MGTN

MGTN: MODULAR GRAPH TRANSFORMER NETWORKS

for Multi-Label Image Classification

We propose a multi-label image classification framework based on graph transformer networks to fully exploit inter-label interactions. The paper presents a modular learning scheme to enhance the classification performance by segregating the computational graph into multiple sub-graphs based on the modularity. The proposed approach, named as Modular Graph Transformer Networks (MGTN), is capable of employing multiple backbones for better information propagation over different sub-graphs guided by graph transformers and convolutions. We validate our framework on MSCOCO and Fashion550K datasets to demonstrate massive improvements for multi-label image classification. Source code and data are at https://github.com/ReML-AI/MGTN.

1 Introduction to the model:

- **MGTN** has configurable building blocks to integrate semantic information ***E*** and topological information ***A*** into visual representation learning. It enables information propagation over multiple sub-graphs guided by graph transformer networks.

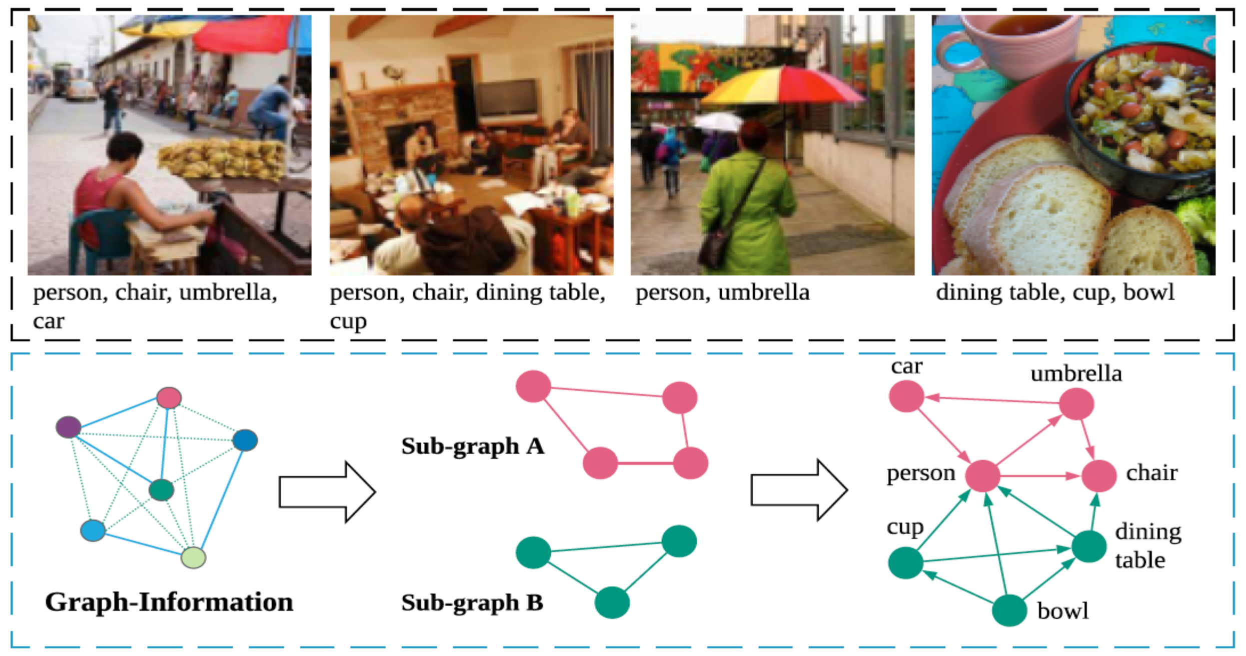

Figure 1: An example of subgraph segregation in which ***"person, chair, umbrella, car"*** and ***"dining table, cup, bowl"*** are in two separate sub-graphs.

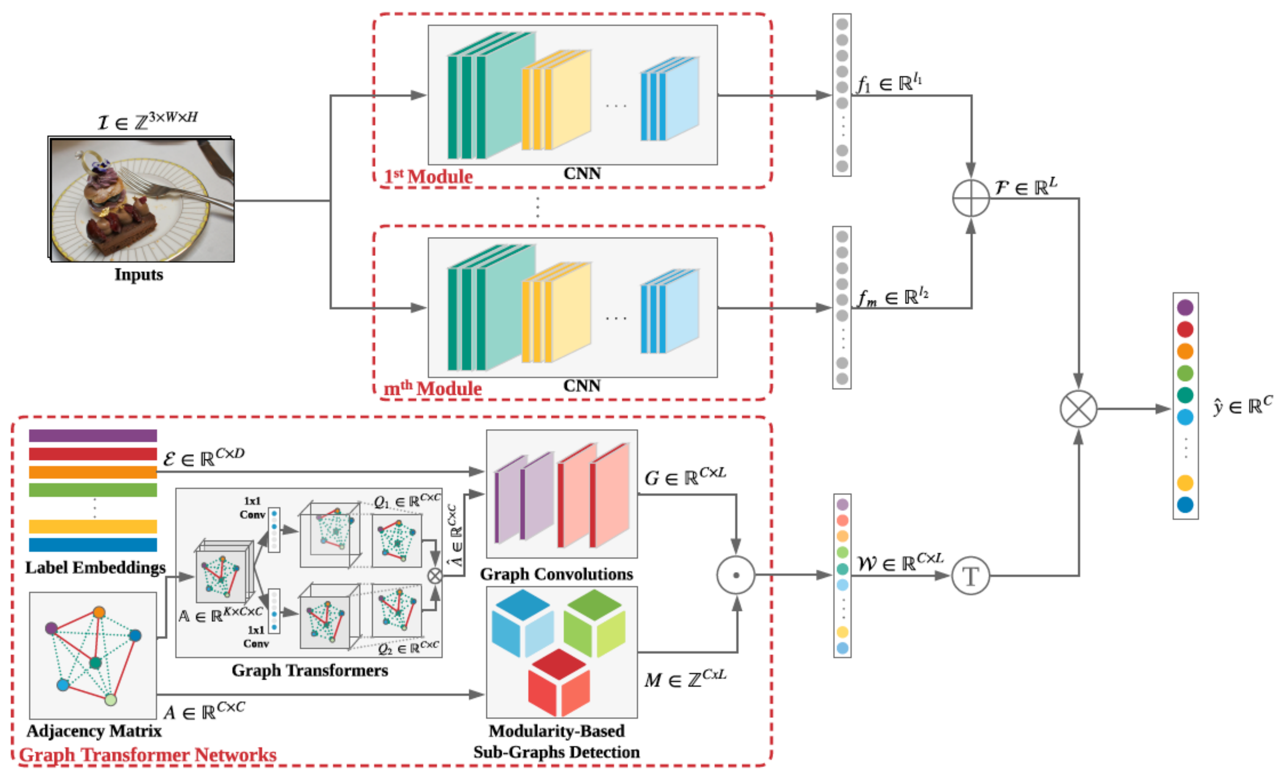

Figure 2: a) Architecture design of Modular Graph Transformer Network (MGTN) support multi-label learning over multiple modules of CNNs for recognising object labels in images.

2 Modularity on MS-COCO:

- We run **Network Analyses** on **MS-COCO** dataset and MGTN's predicted labels on test data. Both analyses reveal the partitions of inter-connected object labels.

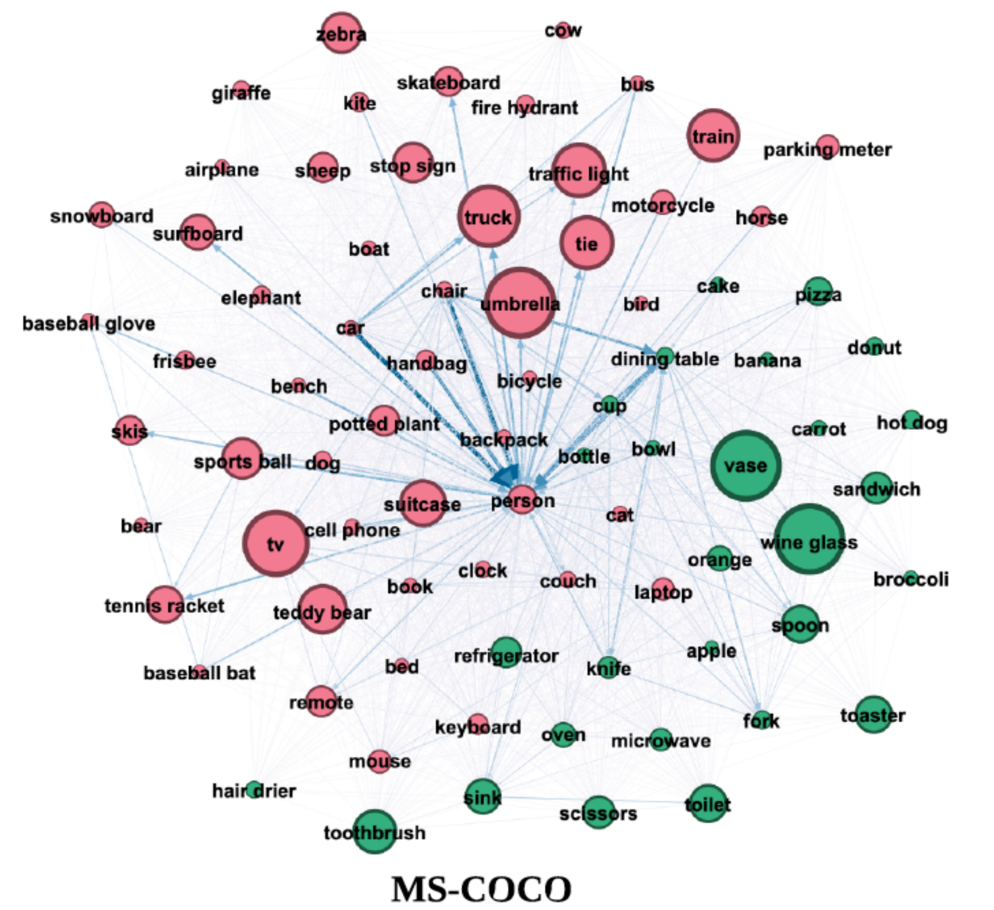

Figure 3: The sizes of the nodes reflect the relative importance of inter-dependent object labels based on the eigenvector centrality measure

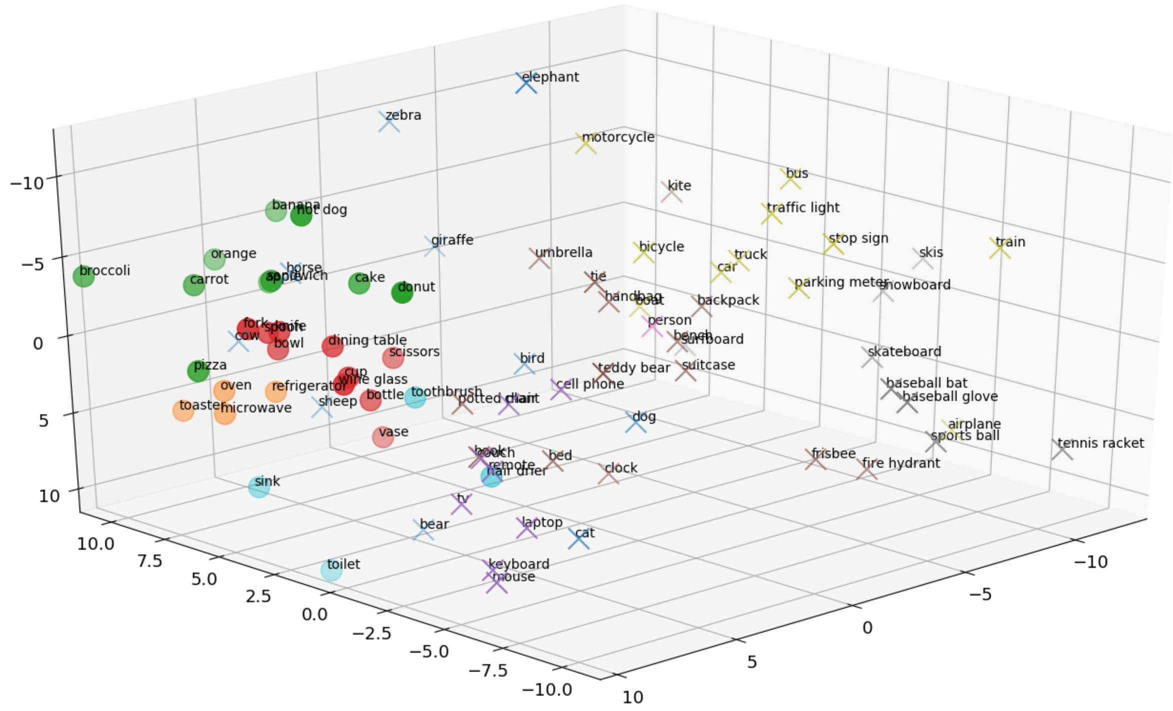

Figure 4: 3D t-SNE visualisation of MGTN's predicted results on the test set of MS-COCO. It shows how good MGTN understands the correlation between labels on unseen images.

3 Performance Evaluation:

- Experiments are exhaustively conducted, and we report the relevant empirical results on two public datasets: MS-COCO and Fashion550K.

METHOD	MAP	CP	CR	CF1
CNN-RNN (WANG ET AL. 2016)	61.2	-	-	-
SRN (ZHU ET AL. 2017)	77.1	81.6	65.4	71.2
BASELINE(RESNET101) (HE ET AL. 2016)	77.3	80.2	66.7	72.8
MULTI-EVIDENCE (GE, YANG, AND YU 2018)	–	80.4	70.2	74.9
ML-GCN (CHEN ET AL. 2019B)	82.4	84.4	71.4	77.4
ML-GCN (RESNEXT50 WITH IMAGENET)	86.2	85.8	77.3	81.3
A-GCN (LI ET AL. 2019)	83.1	84.7	72.3	78.0
KSSNET (WANG ET AL. 2020B)	83.7	84.6	73.2	77.2
SGTN (OUR) (VU ET AL. 2020)	86.6	77.2	**82.2**	79.6
MGTN(BASE)	86.9	**89.4**	74.5	81.3
MGTN(FINAL)	**87.0**	86.1	77.9	**81.8**

Table 1: Performance comparisons on MS-COCO. Our MGTN outperforms all previous approaches with large margins. The multi-learning base model shows significant mAP improvements of 9.4% from the ResNet101 baseline and 3% from KSSNet. The eigenvector based transformation provide MGTN with better learning capabilities with an additional 0.1% in performance.

METHOD	MAP
BASELINE(RESNET50) (INOUE ET AL. 2017)	58.68
STYLENET (SIMO-SERRA AND ISHIKAWA 2016)	53.24
ML-GCN (CHEN ET AL. 2019B)	60.85
A-GCN (LI ET AL. 2019)	61.35
MGTN(FINAL)	**65.10**

Table 2: Performance comparisons on Fashion550K. The results demonstrate the effectiveness of MGTN with significant improvements of 6.4%, 4.2%, and 3.7% in mAP from the baseline(ResNet50), ML-GCN, and A-GCN, respectively.

4 Ablation Study:

- To address that EV-enhancement could help MGTN even learn faster and hence, save more computing power

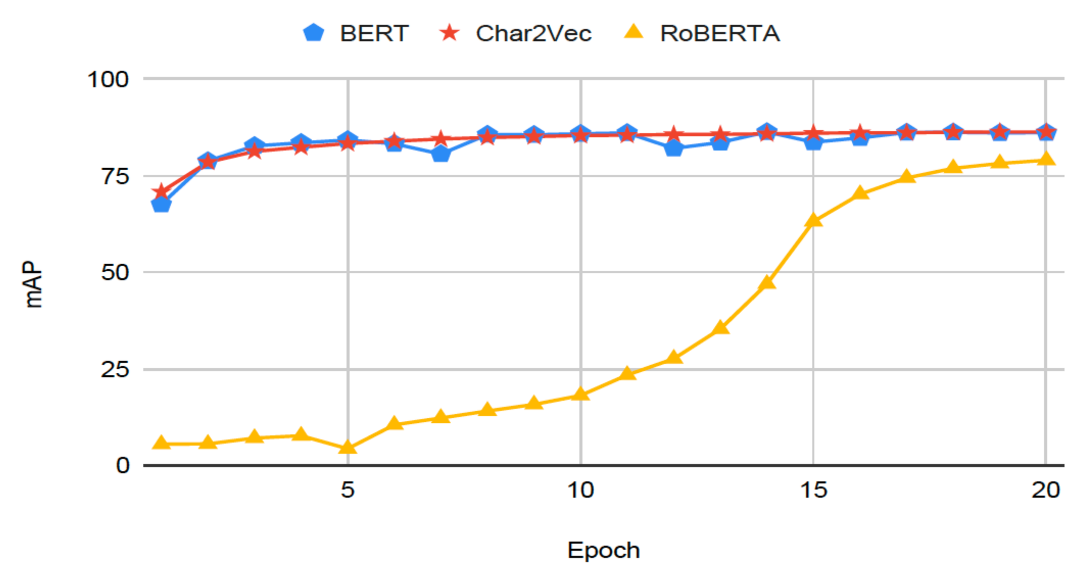

Figure 5: Learning patterns of MGTN with different label embeddings in 20 epochs. The MGTN model with the setting using RoBERTa_{avg_12} label embedding shows a slow learning speed in comparison to others.

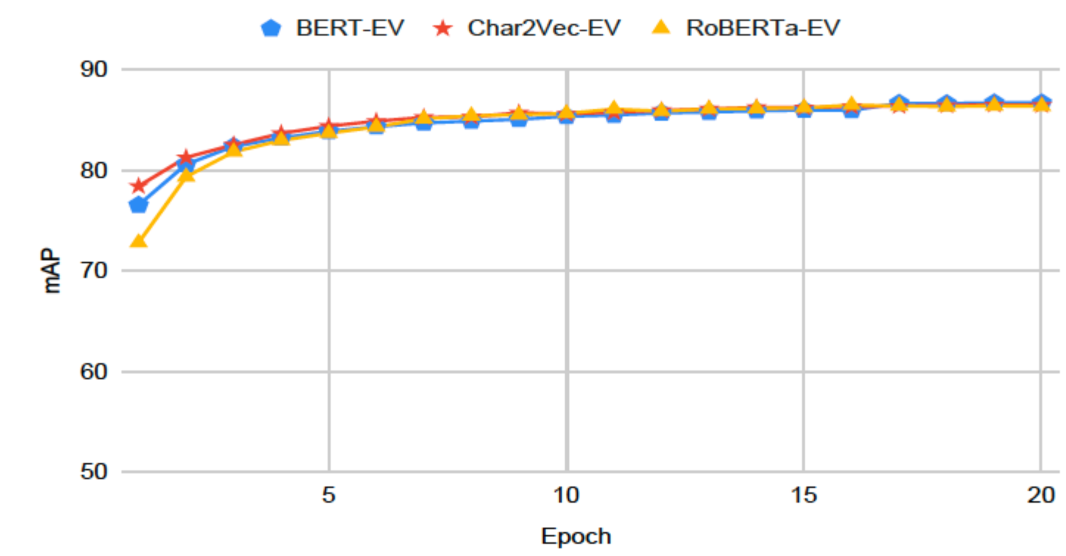

Figure 6: The EV-enhancement for label embedding helps the MGTN's model learn faster, even MGTN with the setting using the $\text{RoBERTa}_{avg\ 12}$ now learns faster. Note: y-axis here is ranged in [50; 100] for visibility.

University of Glasgow

UMEÅ UNIVERSITY

VNU University of Engineering and Technology

Authors: Hoang D. Nguyen (Glasgow), Xuan-Son Vu (UMU), Duc-Trong Le (VNU-UET)

Emails: harry.nguyen@glasgow.ac.uk; sonvx@cs.umu.se; trongld@vnu.edu.vn