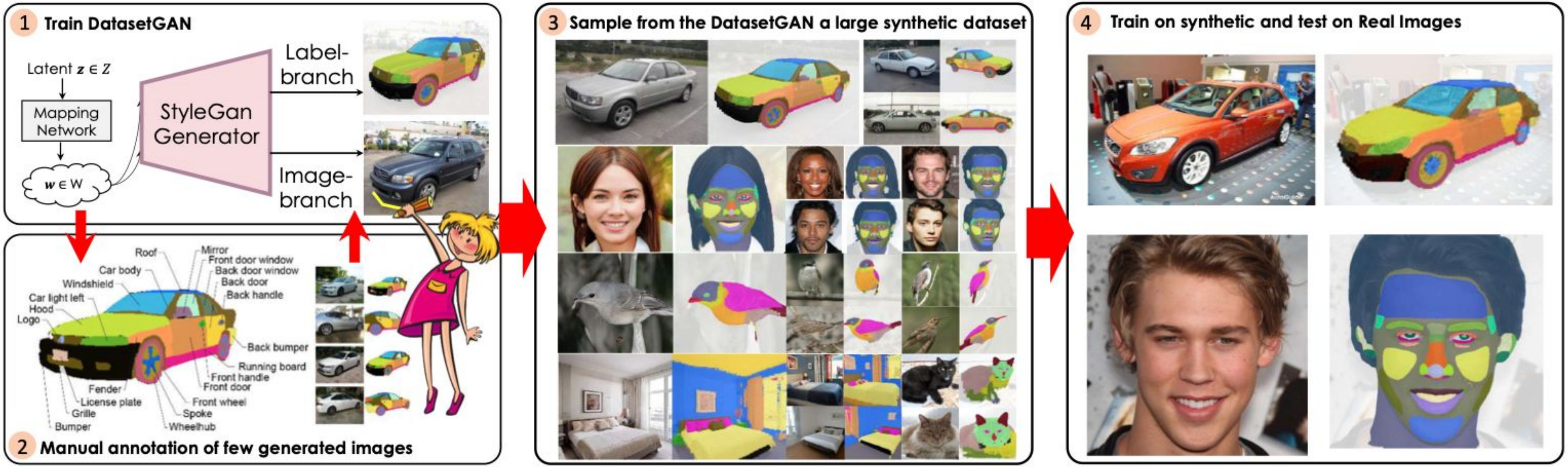
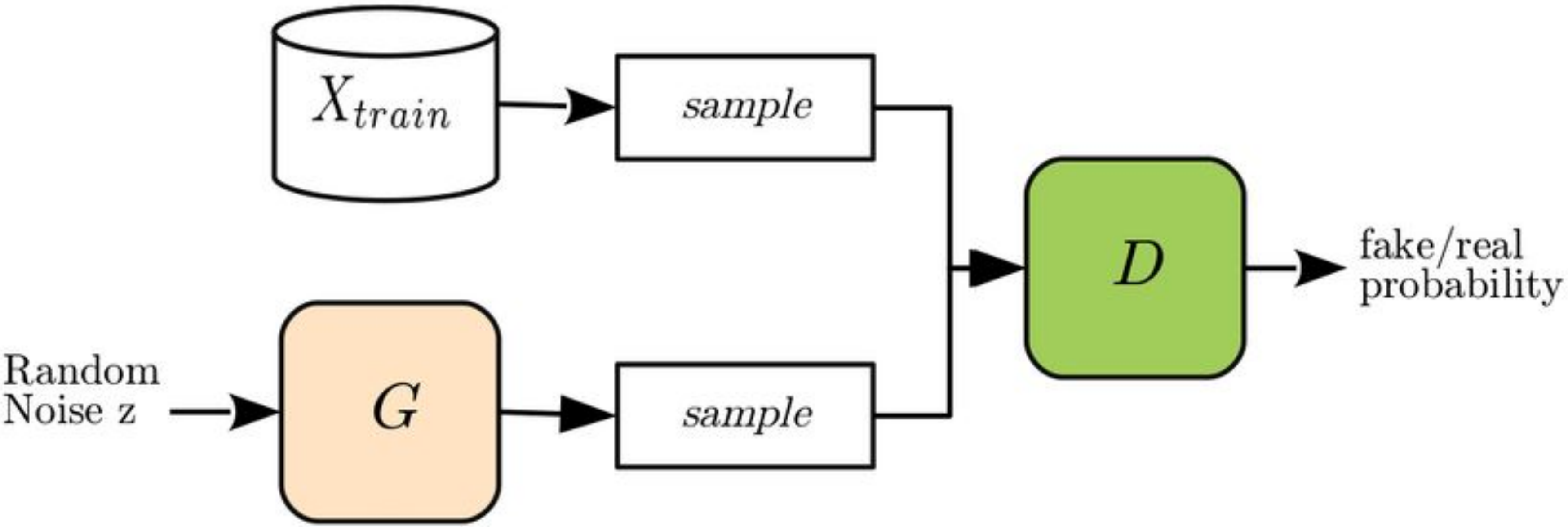
DatasetGAN



Presentation by Slava Pirogov, HSE AMI 2022

GAN reminder

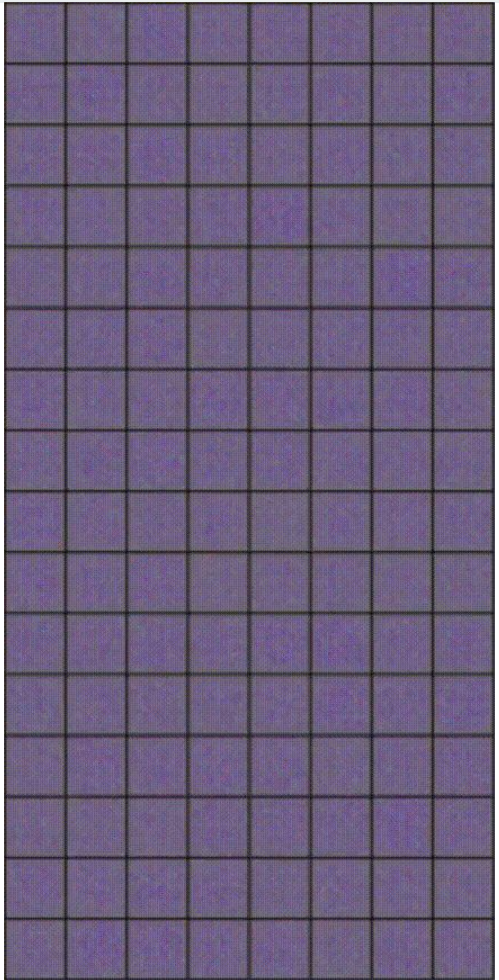


Ian J. Goodfellow 2014

GAN problems

Vanishing Gradients

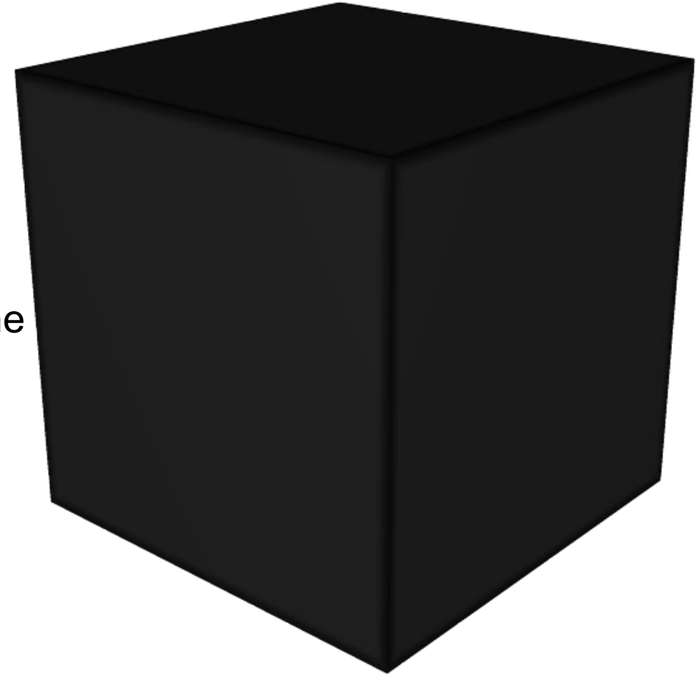
Research



GAN problems

Generator - black box

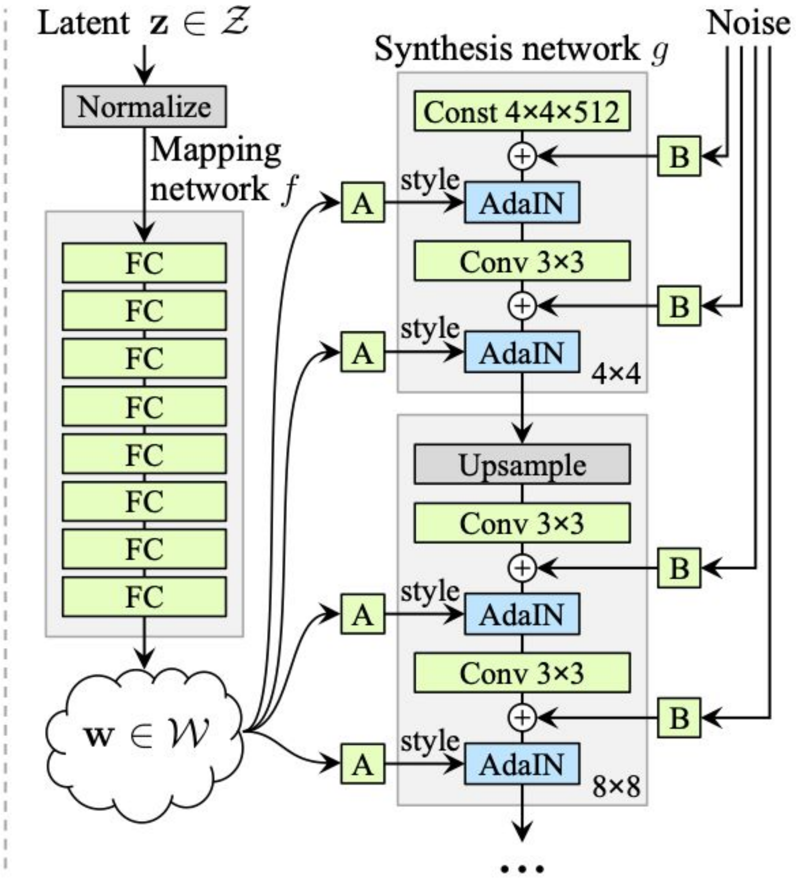
Understanding of various aspects of the image synthesis process is still lacking, the properties of the latent space are also poorly understood



StyleGAN

Redesigned generator architecture in a way that exposes novel ways to control the image synthesis process

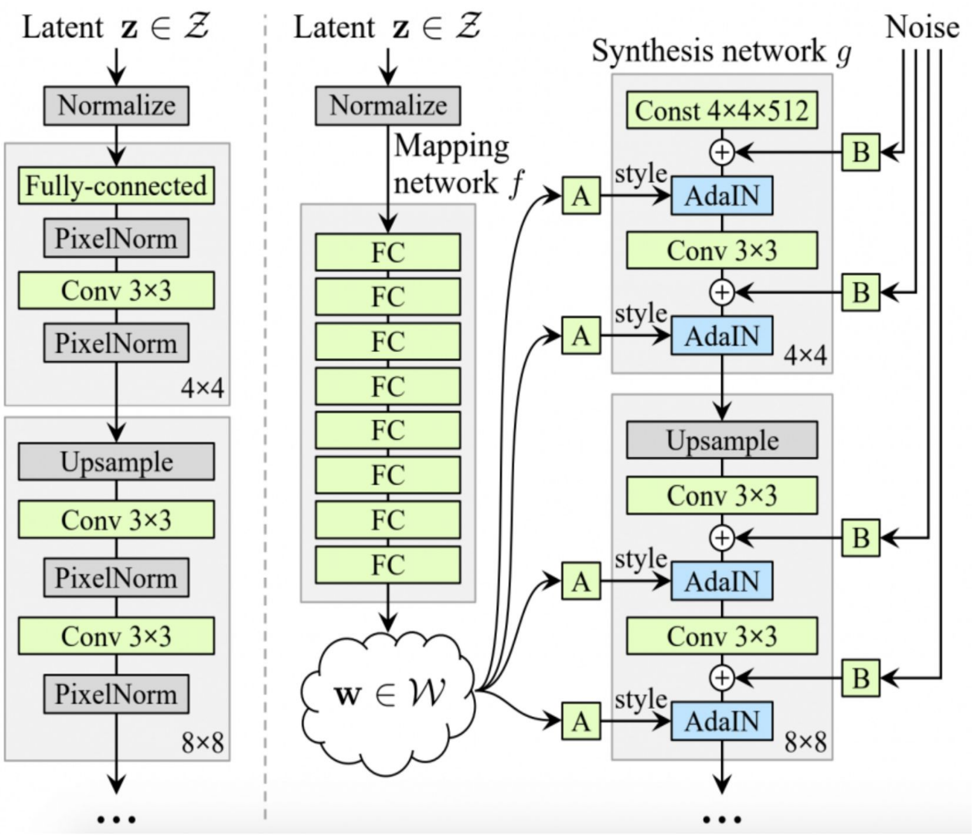
Nvidia, Tero Karras end of 2018



New Generator

StyleGAN

Traditionally the latent code is provided to the generator through an input layer. StyleGAN departs from this design, using MLP with 8 fully-connected layers

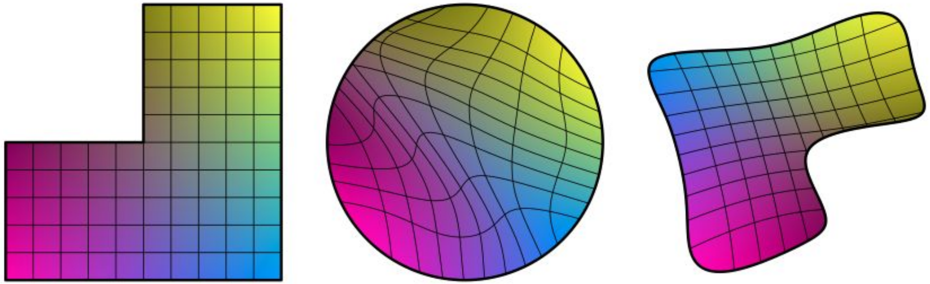


Traditional

Style-based

StyleGAN

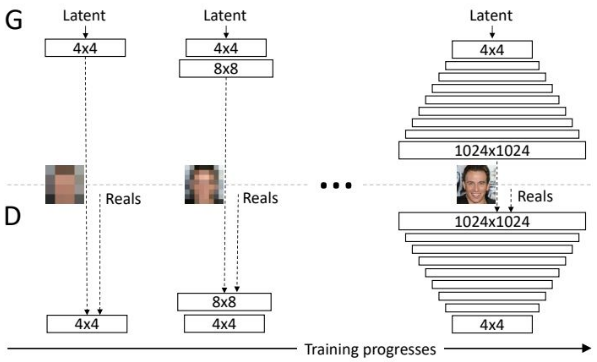
Motivation for using MLP instead of Gaussian noise. For example, on x-axis women and men, on y-axis beard and hair style

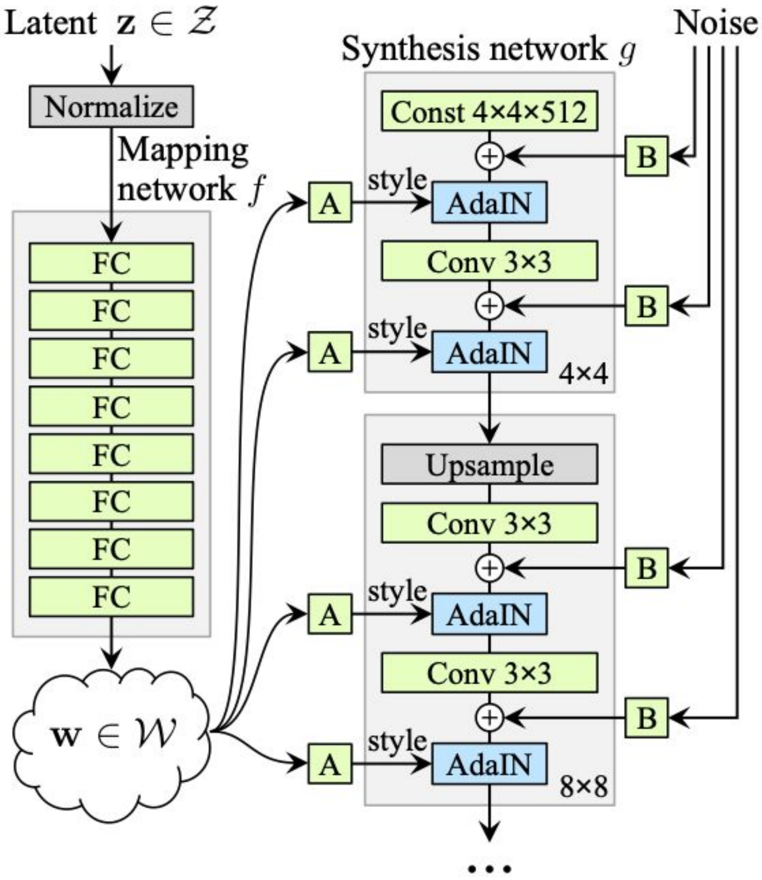


(a) Distribution of features in training set (b) Mapping from $\mathcal{Z}$ to features (c) Mapping from $\mathcal{W}$ to features

StyleGAN

Progressive Growing GAN

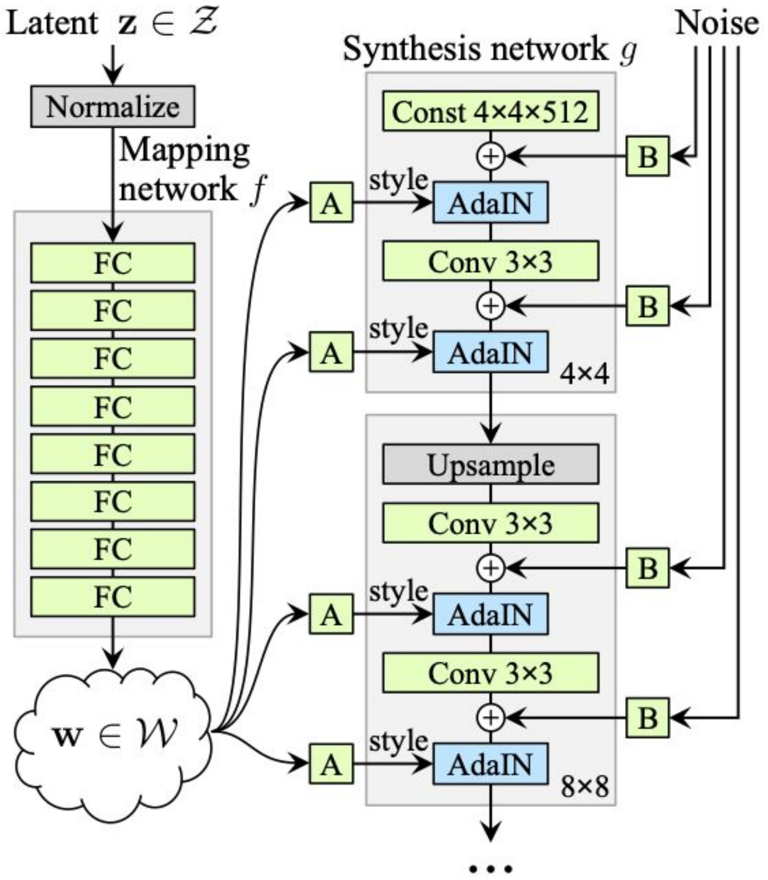




StyleGAN

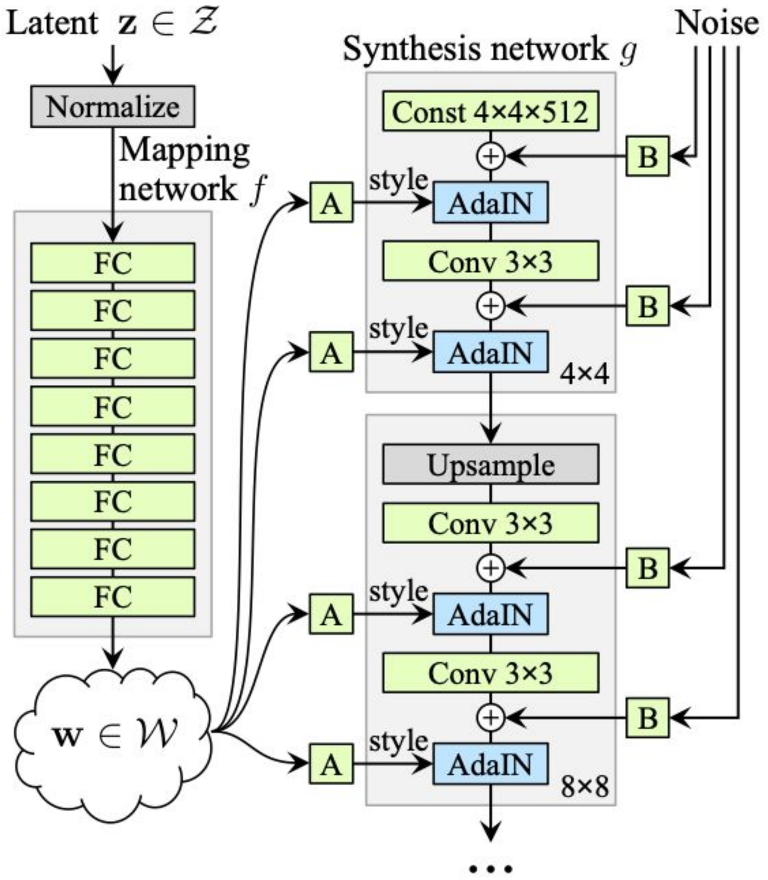
Learned affine transformations (A) specialize w to *styles* that control adaptive instance normalization (AdaIN) operations. The AdaIN operation is defined as

$$\mathrm{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i},$$



StyleGAN

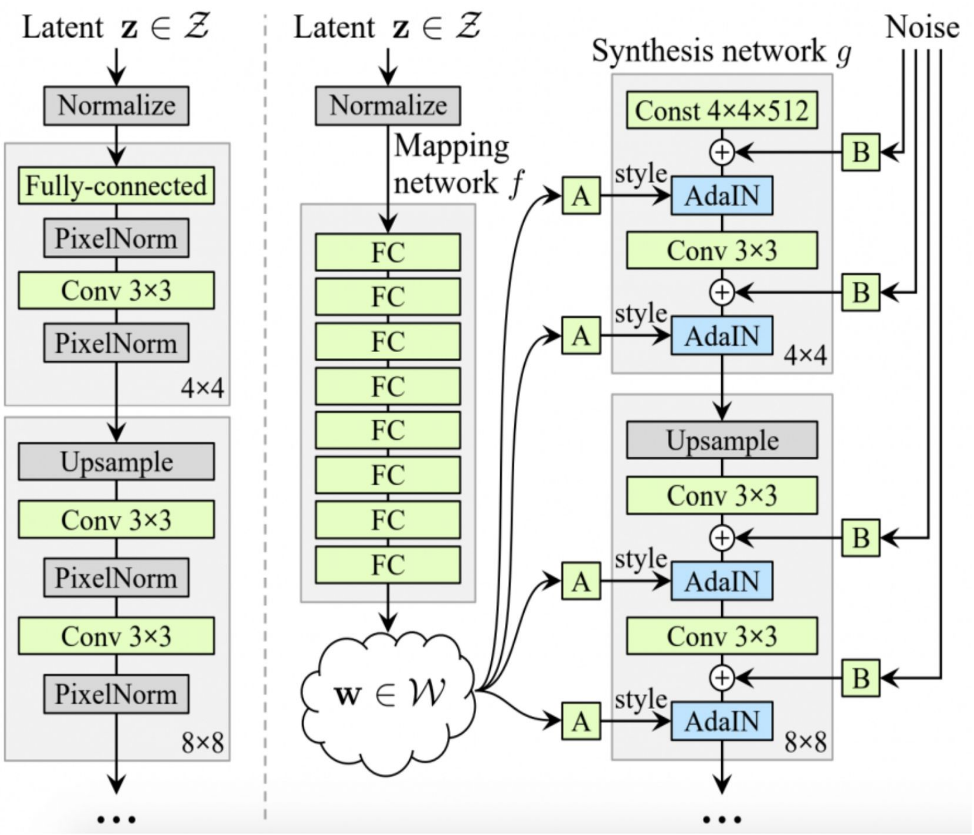
Finally, StyleGAN feeds a Gaussian noise to each layer of the synthesis network



StyleGAN

Here “A” stands for a learned affine transform, and “B” applies learned per-channel scaling factors to the noise input

$$\mathrm{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i}\frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i},$$



Traditional

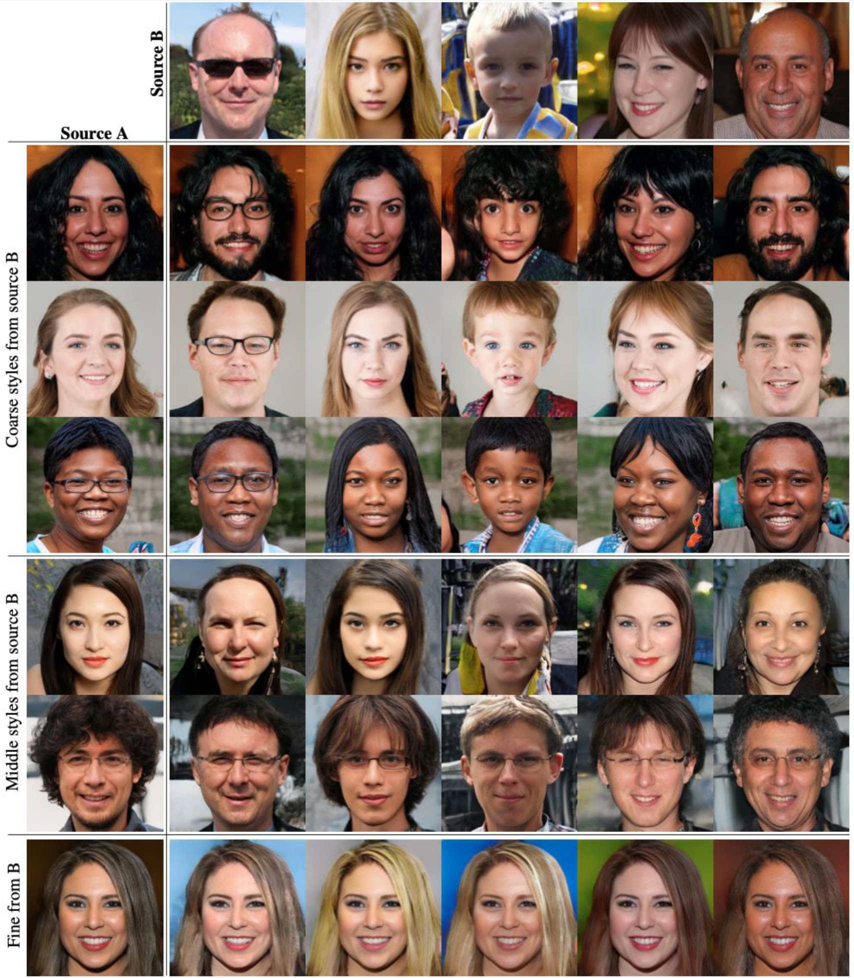
Style-based

Quality

	Method	FID CelebA-HQ	FID FFHQ
A	Baseline Progressive GAN [30]	7.79	8.04
B	+ Tuning (incl. bilinear up/down)	6.11	5.25
C	+ Add mapping and styles	5.34	4.85
D	+ Remove traditional input	5.07	4.88
E	+ Add noise inputs	**5.06**	4.42
F	+ Mixing regularization	5.17	**4.40**

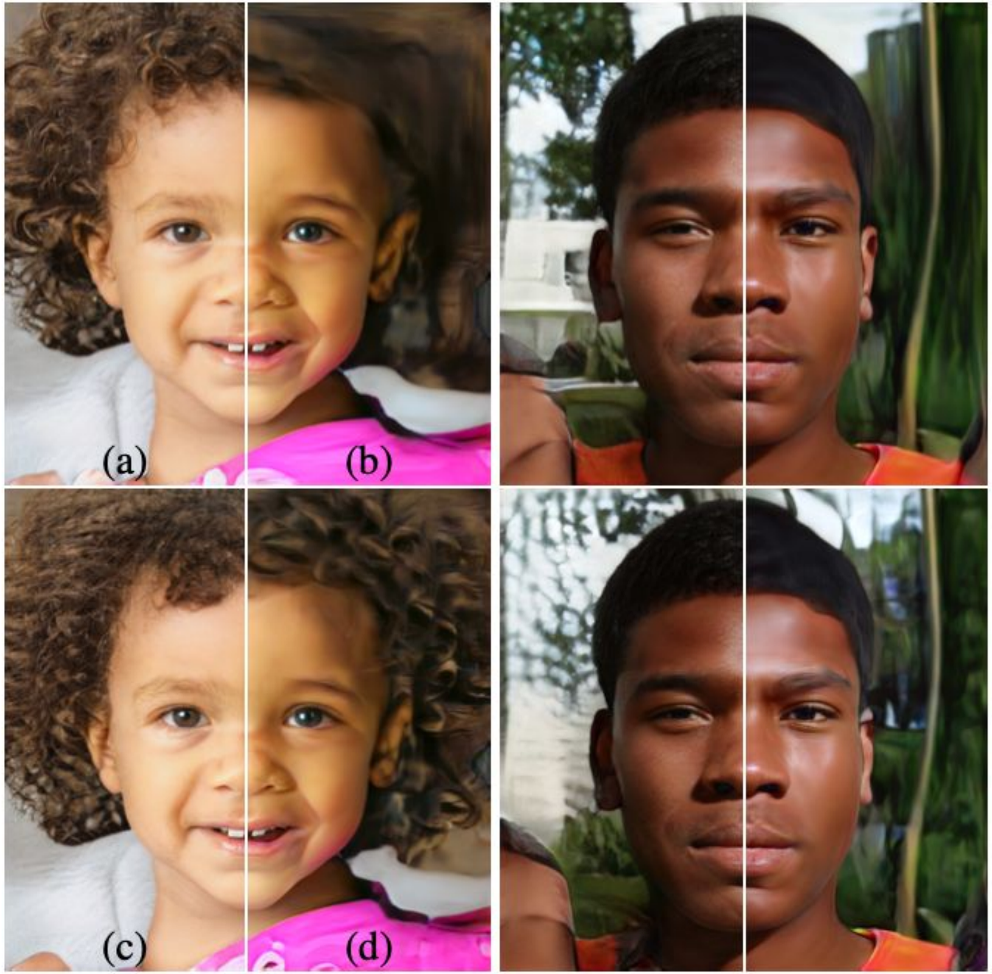
Mixing

Two sets of images were generated from their respective latent codes (sources A and B); the rest of the images were generated by copying a specified subset of styles from source B and taking the rest from source A

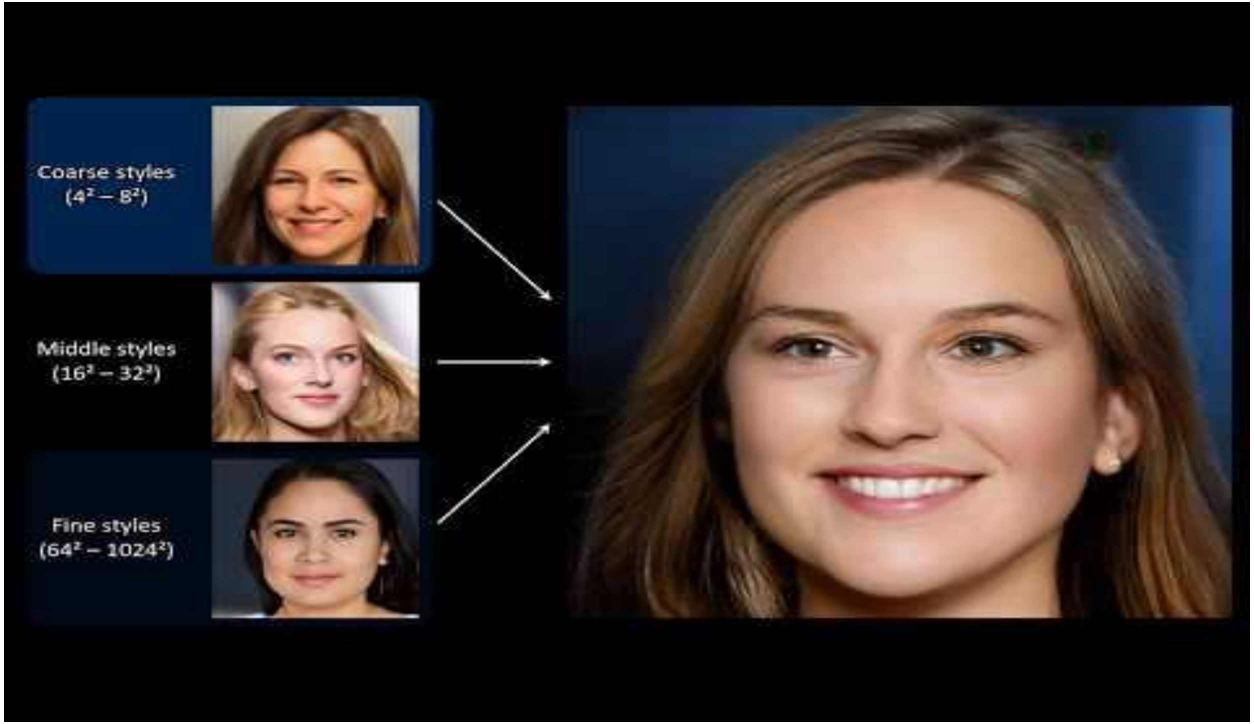


Noise effect

Effect of noise inputs at different layers of our generator (a) Noise is applied to all layers (b) No noise (c) Noise in fine layers only (64^2 – 1024^2) (d) Noise in coarse layers only (4^2 – 32^2)



Examples


Coarse styles
($4^2 – 8^2$)
Middle styles
($16^2 – 32^2$)
Fine styles
($64^2 – 1024^2$)

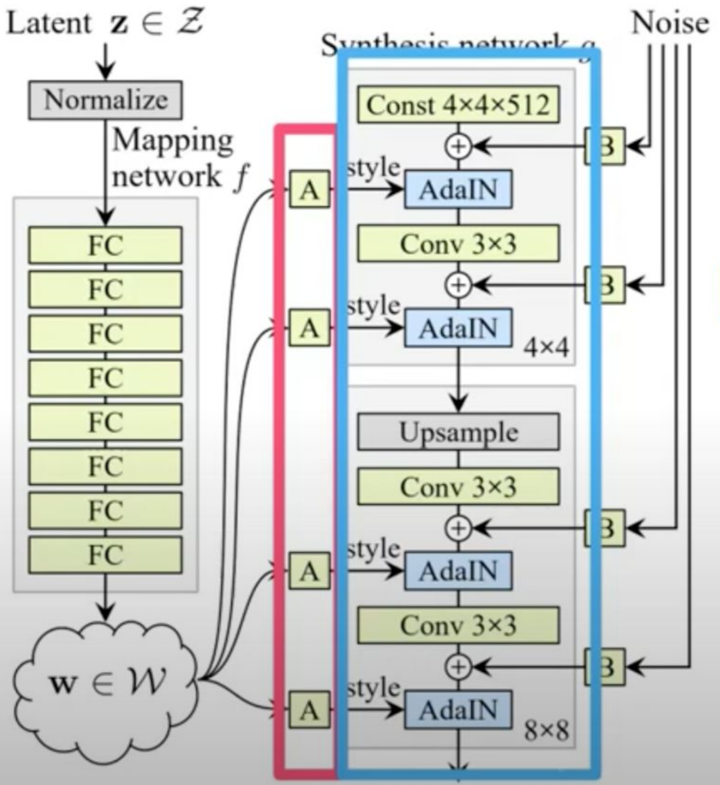
StyleGANs

- StyleGAN2 - new normalization in G, add PPl regularization, change upsample system, … Better results :)
- StyleGAN2 ADA - new augmentations for better results and increasing learning rate
- StyleGAN3 - solving alias problem, old results, paves the way for video generate

StyleGAN summary

- New generator with a lot of details
- Novel ways to control the image synthesis process
- Very good quality
- New datasets with high-quality images at 1024^2 resolution

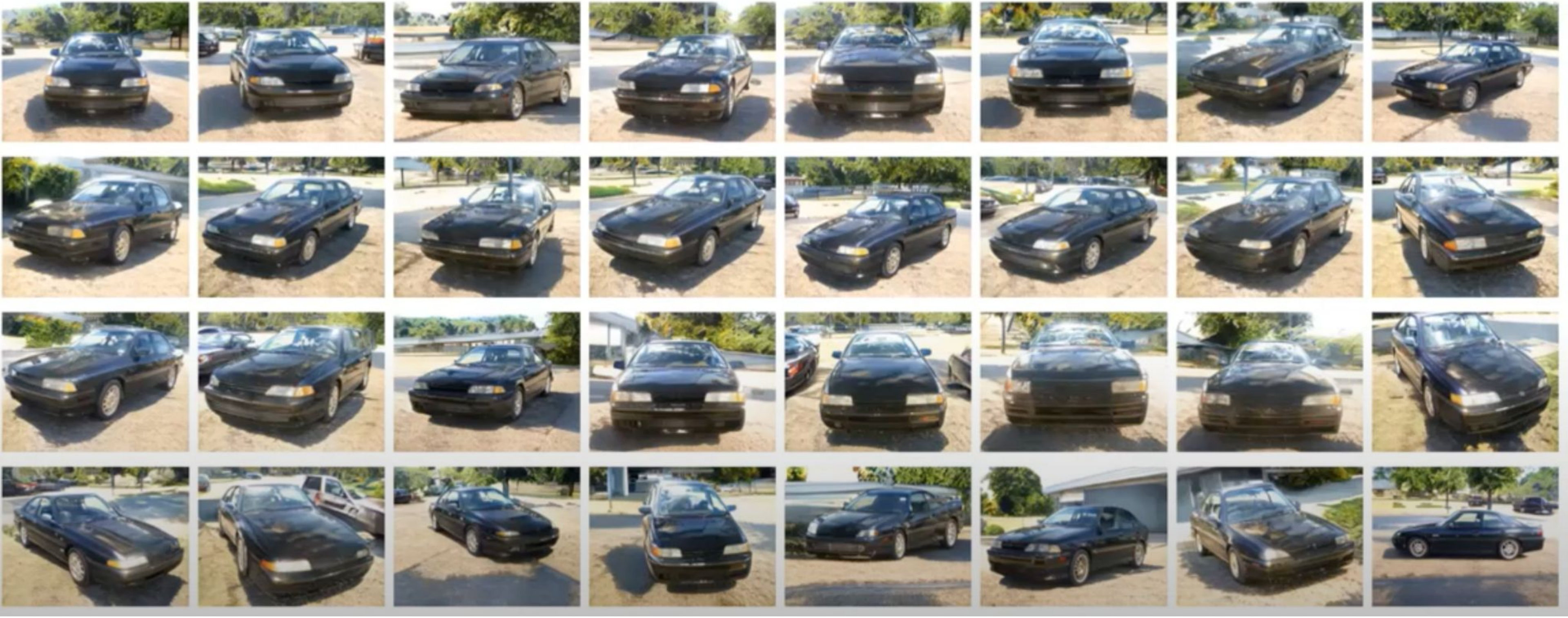
Stylegan again!



"API"

Learned pixel-wise representations

Inspecting StyleGAN's "API"



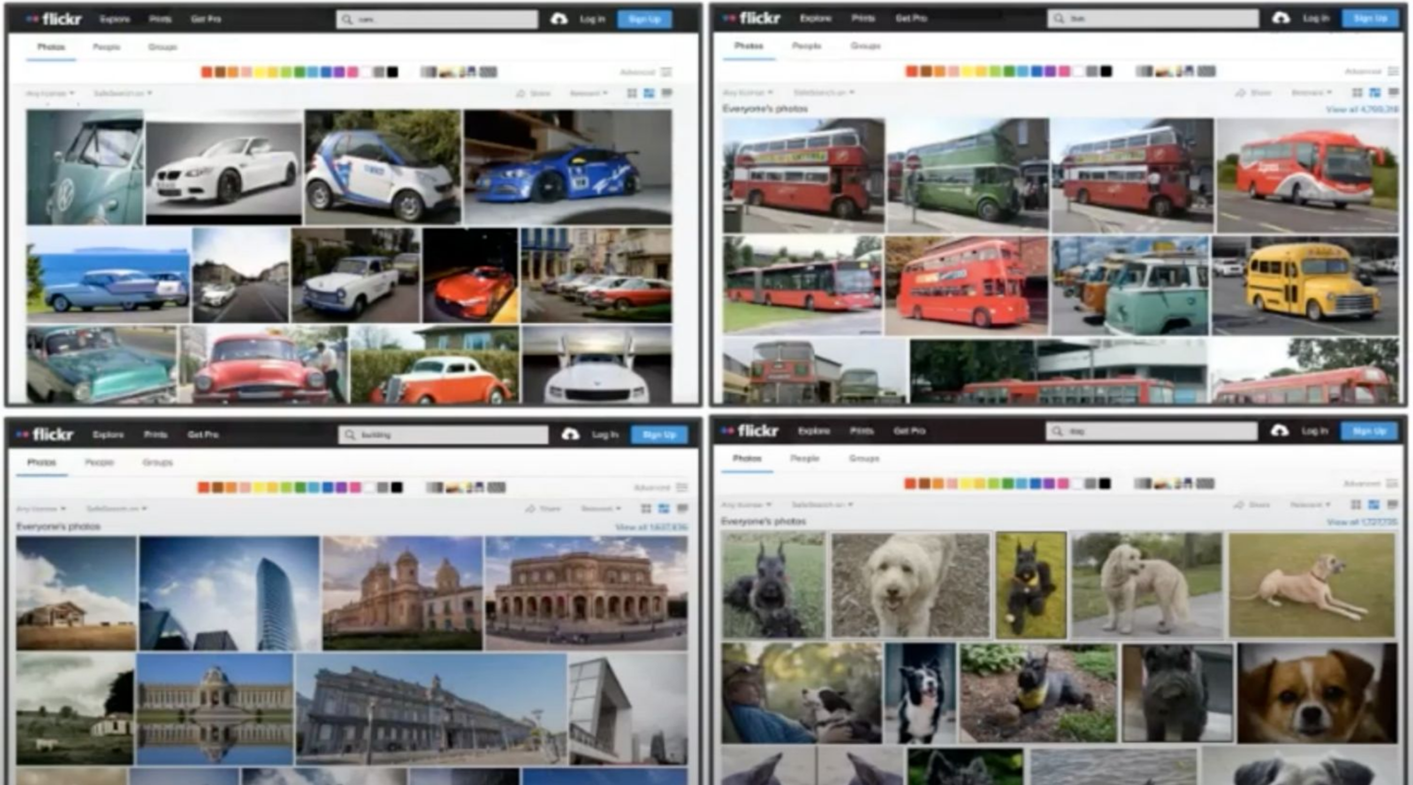
We can very "viewpoint", and keep other content frozen

Inspecting StyleGAN's "API"

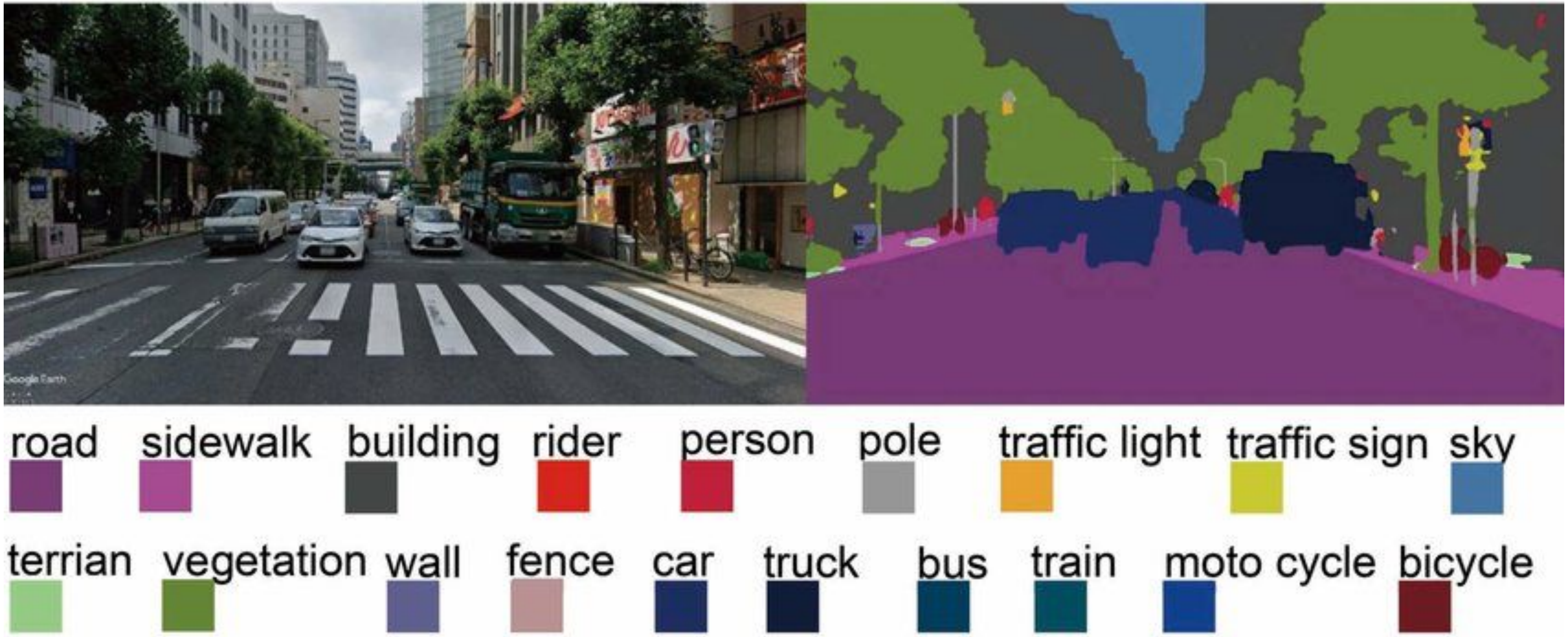
We can keep "viewpoint", and vary other content

Lots of available data: querying Flickr

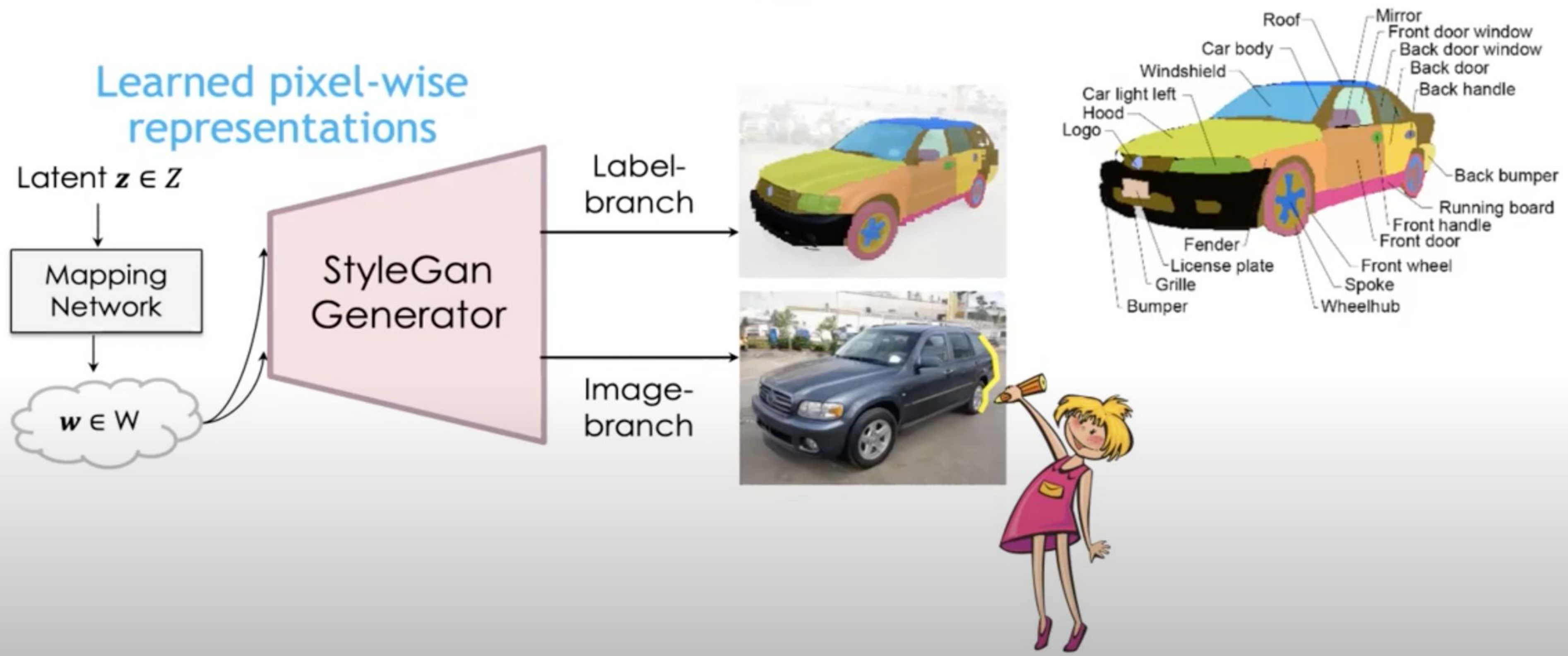
- Car: 1,422,984
- Van: 2,674,607
- Bus: 4,799,402
- Building: 1,638,425
- Bicycle: 3,071,418
- Tricycle: 111,341
- Traffic sign: 150,647
- Dog: 1,731,929
- Human: 3,276,718
- Pedestrian: 659,542
- Person: 507,506
- Skater: 2,454,728
- Skateboard: 1,073,733



Semantic segmentation

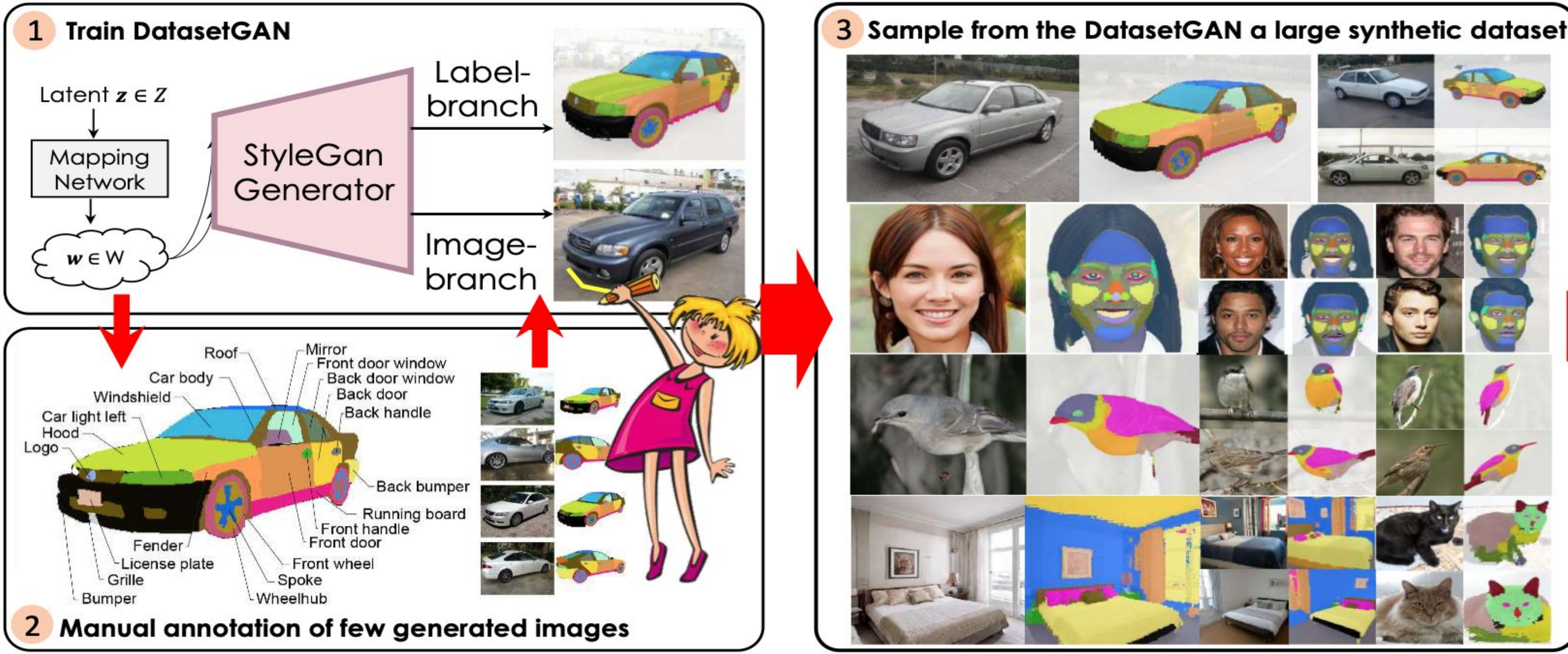


DatasetGAN!



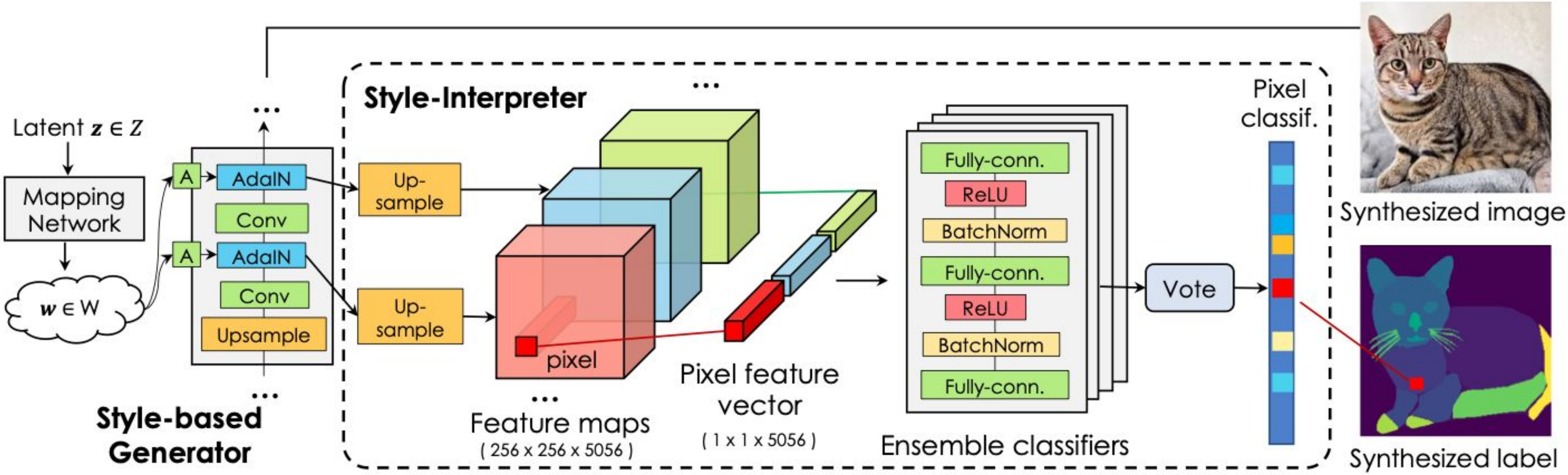
Learned pixel-wise representations
Latent $z \in Z$
Mapping Network
$w \in W$
StyleGan Generator
Label-branch
Image-branch
Roof
Mirror
Front door window
Back door window
Back door
Back handle
Car body
Windshield
Car light left
Hood
Logo
Back bumper
Running board
Front handle
Front door
Fender
License plate
Grille
Bumper
Front wheel
Spoke
Wheelhub

DatasetGAN!



1 Train DatasetGAN
Latent $z \in Z$
Mapping Network
$w \in W$
StyleGan Generator
Label-branch
Image-branch
2 Manual annotation of few generated images
Roof
Mirror
Front door window
Back door window
Back door
Back handle
Car body
Windshield
Car light left
Hood
Logo
Back bumper
Running board
Front handle
Front door
Front wheel
Spoke
Wheelhub
Fender
License plate
Grille
Bumper
3 Sample from the DatasetGAN a large synthetic dataset

Architecture



Synthesized dataset

16
annotated

Synthesized dataset

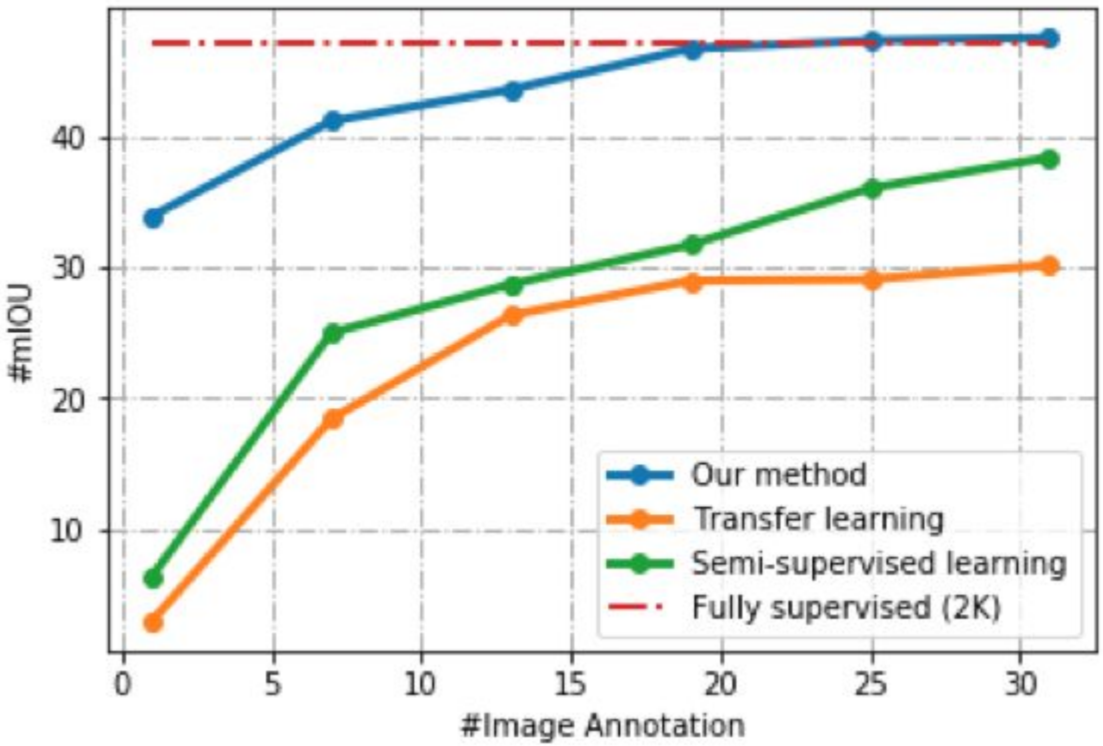
16
annotated

Synthesized dataset

What can we do with a single manually labeled example!

Synthesized dataset

Quality



40
30
20
10
#mIOU
0
5
10
15
20
25
30
#Image Annotation
Our method
Transfer learning
Semi-supervised learning
Fully supervised (2K)

Conclusion

- Annotate your data!
- Image GANs learn geometric and semantically disentangled features
- We can utilize these features for few-shot learning of pixel-wise and 3D tasks

References

- StyleGAN paper https://arxiv.org/abs/1812.04948
- StyleGAN2 paper http://arxiv.org/abs/1912.04958
- DatasetGAN paper https://arxiv.org/abs/2104.06490
- Nvidia, Sanja FIdler - Image GANs for Reducing Pixel-Wise Supervision https://www.youtube.com/watch?v=fWPegVRPb7U