

Background

The ability to find a set of images that are in some way similar to a given image has multiple use cases—from visual search to duplicate product detection to domain-specific image clustering. Google’s Reverse Image Search is an example of a general-purpose image similarity tool that retrieves similar images to a query image. Most modern image similarity tools apply deep learning to quantify the degree of similarity between intensity patterns in pairs of images. This standard approach may not be sufficient, however, when “similarity” must be specific to the business context in which the tool will be used.

A prominent energy firm wanted to create an internal reverse image search capability to apply to a database consisting of images extracted from millions of digital and scanned documents. The image search would allow analysts to quickly retrieve relevant content from within this large library of documents. For example, a user might search for similar seismic imagery across multiple oil fields to find reservoirs with similar seismic profiles in order to compare engineering and productivity records. To pull out valuable insights, an artificially intelligent model would need to efficiently search tens of millions of images to find the small handful most similar to a target image.

The firm engaged Mosaic Data Science to provide deep learning expertise and to develop this capability. Mosaic was tasked with building the image similarity model based on proven expertise in artificial intelligence (AI) consulting.

Analysis

Mosaic’s data science team met with project stakeholders and analytics end users to discuss the project requirements and AI architecture. A key learning from these meetings was that the model would need to incorporate context-specific ideas of similarity. General-purpose image similarity would not meet core business objectives. Mosaic would need to go beyond the standard modeling techniques.

When building an AI model for retrieving images, a classical computer vision approach involves implementing a deep learning autoencoder scheme. In this scheme, images are fed into a deep learning neural network that encodes the image into a simplified representation of hidden nodes (internal nodes within the neural net) and decodes it again to reproduce a close approximation of the original image. By training the model with the objective of minimizing pixel-wise differences between the reproduced image and the original image, the autoencoder learns a reduced (compressed), generalized representation of the images that captures as much information as possible from the input image set. This allows retrieval of “similar” images based on distance measures applied to the compressed representation.

One drawback to the autoencoder method is that it only tends to find similar images based upon broad features in the image set on which the model is trained. If the user wants to find a similar image based upon more subtle or context-specific differences in the image, this classic deep learning approach fails. After an unsuccessful attempt at an autoencoder-based image similarity model, the energy firm engaged Mosaic to implement a more effective strategy.

Considering the downside of the autoencoding method, Mosaic chose to use a more sophisticated approach for finding similar images that was developed by researchers at Northwestern, Google, and

Caltech.¹ This computer vision technique, called deep ranking, attempts to find a rich representation of an image, using deep learning, by training a neural network to predict how likely an external agent is to determine that two images are more similar to each other than they are to a third, dissimilar image. Similarity no longer needs to be quantitatively defined a priori – the model can learn a similarity function to best match the assessments of expert users on a training dataset of image triplets. Similarity can be judged from semantic content (e.g. the blue car is more similar to the blue car image than the red car image in figure 1 - far left column) or from categorical content (e.g. the blue truck is more similar to the blue car than the house is to the blue car in figure 1 - column second from the left).

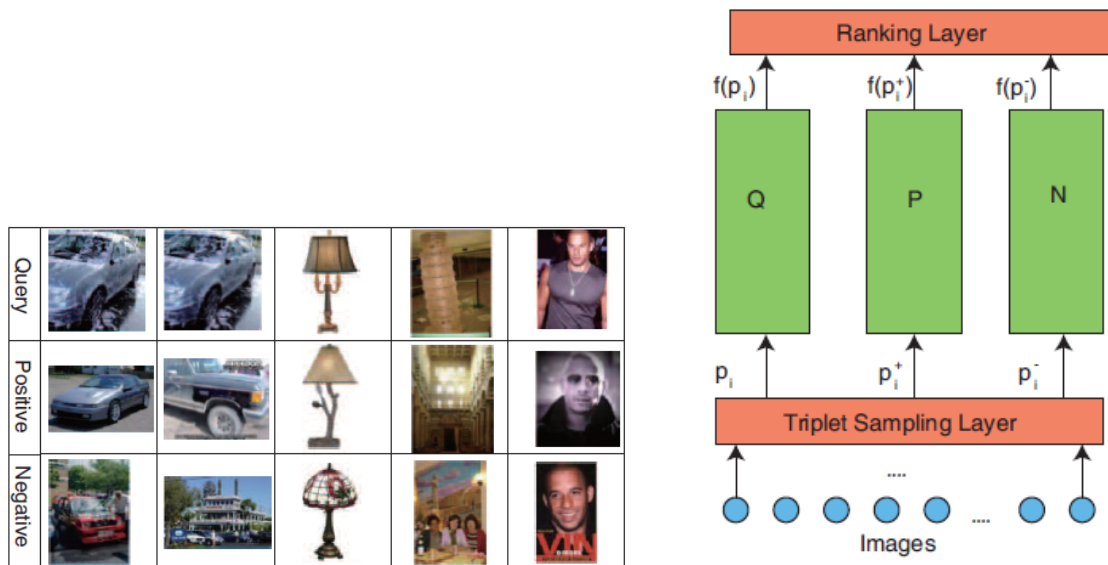


Image 1 – Displaying the triplet strategy.

For this approach to work, Mosaic had to first generate a training image data set containing thousands of examples of similar and dissimilar images. To seed this process, Mosaic leveraged the coarse results from the autoencoder model that had been previously developed to generate candidate triplets. Mosaic then programmed a simple graphical user interface (GUI) to rapidly pick quality training example triplets in a semi-automated way. A user marks a triplet as a good training example in the GUI if the positive image is more similar to the query image than the negative image is to the query.

Results

Mosaic's data science consultants found that the deep ranking method yielded much better results than the autoencoder model. It also allowed the team to focus on improving performance of specific classes of images that were of interest to the end users by providing explicit examples for supervised training of our model. This kind of targeted training is very difficult to achieve using the traditional autoencoder framework.

¹ Wang, Song, Leung, Rosenberg, Wang, Philbin, Chen, and Wu "Learning Fine-grained Image Similarity with Deep Ranking" https://users.eecs.northwestern.edu/~jwa368/pdfs/deep_ranking.pdf

The energy firm is now able to search and query the database based on image similarity. Of the top 10 similar images returned by the algorithm, 80-90% are judged by users to be related to the query image compared to 50% from the baseline autoencoder model. This saves valuable time, allowing users to quickly find images, and their associated documents, related to an image of interest in order to retrieve relevant information or evidence from an otherwise overwhelming document library.

For More Information

Want to learn more? Please contact info@mosaicdatascience.com