C37 Machine Learning to Detect and Localize Forensics-Relevant Features

Sara Mousavi, Knoxville, TN 37909; Audris Mockus, PhD, University of Tennessee, Knoxville, Knoxville, TN 37996-2250; Dawnie W. Steadman, PhD, University of Tennessee, Knoxville, TN 37996; Angela M. Dautartas, PhD, University of Tennessee, Knoxville, TN 37996-1525*

Learning Overview: After attending this presentation, attendees will understand opportunities afforded by machine learning for automatic detection and localization of forensically relevant features in images documenting human decomposition.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by introducing new approaches to automate the annotation of digital collections and by discussing how this can transform research on human decomposition.

Rich forensic content in large digital media collections needs to be extracted in order to advance forensic science; for example, to obtain frequencies of phenomena in the population at large, and, more generally, to make forensic work more effective. Since manual extraction by forensic experts is not feasible for large collections, scientists seek to employ machine learning to detect and localize forensic features in the partially annotated collection of images taken from bodies during their decomposition process at the Anthropology Research Facility (ARF) at the University of Tennessee, Knoxville.

The initial nomenclature and annotation of forensic features in this collection was performed by human decomposition experts and produced approximately 5,000 annotations on 1,000 images so far. These annotations were used to train Convolutional Neural Network (CNN) models and identify areas in the remaining one million unannotated images using these models.

CNN models have revolutionized the field of image analysis, but they require massive numbers of training examples to train accurate models for a specific domain. This study is not aware of any attempts to use them in the forensic domain, in particular in recognizing features of human decomposition.

The forensic community faces several challenges, including uncontrolled natural lighting conditions that vary with weather and seasons, angles and distances at which the photos were taken, different areas of the body that the photos capture, and the difficulty of identifying and localizing the features, even by an expert in human decomposition. Furthermore, there is the need to localize features within an image and, most importantly, thousands of examples of each feature needed to train an accurate CNN model are not available.

Transfer learning, due to the absence of models suitable for forensic features, does not solve the problem. Instead, a hierarchical workflow was constructed that first normalizes the data and equalizes the images to even the distribution of intensity levels. The scale of the images was calculated by using existing CNNs to recognize body parts (using OpenPose), and, finally, sections of the images that do not contain the body were excluded. The preprocessing allows improved training and detection by focusing on specific scale and excluding features that could not be there, for example, signs of scavenging on non-muscular areas of the body, such as skin (scavengers tend to target muscles more than other areas). Existing images are then augmented by moving/resizing the area with annotation and rotating and rescaling annotated images. Mask-RCNN is applied to this normalized and augmented dataset to recognize and localize forensic features.

In preliminary work, highly complex features were recognized, such as signs of scavenging, with 36% accuracy even before fully applying the aforementioned preprocessing techniques and 38% accuracy when normalizing the dataset. The results suggest that CNNs can be successfully applied in the forensic context, even with a limited number of training examples, and that forensic content can be extracted from massive media collections, thus providing novel tools for advancing forensic science. Further work is needed to increase the accuracy of the models and create models for more forensic features.

Machine Learning, Decomposition, Taphonomy