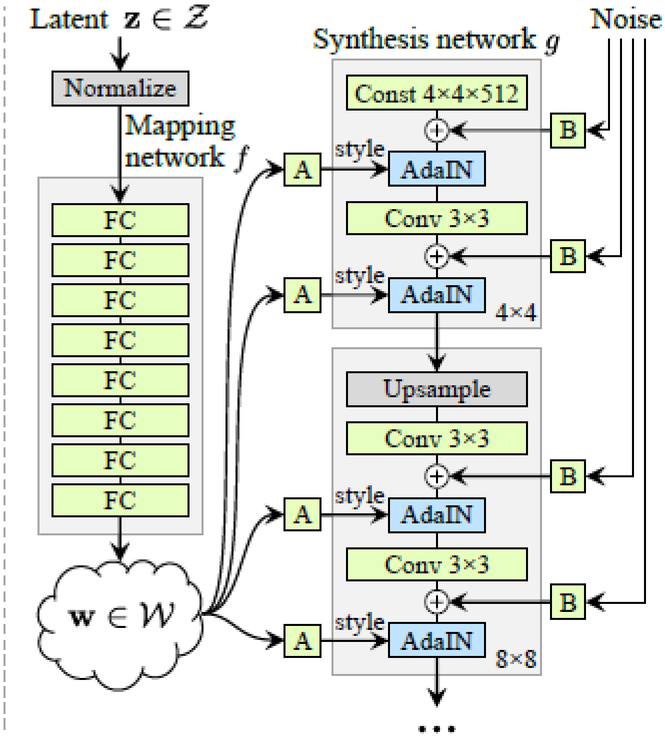
STYLE-GAN SUMMARY

ABSTRACT:

Alternative generator architecture for generative adversarial networks. This new architecture leads to an automatically learned, unsupervised separation of high-level attributes and stochastic variation in the generated images and it enables intuitive, scale-specific control of the synthesis.

Introduction:



(b) Style-based generator

The generator starts from a learned constant input and adjust the "style" of the image at each convolution layer based on the latent code, therefore directly controlling the strength of image features at different scales. Combined with noise injected directly into the network, this architectural change leads to automatic, unsupervised separation of high-level attributes from stochastic variation in the generated images, and enables intuitive scale-specific mixing and interpolation operations. The discriminator and the loss function are not modified.

The generator embeds the input latent code into an intermediate latent space, which has a profound effect on how the factors of variation are represented in the network. The input latent space must follow the probability density of the training data (*this will lead to some degrees of unavoidable entanglement*). The intermediate latent space if free from this restriction and is therefore allowed to be disentangled.

Present two automated metrics for qualifying these aspects of the generator:

- Perceptual path length.
- Linear separability.

PROPERTIES OF THE STYLE-BASED GENERATOR:

- STYLE MIXING:

Mixing Regularization: a given percentage of images are generated using two random latent codes instead of one during training.

Two latent codes, *z1* and *z2*, through the mapping network, and having the corresponding *w1*, *w2* allows to control the styles so that *w1* applies before the crossover point and *w2* after it. This regularization technique prevents the network from assuming that adjacent tyles are correlated.

In other words, to generate an image, it switches from one latent code to another *(style mixing)* at a randomly selected point in the synthesis network.

- Stochastic Variation:

There are many aspects in human portraits that can be regarded as stochastic, such as the exact placement of hairs, stubble, freckles, or skin pores. Any of these can be randomized without affecting our perception of the image as long as they follow the correct distribution.

It is observed that the noise only affects the stochastic aspects, leaving the overall composition and high-level aspects such as identity intact. It is interesting that the effect of noise appears tightly focalized in the network. Appears the hypothesis that at any point in the generator, there is a pressure to introduce a new content as soon as possible, and the easiest way for the network to create a stochastic variation is to rely on the noise provided.

- **Separation of global effects from stochasticity:**

In this style-based generator the style affects the entire image because complete feature maps are scaled and biased with the same values. Therefore, global effects can be controlled coherently. Meanwhile, the noise is added independently to each pixel and is thus ideally suited for controlling stochastic variation. Thus, the network learns to use global and local channels appropriately, without explicit guidance.

DISENTANGLEMENT STUDIES:

The major benefit of the generator architecture is that the intermediate latent space *W* does not have to support sampling according to any *fixed* distribution; its sampling density is induced by the *learned* piecewise continuous mapping *f(z).* This mapping can be adapted to *"unwarp" W* so that the factors of variation become more linear. As such, it is expected the training to yield a less entangled *W* in an unsupervised setting.

- **Perceptual path length:**

As noted by Laine, interpolation of latent-space vectors may yield surprisingly non-linear changes in the image. To quantify this effect, how drastic changes the image undergoes can be measured as the interpolation is performed in the latent space.

As a basis for the metric, a perceptually-based pairwise image distance that is calculated as a weighted difference between two VGG16 embeddings is used. Where the weights are fit so that the metric agrees with human perceptual similarity judgements. If a latent space interpolation is subdivided into linear segments, the total perception length of this segmented path can be defined as the sum of the image distance metric.

- **Linear Separability:**

If the latent space is sufficiently disentangled, it should be possible to find direction vectors that consistently correspond to individual factors of variation. To quantify this effect a metric, that measures how well the latent-space points can be separated into two distinct sets by a linear hyperplane, so that each set corresponds to a binary attribute of the image, is purposed.

CONCLUSIONS

- Traditional GAN generator architecture is in every way inferior to a style-based design.
- This is true due to the purposed metrics and the visual results about how the stochastic and deterministic attributes are understood by the network.

EXTRA INFORMATION

- Same discriminator architectures as for Tensorflow Progressive GANs by Karras et al..
- Resolution dependent minibatch sizes.
- Adam hyperparameters.
- Exponential moving average of the generator.
- Depending on the dataset, enable mirror augmentation.
- The training time is one week on an NVIDIA DGX-1 with 8 Tesla V100 GPUs.