

Detecting fake face images created by both humans and machines

October 29 2018, by Ingrid Fadelli



Image from PGGAN dataset. Credit: Karras et al.

Researchers at the State University of New York in Korea have recently explored new ways to detect both machine and human-created fake images of faces. In [their paper](#), published in ACM Digital Library, the researchers used ensemble methods to detect images created by generative adversarial networks (GANs) and employed pre-processing

techniques to improve the detection of images created by humans using Photoshop.

Over the past few years, significant advancements in image processing and machine learning have enabled the generation of fake, yet highly realistic, images. However, these images could also be used to create fake identities, make [fake news](#) more convincing, bypass image detection algorithms, or fool image recognition tools.

"Fake face images have been a topic of research for quite some time now, but studies have mainly focused on photos made by humans, using Photoshop tools," Shahroz Tariq, one of the researchers who carried out the study told Tech Xplore. "Recently, [a study by Karras et al.](#) showed that a generative adversarial [network](#) (GAN) could produce near realistic human face images. People could use these photos maliciously, for instance, to create fake IDs on the internet."

The focus of the research carried out by Tariq and his colleagues was to detect both computer-generated and human-generated fake photos of faces using deep learning techniques. To do this, they developed a neural network classifier and trained it on a dataset of real and fake images.



Image from CelebA dataset. Credit: Liu et al.

"The neural network classifier learns the discriminating features between the real and fake images by examining a massive database of both fake and real images," Tariq said.

Rather than analysing the metadata of images, this classifier focuses on the image content. In preliminary tests, it achieved remarkable results, detecting both GANs-generated and human-created fake images of faces with 94 percent accuracy.

"Even if computer generated images look highly realistic to the human eye, the neural network classifier was able to spot some minute differences, which allowed it to correctly classify the images," Tariq said. "We also found that fake photos created by humans using

Photoshop tools are much harder to detect, as there are many possible variations."



Real image. Credit: Tariq et al.

In future, the classifier developed by Tariq and his colleagues could help to identify fake images, generated by GANs or by humans using graphics editing software, such as Photoshop. The researchers are now planning to develop their classifier further, training it on more machine and human-generated images.



Photoshopped image. Credit: Tariq et al.

"As methods to generate synthetic images are becoming more sophisticated, photos generated with these methods will become more realistic and it will be harder for the neural network classifier to detect their differences," Tariq explained. "We hence want to keep improving our methods, to better detect such photos."

More information: Shahroz Tariq et al. Detecting Both Machine and Human Created Fake Face Images In the Wild, *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security - MPS '18* (2018). [DOI: 10.1145/3267357.3267367](https://doi.org/10.1145/3267357.3267367)

Progressive growing of GANS for improved quality, stability, and variation. arXiv:1710.10196 [cs.NE] arxiv.org/abs/1710.10196

Large-scale CelebFaces Attributes (CelebA) Dataset.
mmlab.ie.cuhk.edu.hk/projects/CelebA.html

© 2018 Science X Network

Citation: Detecting fake face images created by both humans and machines (2018, October 29)
retrieved 3 February 2023 from
<https://techxplore.com/news/2018-10-fake-images-humans-machines.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.