Adobe Research Extends Disentangled GAN Face Editing

Originally published at https://www.unite.ai/adobe-research-extends-disentangled-gan-face-editing/

It's not difficult to understand why entanglement is a problem in image synthesis, because it's often a problem in other areas of life; for instance, it's far harder to remove turmeric from a curry than it is to discard the pickle in a burger, and it's practically impossible to de-sweeten a cup of coffee. Some things just come bundled.

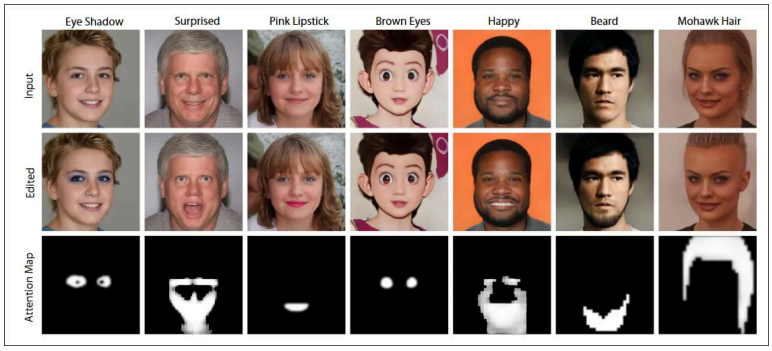
Likewise entanglement is a stumbling block for image synthesis architectures that would ideally like to separate out different features and concepts when using machine learning to create or edit faces (or *dogs*, *boats*, or any other domain).

If you could separate out strands such as *age*, *gender*, *hair color*, *skin tone*, *emotion*, and so forth, you would have the beginnings of real instrumentality and flexibility in a framework that could create and edit face images at a truly granular level, without dragging unwanted 'passengers' into these conversions.

At maximum entanglement (above left), all you can do is change the image of a learned GAN network to the image of another person.

This is effectively using the latest AI computer vision technology to achieve something that was solved by other means over thirty years ago.

With some degree of separation ('Medium Separation' in earlier above image), it's possible to perform style-based changes such as hair color, expression, cosmetic application, and limited head rotation, among others.

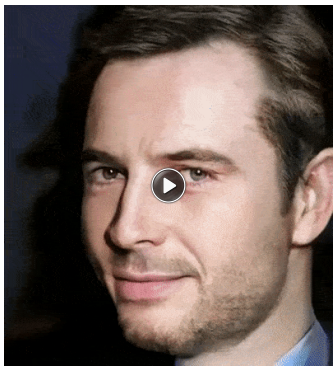


Source: *FEAT: Face Editing with Attention*, February 2022, https://arxiv.org/pdf/2202.02713.pdf

There has been a number of attempts in the last two years to create interactive face-editing environments that allow a user to change facial characteristics with sliders and other traditional UI interactions, while keeping core features of the target face intact when making additions or changes. However, this has proved a challenge due to the underlying feature/style entanglement in the latent space of the GAN.

For instance, the *glasses* trait is frequently enmeshed with the *aged* trait, meaning that adding glasses might also 'age' the face, while ageing the face might add glasses, depending on the degree of applied separation of high-level features (see 'Testing' below for examples).

Most notably, it has been almost impossible to alter hair color and other hair facets without the hair strands and disposition being recalculated, which gives a 'sizzling', transitional effect.

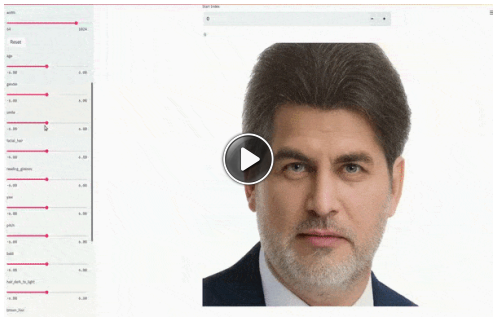


CLICK TO VIEW ANIMATED GIF

Source: InterFaceGAN Demo (CVPR 2020), https://www.youtube.com/watch?v=uoftpl3Bj6w

Latent-to-Latent GAN Traversal

A new Adobe -led paper entered for WACV 2022 offers a novel approach to these underlying issues in a paper entitled *Latent to Latent: A Learned Mapper for Identity Preserving Editing of Multiple Face Attributes in StyleGAN-generated Images*.



CLICK TO VIEW ANIMATED GIF

Supplemental material from the paper Latent to Latent: A Learned Mapper for Identity Preserving Editing of Multiple Face Attributes in StyleGAN-generated Images. *Here we see that base characteristics in the learned face are not dragged into unrelated changes.* ***See full video embed at end of article for better detail and resolution.*** Source: https://www.youtube.com/watch?v=rf_61lIRH0Q

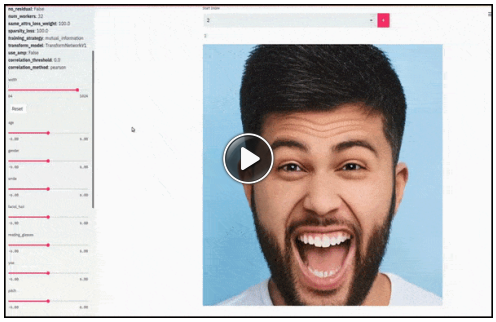
The paper is led by Adobe Applied Scientist Siavash Khodadadeh, together with four other Adobe researchers, and a researcher from the Department of Computer Science at the University of Central Florida.

The piece is interesting partly because Adobe has been operating in this space for some time, and it's tempting to imagine this functionality entering a Creative Suite project in the next few years; but mainly because the architecture created for the project takes a different approach to maintaining visual integrity in a GAN face editor while changes are being applied.

The authors declare:

> *'[We] train a neural network to perform a latent-to-latent transformation which finds the latent encoding corresponding to the image with the changed attribute. As the technique is one-shot, it does not rely on a linear or non-linear trajectory of the gradual change of the attributes.*
>
> *'By training the network end-to-end over the full generation pipeline, the system can adapt to the latent spaces of off-the-shelf generator architectures. Conservation properties, such as maintaining the identity of the person can be encoded in the form of training losses.*
>
> *'Once the latent-to-latent network was trained, it can be reused for arbitrary images without retraining.'*

This last part means that the proposed architecture arrives with the end-user in a finished state. It still needs to run a neural network on local resources, but new images can be 'dropped in' and be ready for altering almost immediately, since the framework is decoupled enough not to need further image-specific training.



CLICK TO VIEW ANIMATED GIF

Gender and facial hair changed as sliders plot random and arbitrary paths through the latent space, not just 'scrubbing between endpoints'. See video embedded at end of article for more transformations at better resolution.

Among the main achievements in the work is the network's ability to 'freeze' identities in the latent space by changing only the attribute in a target vector, and providing 'correction terms' that conserve identities being transformed.

Essentially, the proposed network is embedded in a broader architecture that orchestrates all the processed elements, which pass through pre-trained components with frozen weights that will not produce unwanted lateral effects on transformations.

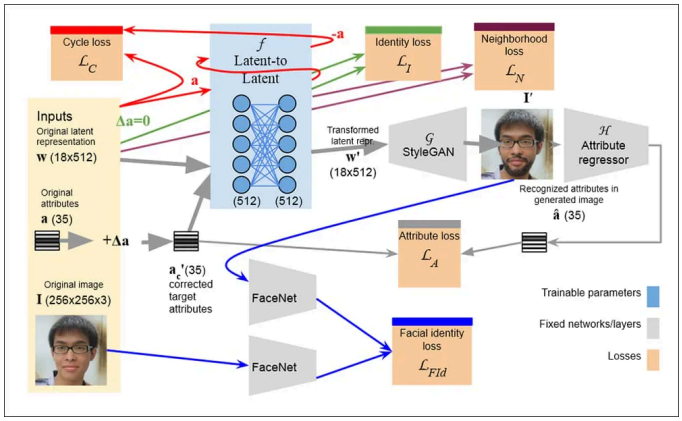
Since the training process relies on triplets that can be generated either by a seed image (under GAN inversion) or an existing initial latent encoding, the entire training process is unsupervised, with the tacit actions of the customary range of labeling and curation systems in such systems effectively baked into the architecture. In fact, the new system uses off-the-shelf attribute regressors:

> *'[The] number of attributes that our network can independently control is only limited by the capabilities of the recognizer(s) – if one has a recognizer for an attribute, we can add it to arbitrary faces. In our experiments, we trained the latent-to-latent network to allow the adjustment of 35 different facial attributes, more than any previous approach.'*

The system incorporates an additional safeguard against undesired 'side-effect' transformations: in the absence of a request for an attribute change, the latent-to-latent network will map a latent vector to itself, further increasing stable persistence of the target identity.

Facial Recognition

One recurring issue with GAN and encoder/decoder-based face editors of the past few years has been that applied transformations tend to degrade resemblance. To combat this, the Adobe project uses an embedded facial recognition network called FaceNet as a discriminator.



Project architecture, see lower mid-left for inclusion of FaceNet. Source: *Latent to Latent: A Learned Mapper for Identity Preserving Editing of Multiple Face Attributes in StyleGAN-generated Images*, OpenAccess.

(On a personal note, this seems an encouraging move toward the integration of standard facial identification and even expression recognition systems into generative networks, arguably the best way forward for overcoming the blind pixel>pixel mapping that dominates current deepfake architectures at the expense of expression fidelity and other important domains in the face generation sector.)

Access All Areas in the Latent Space

Another impressive feature of the framework is its ability to travel arbitrarily between potential transformations in the latent space, at user whim. Several prior systems that provided exploratory interfaces often left the user essentially 'scrubbing' between fixed feature transformation timelines – impressive, but often quite a linear or proscriptive experience.

CLICK TO VIEW ANIMATED GIF

From Improving GAN Equilibrium by Raising Spatial Awareness*: here the user scrubs through a range of potential transition points between two latent space locations, but within the confines of pre-trained locations in the latent space. To apply other kinds of transformation based on the same material, reconfiguration and/or retraining is necessary.* Source: https://genforce.github.io/eqgan/

In addition to being receptive to entirely novel user images, the user can also manually 'freeze' elements that they want to be conserved during the transformation process. In this way the user can ensure that (for instance) backgrounds do not shift, or that eyes are kept open or closed.

Data

The attribute regression network was trained on three networks: FFHQ, CelebAMask-HQ, and a local, GAN-generated network obtained by sampling 400,000 vectors from the Z space of StyleGAN-V2.

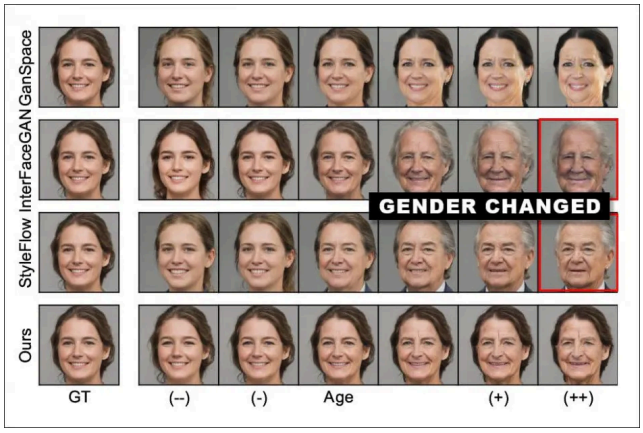
Out-of-distribution (OOD) images were filtered away, and attributes extracted using Microsoft's Face API, with the resulting image-set split 90/10, leaving 721,218 training images and 72,172 test images to compare against.

Testing

Though the experimental network was initially configured to accommodate 35 potential transformations, these were slimmed down to eight in order to undertake analogous testing against the comparable frameworks InterFaceGAN, GANSpace, and StyleFlow.

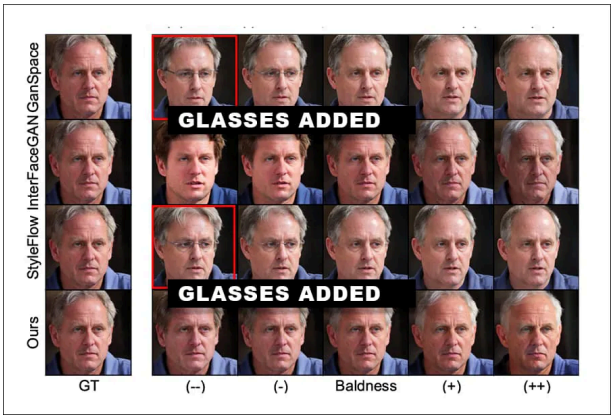
The eight selected attributes were *Age*, *Baldness*, *Beard*, *Expression*, *Gender*, *Glasses*, *Pitch*, and *Yaw*. It was necessary to retool the competing frameworks for certain of the eight attributes that were not provisioned in the original distribution, such as adding *baldness* and *beard* to InterFaceGAN.

As expected, a greater level of entanglement occurred in the rival architectures. For instance, in one test, InterFaceGAN and StyleFlow both changed the gender of the subject when asked to apply *age*:



Two of the competing frameworks rolled a gender change into the 'age' transformation, also changing hair color without direct bidding of the user.

Additionally, two of the rivals found that glasses and age are inseparable facets:



Glasses and hair color change thrown in at no extra charge!

It's not a uniform victory for the research: as can be seen in the accompanying video embedded at the end of the article, the framework is the least effective when trying to extrapolate diverse angles (yaw), while GANSpace has a better general result for *age* and the imposition of *glasses*. The latent-to-latent framework tied with GANSpace and StyleFlow regarding the adding of pitch (angle of head).

ATTRIBUTE	IG	SF	GS	L2L
AGE	0.77	0.65	**0.46**	0.57
BALDNESS	0.53	0.64	0.46	**0.21**
BEARD	0.57	0.55	0.53	**0.28**
EXPRESSION	0.60	0.21	0.23	**0.14**
GENDER	0.54	0.52	0.58	**0.28**
GLASSES	0.59	0.46	**0.24**	0.25
PITCH	0.54	**0.51**	**0.51**	**0.51**
YAW	**0.39**	0.46	0.41	0.46

Results calculated based on a calibration of the MTCNN face detector. Lower results are better.

For further details and better resolution of examples, check out the paper's accompanying video below.

Latent to Latent - WACV 2022

First published 16th February 2022.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More