Transfer Learning for the ICU: Comparing the Performance of YOLOv5 and Faster R-CNN X101-FPN
Ally Nakamura, Kachachan Em Chollattamee, Roshan Swaroop
Mentors: Zane Durante, Ehsan Adeli, Fei Fei Li

Introduction and Background

- Leveraging real-time data and analysis of patients’ conditions 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.
- Clinical 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 (328,000 images, 80 classes)

Fine-tuning data:
- The in-distribution dataset was a vehicles set from OpenImages (427 images).
- The out-of-distribution dataset was a dataset of masks wearing individuals from Kaiser-Society Welfare Foundation (1,419 images).
- The very out-of-distribution dataset was a dataset of red and white blood cells by MontanaDrs on CAPTAIN (514 images).

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. It 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 pre-processing step. We:
- Removed the classification layer of both models
- Randomly initialized a new classification layer with the current output size and appended to both models
- Ran stochastic gradient descent with momentum on the BCE loss, with learning rate of 0.001 and momentum of 0.9

These parameters yielded the following results:

<table>
<thead>
<tr>
<th>Model</th>
<th>BCD</th>
<th>COCO (300)</th>
<th>Mask</th>
<th>Vehicle</th>
<th>BCD (150)</th>
<th>Mask (150)</th>
<th>Vehicle (150)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLOv5</td>
<td>40.6</td>
<td>50.6</td>
<td>64.4</td>
<td>50.6</td>
<td>64.4</td>
<td>50.6</td>
<td>64.4</td>
</tr>
<tr>
<td>Faster R-CNN X101-FPN</td>
<td>44.6</td>
<td>54.6</td>
<td>68.5</td>
<td>54.6</td>
<td>68.5</td>
<td>54.6</td>
<td>68.5</td>
</tr>
</tbody>
</table>

Table 1: Average precision (AP) results for Faster R-CNN (X101-FPN) and YOLOv5 models (last trained on 150 objects and 150 images).

Table 2: Average precision (AP) results for Faster R-CNN (X101-FPN) and YOLOv5 models trained on 300 images and 30 images.

1. How important is the in or out-of-distribution nature of the transfer learning data set?
2. How important is the size of the fine-tuning data set in in-distribution cases?
3. How important is the size of the fine-tuning data set in out-of-distribution cases?
4. How important is the size of the fine-tuning data set in very out-of-distribution cases?
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 it slower training time; however, accuracy is considered more crucial in ICU settings.

References


Sample Results

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 BCD 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 IC data, findings from this project can be used to effectively transfer learning for real clinical use; our results support that even partial datasets can reasonably perform object detection tasks.

Preprocessing

We processed each fine-tuning dataset in the following ways:
- Resized into sub-samples of 25%, 50%, 75%, and 100% of the total image count
- Resized into sub-samples of 30 images and 100 images each
- Resized into sub-samples of 130 objects and 300 objects (measured as amount of label boxes across all images) each

We also resized images to 416x64 by 10%, took care to ensure that subsamples had uniform class distribution, and oriented portrait and landscape photos in the same direction.