C18 Cloth Classification With a Semi-Supervised Generative Adversarial Network (SGAN)

Jette Korthals Altes, BS, Netherlands Forensic Institute, Den Haag 2497GB, NETHERLANDS; Andrea Macarulla, MSc, Netherlands Forensic Institute and University of Amsterdam, Den Haag 2497GB, NETHERLANDS; Owen Corrigan, PhD, Dublin City University, Dublin, IRELAND; Zeno J. Geradts, PhD, Netherlands Forensic Institute, Den Haag, SH 2497 GB, NETHERLANDS*

Learning Overview: After attending this presentation, attendees will be aware of new possibilities for tracking persons based on a description of clothing.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by providing novel ways of searching for certain clothing items on video based on machine learning.

If large amounts of video are collected from different camera systems, often it is necessary to track a person based on clothing, especially if the face is not visible in the images. With that information, investigators can look for the person on Closed-Circuit Television (CCTV) and recordings from other sources in the relevant area. This is a simple task in theory, but tedious in practice, since there may be many hours of video footage with a high number of people in them. The goal of this presentation is to automate this task using machine learning methods.

One practical obstacle to achieving this goal is a lack of labeled training data. A semi-supervised learning method was applied to circumvent the lack of relevant, available datasets. A semi-supervised SGAN was used. This SGAN consists of two networks, a generator, which generates images, and a discriminator, which classifies images into the different classes or as being a generated image. The generator receives feedback from the discriminator, and will thus create more real looking images, and the discriminator will learn to correctly classify the images. The idea is that the discriminator will also learn from the generated images, even though they are not labeled as one of the classes.

New annotations were provided for 280 images of an existing dataset of surveillance footage (CUHK Person Re-identification Data Set). These annotations consist of manually tagged clothing items. New images were created, where each image contained only one clothing item. This was to simplify the task, instead of making it a multi-label classification problem. The ten classes consisting of the most images were selected. These were: a backpack, a bag, boots, a coat, jeans, oxfords, sneakers, a sweater, trousers, and a T-shirt. The SGAN was trained using 25, 50, or 100 images per class. The performance of the SGAN was compared to that of a baseline model, where the generator never got updated, so the generated images dis not resemble the real image. As expected, the baseline performed better the more images it had to train with. However, the SGAN performed worse than the baseline, regardless of the amount of images it had to train with. Visual inspection of the generated images showed that these did not look like the real images; this could be an explanation for the worse performance of the SGAN. With the SGAN, training on 100 images per class, with ten classes, an accuracy of 64% was achieved. This paves the way to classification of clothing in the field of multimedia forensics.

Deep Learning, SGAN, Clothes