Transfer Learning for the ICU: Comparing the Performance of YOLOv5 and Faster R-CNN X101-FPN

Ally Nakamura, Kachachan Em Chotitamnavee, Roshan Swaroop

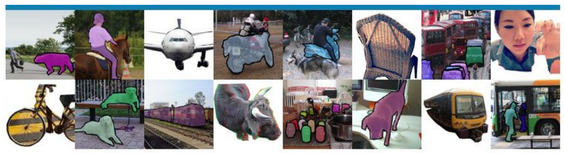
Mentors: Zane Durante, Ehsan Adeli, Fei Fei Li

Introduction and Background

- Leveraging real-time data and analysis of patient's condition can play a critical difference in recovery
- Activity detection in ICUs is currently performed manually by trained personnel, which is expensive and time consuming
- Object detection models can help relieve the monitoring workload of healthcare workers, while providing benefits such as activity-based costing
- ***Problem***:
  - specialized medical data is sparse, thus training data is limited
  - data is also expensive to annotate due to complexity and clearance requirements
  - clinics lack sufficient resources to train large models
- ***Need:*** pre-trained models that train quickly and perform well via transfer learning when fine-tuned on small out-of-distribution data sets.
- ***Goal:*** **determine which pre-trained models perform optimally when**
  - **varying the training data set size**
  - **evaluating in-distribution, out-of-distribution (OOD), and very out-of-distribution data**

Datasets

Pre-training data: both models were pre-trained on COCO (328000 images, 80 classes)



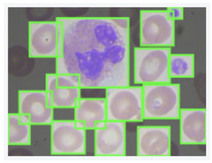
Fine-tuning data:
The in-distribution dataset was a vehicles set from OpenImages (627 images).

The out-of-distribution dataset was a dataset of masks wearing individuals from Eden Social Welfare Foundation (149 images).

The very out of distribution dataset was a dataset of red and white blood cells by akshaylambda on GitHub (364 images).

Vehicles

Masks



Blood Cells

Preprocessing

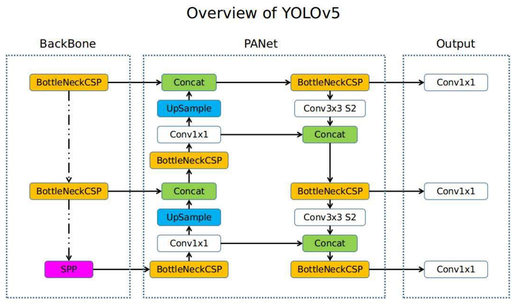
We processed each fine-tuning dataset in the following ways:
- into subsamples of 25%, 50%, 75%, and 100% of the total image count
- into subsamples of 50 images and 100 images each
- into subsamples of 150 objects and 300 objects (measured as amount of labels across all images) each

We also resized images to 416x416, took care to ensure that subsamples had uniform class distribution, and oriented portrait and landscape photos in the same direction.

Object Detector Model Architectures

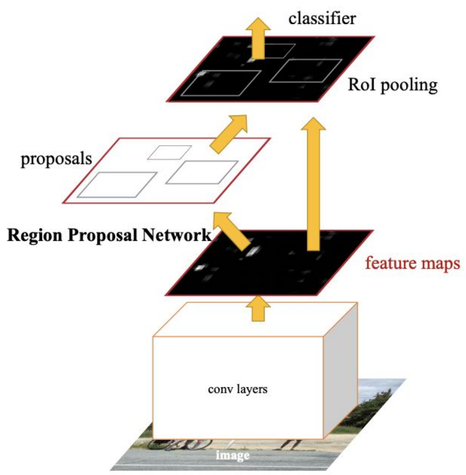
YOLOv5:

YOLOv5 divides images into a grid system, and each grid detects objects within itself. YOLOv5 uses a two-stage object detector, and separates the model into the input, backbone, neck, and head sections. The backbone model is a feature extractor that uses a classification model such as VGG-16.



Faster R-CNN X101-FPN:

The specific Detectron2 model used in this study was the Faster R-CNN X101-FPN developed by Facebook AI which was pre-trained on the 2017 MS COCO dataset. Faster R-CNN is an improvement on Fast R-CNN, which itself is an improvement on R-CNN. X101-FPN has the highest box AP of all the faster R-CNN models that Detectron2 offers, but a relatively high inference time.



Experiment

We ran both the pre-trained YOLOv5 and Faster R-CNN models on each of the 8 subsamples we created in the preprocessing step.
We:
- Removed the classification layer of both models
- Randomly initialized a new classification layer with the correct output size and appended to both models
- Ran stochastic gradient descent with nesterov momentum on the BCE loss, with learning rate of 0.001 and momentum of 0,9

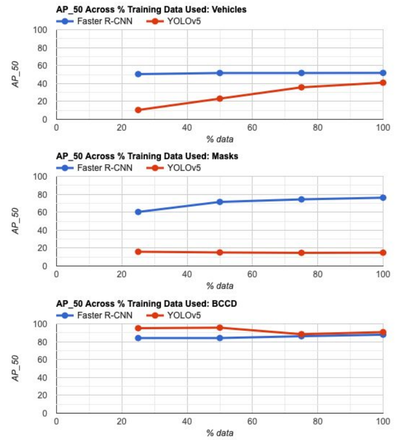
These parameters yielded the following results.

Model	BCCD (300)	Mask (300)	Vehicle (300)	BCCD (150)	Mask (150)	Vehicle (150)
Faster R-CNN X101 FPN	75.25	76.02	50.68	73.12	72.38	48.92
YOLOv5	8.99	3.8	5.48	6.77	14.9	8.6

Table 2. Average precision (AP50) results for Faster R-CNN X101 FPN and YOLOv5 models that were trained on 300 objects and 150 objects.)

Model	BCCD (100)	Mask (100)	Vehicle (100)	BCCD (50)	Mask (50)	Vehicle (50)
Faster R-CNN X101 FPN	79.24	82.11	49.43	77.89	71.03	47.95
YOLOv5	8.99	26.9	4.3	8.2	13.9	6.4

Table 3. Average precision (AP50) results for Faster R-CNN X101 FPN and YOLOv5 models that were trained on 100 images and 50 images.



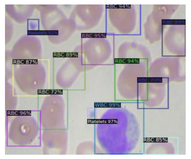
Results

1. **How important is the in or out-of-distribution nature of the transfer learning data set?**
   YOLOv5 and Faster R-CNN performed better for OOD data, suggesting both models are suited for transfer learning to OOD data sets. However, this could be due to the homogeneity of the OOD data compared to the in-distribution data.
2. **How important is the size of the fine-tuning data set in in-distribution cases?**
   The size of the fine-tuning data set matters more for the YOLOv5 model compared to Faster R-CNN. The latter is better suited when in-distribution training data is limited.
3. **How important is the size of the fine-tuning data set in out-of-distribution cases?**
   Both models performed consistently regardless of variations in the training data size. Faster R-CNN performed better in accuracy, indicating it is more suitable for transfer learning for OOD data.
4. **How important is the size of the fine-tuning data in very out-of-distribution cases?**
   Both models performed consistently regardless of variations in the training data size with approximately the same accuracy.
5. **Which model is optimal for use in ICUs?**
   Faster R-CNN due to its consistency across all training sizes and higher overall accuracy. One caveat is its slower training time; however, accuracy is considered more crucial in ICU settings.

Sample Results

Faster R-CNN X101 FPN

Masks

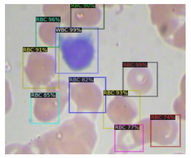


Blood Cells

Vehicles

YOLOv5

Masks



Blood Cells

Vehicles

Conclusions & Future Work

- By varying amounts of training data used in this fine-tuning process, we observed that even partial datasets as small as 25 percent of the overall size can achieve reasonable performance, which validates the path for clinical transfer learning. Medical image data often requires extensive domain knowledge and time to annotate, so these results indicate that even sparse representation of each clinical class may lead to worthwhile results.
- One of the most immediate areas to expand on this project would be further characterizing the differences in which OOD data affects transfer learning. While the BCCD dataset was the furthest OOD, it had a much higher degree of homogeneity in its image samples, so controlling for discrepancies such as these in further experiments is key. In addition, once PAC reaches a critical mass of clinical ICU data, findings from this project can be used to effectively transfer learn for real clinical use; our results support that even partial datasets can reasonably perform object detection tasks.

References

1. Yosinski, J., Clune, J., Bengio, Y., and Lipson, H., "How transferable are features in deep neural networks?", arXiv e-prints, 2014.
2. Morid, Mohammad Amin, Alireza Borjali, and Guilherme Del Fiol. "A Scoping Review of Transfer Learning Research on Medical Image Analysis Using ImageNet." Computers in Biology and Medicine 128 (2021): 104115. https://doi.org/10.1016/j.compbiomed.2020.104
3. [1]Huh, M., Agrawal, P., and Efros, A. A., "What makes ImageNet good for transfer learning?", arXive − prints, 2016.

Stanford MEDICINE | Stanford Partnership in AI-Assisted Care