

Identifying deep network generated images using disparities in color components

September 4 2018, by Ingrid Fadelli



Example of real face image from dataset provided by paper “Progressive Growing of GANs for Improved Quality, Stability, and Variation”. Credit: Karras et al.

Researchers at Shenzhen University have recently devised a method to detect images generated by deep neural networks. Their study, [pre-published on arXiv](#), identified a set of features to capture color image statistics that can detect images generated using current artificial intelligence tools.

"Our research was inspired by the rapid development of image generative models and the spread of generated fake [images](#)," Bin Li, one of the researchers who carried out the study, told *Tech Xplore*. "With the rise of advanced image generative models, such as generative adversarial networks (GAN) and variational autoencoders, images generated by deep networks become more and more photorealistic, and it is no longer easy to identify them with human eyes, which entails serious security risks."

Recently, several researchers and global media platforms have expressed their concern with the risks posed by artificial neural networks trained to generate images. For instance, deep learning algorithms such as generative adversarial networks (GAN) and variational autoencoders could be used to generate realistic images and videos for fake news or could facilitate online frauds and the counterfeit of personal information on social media.

GAN algorithms are trained to generate increasingly realistic images through a trial and error process in which one algorithm generates images and another, the discriminator, gives feedback to make these images more realistic. Hypothetically, this discriminator could also be trained to detect false images from real ones. However, these algorithms

primarily use RGB images as input and do not consider disparities in color components, hence their performance would most likely be unsatisfactory.



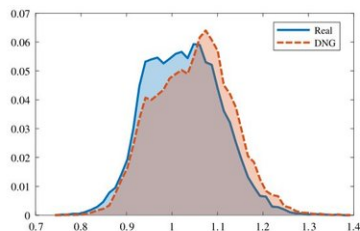
Example of generated face image from dataset provided by paper “Progressive Growing of GANs for Improved Quality, Stability, and Variation”. Credit: Karras et al.

In their study, Li and his colleagues analyzed the differences between images generated by GAN and real images, proposing a set of features that could effectively help to classify them. The resulting method works by analyzing differences in color components between true and generated images.

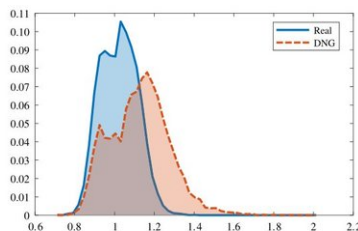
"Our basic idea is that the generation pipelines of real images and generated images are quite different, so the two classes of images should have some different properties," Haodong Li, one of the researchers explained. "In fact, they are coming from different pipelines. For example, real images are generated by imaging devices such as cameras and scanners to capture a real scene, while generated images are created in a totally different way with convolution, connection, and activation from [neural networks](#). The differences may result in different statistical properties. In this study, we mainly considered the statistical properties of color components."

The researchers found that although generated images and real images look alike in the RGB color space, they have markedly different [statistical properties](#) in the chrominance components of HSV and YCbCr. They also observed differences when assembling the R, G, and B color components together.

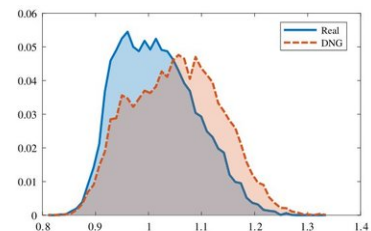
The feature set they proposed, which consists of co-occurrence matrices extracted from the image high-pass filtering residuals of several color components, exploits these differences, capturing disparities in color between real and generated images. This feature set is of low dimension and can perform well even when trained on a small image dataset.



(a) R component in RGB color space



(b) H component in HSV color space



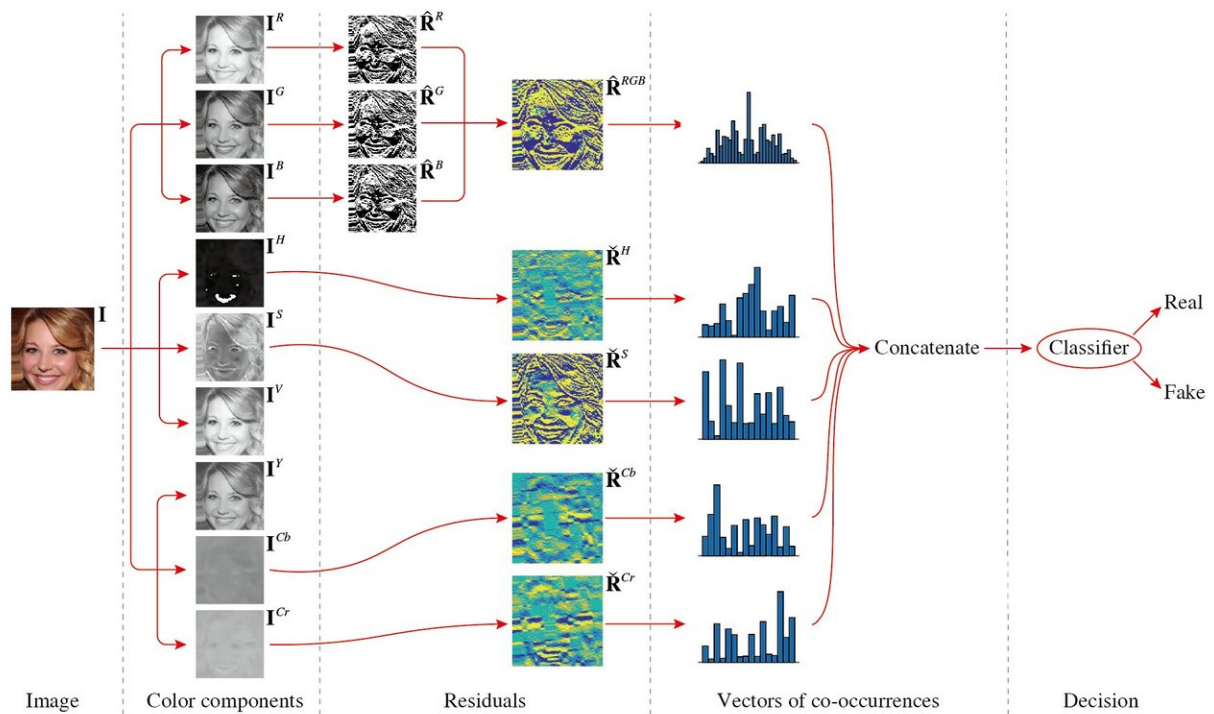
(c) Assembling R, G, and B components

The histograms of image statistics for real images (blue) and deep network generated (DNG) images (red) in different color components. In the R component, the two histograms are largely overlapped. However, in the H component or by assembling R, G, B components together, the histograms are more separable. Credit: Li et al.

Li and his colleagues tested the performance of their method on three image datasets: CelebFaces Attributes, High-quality CelebA, and Labeled Faces in the Wild. Their findings were very promising, with the set of features performing well on all three datasets.

"The most meaningful finding of our study is that images generated by deep networks can be easily detected by extracting features from certain color components, although the generated images may be visually indistinguishable with human eyes," Haodong Li said. "When generated image samples or generative models are available, the proposed features equipped with a binary classifier can effectively differentiate between generated images and real images. When the generative models are unknown, the proposed features together with a one-class classifier can also achieve satisfactory performance."

The study could have a number of practical implications. First, the method could help to identify fake images online.



The overall framework of the proposed method. The input image is first decomposed into different color components, and then the residual of each color component is computed. To calculate the co-occurrences, the R, G, and B components are assembled, while the H, S, Cb, and Cr components are processed independently. Finally, all the co-occurrence vectors are concatenated and fed to a classifier for obtaining the decision result. Credit: Li et al.

Their findings also imply that several inherent color properties of real images have not yet been effectively replicated by existing generative models. In the future, this knowledge could be used to build new models that generate even more realistic images.

Finally, their study proves that different generation pipelines used to produce real images and generated images are reflected in the properties of the images produced. Color components comprise merely one of the

ways in which these two types of images differ, so further studies could focus on other properties.

"In future, we plan to improve the image generation performance by applying the findings of this research to image generative models," Jiwu Huang, one of the researchers said. "For example, including the disparity metrics of color components for real and generated images into the objective function of a GAN model could produce more realistic images. We will also try to leverage other inherent information from the real image generation pipeline, such as sensor pattern noises or properties of color filter array, to develop more effective and robust methods for identifying generated images."

More information: Detection of Deep Network Generated Images Using Disparities in Color Components. arXiv:1808.07276v1. arxiv.org/abs/1808.07276

Progressive Growing of GANs for Improved Quality, Stability, and Variation. arXiv: 1710.10196 [cs.NE]. arxiv.org/abs/1710.10196

© 2018 Tech Xplore

Citation: Identifying deep network generated images using disparities in color components (2018, September 4) retrieved 24 April 2024 from <https://techxplore.com/news/2018-09-deep-network-images-disparities-components.html>

<p>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.</p>
--