

Background: Footwear and Forensic Analysis

In forensic science, shoe prints and outsole characteristics fall into the category of pattern evidence. When a shoe print is found at a crime scene, there are a number of questions that can be asked. For example, an investigator may want to determine the make or model of the shoe that made the print, or possibly tie the shoe print to a specific individual. Another question, more difficult than the first two, is how common the shoe type, or features of that shoe type, are in a local population. This question is more difficult than the first two because it relies not only on the information contained in the found print, but also on characteristics of many other types of shoes. Thus, any sufficient answer to this question requires a way to automatically and efficiently classify many different types of shoes within a common system.

Classification

Visual classification is a complex task that our brains have been trained to do very well. Our eyes detect a large variety of features of an object, including color, shape, and texture, and send that information to our brain. Our brain then learns which combinations of features to associate with a given label, and should be able to apply those rules to future objects with similar characteristics. For example, an orange caterpillar and a baby carrot may be of similar color, shape, and size, but one is distinctly more fuzzy than the other. Thus, our brains learn that when faced with a small, cylindrical orange object, texture becomes an important feature when assigning a label to that object (which keeps us from accidentally ingesting caterpillars).

Convolutional Neural Networks (CNNs)

While our brains are adept at parsing images and classifying the objects within them, the task has proved much more difficult for computers. Convolutional neural networks (CNNs) are a tool for supervised deep-learning that have become standard in recent years for automatic image classification. CNNs use combinations of convolutional and pooling hidden layers to filter raw information into features, which are then fed into densely connected layers which are trained to associate given sets of features with their desired labels. This translation-invariant automated classification mimics the human eye-to-brain classification process and has become one of the most widely used machine learning techniques for image classification.



Pre-trained CNNs

A convolutional neural network consists of