Electronic Letters on Computer Vision and Image Analysis 19(2):9-10, 2020

Distilling Structure from Imagery: Graph-based Models for the Interpretation of Document Images

Pau Riba

Advisors: Josep Lladós and Alicia Fornés
Computer Vision Center, Barcelona, Catalunya
Received 16 October 2020; Accepted 19 December 2020

From its early stages, the community of Pattern Recognition and Computer Vision has considered the importance of leveraging the structural information when understanding images. Usually, graphs have been selected as the adequate framework to represent this kind of information due to their flexibility and representational power able to codify both, the components, objects, or entities and their pairwise relationship. Even though graphs have been successfully applied to a huge variety of tasks, as a result of their symbolic and relational nature, graphs have always suffered from some limitations compared to statistical approaches. Indeed, some trivial mathematical operations do not have an equivalence in the graph domain. For instance, in the core of many pattern recognition applications, there is a need to compare two objects. This operation, which is trivial when considering feature vectors defined in $\mathbb{R}^n$, is not properly defined for graphs.

Along this dissertation the main application domain has been on the topic of Document Image Analysis and Recognition. It is a subfield of computer vision aiming at understanding images of documents. In this context, the structure and in particular graph representations, provides a complementary dimension to the raw image contents.

In computer vision, the first challenge we face is how to build a meaningful graph representation that is able to encode the relevant characteristics of a given image. This representation should find a trade off between the simplicity of the representation and its flexibility to represent the deformations appearing on each application domain. We applied our proposal to the word spotting application where strokes are divided into graphemes which are the smaller units of a handwritten alphabet [5].

We have investigated different approaches to speed-up the graph comparison in order that word spotting, or more generally, a retrieval application is able to handle large collections of documents. On the one hand, a graph indexing framework combined with a votation scheme at node level is able to quickly prune unlikely results [7]. On the other hand, making use of graph hierarchical representations, we are able to perform a coarse-to-fine matching scheme which performs most of the comparisons in a reduced graph representation [6]. Besides, the hierarchical graph representation demonstrated to be drivers of a more robust scheme than the original graph. This new information is able to deal with noise and deformations in an elegant fashion. Therefore, we propose to exploit this information in a hierarchical graph embedding which allows the use of classical statistical techniques [2].

Correspondence to: priba@cvc.uab.cat

Thesis Dissertation ISBN: 978-84-121011-6-4

Recommended for acceptance by Christian Aguilera and Arash Akbarinia
DOI: https://doi.org/10.5565/rev/elcvia.1313
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

Recently, the new advances on geometric deep learning*, which has emerged as a generalization of deep learning methods to non-Euclidean domains such as graphs and manifolds [1], has raised again the attention to these representation schemes. Taking advantage of these new developments but considering traditional methodologies as a guideline, we proposed a graph metric learning framework able to obtain state-of-the-art results on different tasks [4].

Finally, the contributions of this thesis have been validated in real industrial use case scenarios. For instance, an industrial collaboration has resulted in the development of a table detection framework in annonymized administrative documents containing sensitive data. In particular, the interest of the company is the automatic information extraction from invoices. In this scenario, graph neural networks have proved to be able to detect repetitive patterns which, after an aggregation process, constitute a table [3].

FULL ACCESS TO THE DISSERTATION – `https://bit.ly/2FIjyrq`

KEYWORDS – Computer Vision, Pattern Recognition, Graph-based Representations, Graph Indexing, Hierarchical Graphs, Graph Embeddings, Graph Neural Networks, Graph Edit Distance, Table Detection.

References

[1] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[2] Anjan Dutta, Pau Riba, Josep Lladós, and Alicia Fornés. Hierarchical stochastic graphlet embedding for graph-based pattern recognition. *Neural Computing and Applications*, 32:11579–11596, 2020.

[3] Pau Riba, Anjan Dutta, Lutz Goldmann, Alicia Fornés, Oriol Ramos, and Josep Lladós. Table detection in invoice documents by graph neural networks. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 122–127, 2019.

[4] Pau Riba, Andreas Fischer, Josep Lladós, and Alicia Fornés. Learning graph distances with message passing neural networks. In *Proceedings of the International Conference on Pattern Recognition*, pages 2239–2244, 2018.

[5] Pau Riba, Alicia Fornés, and Josep Lladós. Handwritten word spotting by inexact matching of grapheme graphs. In *Proceedings of the International Conference on Document Analysis and Recognition*, pages 781–785, 2015.

[6] Pau Riba, Josep Lladós, and Alicia Fornés. Hierarchical graphs for coarse-to-fine error tolerant matching. *Pattern Recognition Letters*, 134:116–124, 2020.

[7] Pau Riba, Josep Lladós, Alicia Fornés, and Anjan Dutta. Large-scale graph indexing using binary embeddings of node contexts for information spotting in document image databases. *Pattern Recognition Letters*, 87:203–211, 2017.

*`http://geometricdeeplearning.com/`