MATCHING CONVOLUTIONAL NEURAL NETWORKS WITHOUT PRIORS ABOUT DATA

Carlos Eduardo Rosar Kos Lassance[1], **Jean-Charles Vialatte**[1,2], **Vincent Gripon**[1] **& Nicolas Farrugia**[1]
[1]IMT Atlantique, Brest, France: firstname.lastname@imt-atlantique.fr
[2]Cityzen Data, Brest, France: firstname.lastname@cityzendata.com

ABSTRACT

We propose an extension of Convolutional Neural Networks (CNNs) to graph-structured data, including strided convolutions and data augmentation on graph. Our method matches the accuracy of state-of-the-art CNNs when applied on images, without any prior about their 2D regular structure. On fMRI data, we obtain a significant gain in accuracy compared with existing graph-based alternatives.

1 INTRODUCTION

A main asset of Convolutional Neural Networks (CNNs, LeCun et al. (1995)) is that they take advantage of the intrinsic regular 2D structure of the data. However when data lacks regular structure, there is no natural notion of convolutions, stride/pooling or data augmentation.

Methods to adapt CNNs to graphs have abundantly been proposed in the literature, e.g. Bruna et al. (2013); Duvenaud et al. (2015); Defferrard et al. (2016); Vialatte et al. (2016); Atwood & Towsley (2016); Kearnes et al. (2016); Monti et al. (2016); Kipf & Welling (2016); Niepert et al. (2016); Levie et al. (2017); Gilmer et al. (2017); Vialatte et al. (2017); Puy et al. (2017); Simonovsky & Komodakis (2017); Pasdeloup et al. (2017); Du et al. (2017); Velickovic et al. (2017); Hamilton et al. (2017); Tixier et al. (2017); Nikolentzos et al. (2017); Kondor et al. (2018); Li et al. (2018).

Most of the alternative works have failed to match the performance of CNNs when applied to regular domains. In this work we strive towards this goal. We introduce a weight-sharing for convolutional layers that are based on translations inferred upon a graph (given or inferred in case of no structure prior). These translations allow us to define stride and data-augmentation on irregular domains.

Our method is able, without any structure prior, to reach similar performance to state-of-the-art CNNs on CIFAR10. We also demonstrate an accuracy gain on PINES fMRI, a neuroimaging dataset.

2 METHODOLOGY

Our method is based on Pasdeloup et al. (2017), where the authors have introduced a way to infer a graph from training signals, then translations from the obtained graph to design ad-hoc CNNs. Finding translations is an NP-complete problem such that for large graphs it is initially not practical. The authors proposed to find only local approximate translations and use them to move the kernel.

Here we extend this approch and design strided convolutions along graph downscaling, data augmentation and convolutions on downscaled graphs. As local translations can't be used to design the latter two, we design proxies to global translations. The outline of the method is depicted in figure 1.

2.1 TRANSLATIONS

We consider a graph $G = \langle V, E \rangle$ with V the set of vertices, and $E \subseteq \binom{V}{2}$ the set of edges, G is supposed to be connected, as conversely the process can be applied to each connected component.

Definition 1 *A **candidate-translation** is a function $\phi : U \to V$, where $U \subset V$ and such that*
1) *ϕ is* injective*: $\forall v, v' \in U, \phi(v) = \phi(v') \Rightarrow v = v'$,*
2) *ϕ is* edge-constrained*: $\forall v \in U, (v, \phi(v)) \in E$,*
3) *ϕ is* strongly neighborhood-preserving*: $\forall v, v' \in U, (v, v') \in E \Leftrightarrow (\phi(v), \phi(v')) \in E$.*

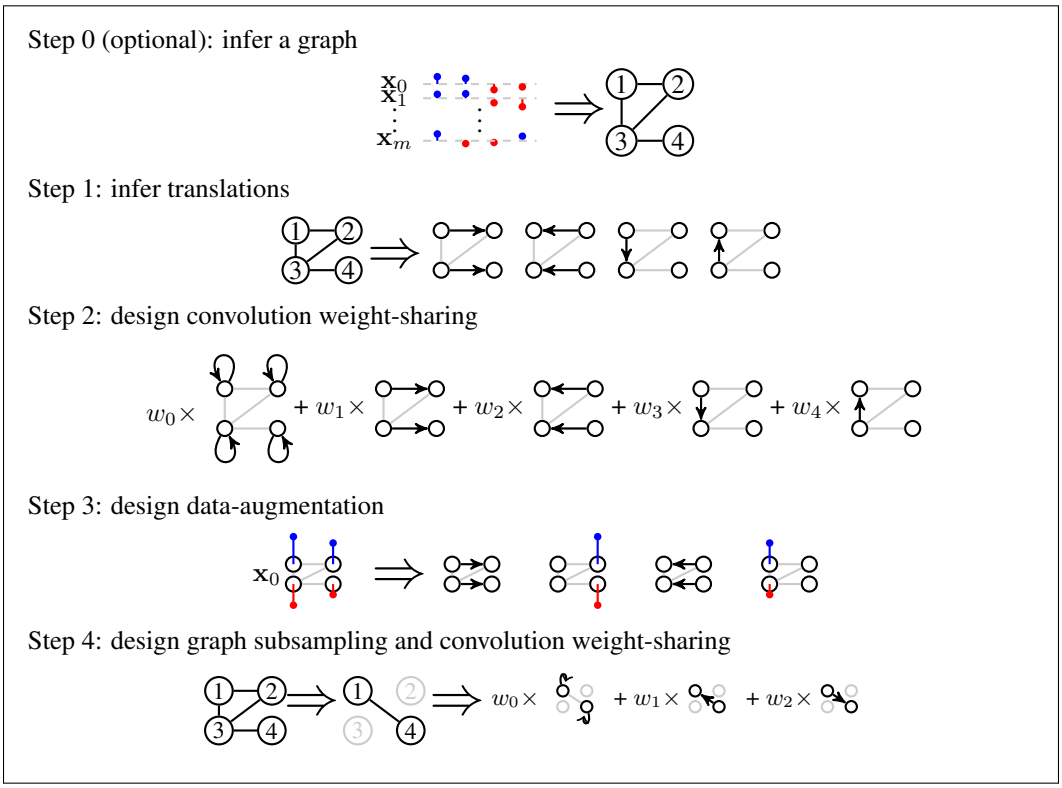

Figure 1: Outline of the proposed method

The loss of ϕ is defined as $|V - U|$. Two candidate-translations ϕ and ϕ' are said to be aligned if $\exists v \in V, \phi(v) = \phi'(v)$. We define $N_r(v)$ as the set of vertices that are at most r-hop away from a vertex $v \in V$.

Definition 2 *A **translation** in G is a candidate-translation such that there is no aligned translation with a strictly smaller loss, or is the identity function.*

Definition 3 *A **local translation** of center $v \in V$ is a translation in the subgraph of G induced by $N_2(v)$, that has v in its definition domain.*

Definition 4 *A family of **proxy-translations** $(\psi_p)_{p=0,..\kappa-1}$ initialized by $v_0 \in V$ is defined algorithmically as follow: (we run several experiments while changing v_0 and keep the best result)*
***1)** We place an indexing kernel on $N_1(v_0)$ i.e. $N_1(v_0) = \{v_0, v_1, ..., v_{\kappa-1}\}$ with $\forall p, \psi_p(v_0) = v_p$,*
***2)** We move this kernel using each local translation ϕ of center v_0: $\forall p, \psi_p(\phi(v_0)) = \phi(v_p)$,*
***3)** We repeat **2)** from each new center until saturation. If a center is being reached again, we keep the indexing that minimizes the sum of losses of the local translations that has lead to it.*

For the sake of simplicity we omitted the case of approximate translations, for details we refer the reader to Pasdeloup et al. (2017).

2.2 Extended Convolution Layers

The operation of the *extended convolution layer* centered on the vertex $v \in V$ is defined as:

$$\mathbf{y}_v = h\left(\sum_{p=0}^{\kappa-1} \mathbf{w}_p \mathbf{x}_{\phi_p(v)} + b\right) \quad \text{with} \quad \begin{cases} \phi_p(v) = \psi_p(v) & \text{if } \psi_p \text{ is defined on } v \\ \phi_p(v) = \perp \notin V & else \end{cases}$$

where $\mathbf{x}_\perp = 0$, h is the activation function, and b is the bias term.

2.3 EXTENDED DATA AUGMENTATION

We use proxy-translations to move training vectors, artificially creating new ones. Note that this type of data-augmentation is poorer than for images since no flipping, scaling or rotations are used.

2.4 EXTENDED DOWNSCALING LAYERS

2.4.1 EXTENDED STRIDED CONVOLUTIONS

Given an arbitrary initial vertex $v_0 \in V$, the set of kept vertices $V_{\downarrow r}$ is defined inductively as follows:
1) $V^0_{\downarrow r} = \{v_0\}$,
2) $\forall t \in \mathbb{N}, V^{t+1}_{\downarrow r} = V^t_{\downarrow r} \cup \{v \in V, \forall v' \in V^t_{\downarrow}, v \notin N_{r-1}(v') \wedge \exists v' \in V^t_{\downarrow r}, v \in N_r(v')\}$.

The output neurons of the extended convolution layer with stride r are $V_{\downarrow r}$.

2.4.2 EXTENDED CONVOLUTIONS ON THE STRIDED GRAPH

Using the proxy-translations, we move a localized r-hop indexing kernel over G. At each location, we associate the vertices of $V_{\downarrow r}$ with weights of the kernel, thus obtaining what we define as induced $_{\downarrow r}$-translations on the set $V_{\downarrow r}$. In other words, when the kernel is centered on v_0, if $v_1 \in V_{\downarrow r}$ is associated with weight index p_0, we obtain $\phi^{\downarrow r}_{p_0}(v_0) = v_1$. Subsequent convolutions at lower scales are defined using these induced $_{\downarrow r}$-translations.

3 EXPERIMENTS

We tested our method on CIFAR-10, a tiny images dataset (Krizhevsky (2009), using a PreActResNet18 (He et al. (2016)) architecture, swapping convolutions and strided convolutions for extended ones. We tested different combinations of graph support and data augmentation. Table 1 summarizes our results. In particular, it is interesting to note that results obtained without any structure prior (91.07%) are only 2.7% away from the reference using classical CNNs on images (93.80%).

Table 1: CIFAR-10 result comparison table.

Support	MLP Lin et al. (2015)	CNN	Grid Graph		Covariance Graph
			Defferrard et al. (2016)	Proposed	Proposed
Full Data Augmentation	78.62%	**93.80%**	85.13%	93.94%	92.57%
Data Augmentation - Flip	—-	92.73%	84.41%	92.94%	91.29%
Graph Data Augmentation	—-	—-	—-	92.81%	**91.07%**

The PINES dataset consists of fMRI scans on 182 subjects, during an emotional picture rating task.(Chang et al. (2015)) To generate our dataset, we fetched individual first-level statistical maps (from `https://neurovault.org/collections/1964/`) for the minimal and maximal ratings. Full brain data was masked on the MNI template and resampled to a 16mm cubic grid, in order to reduce dimensionality of the dataset while keeping a regular geometrical structure. Using a shallow architecture, Table 2 shows that our method improves over CNNs, MLPs and other graph-based networks. The code is available at `http://www.github.com/brain-bzh/MCNN`.

Table 2: PINES fMRI dataset accuracy comparison table.

Graph	None		Neighborhood Graph	
Method	MLP	CNN (kernel 1x1)	Defferrard et al. (2016)	Proposed
Accuracy	82.62%	84.30%	82.80%	**85.08%**

4 CONCLUSION

We proposed a new methodology that extends classical convolutional neural networks to irregular domains represented by a graph. We performed experiments and showed that our method is able to match performance of classical convolutional neural networks on images without explicit knowledge about the underlying regular 2D structure. It also significantly outperforms existing alternatives.

REFERENCES

James Atwood and Don Towsley. Diffusion-convolutional neural networks. In *Advances in Neural Information Processing Systems*, pp. 1993–2001, 2016.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

Luke J Chang, Peter J Gianaros, Stephen B Manuck, Anjali Krishnan, and Tor D Wager. A sensitive and specific neural signature for picture-induced negative affect. *PLoS biology*, 13(6):e1002180, 2015.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, pp. 3837–3845, 2016.

Jian Du, Shanghang Zhang, Guanhang Wu, José MF Moura, and Soummya Kar. Topology adaptive graph convolutional networks. *arXiv preprint arXiv:1710.10370*, 2017.

David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems*, pp. 2215–2223, 2015.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. *arXiv preprint arXiv:1704.01212*, 2017.

William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. *arXiv preprint arXiv:1706.02216*, 2017.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pp. 630–645. Springer, 2016.

Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 30(8): 595–608, 2016.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Risi Kondor, Hy Truong Son, Horace Pan, Brandon Anderson, and Shubhendu Trivedi. Covariant compositional networks for learning graphs. *arXiv preprint arXiv:1801.02144*, 2018.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.

Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

Ron Levie, Federico Monti, Xavier Bresson, and Michael M Bronstein. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. *arXiv preprint arXiv:1705.07664*, 2017.

Ruoyu Li, Sheng Wang, Feiyun Zhu, and Junzhou Huang. Adaptive graph convolutional neural networks. *arXiv preprint arXiv:1801.03226*, 2018.

Zhouhan Lin, Roland Memisevic, and Kishore Reddy Konda. How far can we go without convolution: Improving fully-connected networks. *CoRR*, abs/1511.02580, 2015. URL `http://arxiv.org/abs/1511.02580`.

Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. *arXiv preprint arXiv:1611.08402*, 2016.

Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. Learning convolutional neural networks for graphs. In *Proceedings of the 33rd International Conference on Machine Learning*, pp. 2014–2023, 2016.

Giannis Nikolentzos, Polykarpos Meladianos, Antoine Jean-Pierre Tixier, Konstantinos Skianis, and Michalis Vazirgiannis. Kernel graph convolutional neural networks. *arXiv preprint arXiv:1710.10689*, 2017.

Bastien Pasdeloup, Vincent Gripon, Jean-Charles Vialatte, and Dominique Pastor. Convolutional neural networks on irregular domains through approximate translations on inferred graphs. *arXiv preprint arXiv:1710.10035*, 2017.

Gilles Puy, Srdan Kitic, and Patrick Pérez. Unifying local and non-local signal processing with graph cnns. *arXiv preprint arXiv:1702.07759*, 2017.

Martin Simonovsky and Nikos Komodakis. Dynamic edge-conditioned filters in convolutional neural networks on graphs. *arXiv preprint arXiv:1704.02901*, 2017.

Antoine Jean-Pierre Tixier, Giannis Nikolentzos, Polykarpos Meladianos, and Michalis Vazirgiannis. Classifying graphs as images with convolutional neural networks. *arXiv preprint arXiv:1708.02218*, 2017.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *stat*, 1050:20, 2017.

Jean-Charles Vialatte, Vincent Gripon, and Grégoire Mercier. Generalizing the convolution operator to extend cnns to irregular domains. *arXiv preprint arXiv:1606.01166*, 2016.

Jean-Charles Vialatte, Vincent Gripon, and Gilles Coppin. Learning local receptive fields and their weight sharing scheme on graphs. *arXiv preprint arXiv:1706.02684*, 2017.