2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) | 978-1-6654-1246-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISBI48211.2021.9433956

MGA-NET: MULTI-SCALE GUIDED ATTENTION MODELS FOR AN AUTOMATED DIAGNOSIS OF IDIOPATHIC PULMONARY FIBROSIS (IPF)

Wenxi Yu,[1, 2, 3] *Hua Zhou,*[3] *Youngwon Choi,* [4] *Jonathan G.Goldin,*[1, 2] *Grace Hyun J. Kim* [1, 2, 3]

[1] Center for Computer Vision and Imaging Biomarkers, University of California, Los Angeles, USA
[2]Department of Radiological Sciences, University of California, Los Angeles, USA
[3] Department of Biostatistics, University of California, Los Angeles, USA
[4] Department of Statistics, Seoul National University, Seoul, South Korea

ABSTRACT

We propose a Multi-scale, domain knowledge-Guided Attention model (MGA-Net) for a weakly supervised problem - disease diagnosis with only coarse scan-level labels. The use of guided attention models encourages the deep learning-based diagnosis model to focus on the area of interests (in our case, lung parenchyma), at different resolutions, in an end-to-end manner. The research interest is to diagnose subjects with idiopathic pulmonary fibrosis (IPF) among subjects with interstitial lung disease (ILD) using an axial chest high resolution computed tomography (HRCT) scan. Our dataset contains 279 IPF patients and 423 non-IPF ILD patients. The network's performance was evaluated by the area under the receiver operating characteristic curve (AUC) with standard errors (SE) using stratified five-fold cross validation. We observe that without attention modules, the IPF diagnosis model performs unsatisfactorily (AUC±SE $=0.690 \pm 0.194$); by including unguided attention module, the IPF diagnosis model reaches satisfactory performance (AUC±SE $=0.956 \pm 0.040$), but lack explainability; when including only guided high- or medium- resolution attention, the learned attention maps highlight the lung areas but the AUC decreases; when including both high- and medium- resolution attention, the model reaches the highest AUC among all experiments (AUC± SE $=0.971 \pm 0.021$) and the estimated attention maps concentrate on the regions of interests for this task. Our results suggest that, for a weakly supervised task, MGA-Net can utilize the population-level domain knowledge to guide the training of the network in an end-to-end manner, which increases both model accuracy and explainability.

***Index Terms*—** Attention models, domain knowledge, idiopathic pulmonary fibrosis, medical imaging

1. INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a specific form of progressive, irreversible, and usually lethal lung disease of unknown causes [1]. Making a correct and reliable IPF diagnosis is critical for choosing the appropriate treatment and directly influences patients' survival time. However, IPF diagnosis based on CT scans is a difficult task and is largely subjected to inter-observer variability. To this end, this work aims to develop a deep learning-based automated diagnosis of IPF based on axial chest CT scans.

In recent years, numerous deep learning-based algorithms have achieved great success in various medical imaging tasks, such as segmentation, diagnosis, and detection [2]. The successful application of deep learning systems usually depend on these three requirements: (1) the availability of well-labeled fine-scale data, usually at pixel, regions of interests (RoI), or image slice level; (2) the extent of explainability on where and how the deep learning-based system makes the decision; and (3) the ability to generalize well to a new dataset. To this end, we build an attention-based model that is generally applicable to weakly supervised tasks, where only coarse-level (in our case, CT scan-level) labels are available, to enhance the explainability and generalizability.

Attention mechanisms, originated from natural language processing, have gained research interests to deal with label scarcity, strengthen model generalizability to a new dataset, and encourage long-range dependencies in computer vision [3] [4] [5]. Attention mechanisms are one way to explain which region of the image the network's decision depends on and can be used to improve explainability of deep learning-based systems. Attention mechanisms have recently become popular in the medical imaging domain to solve the research question of segmentation [6], classification [7], detection, and so on.

Attention models fall under two main categories - unguided (without external guidance) and guided (guided by external domain knowledge). The majority of the current work focus on building *unguided* attention mechanisms within different layers of the constructed networks, without providing external guidance of domain knowledge. For example, Schlemper et al. [6] used the coarse features extracted at later layers to guide the training under an attention model, without providing external guidance. Recent work on *guided* attention models include using region-level coarse annotation [7] or binary maps of some RoIs [8] to guide the model in an end-to-end training fashion. In this work, we design an attention model under the guidance of *population-level* domain knowledge, which is less labor-intensive to acquire, compared to the previous work [8] [7].

To summarize, MGA-Net addresses the IPF diagnosis problem by leveraging the multi-scale domain knowledge using a guided attention model. Our contributions are (1) developing an IPF diagnosis model that only uses scan-level weak supervision; (2) incorporating population-level domain knowledge into the training of IPF diagnosis model in an end-to-end manner; (3) enhancing the explainability of deep learning systems at various layers by introducing multi-scale attention mechanisms.

2. METHODS

2.1. Datasets and image preprocessing

Volumetric non-contrast chest HRCT scans with thin slices were retrospectively collected from five studies, including two IPF (N=279) and three non-IPF interstitial lung disease (ILD) cohorts (N=423). HRCT scans underwent an in-house image preprocessing pipeline, including creating lung windows based on Hounsfield units, aligning patients' positions, automatically cropping the scans based on patient's body by canny edge detector, resampling to a uniform cube of size $1 \times 1 \times 1\ mm^3$, resizing to a uniform scale by cubic spline interpolation, and standardizing to a range of [0,1] on a scan level. After preprocessing, each CT scan was resized to a standardized dimension (128, 256, 256). To boost sample size and reduce data dimension, we further resampled a fixed number (in our case, 20) of 3D-volume (dimension: 64, 128, 128) from each scan. The image dimension is represented as (z, x, y) throughout this manuscript, where z-axis is the dimension along the patient's body from apex to base and x-y plane is the axial plane of the HRCT scans.

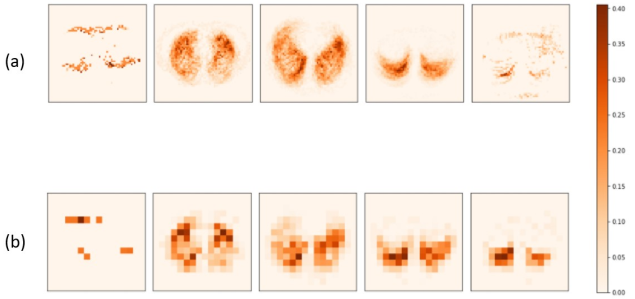


Fig. 1. Population-level domain knowledge at high (a) and medium (b) resolutions. Subplots (a) are produced at the z-axis of 0, 8, 16, 24, 30; Subplots (b) are produced at the z-axis of 0, 2, 4, 5, 6.

2.2. Population-level domain knowledge

In the past ten years, quantitative CT imaging biomarkers have been used as clinical surrogate measures among patients with interstitial lung diseases [9]. These developed measures are sensitive to localized changes and can be used as domain knowledge to guide the training of IPF diagnosis model.

We calculated the marginal probability of getting lung fibrosis (LF_i) and other lung fibrosis (OLF_i) for each voxel location i, among IPF patients based on prior studies [9]. We defined the domain knowledge (D_i) as the maximum of LF_i and OLF_i for each fixed location i: $D_i = \max(LF_i, OLF_i)$). By definition, D_i, LF_i, and OLF_i all range from $[0, 1]$. Domain knowledge (D_i) is later downsampled to two resolution scales: (32,64,64) and (8,16,16) by cubic spline interpolation, as shown in Fig.1.

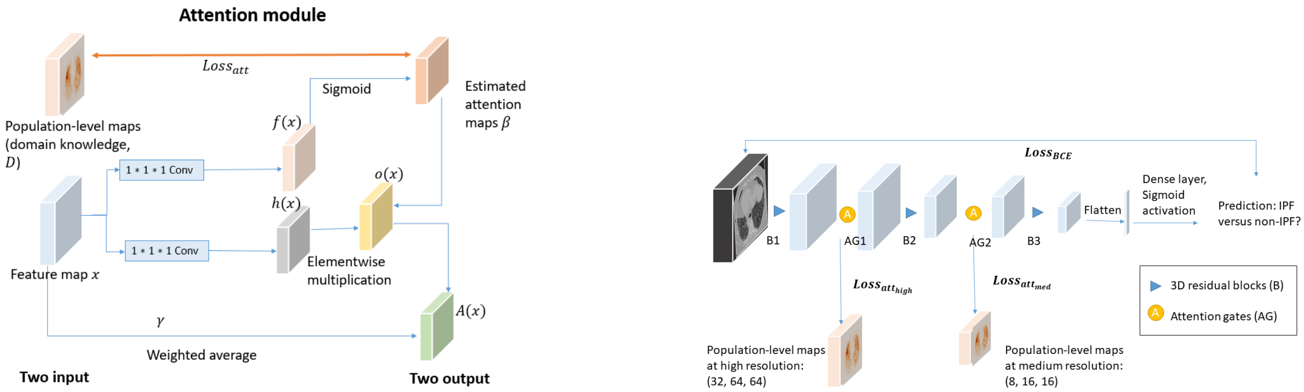
2.3. Attention gates

Inspired by [3], we provide a schematic of the proposed guided attention model in Fig.2 (a). The intermediate feature maps (x) are first transformed into two feature spaces $f(x)$ and $h(x)$ using $1 \times 1 \times 1$ convolutions, where $f(x) = W_f(x)$ and $h(x) = W_h(x)$. Sigmoid function is applied to the feature space $f(x)$ to calculate the attention scores (i.e. estimated attention maps) at location i, β_i, where $\beta_i = \frac{1}{1+e^{(-f(x_i))}}$. We use mean absolute error as the attention-based loss: $Loss_{att} = \frac{\sum_{i=1}^{N} |\hat{\beta}_i - D_i|}{N}$, where $\hat{\beta}_i$ is the marginal estimated attention maps at location i across all training samples, D_i is the domain knowledge map at location i, and N is the number of voxels within the attention maps.

In addition, we further calculate the elementwise multiplication of feature space $h(x)$ and the marginal estimated attention scores $o(x_i) = \hat{\beta}_i \times h(x_i)$. The final output of the attention model is a weighted average of the input intermediate feature maps x and $o(x)$: $A(x_i) = \gamma \times o(x_i) + (1-\gamma) \times x_i$, where γ is a trainable parameter and γ is initialized at zero.

2.4. Overall proposed method: MGA-Net

The overall schematic diagram of MGA-Net is provided in Fig.2 (b). 3D-residual blocks are used as building blocks for our model, which is shown as $B1, B2$, and $B3$. The attention gates are incorporated into the training of the IPF diagnosis model in an end-to-end manner, at two resolution scales, shown as $AG1$ and $AG2$. The overall loss function of the system is composed of a weighted average of two attention-based losses and one diagnosis-based loss: $Loss_{overall} = Loss_{BCE} + \lambda_{high} Loss_{att_{high}} + \lambda_{med} Loss_{att_{med}}$, where $Loss_{BCE}$ is the binary cross entropy for IPF diagnosis, $Loss_{att_{med}}$ is the attention-based loss at a medium resolution, $Loss_{att_{high}}$ is the attention-based loss at a high resolution. λ_{high} and λ_{med} are the relative task importance for the high- and medium- resolution attention model, respectively, with


Attention module
$Loss_{att}$
Sigmoid
Estimated attention maps β
Population-level maps (domain knowledge, D)
$f(x)$
$1*1*1$ Conv
$o(x)$
$h(x)$
$1*1*1$ Conv
Feature map x
Elementwise multiplication
$A(x)$
γ
Weighted average
Two input
Two output
$Loss_{BCE}$
B1
AG1
B2
AG2
B3
Flatten
Dense layer, Sigmoid activation
Prediction: IPF versus non-IPF?
$Loss_{att_{high}}$
$Loss_{att_{med}}$
3D residual blocks (B)
Attention gates (AG)
Population-level maps at high resolution: (32, 64, 64)
Population-level maps at medium resolution: (8, 16, 16)

(a) The proposed attention gates.

(b) Overview of MGA-Net.

Fig. 2. Schematic of the proposed attention gates (a) and the overview of MGA-Net (b). $f(x)$ and $h(x)$ are intermediate feature maps. $o(x)$ is the elementwise multiplication of $h(x)$ and the estimated attention map β. The output of the attention module is $A(x)$. $Loss_{BCE}$ is the binary cross entropy loss for IPF diagnosis, $Loss_{att_{high}}$ and $Loss_{att_{med}}$ are attention-based loss function at a high- and medium- resolutions. AG: attention gates; R: residual blocks.

$\lambda_{high} \geq 0$ and $\lambda_{med} \geq 0$. The hyperparameters (λ_{high} and λ_{med}) are selected based on the performance of the validation set. A systemic evaluation of hyperparameters is underway.

3. EXPERIMENTS AND RESULTS

3.1. Model implementation details

A stratified five-fold cross validation was performed to evaluate the proposed method, where the proportion of IPF and non-IPF patients was fixed across all folds. Five folds were separated at a patient-level. Our results reported in the future sections, including IPF diagnosis and the estimated attention maps, were all based on cases from the testing fold. Initial learning rate was set to be $1e-4$, followed by an exponential decay after 20 epochs of decay rate 0.05. The batch size was set to be 5 and the model trained after 200 epochs was saved for evaluation. Hardware of Tesla V100-SXM2-32GB and Keras framework were used [10].

3.2. Results

Model accuracy: Regarding the scan-level IPF diagnosis, Table 1 summarizes the AUC values with mean and standard deviations across folds using stratified five-fold cross validation, under different scenarios. Without including attention modules (scenario 1), the IPF diagnosis model performs poorly (AUC ± SE = 0.690 ± 0.194). When we include attention modules, but does not guide with domain knowledge (scenario 2.1), the IPF diagnosis model reaches satisfactory model performance (AUC ± SE = 0.956 ± 0.040). Only incorporating guided high- (scenario 2.2) or medium-resolution attention (scenario 2.3) decrease the performance of IPF diagnosis. Our proposal, which includes both high- and medium-resolution attentions (scenario 2.4), is able to reach the highest AUC value (0.971 ± 0.021) among all of the experiments. Notably, our proposal is sensitive to the selection of hyperparameters, i.e. λ_{high} and λ_{med}. When we increase λ_{med} to 10, the AUC decreases to 0.879 ± 0.191.

Table 1. Model results using stratified five-fold cross validation. AUC are reported using mean ± standard deviation across five folds. λ_{high} and λ_{med} are the hyperparameters in the overall loss function, which represent the relative task importance for the two attention modules.

Scenarios	λ_{high}	λ_{med}	AUC
1. No attention	NA	NA	0.690 ± 0.194
2. With attention (Include parameters in the loss function)			
2.1 Unguided attentions	0	0	0.956 ± 0.040
2.2 High resolution only	200	0	0.869 ± 0.123
2.3 Medium resolution	0	1	0.925 ± 0.078
only	0	10	0.927 ± 0.065
2.4 Our proposal (both	200	1	0.971 ± 0.021
high- and medium-	200	10	0.879 ± 0.191
attentions)			

Explainability: We explored the model explainability using one randomly sampled non-IPF patient as an example, shown in Fig.3. Regarding our proposed experiment ($\lambda_{high} = 200, \lambda_{med} = 1$), both attention maps at high- and medium-resolutions highlight the regions of interests and focuses on the peripheral lungs while suppressing background clutter.

4. DISCUSSION AND CONCLUSIONS

In this paper, we presented our multi-scale guided attention network, MGA-Net, which is generally suitable for

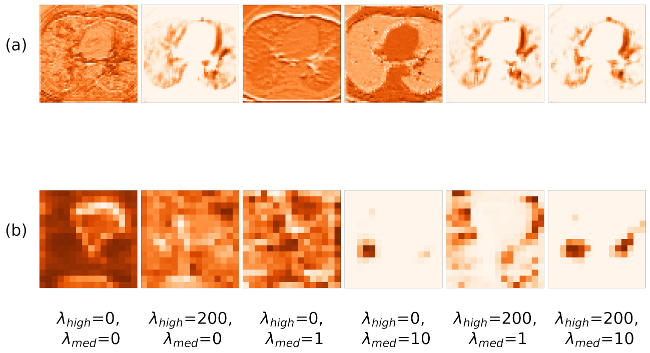

(a)
(b)
λ_{high}=0, λ_{med}=0
λ_{high}=200, λ_{med}=0
λ_{high}=0, λ_{med}=1
λ_{high}=0, λ_{med}=10
λ_{high}=200, λ_{med}=1
λ_{high}=200, λ_{med}=10

Fig. 3. The estimated attention maps at high resolution (row a) and medium resolution (row b) for a randomly selected non-IPF patient under multiple experiments, using the model built in one fold as an example. The randomly selected patient is one of the test cases for all of these experiments.

weakly supervised tasks. We use scan-level IPF diagnosis as the main focus of this paper. Several advantages can be addressed using the MGA-Net. Firstly, population-level domain knowledge is more accessible, whereas acquiring well-labeled medical imaging data is time-consuming and labor-intensive. Guided with population-level domain knowledge in lung boundary and IPF disease location at various resolution scales, we can accomplish satisfactory model performance only using coarse labels for the IPF diagnosis task. Secondly, using attention models at various resolution scales increase model explainability, which is a crucial step for building robustness in the medical imaging domain.

We have demonstrated that MGA-Net is one promising method for both enhancing explainability and increasing the performance of model for the task of automated IPF diagnosis. We find that only including high- or medium- resolution attention (scenario 2.2 and 2.3), the model performance is not comparable to that of including two resolution scales. This may be attributed to the fact that the network exploits different information from different layers; therefore, having two resolution scales let the network focuses on the lung parenchyma from coarse to fine, which can be seen from Fig.3.

5. COMPLIANCE WITH ETHICAL STANDARDS

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the principles of the Declaration of Helsinki.

6. ACKNOWLEDGMENTS

This research is supported by NIH, NHLBI-R21-HL140465. Hua Zhou is supported by grants from the National Human Genome Research Institute (HG006139) and the National Institute of General Medical Sciences (GM053275).

7. REFERENCES

[1] Moisés Selman, Talmadge E King Jr, and Annie Pardo, “Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy,” *Annals of internal medicine*, vol. 134, no. 2, pp. 136–151, 2001.

[2] Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim, “Deep learning applications in medical image analysis,” *Ieee Access*, vol. 6, pp. 9375–9389, 2017.

[3] Minh-Thang Luong, Hieu Pham, and Christopher D Manning, “Effective approaches to attention-based neural machine translation,” *arXiv preprint arXiv:1508.04025*, 2015.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need,” in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[5] Saumya Jetley, Nicholas A Lord, Namhoon Lee, and Philip HS Torr, “Learn to pay attention,” *arXiv preprint arXiv:1804.02391*, 2018.

[6] Jo Schlemper, Ozan Oktay, Michiel Schaap, Mattias Heinrich, Bernhard Kainz, Ben Glocker, and Daniel Rueckert, “Attention gated networks: Learning to leverage salient regions in medical images,” *Medical image analysis*, vol. 53, pp. 197–207, 2019.

[7] Heechan Yang, Ji-Ye Kim, Hyongsuk Kim, and Shyam P Adhikari, “Guided soft attention network for classification of breast cancer histopathology images,” *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1306–1315, 2019.

[8] Yiqi Yan, Jeremy Kawahara, and Ghassan Hamarneh, “Melanoma recognition via visual attention,” in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 793–804.

[9] HJ Kim, DP Tashkin, P Clements, G Li, MS Brown, R Elashoff, DW Gjertson, F Abtin, DA Lynch, DC Strollo, et al., “A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients,” *Clinical and experimental rheumatology*, vol. 28, no. 5 Suppl 62, pp. S26, 2010.

[10] François Chollet et al., “Keras,” `https://keras.io`, 2015.