study made on effects from Inflow and Infiltration in sewage pipes showed that there are some positives to come from having an inflow. These are for example less need for chemicals to treat the water, better control of the groundwater level, higher velocities in pipes to generate a self-cleaning effect, and more inflow water helps to keep the smell down. The negative effects found contains of for example, less efficient water treatment in wastewater treatment plant, requires pipes to be bigger to handle more water, higher energy consumption, higher risk for flooding and blockage, and pipes will age quicker. The conclusion was that even though there might be positives to I/I, the overall effects must be seen as negative [8].

III. METHODOLOGY

A. Study Area



Fig. 1. Kristianstad and a part of Näsby where it is clear how surrounded by water the city is.

The town of Kristianstad (Fig. 1) is built in the middle of a wetland which has its drawbacks. The city is very sensitive to high water levels as the city is located 2.41 m.u.h and Helge river runs straight through. Embankment pumping stations have been built to enable the city to withstand high water levels. The city's embankments are under development and are also being investigated if the existing ramparts need to be raised to cope with future sea level rise. This makes the city very sensitive to future climate change and part of this work is to install smart meters. For the preliminary study in this work, the investigation has been limited to the Näsby region which is the neighborhood where Kristianstad University is located.

Fig. 2 shows the map of the Näsby region. Fig. 3 shows the sewage water pipe network for the same region.



Fig. 2. Map of Näsby in Kristianstad.



Fig. 3. Sewage water pipe system in Näsby.

B. Data Acquisition

The project plans to install 550 different types of sensors all over Kristianstad. At the moment, there is already an installation of 15 ground water level sensors, 12 rain sensors, and 121 sewage water flow sensors. Fig. 4 shows the picture of one of the installed ground water level sensors. All sensors are enabled with wireless communication capability via the LoRaWAN network that is owned by C4 Elnät (who is also a project partner). All sensor data are then assembled and visualized via an IoT data portal. The data could be accessed remotely using scripts to allow real time processing and data analysis.

C. Data Modelling

The additional water that are added to the sewage water systems due to rain, surface drainage, and penetrated ground water, etc., add extra burden to the sewage water systems, and challenges the capacity of both sewage water pipes and the wastewater treatment plants. Therefore, it is important to study the composition of sewage water and how its volume is affected by different factors. In this work, we use a multi-linear regression model to map the relationship between the



Fig. 4. A ground water level sensor with wireless communication.

observed sewage water volumetric flow rate and other sensor data, and use this model to make predictions to the future burden of sewage water systems.

$$\vec{y} = A_0 + A_1 \times \vec{x}_1 + A_2 \times \vec{x}_2 + A_3 \times \vec{x}_3 + \dots + \vec{\epsilon} \quad (1)$$

The multi-linear regression model is formulated in (1). In (1), \vec{y} is the variable to be predicted and it is the future sewage water volumetric flow rate in our case, \vec{x}_i is the data vector sampled by sensor i , $\vec{\epsilon}$ is the error vector, and A_i (for $i = 0, 1, 2, \dots$) are the model parameters.

IV. RESULTS

With linear models, historic data can and should exhibit high correlation with future trend. A flow sensor that is located at the root of the pipe network is selected to allow the observation of a large volume of flow. For the flow data, we have chosen three consecutive weeks that had the same number of rows of data so that we would not have to fill in any missing data which can lead to inaccuracies. Fig. 5 shows the correlation between the flow data from the same sensor but observed in two separate weeks. Fig. 6 shows the predicted flow rate based only on the historical flow rate. The prediction is done on a window of 24 hours and it has a relatively high accuracy.

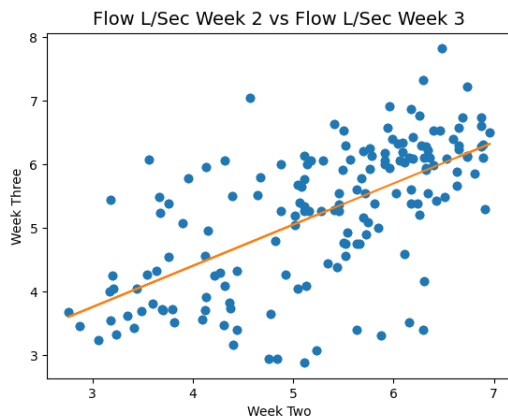


Fig. 5. Scatter diagram of flow L/sec in two consecutive weeks.

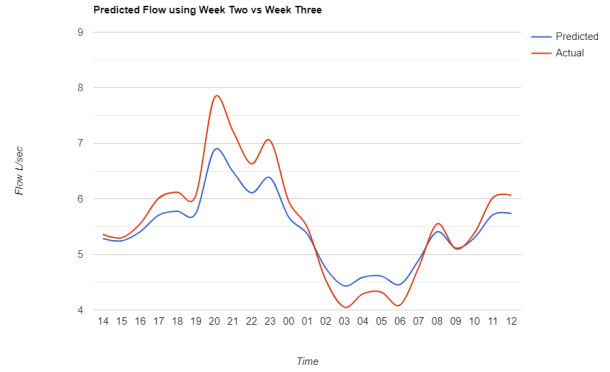


Fig. 6. Predicted flow rate based on historical flow data vs. actual flow rate.

In this study, we have also real-time data from ground water sensors and rain sensors, and historical data about the hourly water consumption pattern. We use these data to predict the volumetric flow rate. Historical data from the last two months are used to train the model. Fig. 7 shows a graph with two curves, with one representing the predicted flow rate from the model and the other showing actual measurements from the flow sensor.



Fig. 7. Predicted flow rate based on data from non-flow sensors vs. actual flow rate.

Fig. 8, 9, and 10 show the correlations between flow data and observations in the other types of sensors. It can be seen that the precipitation data has a high correlation to the flow data, while the ground water level and water usage pattern has relatively lower correlation with the flow data. But this needs to be further investigated.

In the future, it is possible to create a multi-linear regression model to forecast the trends of different sensor readings not only with their own historical data but also with data from other sensors. The model parameters should also be repeatedly trained with the newest data. Cross validation on the prediction results can also be done based on outputs from different models. Anomalies on sensor readings can be detected by comparing the actual data with the predicted data. Anomalies on sensor readings can be coupled to sensor faults, but can

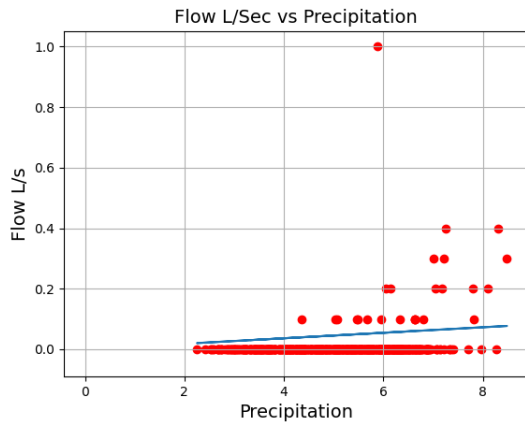


Fig. 8. Scatter diagram of flow L/sec vs precipitation.

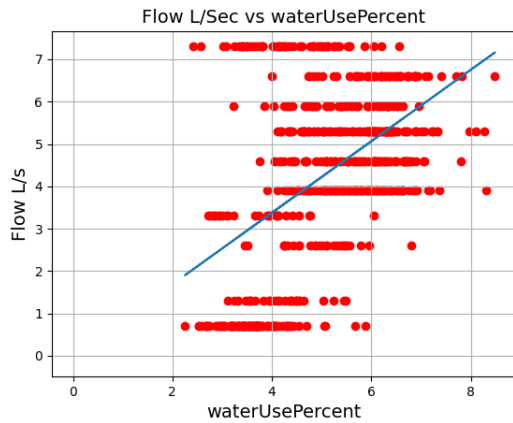


Fig. 9. Scatter diagram of flow L/sec vs water usage.

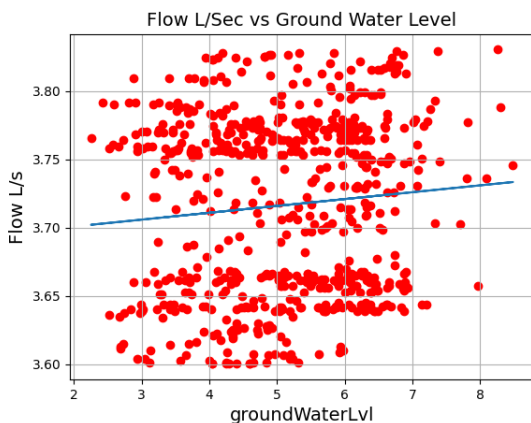


Fig. 10. Scatter diagram of flow L/sec vs Groundwater.

also be due to anomalies at other parts of the system (e.g. leakage).

V. CONCLUSION

In this paper, we have presented a sewage water monitoring project. The project involves the deployment of different types of sensors in the city Kristianstad in southern Sweden to give the decision makers and water engineers a live and comprehensive picture on water situation. A headache in the sewage water systems is the additional water from rain and ground water. Our preliminary study shows that there is correlation between precipitation, groundwater and the water in sewage pipes. We have built a forecast model to the future sewage flow rate using historical flow data, and also data from groundwater, precipitation and household water usage pattern. It is possible to improve the model by integrating more data sources. Once we have gotten a high accuracy forecast model, it is possible to detect anomalies in sewage pipe systems such as water leakage. Decision makers can also rely on data forecast to make important decisions ahead of time to avoid environment hazards, such as drainage of untreated wastewater to the wild.

ACKNOWLEDGMENT

The authors would like to thank Mats Wemmenborn (Kristianstad Municipality) for coordination of this project, and to thank Hans Inge Hansson (Kristianstad Municipality), Johan Holmberg (Atea), Jonas Månsson (Kristianstad Municipality) and many others for assistance of data access and explanation of technical details.

REFERENCES

- [1] N. Kabisch, D. Haase, and A. Haase, "Reurbanisation: A long-term process or a short-term stage?" in *Population, Space and Place*, vol. 25, no. 8, Nov. 2019.
- [2] J. Sjögren, "Consequences of a flood in Kristianstad, Sweden — A GIS-based analysis of impacts on important societal functions," Master degree thesis, Dept. of Physical Geography and Ecosystem Science, Lund University, 2017.
- [3] Transforming our world: the 2030 Agenda for Sustainable Development, United Nations, available at: sdgs.un.org/2030agenda.
- [4] IoT Sweden, available at: iotsverige.se.
- [5] A. Ringqvist, "Utläckage från vattennät – en betydande källa till tillskottsvatten i spillvattennät: Linjär regressionsanalys av VA-data från svenska kommuner," Thesis, Uppsala University, 2021.
- [6] K. L. Keung, C. K. M. Lee, K. K. H. Ng, and C. K. Yeung, "Smart City Application and Analysis: Real-time Urban Drainage Monitoring by IoT Sensors: A Case Study of Hong Kong," in *2018 IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 521-525, 2018.
- [7] Z. Zhang, T. Laakso, Z. Wang, et al. "Comparative Study of AI-Based Methods — Application of Analyzing Inflow and Infiltration in Sanitary Sewer Subcatchments," *Sustainability*, vol. 12, no. 15, 2020.
- [8] A. Ohlin Saletti, "Infiltration and inflow to wastewater sewer systems - A literature review on risk management and decision support," Technical report, Dept. of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden, 2021.