Multi-instance Deep Learning with Graph Convolutional Neural Networks for Diagnosis of Kidney Diseases Using Ultrasound Imaging

Shi Yin[1,2(✉)], Qinmu Peng[1], Hongming Li[2], Zhengqiang Zhang[1], Xinge You[1], Hangfan Liu[2], Katherine Fischer[5,6], Susan L. Furth[4], Gregory E. Tasian[3,5,6], and Yong Fan[2]

[1] School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China
yinshi.wh@gmail.com
[2] Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[3] Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA
[4] Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[5] Department of Surgery, Division of Pediatric Nephrology, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[6] Center for Pediatric Clinical Effectiveness, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

Abstract. Ultrasound imaging (US) is commonly used in nephrology for diagnostic studies of the kidneys and lower urinary tract. However, it remains challenging to automate the disease diagnosis based on clinical 2D US images since they provide partial anatomic information of the kidney and the 2D images of the same kidney may have heterogeneous appearance. To overcome this challenge, we develop a novel multi-instance deep learning method to build a robust classifier by treating multiple 2D US images of each individual subject as multiple instances of one bag. Particularly, we adopt convolutional neural networks (CNNs) to learn instance-level features from 2D US kidney images and graph convolutional networks (GCNs) to further optimize the instance-level features by exploring potential correlation among instances of the same bag. We also adopt a gated attention-based MIL pooling to learn bag-level features using full-connected neural networks (FCNs). Finally, we integrate both instance-level and bag-level supervision to further improve the bag-level classification accuracy. Ablation studies and comparison results have demonstrated that our method could accurately diagnose kidney diseases using ultrasound imaging, with better performance than alternative state-of-the-art multi-instance deep learning methods.

Keywords: Automatic diagnosis · Ultrasound imaging · Graph convolutional neural networks · Multi-instance learning

H. Greenspan et al. (Eds.): CLIP 2019/UNSURE 2019, LNCS 11840, pp. 146–154, 2019.
https://doi.org/10.1007/978-3-030-32689-0_15

1 Introduction

Ultrasound imaging (US) is commonly used in nephrology for diagnostic studies of the kidneys and urinary tract. Anatomic measures of the kidney computed from US data, such as renal parenchymal area, are correlated with kidney function [1], and pattern classifiers built upon US imaging features could aid kidney disease diagnosis [2, 3]. Recent deep learning studies have demonstrated that automated US data analysis could achieve promising performance in a variety of US data analysis tasks, including segmentation and classification [4–6]. However, it remains challenging to automate the kidney disease diagnosis based on clinical 2D US scans since in clinical practice multiple 2D US scans of the same kidney in different views are often collected and the multi-view 2D US scans have heterogeneous appearance, providing partial anatomic information of the kidney, as illustrated by Fig. 1. Therefore, it is desired to develop a clinically useful diagnosis model that is robust to different views of US images of the same kidney.

Multiple instance learning (MIL) is an ideal tool to build a robust classifier on multi-view 2D US scans of the same kidneys by treating multi-views of 2D US scans of the same kidney as multiple instances of a bag and predicting a bag-level classification label [7]. To effectively solve the MIL classification problem, a number of methods have been developed [7]. Among the existing MIL methods, neural network based methods have demonstrated promising performance in a variety of MIL problems, partially due to its end-to-end learning capability [8–11]. Particularly, neural networks could be used to estimate instance-level classification probabilities and fuse them with a log-sum-exp based max operator [8] or a max operator [9] to generate a bag-level classification probability in an end-to-end learning framework. Since the instance-level classification might be affected by instance label instability problem [12], several embedded-space based deep MIL methods have been developed to learn informative features at the instance level, generate a bag mapping with a permutation-invariant MIL pooling operator, and build a bag-level classifier on the embedded-space in an end-to-end learning framework [10, 11]. Particularly, an attention-based MIL pooling has been develop to learn a weighted average of instances [11].

Fig. 1. Multi-view 2D US scans of the same kidney. The images shown on the 1st column have abnormal appearance annotated by radiologists, while others shown on the 2nd and 3rd columns have heterogeneous appearance.

However, the existing deep MIL methods ignore classification labels of instances that are often available in training data and could potentially improve the MIL classification performance if properly integrated, such as those shown on the 1st column of Fig. 1. Furthermore, potential correlation between instances of the same bag has not been well explored in the existing deep MIL methods, which may lead to suboptimal instance-level features. In order to overcome these limitations and further improve deep MIL methods, we develop a novel deep MIL method to learn a deep MIL classification model in an end-to-end learning framework, and apply it to kidney disease diagnosis based on multi-view 2D US images. Particularly, we build a MIL classifier to distinguish kidneys from patients with different kidney diseases based on their multi-view 2D US images. We adopt convolutional neural networks (CNNs) [13] to learn informative US image features, and adopt graph convolutional neural networks (GCNs) [14] as a permutation-invariant operator to further optimize the instance-level CNN features by exploring potential correlation among different instances of the same bag. We adopt the attention-based MIL pooling to learn an optimal permutation-invariant MIL pooling operator in conjunction with learning a bag-level classifier on the embedded space [11]. We further adopt instance-level supervision to enhance the learning of instance features with a focus on instances with reliable labels in the training data. We have validated our method based on clinical 2D US images collected from patients at a local hospital. Extensive comparison and ablation studies have demonstrated that the proposed method could improve the deep MIL methods.

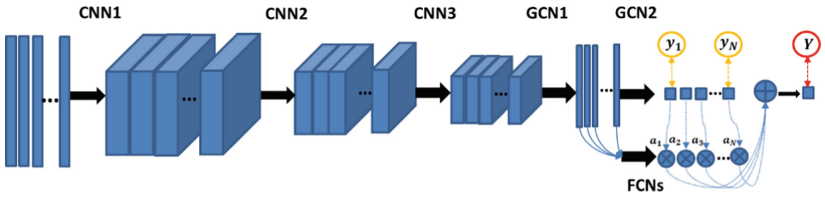

Fig. 2. Network architecture of the proposed deep MIL method. The instance-level supervision is denoted by the yellow circles and the bag-level supervision is denoted by the red circle. (Color figure online)

2 Methodology

We model the kidney disease diagnosis problem based on multi-view 2D US kidney images as a MIL classification problem. Particularly, given kidneys $X_i, i = 1, \ldots, N$, and their 2D US scans in different views, $x_{ik}, k = 1, \ldots, K$, their class label $Y_i = 0$ if all x_{ik} are normal, otherwise $Y_i = 1$. We build a deep MIL network upon recent advances in deep MIL methods that facilitate end-to-end optimization of learning informative features at the instance level, generate a bag mapping with a permutation-invariant MIL pooling operator to embed bags into an embedded-space, and build a bag-level classifier on the embedded-space [10, 11]. As illustrated by Fig. 2, our network consists of CNNs to learn instance-level features from 2D US kidney images, GCNs as permutation-invariant operators to further improve instance-level features, the attention-based MIL

pooling to learn a bag-level classifier using full connected neural networks (FCNs), and instance-level supervision to enhance the instance level feature learning and the bag-level classification.

2.1 Learning Image Features for Instances Using CNNs and GCNs with Instance-Level Supervision

To learn informative image features from 2D US kidney images, we adopt CNNs in conjunction with nonlinear activation functions, as illustrated by Fig. 2. Particularly, each 2D US kidney image, x_{ik}, is first fed into multiple layers of CNNs followed by nonlinear activation functions (in the present study, we use 3 CNN layers and *ReLU* activation). We denote the CNN output of x_{ik} by h_{ik}.

As illustrated by Fig. 1, instances of the same bag are potentially correlated with each other. Such correlation information could not be utilized by the CNNs since they are applied to individual instances of the same bag separately. For modeling such unorganized instances, graph theory-based modeling provides an effective means. By modeling each instance as a graph node, and connecting every pair of instances weighted by their similarity measure, we could model the instances with an undirected graph in a graph convolutional network (GCN) framework [14]. Particularly, new features on the graph nodes could be learned by optimizing weights of GCNs.

Given the CNN features of different instances of the same bag, $h_{ik}, k = 1, \ldots, K$, we build a bag-level graph $G = \{V, E\}$ by treating each instance as a graph node and connecting each pair of nodes with a weight measuring their similarity based on their CNN features. GCNs are then adopted to learn new features on the graph nodes [14]:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} A \tilde{D}^{-\frac{1}{2}} (H^{(l)})^T W^{(l)}), \tag{1}$$

where A with its element denoted by a_{ij} is a symmetric adjacent matrix of the undirected graph $G, \tilde{D}_{ii} = \sum_j a_{ij}$ is its degree matrix, $W^{(l)}$ is a layer-specific trainable weight matrix of GCNs, $\sigma(\cdot)$ is a nonlinear activation function, $H_i^{(l)} = \{h_{i1}^l, h_{i2}^l, \ldots, h_{iK}^l\}, h_{ik}^l \in R^F$ is a set of node features obtained by the l^{th} GCN layer, $H_i^{(l+1)} = \{h_{i1}^{l+1}, h_{i2}^{l+1}, \ldots, h_{iK}^{l+1}\}, h_k^{l+1} \in R^M$ is a set of new nodal features obtained by the $(l+1)^{\text{th}}$ GCN layer, K is the number of nodes, F and M are the numbers of features on each node obtained by the l^{th} and the $(l+1)^{\text{th}}$ GCN layers, respectively.

In the present study, we adopt a Euclidean distance function to obtain the adjacency matrix based on the input feature $H^{(l)}$:

$$a_{ij} = exp\left(-\left\|h_i^l - h_j^l\right\|^2\right). \tag{2}$$

To guide the feature learning, an instance-level supervision is adopted. Particularly, instances with reliable positive labels and all negative instances are used to optimize the feature learning using a softmax loss function. For a two-layer GCN, its forward model takes the simple form:

$$Z = f\left(H^l\right) = A_2 \mathrm{ReLU}\left(A_1 (H^{(0)})^T W^0\right) W^1, \tag{3}$$

where $W^{(0)} \in R^{F \times M}$ and $W^{(l)} \in R^{M \times 2}$ are GCN weight matrices, $H^{(0)}$ is the input CNN features, $A_{i=1,2}$ is the i^{th} layer adjacency matrix which is computed based on the i^{th} layer input features. The second GCN layer yields the instance-level feature $Z^T = \{z_1, z_2, \ldots, z_K\}$ with $z_k \in R^2$, and the instance-level classification probability $P^T = \{p_1, p_2, \ldots p_K\}$ is obtained by applying a row-wise softmax activation function.

2.2 Attention-Based MIL Pooling Layer with a Gating Mechanism

Once we obtain the instance-level features, we aggregate them to obtain an embedded-space representation using a MIL pooling operator. Instead of adopting simple mean or max MIL pooling, we adopt a gated attention-based MIL pooling layer [11]. Particularly, the attention-based MIL pooling layer learns a weighted average operator to aggregate instance features with a gating mechanism. Given a bag of K instances with GCN features $H^{l+1} = \left\{h_1^{l+1}, h_2^{l+1}, \ldots, h_K^{l+1}\right\}, h_k^{l+1} \in R^M$, gated attention-based MIL pooling weight a_k is computed as

$$a_k = \frac{\exp\{w^T\left(\tanh\left(\boldsymbol{V}\left(h_k^{l+1}\right)^T\right) \bullet \mathrm{sigm}\left(\boldsymbol{U}\left(h_k^{l+1}\right)^T\right)\right)}{\sum_{j=1}^{K} \exp\left\{w^T\left(\tanh\left(\boldsymbol{V}\left(h_j^{l+1}\right)^T\right) \bullet \mathrm{sigm}\left(\boldsymbol{U}\left(h_j^{l+1}\right)^T\right)\right)\right\}}, \tag{4}$$

where $w \in R^{L \times 1}$ and $\boldsymbol{V}, \boldsymbol{U} \in R^{L \times M}$ are parameters to be optimized, $\bullet$ is an element-wise multiplication, sigm $(\cdot)$ is the sigmoid non-linear activation function, and tanh $(\cdot)$ is used as the gating mechanism. So, the embedded-space representation of bag Z_X is defined as:

$$Z_X = \sum\nolimits_{k=1}^{K} a_k z_k. \tag{5}$$

Once the embedded-space representation of bags is obtained, a softmax operation is applied to obtain the bag positive score P_X.

2.3 Jointly Training the Instance-Level and Bag-Level Loss Functions

Once the bag positive score P_X is obtained, the bag-level loss function is defined as:

$$L_X = -\{Y ln P_X + (1 - Y) \ln(1 - P_X)\}, \tag{6}$$

To utilize information of instances with reliable labels, we also generate instance-level classification results by optimizing a cross-entropy loss function.

$$L_M = -\sum\nolimits_{n \in N_Y} \sum\nolimits_{c=1}^{2} y_{nc} ln p_{nc}, \tag{7}$$

where p_{nc} is the classification probability of an instance, y_{nc} is its ground truth classification label, and N_Y is the set of node indices that have reliable classification labels in a bag X. Finally, an overall loss function is defined as:

$$L = L_M + L_X. \tag{8}$$

3 Experimental Results

3.1 Clinical US Kidney Scans

We evaluated our method based on a data set of clinical US kidney scans of kidney patients collected at the Children's Hospital of Philadelphia (CHOP). The work described has been carried out in accordance with the Declaration of Helsinki. The study has been reviewed and approved by the institutional review board.

Participants were randomly sampled from two patient groups. Particularly, one group of the patients were children with mild hydronephrosis (MH) which does not affect the echogenicity, growth, or function of the affected or contralateral normal kidney. The other group of the patients were children with Congenital anomalies of the kidneys and urinary tract (CAKUT), with varying degrees of increased cortical echogenicity, decreased corticomedullary differentiation, and hydronephrosis. All images were obtained for routine clinical care. The first US scans after birth were used, and all identifying information was removed. In total, we obtained 105 MH patients with 2246 scans and 120 CAKUT patients with 2687 scans. All the MH scans were labeled as negative instances with reliable classification labels, all CAKUT scans were labeled as positive instances, and 335 of CAKUT scans with noticeable abnormality from different patients were deemed as instances with reliable classification labels. All the US scans were resized to have a spatial resolution of 321 $\times$ 321, and their image intensities were linearly scaled to [0, 255].

3.2 Implementation Details

Our network consisted of 3 layers of CNNs, and their numbers of channels were set to 128, 64 and 32 respectively. All the CNNs had the same kernel sizes of 5 $\times$ 5 and the same stride sizes of 2. Our GCNs had 2 layers, and their numbers of hidden features were set to 64. In the attention-based MIL pooling network, the number of hidden

Table 1. Comparison results of different versions of the proposed method (mean±std).

Method	Accuracy (mean±std)
Bag-level MILNN	0.852 ± 0.058
Bag-level MILNN+attention	0.852 ± 0.016
Bag-level MILNN+GNN+attention	0.869 ± 0.000
Bag-level MILNN+GNN+attention+all instance supervised	0.869 ± 0.064
Proposed	0.886 ± 0.032

nodes L was set to 64. The learning rate was 0.0001 and batch size was set as 6. The maximum number of iteration steps was set to 20000. All the methods were implemented using TensorFlow and executed on a GeForce GTX 6.00 GB GPU.

3.3 Ablation Studies and Comparisons with Alternative Methods

We compared the proposed network with its degraded versions to investigate how GCNs, the attention-based MIL pooling, and the instance-level supervision contribute to the overall classification based on validation datasets. All the networks had the same number of parameters. Particularly, we first implemented the proposed network with only the bag-level loss function (Bag-level MINN), but without the GCNs (replaced the GCNs with FCNs having the same hidden nodes), the attention-based MIL pooling, and the instance-level loss. Then, Bag-level MINN was enhanced by adding the attention-based MIL pooling (Bag-level MINN+attention"), the GCNs (Bag-level MINN+GNN+attention), and the instance-level supervision (Bag-level MINN+GNN+attention+all instance supervised). In the implementation of Bag-level MINN+GNN+attention+all instance supervised, all instances of the positive bags were labelled as positive.

In the ablation studies, we randomly selected 79 MH and 99 CAKUT patients as a training data set, random 45 subjects from the remaining dataset were used as a validation data set. This procedure was repeated twice to estimate the classification performance of different versions of the proposed method. Their classification results are summarized in Table 1, demonstrating that GCNs, the attention-based MIL pooling, and the instance-level supervision based on instances with reliable labels could improve the MIL classification performance. Particularly, these results also indicated that the instance-level supervision based on all instance might be affected by the instance label instability problem [12].

We further evaluated our method and compared it with state-of-the-art MIL methods, including CNN based instance level classification with max MIL pooling (minet) [9], embedded-space based deep MIL method with mean (Minet+mean) MIL pooling [10], as well as embedded-space based deep MIL with an attention-based MIL pooling (Gated-Attention) [11]. All the deep MIL methods under comparison had the same CNNs with the same numbers of parameters. In the minet, we labelled all instances of the positive bags as positive instances. The classification performance of these methods were estimated using 5-fold cross-validation. All the classification results are summarized in Table 2. These results further demonstrated that our method could improve the classification performance of the state-of-the-art MIL methods.

Table 2. Comparison results of different MIL methods (mean±std)

Method	Accuracy	Sensitivity	Specificity
Minet	0.6488 ± 0.0852	0.5917 ± 0.1037	0.7143 ± 0.1683
Minet	0.8044 ± 0.0656	0.8250 ± 0.1229	0.7809 ± 0.1372
Gated-Attention	0.8222 ± 0.0471	0.8083 ± 0.0228	0.8381 ± 0.0865
Proposed	0.8489 ± 0.0365	0.8582 ± 0.0697	0.8381 ± 0.0865

Finally, we adopted Grad-CAM to identify informative image regions for the classification [15]. Figure 3 shows Grad-CAM maps of two randomly selected testing subjects with CAKUT. Particularly, instances with relatively larger weights learned by the attention-based MIL pooling are shown from left to right, indicating that our method could capture clinically meaningful image features.

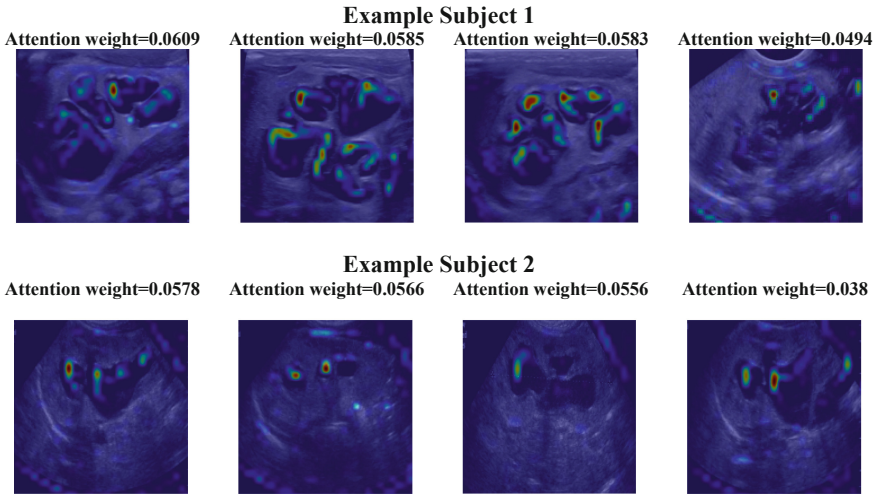

Fig. 3. Two examples of the multi-instance Grad-CAM maps of abnormal subjects with relatively lager attention weights. The largest weight across all test subjects was about 0.06.

4 Conclusions

In this study, we develop a novel multi-instance deep learning method to build a robust classifier to aid kidney disease diagnosis using ultrasound imaging. Our method is built upon recent advance in deep MIL on the embedded-space [10, 11] with novel components, including the GCNs to optimize the instance-level features learned by CNNs and the integrated instance-level and bag-level supervision to improve the classification. Extensive ablation studies and comparison experiments have demonstrated that our method could improve state-of-the-art deep MIL methods for the kidney disease diagnosis. Our future work will be devoted to automatic network architecture optimization and extensive validation of the proposed method based on different data sets.

Acknowledgements. This work was supported in part by the National Institutes of Health (DK117297 and DK114786); the National Center for Advancing Translational Sciences of the National Institutes of Health (UL1TR001878); the National Natural Science Foundation of China (61772220 and 61473296); the Key Program for International S&T Cooperation Projects of China (2016YFE0121200); the Hubei Province Technological Innovation Major Project (2017AAA017 and 2018ACA135); the Institute for Translational Medicine and Therapeutics' (ITMAT) Transdisciplinary Program in Translational Medicine and Therapeutics, and the China Scholarship Council.

References

1. Cost, G.A., Merguerian, P.A., Cheerasarn, S.P., Shortliffe, L.M.D.: Sonographic renal parenchymal and pelvicaliceal areas: new quantitative parameters for renal sonographic followup. J. Urol. **156**, 725–729 (1996)
2. Sharma, K., Virmani, J.: A decision support system for classification of normal and medical renal disease using ultrasound images: a decision support system for medical renal diseases. Int. J. Ambient Comput. Intell. **8**, 52–69 (2017)
3. Subramanya, M.B., Kumar, V., Mukherjee, S., Saini, M.: SVM-based CAC system for B-mode kidney ultrasound images. J. Digit. Imaging **28**, 448–458 (2015)
4. Liu, S., et al.: Deep learning in medical ultrasound analysis: a review. Engineering **5**, 261–275 (2019)
5. Yin, S., et al.: Subsequent boundary distance regression and pixelwise classification networks for automatic kidney segmentation in ultrasound images. arXiv preprint arXiv:1811.04815 (2018)
6. Yin, S., et al.: Fully-automatic segmentation of kidneys in clinical ultrasound images using a boundary distance regression network. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1741–1744 (2019)
7. Amores, J.: Multiple instance classification: review, taxonomy and comparative study. Artif. Intell. **201**, 81–105 (2013)
8. Ramon, J., De Raedt, L.: Multi instance neural networks (2000)
9. Zhou, Z.-H., Zhang, M.-L.: Neural networks for multi-instance learning. In: Proceedings of the International Conference on Intelligent Information Technology, Beijing, China, pp. 455–459 (2002)
10. Wang, X., Yan, Y., Tang, P., Bai, X., Liu, W.: Revisiting multiple instance neural networks. Pattern Recogn. **74**, 15–24 (2018)
11. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: Jennifer, D., Andreas, K. (eds.) Proceedings of the 35th International Conference on Machine Learning, PMLR, Proceedings of Machine Learning Research, vol. 80, pp. 2127–2136 (2018)
12. Cheplygina, V., Sørensen, L., Tax, D.M.J., de Bruijne, M., Loog, M.: Label stability in multiple instance learning. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 539–546. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24553-9_66
13. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
14. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
15. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. arXiv:1610.02391 (2016)