Boston Scientific: Avoiding the pitfalls of transfer learning

Summary

Transfer learning involves re-purposing a model originally trained for one task and use if for a different task. In this sense, the 'knowledge' gained while solving one problem can be applied to a related problem e.g., a model for recognizing cars can be applied to recognizing trucks.[1]

This can save some work and help reduce training costs, but it can also come with an outsized (up to 45%) drop in performance, so there are limited applications where it makes sense to do. However, Boston Scientific's experience shows how high performance can still be achieved with transfer learning by "stepping down" the problem, enabling them to achieve model performance of over 99% and labor savings of over $5m.

Opportunity: Stent inspection is critical to patient safety, and it's costly

Boston Scientific produces stents for a range of surgical applications (e.g., interventional cardiology), and needs to inspect them for defects such as broken links or surface imperfections. Accurate inspections are critical for successful clinical outcomes, and so escape rates (the proportion of defective parts that might slip through the cracks) are regulated by the US Food and Drug Administration based on the risk to patients.

Traditionally, much of the inspection has been done by human experts, but this isn't optimal. Eric Wespi, a Data Science manager at Boston Scientific commented that "Human visual inspection is often slow, expensive, and can present unwanted quality risks". This makes intuitive sense; people typically don't perform well on tasks that require focused attention, for a long period of time, where the probability of an event is very infrequent. Moreover, their judgements will vary person to person, and experts' time is expensive too; Boston Scientific has around 3000 experts doing inspections at a cost of several million dollars each year.

Challenge: Image classification requires a lot of data for training

Boston Scientific had already implemented an automated rules-based system to capture common issues (using e.g., dimensional measurements), and had tuned it to be conservative (i.e., with a negligible false-negative rate). However, the false-positive rate of 5-10% was still too high; too many good parts were being flagged as defective for human inspectors.

---

[1] https://en.wikipedia.org/wiki/Transfer_learning

Convolutional Neural Networks (CNNs) are particularly well-suited to image classification, but such models require an enormous amount of data to train. In many cases (particularly for newer and rarer defects) the team didn't have enough data to train these models from scratch; collecting or generating this data would be impractical and the cost prohibitive.

Solution: Transfer learning applied to a scaled-down problem

The team wondered if they could do better by starting with a pre-trained model. They applied the following approach:

1. **Scale down the problem:** For each defect, inspection could be segmented into smaller and narrower tasks (e.g., "does this portion of the image contain a link?" and "is this link broken or not?").
2. **Customize existing models:** Several open-source CNNs were used e.g., VGG16, EfficientNet (B0 through B7), Mask R-CNN, YOLOv3, ResNet-50, Inception-v3. The team started with pre-trained weights from literature, customized the last couple of network layers and then re-trained the models using their own data.
3. **Test data requirements:** The team found that they needed very little (e.g., 100-1000 examples of each defect and 50-60k examples of non-defective stents) to exceed human level performance.

Note: To improve the robustness of the models, they also augmented the training data by generating additional examples through perturbation (these could be simple adjustments that should not impact classification e.g., brightness adjustments or the addition of noise).

All the work was completed within a relatively modest budget of $50k (model training was quick and cheap, taking 1-2s per image across 9 models, and 2-10 hours to train each model on a single GPU), and a small team of ~3 people.

Outcomes: Dramatic model performance and reduced labor costs

The resulting accuracy was above 90% for all models, with even smaller networks (e.g., VGG16) performing well for simple problems. Accuracy increases for more sophisticated models and with more data e.g., EfficientNet can achieve up to 97% for a B0 network with 100 examples and above 99% for a B7 network with 1000 examples.

This level of performance is not what we might typically expect with transfer learning. Performance usually drops significantly, requiring more data to offset the deficit. In this case, applying the existing models to a simpler problem appears to have obviated that need.

Deploying the 9 models enabled an equivalent of $5m in direct labor savings from the reduction in parts being flagged for human inspection, and the opportunity to re-assign several experts to other high-value projects.

In summary, Boston Scientific's experience suggest that transfer learning works well with the right conditions:

- There is an existing generic model that can be leveraged (in the case of image processing tasks, the early layers of such networks seem to be high transferable even when the task is notably different).
- The usual drop in performance from transfer learning can be obviated by using the system on a simpler problem and doing fine-tuning on the network