SUPERVISED CONTRASTIVE LEARNING FOR IMPROVED VIEW LABELING IN PEDIATRIC RENAL ULTRASOUND VIDEOS

Stanley Z. Hua[⋆] *Irene Y. Chen*[†] *Alex X. Lu*[†] *Lauren Erdman*[⋆]

[⋆] The Hospital for Sick Children
[†] Microsoft Research

Manual capture of ultrasound (US) images at a specific view can be labor intensive because of the need for clinical expertise. We automate the task of view labeling of pediatric renal videos, specifically side of body and anatomical plane classification. Earlier work shows that pretraining with ImageNet is beneficial for US view-labeling [1]. Here, we investigate if self-supervised (SSL) pre-training can improve on ImageNet pretraining for US view labeling. We demonstrate that SSL pretraining is beneficial for automatic view labeling within-institution data. Furthermore, label-dependent contrastive pretraining extends model generalization to US video data from an external institution.

1. DATA

Following ethical review (REB 1000053438), renal US video data was collected from the Hospital for Sick Children. Data from 69 patients (204 videos) was used for training/validation. Unseen data from 30 patients (288 videos) was used for evaluation. To test for out-of-distribution performance, data from 37 patients (72 videos) was collected at Lucile Packard Children's Hospital. We labeled each image with side (Left/Right/Center) and plane (Sagittal/Transverse/Mid).

2. METHODS AND RESULTS

Models with no pre-training showed quick overfitting and poor validation performance (Table 1), possibly due to an insufficient amount of labeled data. To address this challenge, one common approach is fine-tuning from pretrained models. We investigated self-supervised learning and transfer learning. For SSL, we implemented an established contrastive method (MoCo) [2], and for transfer learning, we used models pretrained on ImageNet. We observed that both strategies are effective in boosting performance on same-institution data. However, both methods drastically decline in performance when generalizing to an external institution compared to supervised training without pretraining.

Based on these results, we hypothesized that injecting supervision during self-supervised pretraining would promote more robust feature learning and generalization. To do this, we produced a novel adaptation of the MoCo method that incorporates supervised label data (MoCo [SL, SV] in Table 1). Briefly, since MoCo applies augmentations to a positive pair of "similar" images for learning, we sample these positive pairs from images sharing the same label in the same video. We observe that this method not only improves performance on the same-institution datasets, but also generalizes these performance improvements to external institutions.

Methods	Side		Plane	
	Internal	External	Internal	External
No Pretraining	58.76	51.96	70.66	70.23
ImageNet	**66.64**	23.01	**85.2**	26.93
MoCo [SI, Any]	**65.5**	31.04	77.27	75.09
MoCo [SI, SV]	**65.19**	31.29	78.57	77.43
*MoCo [SL, SV]	64.12	**62.77**	75.96	**82.62**

Table 1. Side and Plane Classification Accuracies for Internal or External Institution Data. (SI=*Same-Image*, SL=*Same-Label*, SV=*Same-Video*). *[SL, SV] is the only pretraining method with supervision. Best equivalent models are in **bold**.

3. CONCLUSION

SSL-pretraining improves view labeling performance on same-institution data while on external institution data, some supervision appears to be key for learning image features that are both useful for the view labeling task and that generalize beyond the training institution.

4. REFERENCES

[1] Phillip M. Cheng and Harshawn S. Malhi, "Transfer learning with convolutional neural networks for classification of abdominal ultrasound images," *Journal of Digital Imaging*, vol. 30, no. 2, pp. 234–243, 2016.

[2] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick, "Momentum contrast for unsupervised visual representation learning," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.