Navtech: The new platforms being created by deep learning

Summary

Some innovations are only economically feasible where a centralized provider can make the upfront investments required and spread the costs over a wide enough customer base. This is increasingly true for machine-learning, particularly where the tasks being automated are perception-based (e.g., image recognition), because these can require significant data and compute resources to develop and maintain. For many individuals and businesses, this technology is simply out of reach.

Navtech identified an opportunity to bring advanced computer vision to individual diamond retailers across the globe, by creating a model and delivering it as a service. Doing so is a win-win and is an excellent example of where AI can not only deliver improved efficiency and performance, but also unlock new capabilities and business models in the process.

Opportunity: Digitizing catalogs helps increase sales

India alone has an estimated 300,000 diamond jewelry retailers. Many of these are smaller companies, with limited inventory capacity, and typically increase their offerings through custom made jewelry. Relative to offering their own inventory alone, adding custom-made jewelry options can result in a doubling of conversion rate (i.e., the twice as many customers who find something they like, and make a purchase).

Visual catalogs are a key part of the sales process. Each retailer maintains one for their own inventory and supplements it with images of other jewelry as inspiration for customers looking for bespoke pieces. Traditionally these have been physical booklets or magazines, but they can only present a limited number of items, can't be refreshed often, and are slow to read through.

Digital catalogs remove many of these constraints but introduce other challenges. Staff compile images from various sources (e.g., photographs of inventory, from the web, from manufacturers' catalogs), and categorize them manually into folders. The process is slow (30-60 seconds per image, up to a million images), prone to error (e.g., there are lots of duplicate images and it's difficult to remember what you've seen already), and results in only very high-level categorization (e.g., rings vs necklaces). Ideally there would be a way to automatically create catalogs and enable customers to search based on additional criteria.

Challenge: Computer vision is too expensive for individual retailers

Computer vision systems that leverage deep learning to classify images could help improve speed and accuracy, but the reality is that they are out of reach for most retailers. Deep learning is resource intensive; requiring enormous amounts of data and compute both to train and implement. The system may not be used frequently enough to justify that expense, particularly when the comparative cost of labor (i.e., for manual classification) is low (e.g., in India, retail worker salaries might start at around $100 / month). The business case for computer vision, for any one retailer, is thus unlikely to be attractive.

Solution: Build once and provide as a service

Dr. M.I.M. Loya, General Manager of Emerging Technologies at Navtech had an idea; build a computer vision system and offer it as a service. This would be a win-win; Navtech had the resources to make the initial investment and offer access to the system for a modest fee, and retailers would benefit from low-cost access to the system on an ongoing basis.

Navtech selected three attributes for their pilot and created a deep learning model for each:

- For **product category** (e.g., rings, bracelets) and **style**, they used a VGG16[1] network to classify images. The open-source ImageNet-trained backbone was fine-tuned by custom training the head, first and second layers of the network.
- For **diamond cut** (e.g., round, square), they instead used Mask RCNN for object detection and classification (having achieved only 55-56% accuracy using VGG16). The training data for this model was labelled by an intern, who manually drew a polygonal mask around the shape of each diamond.

Outcomes: Broader access to cutting edge ML at manageable cost

The system enables retailers to build larger digital catalogs much more rapidly. It can classify up to 100 images per minute (vs. 1-2 with manual classification), at an accuracy of 90-93% for product category and style and 85-86% for diamond cut (vs. 80% for manual labelling). The team was also able to achieve this using a relatively small data set (only 3000 labelled images for each model), which is somewhat surprising. Navtech did perform some post-processing, where the results of one model (e.g., style) were used to increase the confidence of predictions for another model (e.g., diamond cut), but it's possible that the high leverage from this small dataset was also due to a step-down in problem complexity relative to ImageNet (images of jewelry vary less than images of cats, say).

[1] VGG16 (also called OxfordNet) is a convolutional neural network architecture named after the Visual Geometry Group from Oxford. https://blog.keras.io/how-convolutional-neural-networks-see-the-world.html

This experience illustrates the cost-benefit tradeoffs of deep learning, where some use cases can only be enabled by larger, centralized players with the ability to serve wider audiences. Such systems, because they are delivered at scale, must be delivered as part of a wider product and service architecture, underpinned by traditional software development, but are much more cost-effective.