C25 Analysis of "Deepfakes" Creation and Detection: Video and Image Fabrication Using Deep Learning

Jeff M. Smith, MS, National Center for Media Forensics, CU Denver, Denver, CO 80204; Catalin Grigoras, PhD, Denver, CO 80202*

Learning Overview: After attending this presentation, attendees will have a better understanding of how deep learning can be used to fabricate faces in images and videos and the challenge facing forensic examiners in the analysis of these videos.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by shedding light on how "deepfakes" face swaps are created, including the software and hardware requirements, and help in understanding the real threats facing examiners in forensics and intelligence.

Artificial Intelligence (AI) presents a boon of technological capabilities in many areas that will affect lives for years to come. With today's research in driverless cars, big data, multimedia authentication, and facial recognition technology, it is clear that AI and machine learning will be some of the underlying technologies relied upon in the future. The recent trend to use deep learning to train models for face swapping has made an impact on how the public perceive the future of media, the news, and the entertainment industry. In late 2017, reddit® Inc. website user "deepfakes" posted well-known celebrities' faces swapped onto adult film content using these machine learning techniques. Since then, non-consensual content of this nature has been banned from the reddit® website and many other well-known sites, but Safe For Work (SFW) image and video deepfakes have been very popular on the internet since then. This has led to a shift in the public perception of the trustworthiness of media and the potential spread of fake news in the future becoming a real threat. Deepfakes viral videos demonstrate to the general public that a technology can be used to create convincing false content videos. Fake videos of one person saying what is not true content and not actually what the original person had said with the appearance of accurate facial expressions and mouth movements is a prospect that has the public concerned.

During this presentation, many examples will be shown. With this research, the realities of producing deepfake face swaps (Figure 1) using current technology will be discussed, including the Python machine learning libraries and Graphics Processing Unit (GPU) hardware processing that make it possible. Limitations will be explored as well as the potential advantages of using deep learning for video content creation over traditional manual or automated copy/paste methods. Finally, a discussion of the forensic detection of deepfakes fabricated videos in a forensic setting (Figure 2) will be covered.

This material is based on research sponsored by Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL) under agreement number FA8750-16-2-0187. The United States government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon.

Figure 1: Frame of video face swapped (center) using deep learning given the source face (right) swapped onto the original face and body (left).

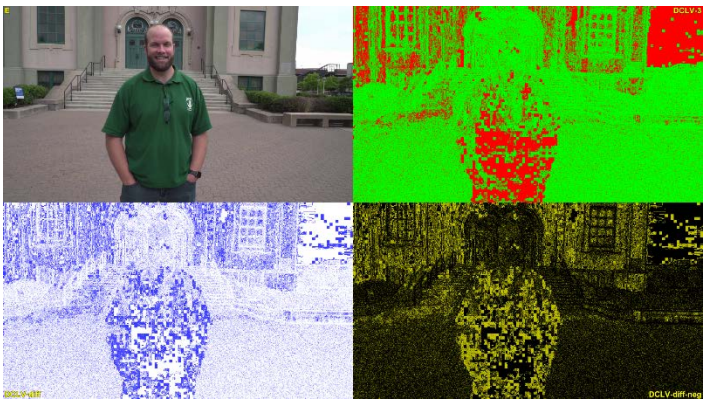

Figure 2: Video authenticity plots of a deepfakes face swapped video frame.