IML-GCN: Improved Multi-Label Graph Convolutional Network for Efficient yet Precise Image Classification

https://cvi2.uni.lu/

Inder Pal Singh, Oyebade Oyedotun, Enjie Ghorbel, Djamila Aouada
Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg
inder.singh@uni.lu, oyebade.oyedotun@uni.lu, enjie.ghorbel@uni.lu, djamila.aouada@uni.lu

AAAI-22
Fonds National de la Recherche Luxembourg

Introduction

In this paper, we propose IML-GCN as a precise and efficient framework for multi-label image classification. Although previous approaches have shown great performance, they usually make use of very large architectures. To handle this, we propose to combine the small version of a newly introduced network called TResNet[2] with an extended version of Multi-label Graph Convolution Networks (ML-GCN)[1]; therefore, ensuring the learning of label correlation while reducing the size of the overall network.

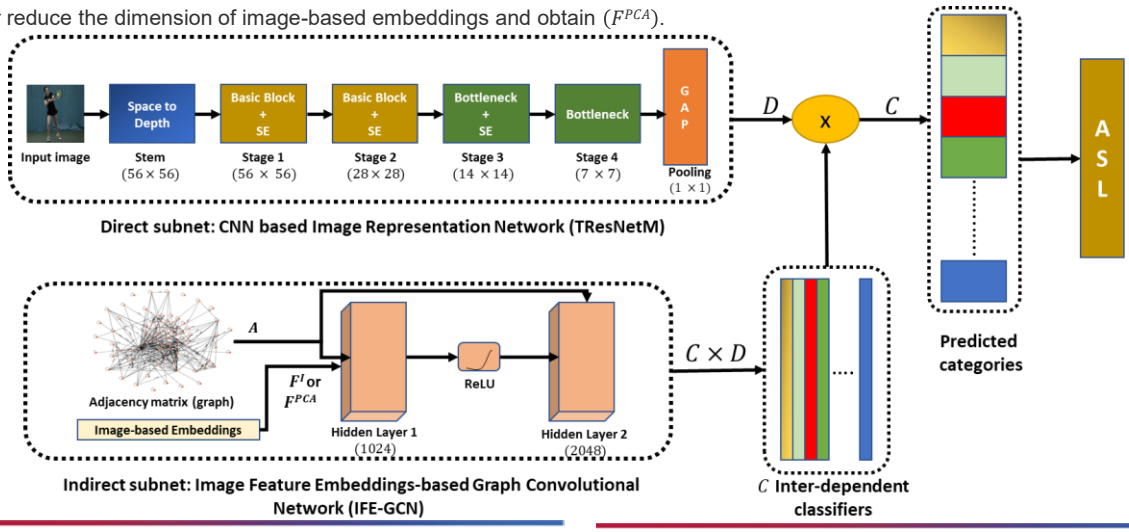
Motivation

Despite the impressive performance in existing graph-based approaches, two main limitations can be noted:

- the backbone network for image representation is very deep (ResNet-101)[3] and therefore naturally induces a heavy architecture leading to high memory computation and consumption.
- it uses word embeddings generated by Glove which were originally introduced for the task of NLP and are not optimal in context of multi-label image classification.

Proposed Approach

- We propose to generate node embeddings from a trained CNN's penultimate layer i.e., typically GAP layer and represent them using F_I.
- We replace the word embeddings by our proposed novel image-based embeddings (F^I) in order to give more robustness to the graph nodes showing its relevance in terms of precision and model size with respect to the state-of-the-art.
- Furthermore, as improving the signal-to-noise of the input can facilitate the learning of robust representations by the GCN, we use PCA to further reduce the dimension of image-based embeddings and obtain (F^{PCA}).



Results

Method	#Parameters (Millions)	mAP (MS-COCO)	mAP (VG-500)
CNN-RNN	66.2	61.2	-
SRN	48.0	77.1	-
ResNet101[3]	44.5	77.3	30.9
ML-GCN[1]	44.9	83.0	32.6
SSGRL	92.2	83.8	-
KGGR	45.0	84.3	-
C-Tran*	120.0*	85.1	38.4*
ASL (TRESNET-M)[2]	29.5	81.8	33.6
ASL (TRESNET-L)[2]	53.8	86.6	-
IML-GCN (Ours with FI)	33.5	85.9	34.0
IML-GCN (Ours with FPCA)	31.5	86.6	34.5

*The model size is roughly 273% bigger than our proposed model

Conclusion

In this paper, we propose to use novel image-based embeddings, which are more adapted in an image classification context, as node features for the GCN in a CNN-GCN framework for multi-label image classification. In this context, by exploiting the latest advancement in CNN, we propose IML-GCN that achieves high precision while reducing the size of the network.

Acknowledgements

This work was funded by the Luxembourg National Research Fund (FNR), under the project reference BRIDGES2020/IS/14755859/MEET-A/Aouada.

References

- [1] Chen, Z.-M.; Wei, X.-S.; Wang, P.; and Guo, Y. *Multi-label image recognition with graph convolutional networks.* CVPR 2019.
- [2] Ridnik, T.; Lawen, H.; Noy, A.; Ben Baruch, E.; Sharir, G.; and Friedman, I. TResNet: *High Performance GPU Dedicated Architecture*. WACV 2021.
- [3] He, K.; Zhang, X.; Ren, S.; and Sun, J. *Deep Residual Learning for Image Recognition*. CVPR 2016