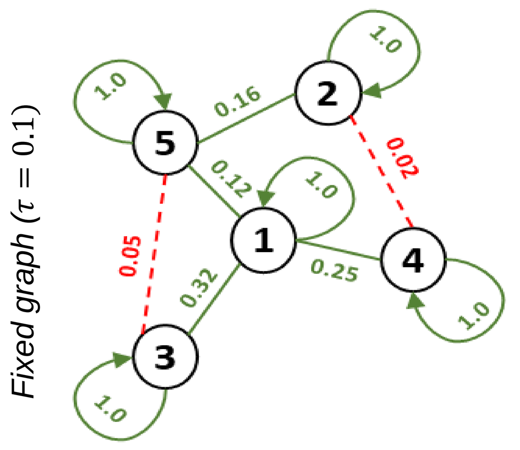
MULTI LABEL IMAGE CLASSIFICATION USING ADAPTIVE GRAPH CONVOLUTIONAL NETWORKS (ML-AGCN)

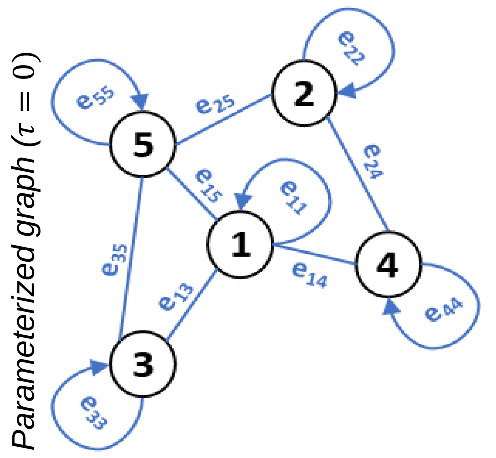
Inder Pal Singh, Enjie Ghorbel, Oyebade Oyedotun, Djamila Aouada

Interdisciplinary Center for Security, Reliability and Trust (SnT), University of Luxembourg

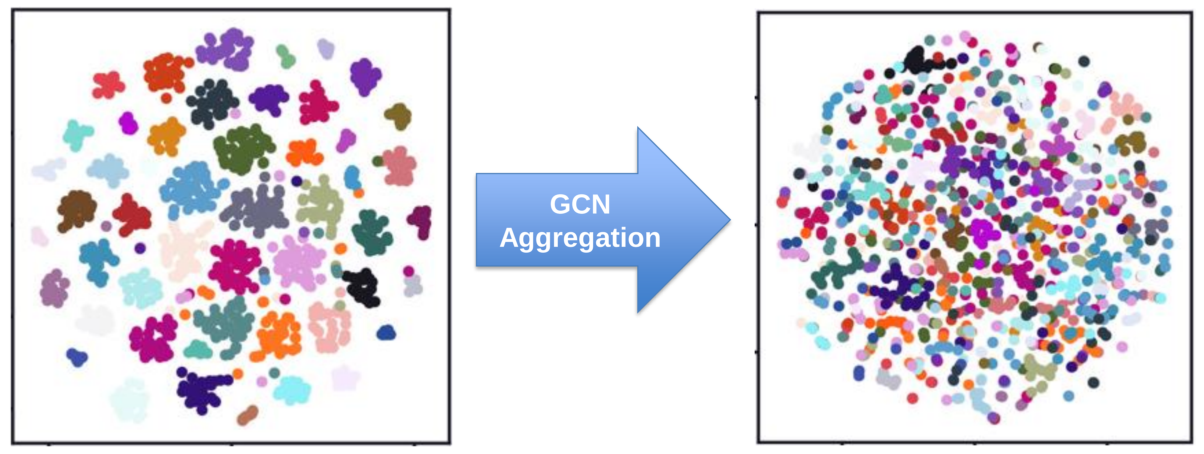
Introduction



Existing graph-based approaches



Our proposed ML-AGCN



Node features *(before GCN)*

Node features *(after GCN)*

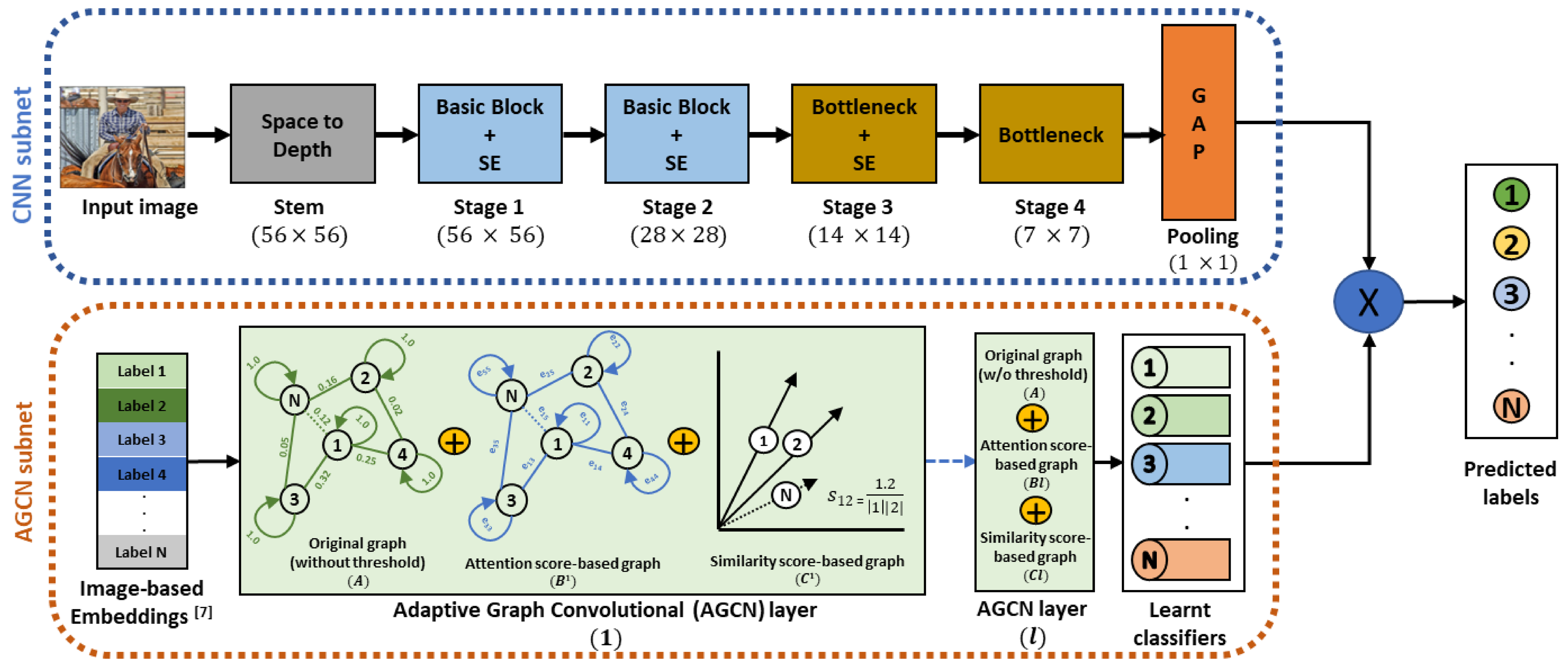
Graph-based approaches [1,2] are successful in multi-label image classification. However, they :

- need a **fixed** label graph based on the label co-occurrences [1],
- select **empirically** a threshold [1,2],
- **loose** the node feature similarity during the GCN aggregation [3],
- achieve a high accuracy at the cost of a **large** model.

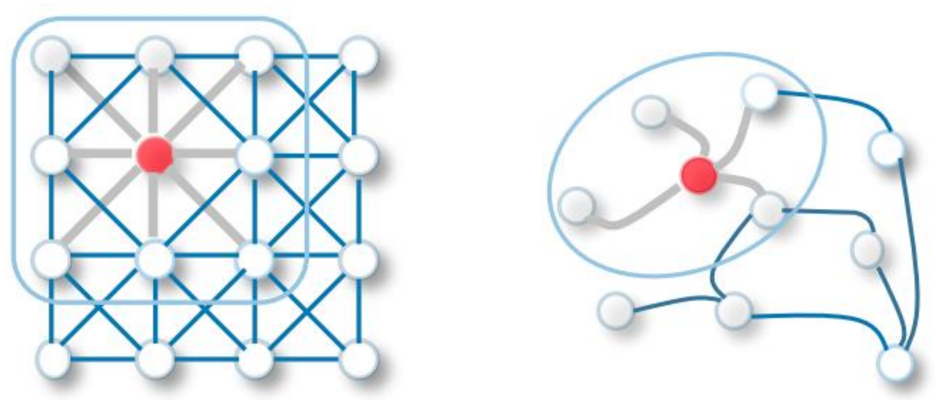
Contributions

- The graph topology is learnt **adaptively** in an **end-to-end** manner by incorporating:
  - → an attention-based mechanism learning the **importance** of each node pair
  - → a similarity-based mechanism for **preserving** the node feature similarity.
- An effective approach with a reduced number of model parameters.

ML-AGCN: Architecture Overview



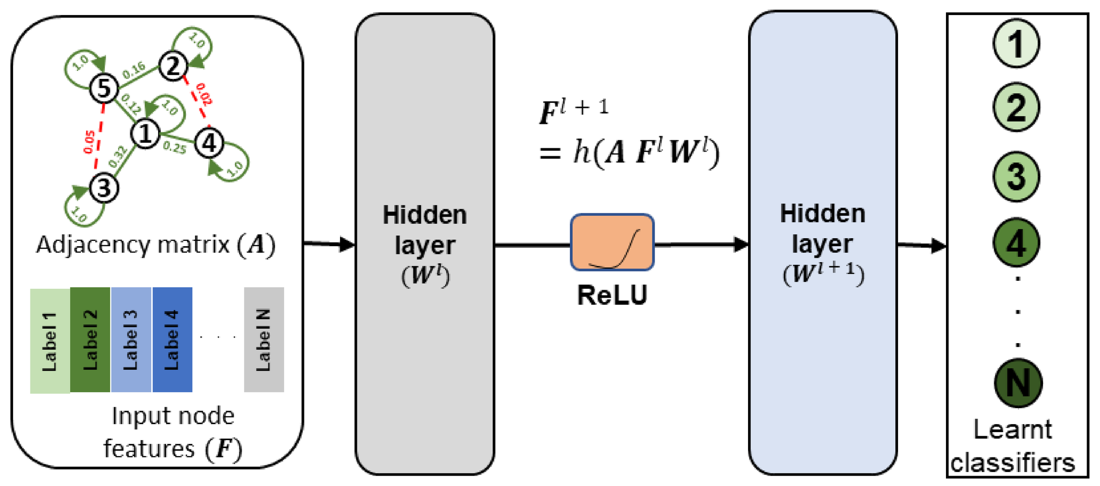
Graph Convolutional Networks



- GCN is an extension of CNN to graphs [1].
- Node features are leaned from each hidden layer as follows;

$$\boldsymbol{F}^{l+1} = h(\boldsymbol{A}\,\boldsymbol{F}^{l}\,\boldsymbol{W}^{l}),$$

$\boldsymbol{A}$: the adjacency matrix, $\boldsymbol{W}^{l}$: the learned weight matrix of the l^{th} layer, $\boldsymbol{F}^{l}$: node features from the l^{th} layer.



Proposed Approach

Two additional adaptive graphs $(\boldsymbol{B}\boldsymbol{l})$ and $(\boldsymbol{C}^{l})$ are proposed, such that;

$$\boldsymbol{F}^{l+1} = h((\boldsymbol{A} + \boldsymbol{B}^{l} + \boldsymbol{C}^{l})\,\boldsymbol{F}^{l}\,\boldsymbol{W}^{l})$$

- $\boldsymbol{B}^{l} = (b^{l}_{ij})_{i\,j\,\in\,\mathcal{L}}$: quantifies the connectivity importance of each node pair [4] as follows;

$$\alpha_{ij}^{(l)} = \frac{\exp\left(e_{ij}^{(l)}\right)}{\Sigma_{k\in\mathcal{N}(i)}\exp\left(e_{ik}^{(l)}\right)},$$

where $e_{ij} = LeakyReLU\left(a^{(l)T}\ \left(\boldsymbol{W} f_i^{(l)} || \boldsymbol{W} f_j^{(l)}\right)\right)$ and || denotes the concatenation.

Self-importance is estimated;

$$b_{ij}^{(l)} = \begin{cases} \alpha_{ij}^{(l)} + \max_{\ k\in\mathcal{L}}\left(\alpha_{ik}^{(l)}\right), & \text{if } i = j \\ \alpha_{ij}^{(l)}, & \text{if } i \neq j \end{cases}.$$

- $\boldsymbol{C}^{l} = (c^{l}_{ij})_{i\,j\,\in\,\mathcal{L}}$: computes the cosine similarity between each node feature pair as follows;

$$c_{ij}^{(l)} = \frac{f_i^{(l)}.f_j^{(l)}}{|f_i^{(l)}||f_j^{(l)}|}.$$

Comparative Analysis of Performance

MS-COCO: Comparison with the state-of-the-art

Method	# params (millions)	mAP	CP	CR	CF1	OP	OR	OF1
CNN-RNN	66.2	61.2	-	-	-	-	-	-
SRN	48.0	77.1	81.6	65.4	71.2	82.7	69.9	75.8
ResNet101	44.5	77.3	80.2	66.7	72.8	83.9	70.8	76.8
ML-GCN [1] (1-L)*	43.1	80.9	82.9	69.7	75.8	84.8	73.6	78.8
IML-GCN [2] (1-L)*	**29.5**	81.3	81.3	72.2	76.0	86.7	77.9	82.1
ASL (TRESNET-M)	**29.5**	81.8	82.1	72.6	76.4	83.1	76.1	79.4
ML-GCN [1]	44.9	83.0	85.1	72.0	78.0	85.8	75.4	80.3
SSGRL	92.2	83.8	**89.9**	68.5	76.8	**91.3**	70.8	79.7
KGGR	45.0	84.3	85.6	72.7	78.6	87.1	75.6	80.9
C-Tran	120.0	85.1	86.3	74.3	79.9	87.7	76.5	81.7
ASL (TRESNET-L)	53.8	86.6	87.4	76.4	81.4	88.1	79.2	81.8
IML-GCN [2]	31.5	86.6	78.8	**82.6**	80.2	79.0	**85.1**	81.9
Ours: ML-AGCN (1-L)*	29.9	86.7	79.6	82.4	80.7	79.8	84.5	82.1
Ours: ML-AGCN	35.9	**86.9**	86.2	78.3	**81.7**	87.2	80.7	**83.8**

VG-500: Comparison with the state-of-the-art

Method	# params (millions)	mAP
IML-GCN [2] (1-L)*	30.6	17.01
ResNet101	44.5	30.9
ML-GCN [1]	44.9	32.6
ASL (TResNet-M)	**29.5**	33.6
IML-GCN [2]	32.1	34.5
SSGRL	92.2	36.6
C-Tran†	120†	38.4†
Ours: ML-AGCN (1-L)*	32.7	**37.9**
Ours: ML-AGCN	37.4	37.1

*Graph-based approaches with only 1 hidden layer

†The model is roughly 120 times higher than our proposed model

Conclusion

- We propose ML-AGCN that **adaptively** learns the label graph topology in an **end-to-end** manner for multi-label image classification.
- ML-AGCN achieves **competitive** results in terms of **model size** and **accuracy** with respect to current state-of-the-art.

References

[1] Chen, Zhao-Min, et al. "Multi-label image recognition with graph convolutional networks." *CVPR 2019*.

[2] Singh, Inder Pal, et al. "IML-GCN: Improved Multi-Label Graph Convolutional Network for Efficient yet Precise Image Classification." *DLG:AAAI 2022.*

[3] Jin, Wei, et al. "Node similarity preserving graph convolutional networks." *WSDM. 2021.*

[4] Veličković, Petar, et al. "Graph attention networks." *arXiv preprint arXiv:1710.10903 (2017).*

Acknowledgement

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

https://cvi2.uni.lu/