Exposing GAN-generated Faces Using Inconsistent Corneal Specular Highlights

Shu Hu[1], Yuezun Li[2], and Siwei Lyu[1]

[1]Media Forensic Lab, Department of Computer Science and Engineering, University at Buffalo, SUNY, USA

[2]The Department of Computer Science and Technology, Ocean University of China, China

Poster number: 1634

Problem Description

Can you tell which are fake?

Figure 1: Image source: real images (the first row): the Flickr-Faces-HQ (FFHQ) dataset. GAN-synthesized images (the second row): http://thispersondoesnotexist.com

Generative adversarial network (GAN) models can now be used to create highly realistic synthetic images of human faces down to the minuscule details. Such fake images have put our trust in online visual media on the line, corroborated by the recent reports (e.g., from the Verge, CNN, Reuters) of GAN synthesized faces being used as profile images of fake social media accounts.

How to identify the fake images?

Motivation

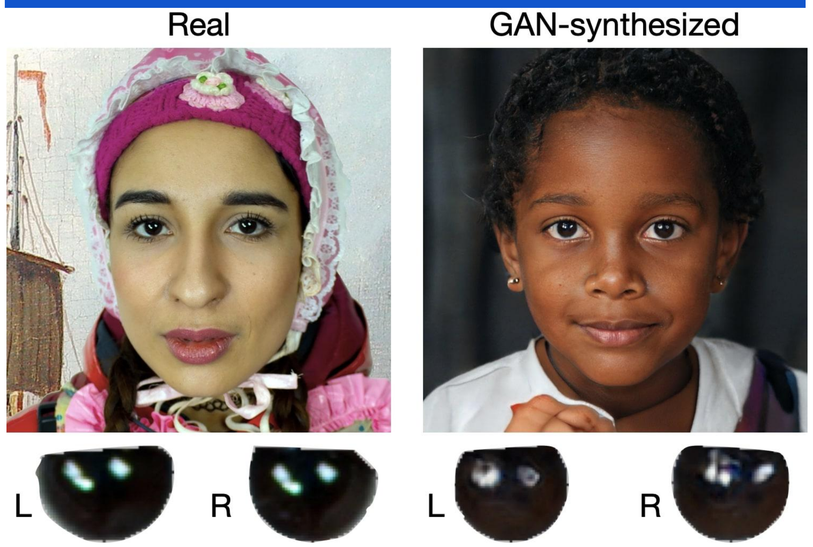

Figure 2: Corneal specular highlights for a real human face (left) and a GAN-synthesized face (right). The corneal regions are isolated and scaled for better visibility.

Note that the corneal specular highlights for the real face have strong similarities while those for the GAN-synthesized face are different.

Background

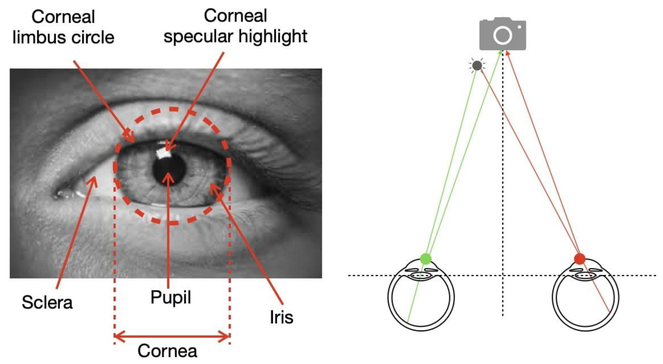

Figure 3: (left) Anatomy of a human eye. (right) The portrait setting with the corneal specular highlights.

The cornea has a spherical shape and its surface exhibits mirror-like reflection characteristics, which generates the corneal specular highlights when illuminated by light emitted or reflected in the environment at the time of capture.

Method

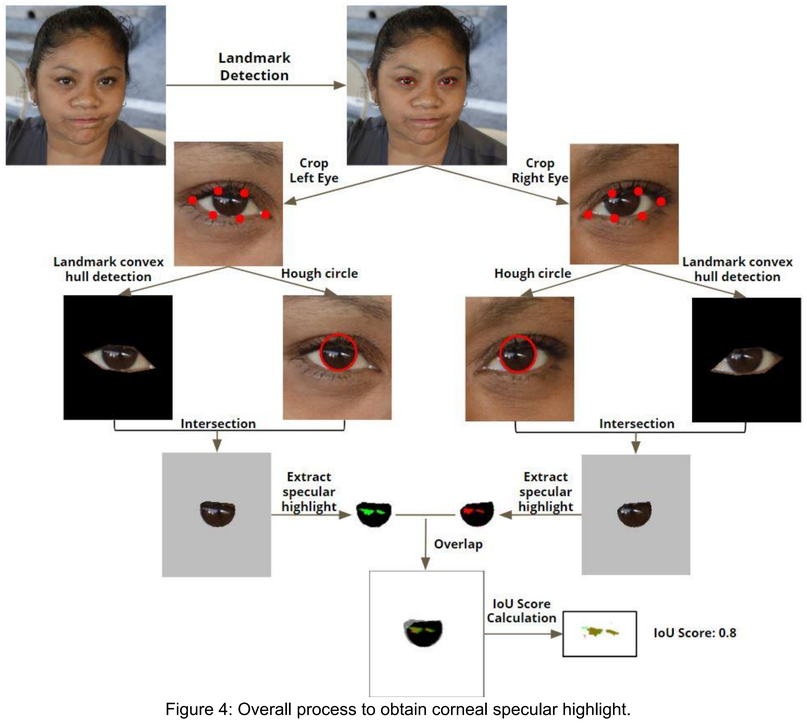

Figure 4: Overall process to obtain corneal specular highlight.

We align the extracted corneal specular highlights of the two eyes (denoted as R_L and R_R) with a translation, and use their *Intersection over Union* (IoU) scores, $\frac{|R_L \cap R_R|}{|R_L \cup R_R|}$, as a similarity metric.

Experiments (Consistency Analysis)

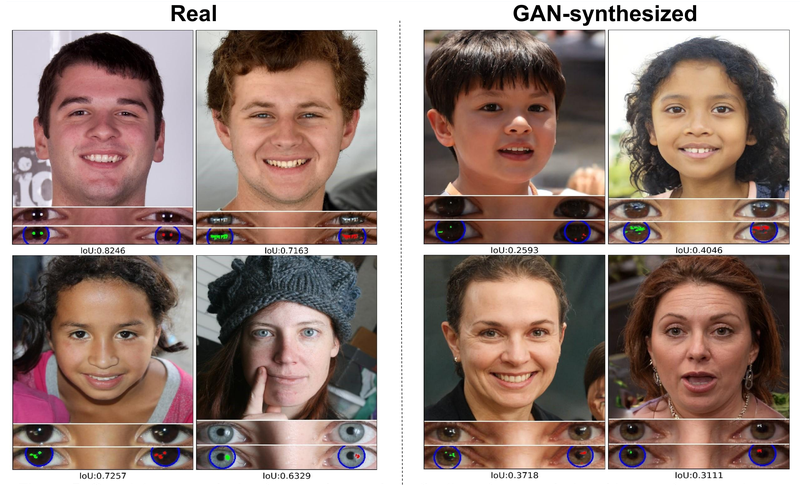

Figure 5: Consistency analysis of corneal specular reflections for sampled real human faces and sampled GAN synthesized faces.

Experiments (Numerical Analysis)

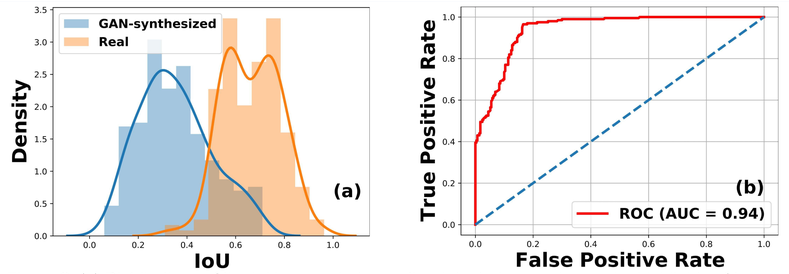

Figure 6: (a) Distributions of the IoU scores between the detected corneal specular highlights of two eyes for real and GAN synthesized faces. (b) The *receiver operating characteristic* (ROC) curve based on the IoU scores.

Conclusion & Future Work

In this work, we show that GAN synthesized faces can be exposed with the inconsistent corneal specular highlights between two eyes. Our method has several limitations.

- We only compare pixel differences without considering inconsistencies in geometry and scene.
- Also, we may have false positives when the portrait setting is not obeyed, e.g, when a light source is very close to the subject or a peripheral light source that is not visible in both eyes. It does not apply to images where specular patterns are not present.

In the future, we will investigate these aspects and further improve the effectiveness of our method.

Project page: https://bit.ly/3rfnWjB. UB NEWS: https://bit.ly/3tOLYUh.

Acknowledgments. This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001120C0123 and the Center for Identification Technology Research and the National Science Foundation under Grant No. 1822190.