PhD thesis subject

Deep Graph Neural Networks for Visual Scene Recognition

Deep learning is currently witnessing a major interest in computer vision and different related fields [1-5,33]. Its principle consists in training multi-layered neural networks by designing suitable architectures and optimizing their parameters [6]. In particular, convolutional networks are well studied and aim at extracting features that gradually capture low-to-high semantics of visual patterns. Early convolutional networks were dedicated to regular (grid-like) scenes including images where convolutions are achieved by shifting equivariant filters and measuring their responses across different image locations. However, scenes sitting on top of irregular domains (such as skeletons in action recognition or regions in object detection) require extending convolutional networks to unstructured data (namely graphs) [7-10]; indeed, while shifting filters across regular grids is a straightforward and a well-defined operation, its extension to irregular domains (i.e., graphs with heterogeneous topological properties) is generally ill-posed. Motivated by the success of deep learning in computer vision, graph convolutional networks (GCNs) are currently emerging for different use-cases and applications [11,34]. The common ground of these networks consists in aggregating node representations prior to apply filters on the resulting node aggregates [11-16]. Two categories of GCNs are known in the literature: the first one, dubbed as spatial [17-22,32], achieves convolution by locally averaging representations through nodes and their neighbors before applying convolutions using inner products. The second category, known as spectral [23-25,7,8,10,26-29], proceeds differently by first mapping filter and input graph signals using the eigen-decomposition of their Laplacians, then achieving filtering in the resulting spectral domain prior to back-project the filtered signal onto the input graph domain [30-31]. While spectral GCNs make convolutions well-defined compared to spatial GCNs, their downside resides in the non-localized aspect of the learned filters and also in the high complexity of Laplacian eigen-decomposition.

Considering the aforementioned issues, the goal of this thesis subject is to devise highly effective and also efficient GCNs for the task of visual scene recognition. In our targeted solutions, graphs will be used to model scene parts together with their spatial, temporal and semantic interactions. Concepts (and their combinations) will also be described with graphs where nodes correspond to individual classes and links correspond to their interactions. In contrast to most of the existing solutions, where nodes/edges in graphs are handcrafted, we will consider in this thesis subject an «end-to-end» training process that infers both nodes and their interactions, prior to learn the underlying GCNs. Other aspects will be addressed including attention mechanisms as well as transformer networks [35,36] that help designing convolutions with topologically variant filter supports. We will also consider spectral GCNs that make convolutions through the graph Fourier transform principled and well defined; nevertheless, the relevance of these convolutions relies on the adequacy of the used Laplacian operators which are usually handcrafted. The latter are not able to capture all the relationships between nodes as their setting is agnostic to the targeted tasks. For instance, in skeleton-based action recognition, pre-existing node-to-node relationships capture the intrinsic anthropometric aspects of individuals which are necessary for their identification, while other relationships, yet to infer, about their dynamics are necessary in order to recognize their actions. Put differently, depending on the task at hand, connectivity in Laplacian operators should be appropriately learned by including not only the available (intrinsic) node-to-node connections in graphs but also their inferred (extrinsic) relationships [37,38]. Moreover, the consistency of the learned Laplacian operators is also critical and requires adapting the domains of these operators to the input graphs [23]. Finally, all these aspects will be investigated in the context of visual scene recognition including image/video classification and segmentation.

Keywords. Deep machine learning, graph convolutional networks, visual scene recognition, image classification, video action recognition.

Thesis Director: Hichem Sahbi, CNRS Researcher, HDR, LIP6 Lab, Sorbonne University (**contact:** hichem.sahbi@lip6.fr).

PhD Student Background. We are seeking a highly motivated PhD candidate, with a preferred background in applied mathematics or computer science with more emphasis on statistics, machine learning and/or image processing/computer vision, and familiar with existing machine learning tools and programming platforms.

Related bibliography

[1] M. Jiu and H. Sahbi, “Nonlinear deep kernel learning for image annotation,” IEEE Transactions on Image Processing, vol. 26(4), 2017.

[2] H. Sahbi. Coarse-to-fine deep kernel networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1131-1139, ICCV - CEFRL W, 2017

[3] R. Girshick. Fast R-CNN. In ICCV, pages 1440–1448, 2015

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In CVPR, pages 770–778, June 2016.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In NIPS, pages 1097–1105, 2012.

[6] I. Goodfellow, Y. Bengio, and A. Courville. Deep learning. MIT press, 2016.

[7] J. Bruna, W. Zaremba, A. Szlam, Y. LeCun. Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203, 2013.

[8] M. Defferrard, X. Bresson, P. Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In NIPS, 3844-3852, 2016.

[9] W. Huang, T. Zhang, Y. Rong, J. Huang. Adaptive sampling towards fast graph representation learning. In NIPS. pp. 4558-4567, 2018.

[10] T.N. Kipf, M. Welling. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.

[11] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, M.M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In CVPR, pp. 5115–5124, 2017.

[12] W. Hamilton, Z. Ying, J. Leskovec. Inductive representation learning on large graphs. In NIPS. pp. 1024–1034, 2017.

[13] J. Atwood, D. Towsley. Diffusion-convolutional neural networks. In NIPS, pp. 1993–2001, 2016.

[14] H. Gao, Z. Wang, S. Ji. Large-scale learnable graph convolutional networks. In the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1416–1424. ACM, 2018.

[15] M. Niepert, M. Ahmed, K. Kutzkov. Learning convolutional neural networks for graphs. In ICML, pp. 2014-2023, 2016.

[16] J. Zhang, X. Shi, J. Xie, H. Ma, I. King, D.Y. Yeung. Gaan: Gated attention networks for learning on large and spatiotemporal graphs. ArXiv:1803.07294, 2018.

[17] H. Sahbi. “Kernel-based Graph Convolutional Networks”. In the Proceedings of IAPR ICPR. 2021.

[18] M. Gori, G. Monfardini, F. Scarselli. A new model for learning in graph domains. In IEEE IJCNN, vol. 2, pp. 729–734, 2005.

[19] A. Micheli. Neural network for graphs: A contextual constructive approach. IEEE TNN 20(3), 498-511, 2009.

[20] F. Scarselli, M. Gori, A.C. Tsoi, M. Hagenbuchner, G. Monfardini. The graph neural network model. IEEE TNN 20(1), 61–80, 2008.

[21] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P.S. Yu. A comprehensive survey on graph neural networks. ArXiv:1901.00596, 2019.

[22] W. Hamilton, Z. Ying, J. Leskovec. Inductive representation learning on large graphs. In NIPS. pp. 1024–1034, 2017.

[23] A. Mazari and H. Sahbi. MLGCN: Multi-Laplacian Graph Convolutional Networks for Human Action Recognition. In BMVC. 2019.

[24] C. Li, Q. Zhong, D. Xie, and S. Pu. Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation. arXiv preprint arXiv:1804.06055, 2018.

[25] M. Henaff, J. Bruna, and Y. LeCun. Deep convolutional networks on graph-structured data. arXiv preprint arXiv:1506.05163, 2015.

[26] R. Levie, F. Monti, X. Bresson, M.M. Bronstein. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. IEEE Transactions on Signal Processing 67(1), 97–109, 2018.

[27] J. Chen, T. Ma, C. Xiao. Fastgcn: fast learning with graph convolutional networks via importance sampling. arXiv preprint arXiv:1801.10247, 2018.

[28] Z. Chenyi and Q. Ma. Dual graph convolutional networks for graph-based semi-supervised classification. Proceedings of WWW, 2018.
[29] W. Huang, T. Zhang, Y. Rong, J. Huang. Adaptive sampling towards fast graph representation learning. In NIPS. pp. 4558-4567, 2018.

[30] D. Slepian. Some comments on Fourier analysis, uncertainty and modeling. In Society for Industrial and Applied Mathematics (SIAM review), 1983.

[31] F. Chung. Spectral graph theory. American Mathematical Soc.. 1997.

[32] H. Sahbi, J.-Y. Audibert, and R. Keriven. Context-dependent kernels for object classification. In IEEE Pattern Analysis and Machine Intelligence, TPAMI, 2011.

[33] He, Kaiming, et al. ”Mask r-cnn.” In Proceedings of International Conference on Computer Vision, 2017.

[34] Zhang, Ziwei, Peng Cui, and Wenwu Zhu. ”Deep learning on graphs: A survey.” IEEE Transactions on Knowledge and Data Engineering, 2020.

[35] Ashish Vaswani et al. Attention Is All You Need, arxiv, 2017.

[36] Petar Veličković et al. Graph Attention Networks. In the proceedings of ICLR 2018.

[37] H. Sahbi. Learning Chebyshev Basis in Graph Convolutional Networks for Skeleton-based Action Recognition, arXiv preprint, april, 2021.

[38] H. Sahbi. Skeleton-based Hand-Gesture Recognition with Lightweight Graph Convolutional Networks, arXiv preprint, april, 2021.