Supplementary Material - CNN Attention Guidance for Improved Orthopedics Radiographic Fracture Classification

Zhibin Liao, Kewen Liao, Haifeng Shen, Marouska F. van Boxel, Jasper Prijs, Ruurd L. Jaarsma, Job N. Doornberg, Anton van den Hengel, Johan W. Verjans

In this supplementary material, we show the experiment results of the proposed method on the ResNet-101 [1] and various attentive backbones that demonstrate a consistent performance across the experimented models.

I. ResNet-101

The results using the ResNet-101 backbone are displayed in Table I. In summary, we found the baseline performances (*i.e.*, classification models without attention guidance) are improved from the respective ResNet-50 [1] counterparts (scaphoid (s-512) study-wise is up from $76.4\% \pm 0.9\%$ to $79.6\% \pm 1.5\%$, $p < 0.01$, ankle (Test Set) from $69.0\% \pm 2.8\%$ to $73.4\% \pm 4.0\%$, $p = 0.08$), showing that the larger/deeper neural network is able to generalize better. On the other hand, the attention guidance models are still able to provide consistent improvements over the ResNet-101 baseline results, which demonstrates the effectiveness of the proposed method. In addition, the best results collected from the ResNet-101 backboned models (scaphoid $85.8\% \pm 1.3\%$ using Scribble at $\lambda = 0.1$, and ankle $79.6\% \pm 1.1\%$ using Segm. at $\lambda = 1$) are not significantly better than those from the paper's ResNet-50 models (scaphoid $84.2\% \pm 0.8\%$ using BBOX at $\lambda = 1$, $p = 0.06$, and ankle $80.6\% \pm 3.8\%$ using Segm. at $\lambda = 0.1$, $p = 0.60$).

TABLE I
THE SCAPHOID AND ANKLE FRACTURE EXPERIMENT RESULTS IN CLASSIFICATION ACCURACY (%), USING THE RESNET-101 BACKBONE WITH $\mathbf{512} \times \mathbf{512}$ INPUT SIZE.

Methods	Scaphoid					
	Image-wise			Study-wise		
Baseline	76.9 ± 1.0			79.6 ± 1.5		
Attn. Guid.	Scribble	BBOX	Segm.	Scribble	BBOX	Segm.
λ = 0.01	78.2 ± 1.2	77.0 ± 2.2	77.3 ± 1.5	83.6 ± 1.3	81.6 ± 3.4	81.6 ± 2.6
λ = 0.1	79.7 ± 1.1	$\mathbf{78.5 \pm 0.8}$	$\mathbf{79.0 \pm 1.6}$	$\mathbf{85.8 \pm 1.3}$	82.8 ± 1.6	83.4 ± 2.6
λ = 1	$\mathbf{79.9 \pm 1.3}$	78.3 ± 1.0	78.7 ± 1.0	84.8 ± 1.9	$\mathbf{83.0 \pm 1.4}$	$\mathbf{85.0 \pm 1.2}$
λ = 5	73.7 ± 1.0	77.9 ± 1.1	77.5 ± 1.1	78.4 ± 2.2	82.2 ± 0.8	82.4 ± 2.0
	Ankle					
	Image-wise			Study-wise		
Baseline	69.1 ± 3.1			73.4 ± 4.0		
Attn. Guid.	Scribble	BBOX	Segm.	Scribble	BBOX	Segm.
λ = 0.1	71.0 ± 1.9	71.2 ± 0.8	69.2 ± 1.5	76.8 ± 2.3	76.6 ± 2.5	75.0 ± 1.7
λ = 0.5	$\mathbf{71.9 \pm 1.8}$	$\mathbf{72.3 \pm 1.6}$	70.5 ± 2.9	$\mathbf{77.4 \pm 1.1}$	$\mathbf{79.2 \pm 1.5}$	76.0 ± 4.6
λ = 1	70.2 ± 2.7	71.3 ± 2.5	$\mathbf{73.6 \pm 0.7}$	76.2 ± 1.8	77.6 ± 1.3	$\mathbf{79.6 \pm 1.1}$

II. Attentive Backbones

The results of all three types of human guidance (*i.e.*, scribble, bounding box, and segmentation) on the attentive backbones (*i.e.*,residual attention [2], self attention [3], and pyramid attention [4]) can be found in Table II. We use the ResNet-50 backbone for these attentive models and set input size to 512. The λ values are set as 1 for the scaphoid experiments and 0.5 for the fracture experiments as they were the optimal values from the previous experiments (*i.e.*, models backboned by the non-attentive vanilla ResNet-50) in the main paper.

In summary, these experiment results convey the similar observation, *i.e.*, with the use of human guidance, the classification results are much improved. On the scaphoid (s-512) dataset, the best classification performance is observed with the combination of self attention + scribble type guidance ($84.6\% \pm 1.7\%$), on par with the previous optimal result ($84.2\% \pm 0.8\%$, vanilla ResNet50 + BBOX). On the ankle test set, the optimal result is $81.2\% \pm 1.3\%$ (self attention + Segm.), which is also on par with the previous optimal result ($80.4\% \pm 3.9\%$, vanilla ResNet50 + BBOX). The only exception is with the pyramid attention on the ankle test set, where the model performs significantly worse both with and without human guidance. This may be a result of the backbone model in which the addition of pyramid attention mechanism significantly changed the energy landscape of the optimization on the ankle dataset.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[2] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, *et al.*, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156–3164.
[3] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*, PMLR, 2019, pp. 7354–7363.
[4] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," *arXiv preprint arXiv:1805.10180*, 2018.

TABLE II

THE HUMAN-GUIDED ATTENTION EXPERIMENT RESULTS WITH THE USE OF THE ATTENTIVE BACKBONES, *i.e.*, RESIDUAL ATTENTION [47], SELF ATTENTION [48], AND PYRAMID ATTENTION [49]. THE NO GUIDANCE ENTRY REPEATS THE RESULTS FOR THE RESPECTIVE MODELS IN TABLE III AND IV IN THE MAIN MANUSCRIPT FOR CONVENIENCE.

Dataset	Scaphoid						Ankle					
Backbone	Residual Attn.		Self Attn.		Pyramid Attn.		Residual Attn.		Self Attn.		Pyramid Attn.	
	Image	Study	Image	Study	Image	Study	Image	Study	Image	Study	Image	Study
No Guid.	75.3 ± 3.4	78.8 ± 2.4	75.5 ± 1.7	79.2 ± 1.8	75.7 ± 0.6	79.2 ± 2.7	62.5 ± 1.6	65.0 ± 1.2	68.9 ± 2.5	74.6 ± 3.1	$\mathbf{55.4 \pm 6.2}$	$\mathbf{59.4 \pm 9.6}$
Scribble	77.3 ± 0.8	82.6 ± 2.7	$\mathbf{78.7 \pm 1.0}$	$\mathbf{84.6 \pm 1.7}$	77.3 ± 1.0	81.0 ± 1.4	72.7 ± 2.4	$\mathbf{79.8 \pm 1.6}$	71.9 ± 3.0	78.0 ± 4.4	50.7 ± 2.3	51.4 ± 1.5
BBOX	$\mathbf{78.9 \pm 1.0}$	82.0 ± 2.4	77.6 ± 1.4	83.6 ± 1.5	77.7 ± 0.9	82.2 ± 1.8	71.3 ± 1.3	78.0 ± 1.6	72.3 ± 2.1	77.8 ± 2.2	53.2 ± 4.0	55.0 ± 5.4
Segm.	77.6 ± 1.8	$\mathbf{83.4 \pm 0.6}$	77.2 ± 1.9	83.0 ± 3.1	$\mathbf{79.4 \pm 1.0}$	$\mathbf{83.4 \pm 1.8}$	$\mathbf{73.2 \pm 1.9}$	79.2 ± 1.9	$\mathbf{74.0 \pm 2.1}$	$\mathbf{81.2 \pm 1.3}$	53.4 ± 5.0	55.2 ± 4.8