2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)
Venice, Italy, April 8-11, 2019

MULTI-SIZE COMPUTER-AIDED DIAGNOSIS OF POSITRON EMISSION TOMOGRAPHY IMAGES USING GRAPH CONVOLUTIONAL NETWORKS

Xuandong Zhao[1], Xiang Li[2*], Ning Guo[2], Zhiling Zhou[2], Xiaxia Meng[2], Quanzheng Li[2]*
[1] College of Computer Science, Zhejiang University, Hangzhou, China. [2] Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA. *Joint first authors.

ABSTRACT

Recent advancement of deep learning-based algorithms has greatly improved the field of medical image analysis and computer-aided diagnosis/prognosis. Convolutional Neural Network (CNN) has shown superior accuracy and generalizability in performing prediction/classification tasks, thanks to its good utilization of the grid-like structure of input images in Euclidean space. In practice, one of the challenges in using classical CNN is the multi-size nature of medical images, which is especially prominent when the input images are from specific target region of interest (ROI) (e.g. tumor). Image sizes of those ROIs can vary a lot across patients, making the images difficult to be analyzed by CNNs where constant-sized inputs are expected. To address this challenge, we propose the Deep Voxel-Graph Convolution Network (DVGCN). DVGCN represents input images as their affinity graph and performs graph convolution to extract discriminative features. It then utilizes a sortpooling layer to sort the nodes and unifies the feature size used for prediction across images, thus solves multi-size challenge without explicitly resizing images. DVGCN is tested on 3D Positron-Emission Tomography (PET) images to predict the patient's cancer staging, its performance is compared with classical 3DCNN (with image padding) and radiomics models.

Index Terms—graph convolutional network, multi-size image analysis, computer-aided diagnosis

1. INTRODUCTION

Computer-aided diagnosis/prognosis based on images has been the central focus and ultimate goal of medical image analysis in both research and clinical practice. Convolutional Neural Networks (CNNs) [1] have rapidly become the method of choice for analyzing medical images, and obtained promising performance in various applications [2]. The algorithmic foundation of CNN lies in its capability of learning complex, low-to-high level representation of the input images based on spatial contextual localization defined on regular voxel grid. Classical CNN-like frameworks take advantage of several good properties of such Euclidean space, while at the same time being limited by its spatial regularity. One major yet often underestimated challenge in using CNNs is the multi-size nature of medical images. The challenge is especially prominent when performing computer-aided diagnosis/prognosis on a specific target region of interest (ROI) (e.g. tumor), where the region can be of varying sizes across patients. As common CNN needs to take constant-sized images as input in order to ensure that is convolutional kernels will extract image features of the same lengths, when dealing with multi-size images the network needs to either adopt patch-based or image padding approach (i.e. smaller images are padded by empty pixels/voxels to become larger images). Patch-based network suffers from localized view of the images (limited by patch size) and reduced resolution (limited by patch stride). Image padding, on the other hand, could be a viable solution yet with its own limitation due to imbalanced feature extraction caused by size variation, which will be investigated in detail later in this work.

To address the challenge of multi-size image input, while inspired by the flexibility of graph-based image analysis (e.g. graph cutting for segmentation [3]), we propose the Deep Voxel-Graph Convolution Network (DVGCN). Graph Convolution Networks (GCN) extends the classical CNNs to non-Euclidean space [4], utilizing either graph spectral [5] or propagation of features among neighborhood nodes [6]. DVGCN represents images as their graph counterparts by constructing the affinity graphs [7] from images. It then performs node-wise graph convolution [8] to extract graph features in a similar fashion as in classical convolutional filters, with added *SortPooling* operation [9]. We apply the proposed DVGCN on 3D Positron-Emission Tomography (PET) images for the diagnosis of binary patient labels of *lung cancer staging*. For performance evaluation, we also implement: 1) 3DCNN (with image padding) which is the current standard approach for image-based classification tasks; and 2) Radiomics-based model [10], which is the commonly applied approach in oncology and radiology [11].

2. MATERIALS AND METHODS

2.1. Affinity Graph Representation

To represent multi-size PET images as graphs, we construct the undirected graph G based on grouping algorithm [12]. Each voxel in the image (to the total number of n) is represented by a node in a graph. Based on the similarity across voxels as well as their distance, nodes are connected by edges with corresponding weights. In this work, the edge weight $W_{i,j}$ between voxel i and j within a neighborhood of (5,5,5) is defined based on the intensity similarity term and spatial proximity term:

$$W_{ij} = \lambda e^{\left(-\frac{r(i,j)}{\sigma_R}\right)} e^{\left(-\frac{\|p(i)-p(j)\|^2}{\sigma_W}\right)}. \quad (1)$$

In Eq.1 $\|\cdot\|$ is the Euclidian norm, $r(i,j)$ is the distance between voxel i and j, $p(i)$ is voxel value of a node. λ is set to $p(i)+p(j)$, reflecting the fact that in PET images voxels with higher values are more important (indicating higher uptake). σ_R, σ_W are constants to weight the importance between intensity similarity term and spatial proximity term, which are set to 1 (i.e. same importance) in this work. Weights calculated for each edge in the graph are then thresholded, so that edges with weights lower than 20% percentile are removed (i.e. only top 20% edges in the graph are retained). Weight of the edge between voxel pairs that are not within the (5,5,5) neighborhood (i.e. far away from each other) is set to zero. Sample illustrations of graph representation with various sizes are shown in Fig. 1.

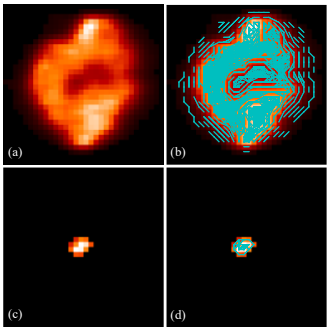

Figure. 1. (a) Visualization of a 2D slice of PET image from one sample subject with larger ROI, and (b) graph representation of it with edges colored in blue connecting each node. (c) Visualization of a 2D slice of PET image from another sample subject with smaller ROI, and (d) graph representation of it.

For each node in the affinity graph, we also define extra features upon it in addition to its voxel value, including node degree, mean affinity value (by averaging edge weights $W_{i,j}$ connecting to this node), and heterogeneity value centered at this node [13]. Node degree and mean affinity value characterize importance of the given node in the graph. Heterogeneity feature, which is calculated based on local variations in voxel image values, characterizes local texture features of the node [13].

2.2. Deep Voxel-Graph Convolution Network

By representing the given 3D images with various sizes by their corresponding affinity graphs, we can then perform deep graph convolution and classification through the proposed model ("Deep Voxel-Graph Convolution Network", or DVGCN). DVGCN firstly performs graph-based convolution to extract substructure features of each vertex [8], similar to traditional convolution operations defined in Euclidean space where local-to-global image features are extracted. Specifically, given an affinity graph matrix $G \in \mathbb{R}^{n\times n}$ with n nodes, and the node feature matrix $X^t \in \mathbb{R}^{n\times c}$ with c number of attributes defined on each node, the graph convolution operation between the t-th and $t+1$th layer is defined as:

$$Z^{t+1} = f\left(\tilde{D}^{-1}\tilde{G}X^tW^t\right), \quad (2)$$

where $\tilde{G} = G + I$ is the graph adjacency matrix with added self-loops, $\tilde{D}$ is the diagonal degree matrix of $\tilde{G}$, $W^t \in \mathbb{R}^{c\times c'}$ are the trainable graph convolution parameters which are shared among all vertices, f is the nonlinear activation function (rectified linear units ReLU is used for DVGCN), and $Z^{t+1} \in \mathbb{R}^{n\times c'}$ is the output of t-th layer containing c' number of attributes for each node. In Eq. 2, information (attributes) in each node X^t is firstly transformed by the trainable parameters W. The transformed information is then normalized by the node degrees defined in $\tilde{D}$ and propagated to the neighbors of this node as well as itself based on $\tilde{G}$. Finally, pointwise nonlinear activation function f is applied to the new information (attributes) of the graph. In DVGCN we further use *SortPooling* layer [9] inspired by the Weisfeiler-Lehman (WL) algorithm defined in [14], which sorts nodes based on graph topology. Node ordering by WL algorithm in DVGCN is consistent across graphs (i.e. nodes in two graphs will be assigned by similar order if they have similar graph structure), thus enables analysis of different graphs without explicit correspondence of their nodes. The output of *SortPooling* is pooled features from k number of nodes in the graph, where k is a user-defined integer. If $n>k$, *SortPooling* will delete the last (n-k) rows from the feature matrix; if $n<k$, it will add (k-n) number of zero rows into the feature matrix. In this way, we truncate/extend the size of extracted features from n to k, unifies input into next classification layer to the same size of k, and effectively solves the multi-size problem of the input images. After *SortPooling*, we use 1-D convolutional layers to sequentially filter the features and learn local patterns of the node sequence. Finally, several fully connected layers followed by *softmax* is used to make the binary prediction of labels.

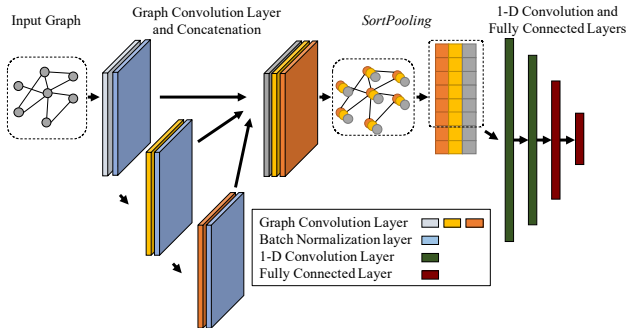

Figure. 2. Illustration of the DVGCN analytics pipeline. In the network used for this study, outputs from three graph convolution layers are concatenate and feed into the *SortPooling* layer. The sorted and pooled features are then passed to 1-D convolution and fully connected layers to predict the binary label of the input image.

2.3. Alternative Methods: 3D convolution neural networks (3DCNN) and Radiomics Models

To evaluate the performance of DVGCN against other commonly-applied methods for image-based computer aided diagnosis , we implement and test the 3D convolution neural networks (3DCNN), as well as radiomics models, on the same dataset. As 3DCNN needs constant-sized input, images will be padded to the same size before analysis. The 3DCNN used in this work takes 3D PET image of size 36*32*24 as input. It consists of five 3D convolution layers and two fully connected layers for the prediction, following a similar network structure used in our previous work [15]. We also implement a classification model based on radiomics features, which is the commonly applied approach in oncology and radiology [11]. We firstly extract radiomics features from ROIs following guidelines in [16], including intensity distribution, spatial patterns (texture), tumor geometry, etc. The extracted features are then used to train a three-layer neural network for predicting the final labels of each patient.

3. RESULTS

We use the pubic available Lung Cancer dataset [17] from the Cancer Imaging Archive (TCIA) [18] for the model development and validation. The dataset contains 135 patients with clinical information, PET images, and the corresponding ROI (region of interest) definition for the potential tumor regions. Images within the ROI are used as input for DVGCN and radiomics analysis. The same images are padded to the same size to be used as input for 3DCNN. Sizes of ROI vary a lot across different subjects, from 140 voxels to 25296 voxels. Visualizations of the largest and the smallest ROI (on 2D slices) can be found in Fig. 1. We use the label of *Cancer Staging* by *American Joint Committee on Cancer (AJCC)* from the patient information as prediction output. *AJCC Staging* is a classification system developed by the American Joint Committee on Cancer for describing the extent of disease progression in cancer patients [19]. Patients with cancer will be assigned a number (I~IV) reflecting the development of cancer. In this work patients of stage I and II are marked as "early stage/0", and those of stage III and IV are marked as "late stage/1".

For training and testing all the three models, in each experiment we randomly select 100 subjects for training and 35 for testing. The experiment is repeated for 10 times and we use the averaged result across 10 experiments to evaluate the network performance. As number of samples with label of "0" or "1" differs a lot for *AJCC Staging* labels, resulting in the majority class (i.e. label with much more samples) and minority class, we adapt imbalance adjustment strategy [20] for balancing the training set. Specifically, during the network training, every time a different bootstrap sample is drawn from the training set to fit the network, samples of minority class will be copied and used in every partition, while samples of majority classes will be randomly sampled without replacement.

Network structure of DVGCN used for the experiment contains four graph convolution layers, each of the first three layers have 32 output channels, while the fourth layer has a single output channel. Every convolutional layer is followed by batch normalization. Outputs from all the four graph convolution layers, to the total of 97 output channels, are sorted by pooled by the *SortPooling* layer. Parameter *k* of the *SortPooling* is set so that 10% of the input graphs will have nodes more than *k*, thus their nodes will be truncated. For the rest 90% of the graphs, extra nodes with zero values will be padded to them. Two fully-connected layers with dropout is connected to *SortPooling* layer to get the final prediction label for each input image.

Table 1. Performance of DVGCN, DVGCN without graph information (DVGCN*), DVGCN without extra node feature (DVGCN**), 3DCNN and radiomics model for predicting patients' labels of *AJCC Staging*.

Label	Method	Accuracy	Sensitivity	Specificity
AJCC Staging	DVGCN	**0.837**	**0.861**	0.813
	DVGCN*	0.792	0.692	0.891
	DVGCN**	0.760	0.792	0.729
	3DCNN	0.711	0.642	0.779
	Radiomics	0.821	0.662	**0.981**

Performance of DVGCN, 3DCNN and Radiomics model on the lung cancer PET images, measured by accuracy/ sensitivity/specificity of the predictions, are listed in Table 1. It should be noticed that testing data is balanced in a similar approach as for training data. It can be found that DVGCN outperforms both 3DCNN and radiomics models in the prediction of patient labels. We hypothesize that improvement of DVGCN over 3DCNN can be caused by: 1) Topology of the affinity graphs can be a useful addition to the image itself. More importantly, the *SortPooling* operation of DVGCN effectively solves the multi-size problem through sorting of the nodes. To validate this, we apply DVGCN for the same prediction but without graph information (i.e. set the affinity graphs of all images to empty graph). Result are denoted by DVGCN* in the table, showing decrease of prediction accuracy of DVGCN. 2) In DVGCN, values in each graph nodes contains rich graph and image features (node degree, node mean affinity, image heterogeneity) in addition to the voxel values. These features can be helpful in its prediction of patient labels. To validate, we remove these extra features (node degree, mean affinity and heterogeneity) from the nodes and only retain the voxel values. Results are donated by DVGCN** in the table, also show performance decrease comparing with DVGCN model with full features.

4. CONCLUSION

To address the challenge of multi-size image analysis for computer-aided diagnosis, in this work we develop and implement a graph-based deep learning model Deep Voxel-Graph Convolution Network (DVGCN). The proposed model shows good performance on the testing dataset consisting of 3D Positron-Emission Tomography (PET) images and their patient labels, with good flexibility of incorporating more image/graph features, and the advantage of utilizing topological structure of the images. Currently the method is only used for predicting one type of patient label (i.e. *staging*), while we are formulating a multi-task framework for simultaneous prediction of multiple patient labels. We are further investigating patterns in the extracted graph features, in order to identify key factors in the graphs that are affecting the prediction results, which will be valuable to radiologists and physicians. We are also testing DVGCN on more datasets consisting of multi-modal 2D/3D images for its generalizability in different application scenarios. Source code of DVGCN is available at our GitHub page: https://github.com/XuandongZhao/DVGCN.

5. ACKNOWLEDGEMENT

This work was supported by the National Institutes of Health under grant RF1AG052653, P41EB022544 and C06 CA059267. We also acknowledge MGH & BWH Center for Clinical Data Science for support of computational resources.

5. REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*: MIT Press, 2016.

[2] J. H. Thrall, X. Li, Q. Li, C. Cruz, S. Do, K. Dreyer, and J. Brink, "Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success," *Journal of the American College of Radiology,* vol. 15, no. 3, pp. 504-508, 2018.

[3] D. M. Greig, B. T. Porteous, and A. H. Seheult, "Exact Maximum A Posteriori Estimation for Binary Images," *Journal of the Royal Statistical Society. Series B (Methodological),* vol. 51, no. 2, pp. 271-279, 1989.

[4] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric Deep Learning: Going beyond Euclidean data," *IEEE Signal Processing Magazine,* vol. 34, no. 4, pp. 18-42, 2017.

[5] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral Networks and Locally Connected Networks on Graphs," *arXiv:1312.6203*, 2013.

[6] J. Atwood, and D. Towsley, "Diffusion-Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NIPS), 2016.

[7] X. Zhu, C. C. Loy, and S. Gong, "Constructing Robust Affinity Graphs for Spectral Clustering." pp. 1450-1457.

[8] H. Gao, Z. Wang, and S. Ji, "Large-Scale Learnable Graph Convolutional Networks," in Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 1416-1424.

[9] M. Zhang, Z. Cui, M. Neumann, and Y. Chen, "An End-to-End Deep Learning Architecture for Graph Classification."

[10] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. P. M. van Stiphout, P. Granton, C. M. L. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. W. L. Aerts, "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer,* vol. 48, no. 4, pp. 441-446, 2012.

[11] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology,* vol. 278, no. 2, pp. 563-577, 2016.

[12] S. Jianbo, and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 888-905, 2000.

[13] N. Guo, R. Yen, G. E. Fakhri, and Q. Li, "SVM based lung cancer diagnosis using multiple image features in PET/CT," in IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2015, pp. 1-4.

[14] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48, New York, NY, USA, 2016, pp. 2014-2023.

[15] X. Li, A. Zhong, M. Lin, N. Guo, M. Sun, A. Sitek, J. Ye, J. Thrall, and Q. Li, "Self-paced Convolutional Neural Network for Computer Aided Detection in Medical Imaging Analysis," *Machine Learning in Medical Imaging: 8th International Workshop, MLMI 2017, Held in Conjunction with MICCAI 2017, Quebec City, QC, Canada, September 10, 2017, Proceedings*, Q. Wang, Y. Shi, H.-I. Suk and K. Suzuki, eds., pp. 212-219, Cham: Springer International Publishing, 2017.

[16] S. Tan, S. Kligerman, W. Chen, M. Lu, G. Kim, S. Feigenberg, W. D. D'Souza, M. Suntharalingam, and W. Lu, "Spatial-Temporal FDG-PET Features for Predicting Pathologic Response of Esophageal Cancer to Neoadjuvant Chemoradiation Therapy," *International Journal of Radiation Oncology Biology Physics,* 85-5, pp. 1375-1382, 2013.

[17] M. Vallières, C. R. Freeman, S. R. Skamene, and I. E. Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in Medicine & Biology,* vol. 60, no. 14, pp. 5471, 2015.

[18] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging,* vol. 26, no. 6, pp. 1045-1057, 2013.

[19] C. M. Balch, J. E. Gershenwald, S.-j. Soong, J. F. Thompson, M. B. Atkins, D. R. Byrd, A. C. Buzaid, A. J. Cochran, D. G. Coit, S. Ding, A. M. Eggermont, K. T. Flaherty, P. A. Gimotty, J. M. Kirkwood, K. M. McMasters, M. C. Mihm, D. L. Morton, M. I. Ross, A. J. Sober, and V. K. Sondak, "Final Version of 2009 AJCC Melanoma Staging and Classification," *Journal of Clinical Oncology,* vol. 27, no. 36, pp. 6199-6206, 2009.

[20] X. Zhu, H.-I. Suk, S.-W. Lee, D. Shen, and I. the Alzheimer's Disease Neuroimaging, "Subspace Regularized Sparse Multi-Task Learning for Multi-Class Neurodegenerative Disease Identification," *IEEE transactions on bio-medical engineering,* vol. 63, no. 3, pp. 607-618, 2016.