Video generation based on StyleGAN

As denoted on the page of Sep. 7 - Sep.11, method a) and method b) are based on images generated through StyleGAN, so generally speaking it's like playing around with latent codes of the images. I found two ways to achieve this: latent editing and interpolation.

1. Latent Space Editing

In the StyleGAN Study Note I mentioned that traditional latent codes is disentangled into dlatent space (i.e. intermediate latent space W') for better control and interpretability of the generation process. Thus, we could get use of the feature to decide on what each frame should be like based on the dlatent space.

1.1 Feature Editing

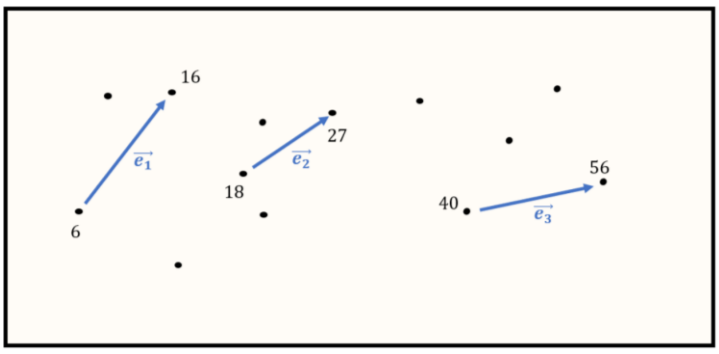
The editing of features of face is achieved through computing the relationship between change of certain feature (e.g. how much angle the face turns) and the change of the value of dlatent.

+rotation_angle	int32
+yaw	double
+pitch	double
+roll	double
+expression	uint32
+expression_probability	double
+faceshape	object[]
++type	string
++probability	double
+gender	string
+gender_probability	double
+glasses	uint32
+glasses_probability	double

The figure on the right side shows some examples of such features. And then how could we compute such a relation between changes?

We could actually represent the relationship e as a vector such that:

$$\hat{e} = \frac{\sum_{i=1}^{n} \overrightarrow{e_i}}{n} = \frac{\sum_{i=1}^{n} random_{p,q}\left(\frac{dlatent_p - dlatent_q}{label_p - label_q}\right)}{n}$$



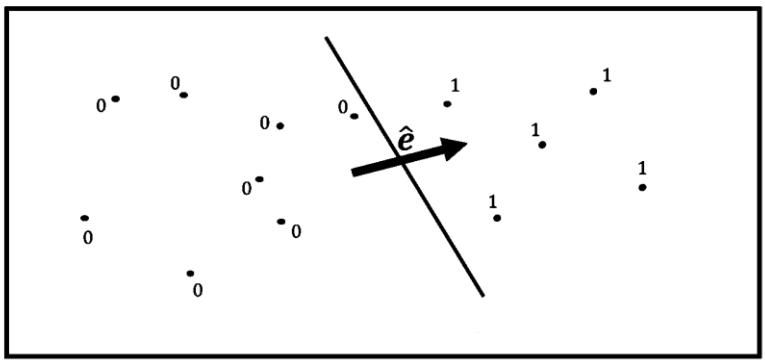
dlatent distribution

where the labels represents the features of the faces.

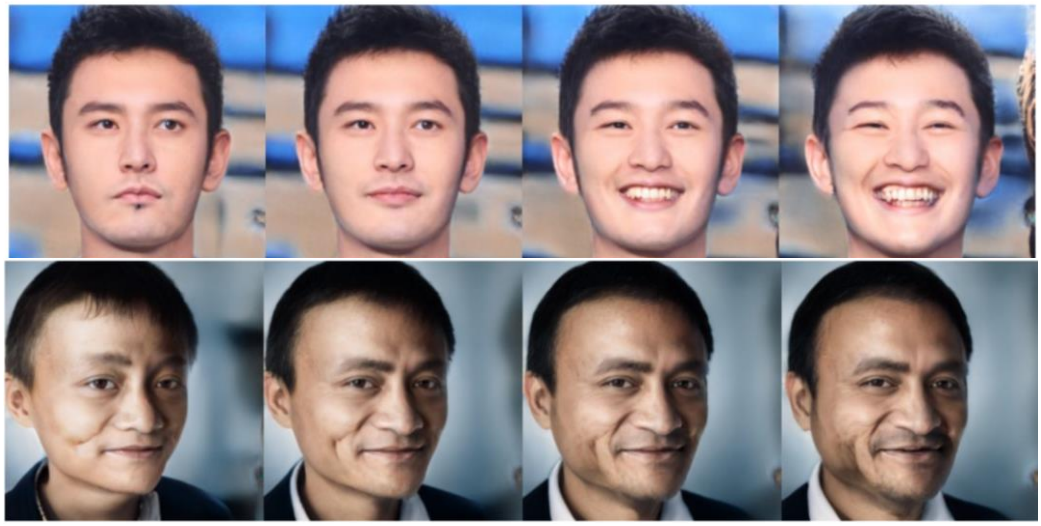
To further improve the efficiency of computation, we could apply logistic regression where we

find a vector w to maximum P(wx + b = y) with y = 0 for all latenti < median of label and y = 1 for all latenti > median of labels.

$$\hat{e} = \vec{w} = argmax_w P(\vec{w} \cdot x + \vec{b} = y), y \in \{0,1\}$$

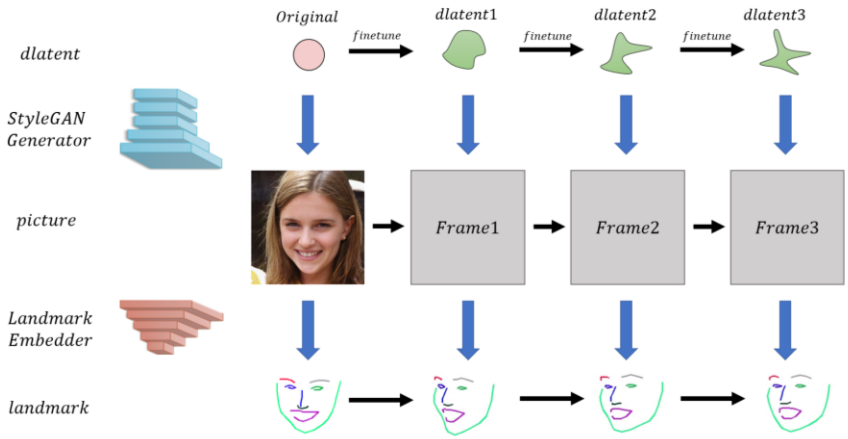


With the vector w computed for each of the feature, we could apply linear operation on the vector w and original dlatent. To control the changing extent, we could multiply the vector with a certain coefficient. Thus, by multiplying with a sequence of coefficients, we could get a smooth change of the images.



1.2 Editing based on landmarks

This is a possible future improvement of the feature editing. Instead of deciding the changing process by users, we could encode the landmarks (simple contour line of a face) of aimed output and then tune the dlatent of original image towards the dlatet if the aimed output through the loss of landmraks.



2. Interpolation

Although StyleGAN makes many improvements based on ProGAN, it removes the config of interpolation video generation. From [2], I learned about three general ways of interpolation video generation. The following description is copied from the website:

- A standard interpolation video, which is simply a random walk through the latent space, modifying all the variables smoothly and animating it. [2]



- A coarse “style mixing” video; a single “source” face is generated & held constant; a secondary interpolation video, a random walk as before is generated; at each step of the random walk, the ‘coarse’/high-level ‘style’ noise is copied from the random walk to overwrite the source face’s original style noise. For faces, this means that the original face will be modified with all sorts of orientations & facial expressions while still remaining recognizably the original character. [2]



- A “fine” style mixing video; in this case, the style noise is taken from later on and instead of affecting the global orientation or expression, it affects subtler details like the precise shape of hair strands or hair color or mouths. [2]

- *Circular interpolations* are another interesting kind of interpolation, written by snowy halcy, which instead of random walking around the latent space freely, with large or awkward transitions, instead tries to move around a fixed high-dimensional point doing: “binary search to get the MSE to be roughly the same between frames (slightly brute force, but it looks nicer), and then did that for what is probably close to a sphere or circle in the latent space.” [2]

There are also many other implementations of interpolation video. With transfer learning, we could apply the interpolation between any of categories of images and thus make the style mixing. However, it is also found the more complex the categories are, the poorer performance the StyleGAN gives. Thus, future work might be focused on BigGAN for more complicated categories of images interpolation.

I also found that interpolation is actually quite similar to editing through landmarks, except that one is based on the loss of landmarks and the other is directly based on dlatent. Both are actually playing around the latent codes in nature.

3. References

[1] a312863063. “Note_StyleGAN.” http://www.seeprettyface.com/research_notes.html, 21 Nov. 2019. Web. 23 Sep. 2020

[2] gwern. “Making anime faces with Style GAN” https://www.gwern.net/Faces, 26 Jul. 2020. Web. 23 Sep. 2020