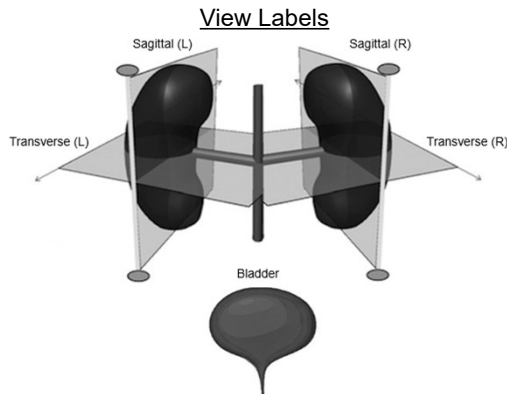


Supervised Contrastive Learning for Improved View Labeling in Pediatric Renal Ultrasound Videos

Stanley Z. Hua^{1,2}, Irene Y. Chen³, Alex X. Lu³, Lauren Erdman^{1,2,4}

University of Toronto¹, The Hospital for Sick Children², Microsoft Research³, Vector Institute⁴

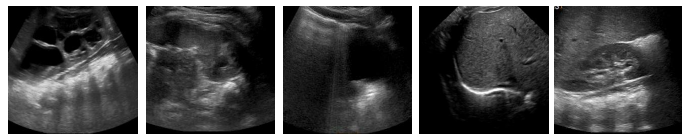
Automated view labeling on **pediatric renal ultrasound videos**
is an unexplored, difficult yet valuable task.



Key Challenges

1. Lack of labeled data (*within and across institutions*)
2. Noisiness of ultrasound images
3. Softness of view labels on videos

Example Images



Left Kidney (Sagittal) Left Kidney (Transverse) Bladder Right Kidney (Transverse) Right Kidney (Sagittal)

To overcome these challenges, we propose **pretraining**
with **self-supervised MoCo** to learn more **robust features** [1].

Step 1. Pre-training

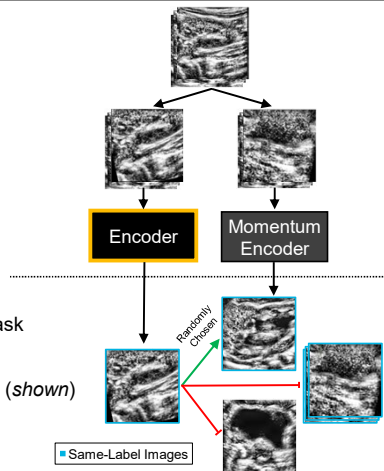
a. Random Augmentation

b. Extract CNN Features

c. **Unsupervised** Contrastive Task

or

c. **Supervised** Contrastive Task (*shown*)



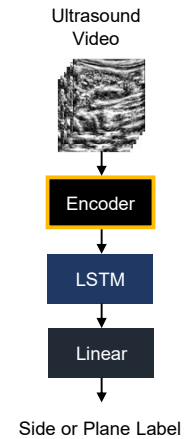
Step 2. Fine-tuning

a. Random Augmentation (*Optional*)

b. Extract CNN Features

c. Aggregate Features over Time

d. Classify Side / Plane



Self-supervised MoCo pretraining and **augmentations with supervision**
extend **model generalization** to data from an **external institution**.

Table 1. Side/Plane Classification Accuracies (for Internal or External Institution Data). Methods are **fine-tuned without augmentations**.

Methods	Side		Plane	
	Internal	External	Internal	External
No Pretraining	57.77	49.97	71.57	62.54
ImageNet	62.10	21.55	81.24	22.19
MoCo [Unsupervised, All]	66.04	33.02	79.76	59.20
MoCo [Unsupervised, Same-Video]	66.10	55.42	78.34	64.98
MoCo [Supervised, Same-Video]	61.96	72.03	76.45	78.19

Table 2. Side/Plane Classification Accuracies (for Internal or External Institution Data). Methods are **fine-tuned with augmentations**.

Methods	Side		Plane	
	Internal	External	Internal	External
No Pretraining	63.72	58.69	75.77	69.79
ImageNet	68.00	54.27	80.81	56.06
MoCo [Unsupervised, All]	62.93	69.4	76.64	79.73
MoCo [Unsupervised, Same-Video]	67.40	56.32	78.66	80.31
MoCo [Supervised, Same-Video]	62.30	60.68	75.71	80.82

Key Takeaways

1. **Augmentations paired with supervision** contribute significantly to out-of-domain generalization.
2. We find that combining **self-supervised pretraining + augmentations paired with supervised labels** yield the greatest out-of-domain generalization.

REFERENCES

- [1] 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.
- [2] 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.