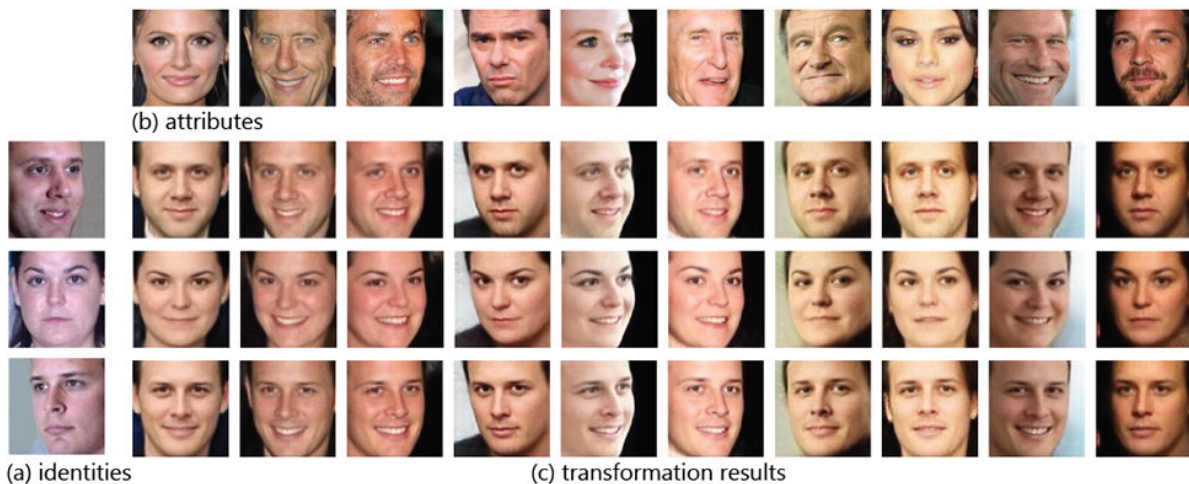


# Insightful research illuminates the newly possible in the realm of natural and synthetic images

June 18 2018

---



Credit: Microsoft

A pair of groundbreaking papers in computer vision open new vistas on possibilities in the realms of creating very real-looking natural images and synthesizing realistic, identity-preserving facial images. In CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training, presented this past October at ICCV 2017 in Venice, the team of researchers from Microsoft and the University of Science and Technology of China came up with a model for image generation based on a variational autoencoder generative adversarial network capable of

synthesizing natural images in what are known as fine-grained categories. Fine-grained categories would include faces of specific individuals, say of celebrities, or real-world objects such as specific types of flowers or birds.

The researchers – Dong Chen, Fang Wen and Gang Hua of Microsoft, Jianmin Bao, an intern at Microsoft Research, along with Houqiang Li of China's University of Science and Technology – in looking at how to better build effective generative models of natural images were grappling with a key problem in computer vision: how to generate very diverse and yet realistic images by varying a finite number of latent parameters as related to the natural distribution of any image in the world. The challenge lay in coming up with a generative model to capture that data. They opted for an approach using generative adversarial networks combined with a variational auto-encoder to come up with their learning framework. The approach models any image as a composition of label and latent attributes in a probabilistic model. By varying the fine-grained category label (say, "oriole" or "starling" for specific bird types, or the names of specific celebrities) that would be fed into the generative model, the team was able to synthesize images in specific categories using randomly drawn values with respect to the latent attributes. It's only recently that this kind of deep learning makes possible the modeling of the distribution of images of specific objects out in the world, allowing us to draw from that model to basically synthesize the image, explained Gang Hua, principal researcher at Microsoft Research in Redmond, Washington.

"Our approach has two novel aspects," said Hua. "First, we adopted a cross-entropy loss for the discriminative and classifier network but opted for a mean discrepancy objective for the generative network." The resulting asymmetric loss function and its effect on the machine learning aspects of the framework were encouraging. "Asymmetric loss actually makes the training of the GANs more stable," said Hua. "We designed

an asymmetric loss to address the instability issue in training of vanilla GANs that specifically addresses numerical difficulties when matching two non-overlapping distributions."

The other innovation was adopting an encoder network that could learn the relationship between the latent space and use pairwise feature matching to retain the structure of the synthesized images.

Experimenting with [natural images](#) – genuine photographs of real things found in nature such as faces, flowers and birds, the researchers were able to show that their machine learning models could synthesize recognizable images with an impressive variety within very specific categories. The potential applications cover everything from image inpainting, to data augmentation and better facial recognition models.

"Our technology addressed a fundamental challenge in image generation, that of the controllability of [identity](#) factors. This allows us to generate images as we want them to look. said Hua."

## **Synthesizing faces**

How do you take the power to synthesize realistic images of flowers or birds a step further? You look at human faces. Human faces, when taken in the context of identity, are among the most sophisticated images that can be captured in nature. In *Toward Open-Set Identity Preserving Face Synthesis*, presented this month at CVPR 2018 in Salt Lake City, the researchers developed a GAN-based framework that can disentangle the identity and attributes of faces, with attributes including such intrinsic properties as the shapes of noses and mouths or even age, as well as environmental factors, such as lighting or whether makeup was applied to the face. While previous identity preserving face synthesis processes were largely confined to synthesizing faces with known identities that were already contained in the training dataset, the researchers developed

a method of achieving identity-preserving face synthesis in open domains – that is, for a face that fell outside any training dataset. To do this, they landed upon a unique method of using one input images of a subject that would produce an identity vector and combined it with any other input face image (not of the same person) in order to extract an attribute vector, such as pose, emotion or lighting. The identity vector and the attribute vector are then recombined to synthesize a new face for the subject featuring the extracted attribute. Notably, the framework does not have to annotate and categorize the attributes of any of the faces in any way. It is trained with an asymmetric loss function to better preserve the identity and stabilize the machine learning aspects. Impressively, it can also effectively leverage massive amounts of unlabeled training face images (think random facial images) to further enhance the fidelity or accuracy of the synthesized [faces](#).

One obvious consumer application is the classic example of the photographer's challenge of taking a group photo that includes dozens of subjects; the common objective is the elusive ideal shot in which all subjects are captured eyes open and even smiling. "With our technology, the great thing is that I could literally render a smiling face for each of the participants in the shot!" exclaims Hua. What makes this utterly different from mere image editing, says Hua, is that the actual identity of the face is preserved. In other words, although the image of a smiling participant is synthesized – a "moment" that did not in fact occur in reality, the face is unmistakably that of the individual; his or her identity has been preserved in the process of altering the image.

Hua sees many useful applications that will benefit society and sees constant improvements in image recognition, video understanding and even the arts.

Provided by Microsoft

Citation: Insightful research illuminates the newly possible in the realm of natural and synthetic images (2018, June 18) retrieved 23 April 2024 from <https://phys.org/news/2018-06-insightful-illuminates-newly-realm-natural.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.