Guest Editorial Introduction to the Special Section on Graphs in Vision and Pattern Analysis

In THE real world, data have diverse structures. While some data, e.g., digital images, have a regular and grid-like structure, a vast majority of data do not. To handle this issue, data is usually represented by a graph. Graph has been an important and frequently studied structure in computer vision for decades, owing to its great ability to model between-object relationships and flexibility in representation learning. This fact is especially true in recent years, as the advance of deep learning and GPU parallelization makes it feasible to train neural networks with graph-structured data. It then brings about a research boom in graph neural networks (GNNs) in the communities of computer vision, pattern recognition, computer graphics, machine learning, and multimedia computing.

The goal of this special section is to provide a platform to summarize what we have achieved and where we are moving toward this methodological research direction. It gathers the latest advances in learning with graph-structured data in computer vision, as well as interdisciplinary efforts on graph-based pattern analysis in sociology, physics, chemistry, finance, biology, *etc.* As guest editors of the special section on Graphs in Vision and Pattern Analysis, we were glad to receive 24 submissions. Among them, 15 papers were accepted in this special section. The accepted papers are grouped into 3 categories: 1) Graph Network Design; 2) Graph Network Applications; and 3) Graph Network Properties, which will be detailed in the following.

I. Graph Network Design

The paper "Second-order Pooling for Graph Neural Networks" by Zhengyang Wang and Shuiwang Ji, uses second-order pooling as graph pooling to solve the challenges due to the variable sizes and isomorphic structures of graphs. Two novel and effective graph pooling approaches are proposed, namely, bilinear mapping and attentional second-order pooling. Thorough experiments on graph classification tasks demonstrate the effectiveness and superiority of the proposed methods.

The paper "Global Context Networks" by Yue Cao, Jiarui Xu, Stephen Lin, Fangyun Wei, and Han Hu, proposes a simplification of the Non-Local block, named global context (GC) block, in which a query-independent attention map is explicitly used for all query positions. In addition, they replace the one-layer transformation function of the non-local block with a two-layer bottleneck, which enables GCNet to significantly reduce the computation burden but maintain accuracy. Their experiments show that GCNet outperforms non-local networks on major benchmarks for various recognition tasks.

The paper "CCNet: Criss-Cross Attention for Semantic Segmentation" by Zilong Huang, Xinggang Wang, Yunchao Wei, Lichao Huang, Humphrey Shi, Wenyu Liu, and Thomas S. Huang, proposes a Criss-Cross Network (CCNet) for obtaining full-image contextual information in a very effective and efficient way. To capture the full-image dependencies, CCNet sequentially stacks two criss-cross attention modules on feature maps. Experiments show that CCNet achieves leading performance on segmentation-based benchmarks while remain high computational efficiency.

The paper "Revisiting 2D Convolutional Neural Networks for Graph-based Applications" by Yecheng Lyu, Xinming Huang, and Ziming Zhang, aims to bridge the gap between graph neural networks and convolutional neural networks. Two graph-to-grid mapping schemes are proposed, namely, graph-preserving grid layout (GPGL) and Hierarchical GPGL (H-GPGL), to enable the use of CNNs as the backbone for graph-based applications. The empirical success of GPGL is demonstrated on graph classification with small graphs, and that of H-GPGL is demonstrated on 3D point cloud segmentation with large graphs.

The paper "DeepGCNs: Making GCNs Go as Deep as CNNs" by Guohao Li, Matthias Muller, Guocheng Qian, Itzel C. Delgadillo, Abdulellah Abualshour, Ali Thabet, and Bernard Ghanem, adapts concepts that were successful in training deep convolution neural networks, in particular, residual connections, dense connections, and dilated convolutions to successfully train very deep graph networks. This work shows how these concepts can be incorporated into a graph framework and quantifies the effect of these additions by extensive experiments on point cloud and biological graph data.

II. Graph Network Applications

The paper "Learning Multi-View Interactional Skeleton Graph for Action Recognition" by Minsi Wang, Bingbing Ni, and Xiaokang Yang, proposes Multi-View Interactional Graph Network to better address the intrinsic multi-level spatial skeleton context in the action recognition task. It constructs, learns, and infers multi-level spatial skeleton context, including view-level (global), group-level, and joint-level (local) context in a unified way. Through rigorous experiments, the authors show that the proposed method achieves impressive performance on large-scale benchmarks.

Digital Object Identifier 10.1109/TPAMI.2023.3259779

The paper "HiGCIN: Hierarchical Graph-based Cross Inference Network for Group Activity Recognition" by Rui Yan, Lingxi Xie, Jinhui Tang, Xiangbo Shu, and Qi Tian, exploits the latent spatiotemporal dependencies among feature nodes to capture the spatiotemporal co-occurrence of group activities. In a Hierarchical Graph-based Cross Inference Network (HiGCIN), three levels of information, including body-region level, person level, and group-activity level, are constructed and learned. To achieve this, they first present a generic Cross Inference Block (CIB) to concurrently capture the latent spatiotemporal dependencies among body regions and persons. Then based on CIB, Body-regions Inference Module and Persons Inference Module are designed to extract and refine features for each person and to further explore the spatiotemporal dependencies among personal features, respectively.

The paper "Learning Graph Convolutional Networks for Multi-Label Recognition and Applications" by Zhao-Min Chen, Xiu-Shen Wei, Peng Wang, and Yanwen Guo, proposes Graph Convolutional Networks-based solutions to explicitly model the label correlations for multi-label recognition. The key idea is to build a directed graph over the classes, where the nodes can represent some types of class-relevant features and the weights of the edges capture the dependencies between classes. Following this idea, they propose Classifier Learning GCN, in which class-level semantic representations are mapped into classifiers that maintain the inter-class topology, and Prediction Learning GCN, which encodes label-aware features into inter-dependent image-level prediction scores.

The paper "Combinatorial Learning of Robust Deep Graph Matching: an Embedding based Approach" by Runzhong Wang, Junchi Yan, and Xiaokang Yang, adopts a deep graph embedding network for learning graph matching. To do so, they transform the graph-matching problem into a linear assignment task by utilizing the graph convolutional network to extract the graph structures into node-wise feature vectors. They claim that the embedded node features contain structural information around the node such that even higher-order (beyond second-order) information can be incorporated into the matching procedure, and in this way, the model circumvents the notoriously challenging QAP problem.

The paper "Learning Multi-Attention Context Graph for Group-Based Re-Identification" by Yichao Yan, Jie Qin, Bingbing Ni, Jiaxin Chen, Li Liu, Fan Zhu, Wei-Shi Zheng, Xiaokang Yang, and Ling Shao, focuses on employing context information for identifying groups of people, i.e., group re-identification, instead of single person re-identification (re-id). They propose a novel unified framework, namely MACG, for both group re-id and group-aware person re-id tasks. In addition, a multi-level attention mechanism is developed to capture both intra- and inter-group contexts. It verifies that the context information exploited by GNNs largely enhances person (node)-level and group (graph)-level representation capabilities, and thus, benefits both re-id and group re-id.

The paper "Fashion Retrieval via Graph Reasoning Networks on a Similarity Pyramid" by Yiming Gao, Zhanghui Kuang, Guanbin Li, Ping Luo, Yimin Chen, Liang Lin, and Wayne Zhang, focuses on the task of fashion retrieval. They propose a novel Graph Reasoning Network (GRNet) on a similarity pyramid, which learns similarities between a query and a gallery cloth by using both global and local representations at multiple scales. The similarity pyramid is represented by a graph of similarity, in which nodes represent similarities between clothing components at different scales, and the final matching score is obtained by message passing along edges. Their experiments show that GRNet obtains new state-of-the-art results on two challenging benchmarks and outperforms competitors by a large margin.

III. Graph Network Properties

The paper "Structured Knowledge Distillation for Dense Prediction" by Yifan Liu, Changyong Shu, Jingdong Wang, and Chunhua Shen, focuses on structured knowledge distillation and transfers the structure information with two schemes: pair-wise distillation and holistic distillation. The pair-wise distillation distills the pairwise similarities by building a static graph and holistic distillation uses adversarial training to distill holistic knowledge. The authors demonstrate that the proposed knowledge distillation approaches achieve better performance in transferring structure information by conducting experiments on three dense prediction tasks.

The paper "LogicENN: A Neural Based Knowledge Graphs Embedding Model with Logical Rules" by Mojtaba Nayyeri, Chengjin Xu, Jens Lehmann, and Hamed Shariat Yazdi, addresses the challenge of including rules in knowledge graph embedding models and presents a new neural-based embedding model named LogicENN. It learns every ground truth of encoded rules in a knowledge graph and outperforms the state-of-the-art models in link prediction.

The paper "Fourier-based and Rational Graph Filters for Spectral Processing" by Giuseppe Patane proposes novel Fourier-based and rational graph filters for graph processing, which generalizes the notion of polynomial filters and the Fourier transform to non-Euclidean domains. To better evaluate discrete spectral Fourier-based and wavelet operators, the author introduces a spectrum-free approach, which solves a small set of sparse, symmetric, and well-conditioned linear systems. The link between spectral operators, wavelets, and filtered convolution with integral operators induced by spectral kernels is also studied.

The paper "Co-embedding of Nodes and Edges with Graph Neural Networks" by Xiaodong Jiang, Ronghang Zhu, Pengsheng Ji, and Sheng Li, focuses on the task of handling information in edge features. They propose Convolution with Edge-Node Switching graph neural network (CensNet) for learning tasks in graph-structured data with both node and edge features. CensNet switches the role of nodes and edges by using the line graph of the original undirected graph and employs two forward-pass feature propagation rules on the graph and its line graph to alternatively update the node and edge embeddings. They claim that the model can learn node and edge embeddings effectively and leads to significant performance gain in four graph learning tasks.

SONG BAI, *Guest Editor*
University of Oxford
OX1 2JD Oxford, U.K.
E-mail: songbai.site@gmail.com

PHILIP H.S. TORR, *Guest Editor*
University of Oxford
OX1 2JD Oxford, U.K.
E-mail: philip.torr@eng.ox.ac.uk

RANJAY KRISHNA, *Guest Editor*
Stanford University
Stanford, CA 94305, USA
E-mail: ranjaykrishna@cs.stanford.edu

FEI-FEI LI, *Guest Editor*
Stanford University
Stanford, CA 94305, USA
E-mail: feifeili@cs.stanford.edu

ABHINAV GUPTA, *Guest Editor*
Carnegie Mellon University
Pittsburgh, PA 15213, USA
E-mail: abhinavg@cs.cmu.edu

SONG-CHUN ZHU, *Guest Editor*
University of California
Los Angeles, CA 90095, USA
E-mail: sczhu@stat.ucla.edu