JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Network-free, unsupervised semantic segmentation with synthetic images

Qianli Feng, Raghudeep Gadde, Wentong Liao, Eduard Ramon, Aleix Martinez

Amazon

THU-PM-286

Network-free, unsupervised semantic segmentation with synthetic images

Qianli Feng, Raghudeep Gadde, Wentong Liao, Eduard Ramon, Aleix Martinez

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

1. Introduction

Problem: Segmentation on synthetic images

Contribution: we made A key observation that, in StyleGAN2

Across style mixings, pixels categorized together, change together.

(Long version) The correlation of a set of pixels belonging to the same semantic segment do not change when generating synthetic variants of an image using the style mixing approach

From this observation, we proposed a novel segmentation method that

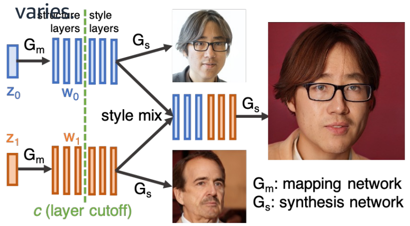
1. Does not need to train a new network
2. Unsupervised
3. Highly accurate

So why do we care? It is because:

1. This method don't need to be re-trained for every new generator deployed.
2. Fewer manual annotations, lower costs.
3. Can be used for downstream supervised training.

2. Background - Style mixing

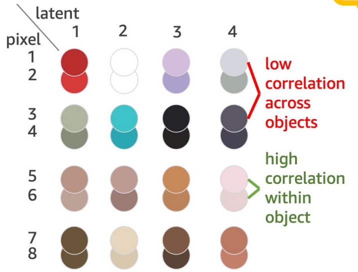
StyleGAN2 uses "style mixing" to create style variants of a synthetic image, where the shape is preserved, but colors/styles varies.



Although it is developed for StyleGAN2, It can work on real images as long as style variants can be produced, e.g., with ControlNet.

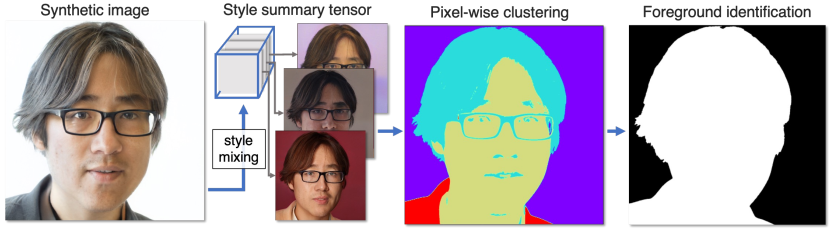
3. Key observation

(a) Style mixing



(b) Style summary tensor

4. Method Design

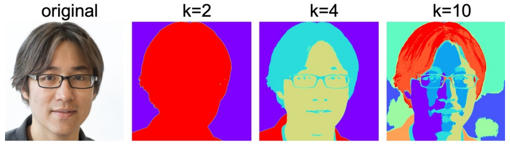


For a given input image, our algorithm has the following steps

1. Generate N style variants
2. Concatenate style variants, get NxCxHxW tensor
3. Cluster HxW pixels into k clusters using flattened N-C dimensions
4. Perform foreground identification for fg/bg task

Our method can be extended to object/instance segmentation with the help of a detector. Simply do the detection first, then run through steps 1-4.

For a foreground / background task, k = 2 works the best.



5. Results - Quantitative

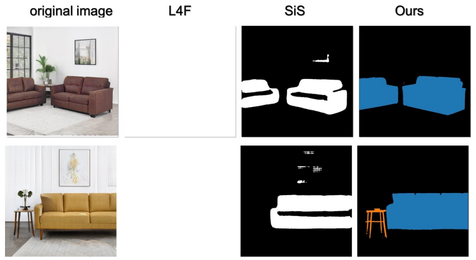
Methods	# manual gt	IOU fg	IOU bg	mIOU fg/bg	Trimap IOU fg	Trimap IOU bg	Trimap mIOU fg/bg
U-net [23]	1000	0.95	0.87	**0.91**	0.53	0.45	**0.49**
w/ DatasetGAN [28]	16	0.90	0.79	0.84	0.43	0.39	0.41
w/ L4F [1]	0	0.92	0.82	0.87	0.43	0.38	0.41
w/ SiS [21]	0	0.92	0.80	0.86	0.45	0.33	0.39
w/ Ours	0	0.92	0.82	**0.87**	0.42	0.43	**0.42**

Table 3. Using synthetic data as training data for image segmentation. Trained on images generated from FFHQ model, test on CelebA-Mask-HQ (real data). The supervised segmentation method is DeepLabV3. All synthetic data performances are trained from scratch using synthetic data only. Trimap width is 3 pixels.

	LSUN-Horse		DeepRoom-livingroom			
Methods	IOU (horse-fg/bg)	mIOU	IOU (sofa-fg/bg)	mIOU	IOU (table-fg/bg)	mIOU
L4F [1]	0.51/0.73	0.62	×	×	×	×
SiS [21]	0.44/0.78	0.61	×	×	×	×
Ours	0.64/0.89	**0.77**	**0.88/0.97**	**0.93**	**0.14/0.96**	**0.55**

Table 2. Semantic segmentation performance on LSUN-horses, and DeepRoom-livingroom datasets, all with synthetic images and DeepLabV3 as psuedo ground-truth.×: method not easily extendable to segment the target class.

6. Results – Qualitative



7. Conclusion

We proposed a novel segmentation method on synthetic images that 1. does not need to train a new network, 2. Unsupervised 3. Highly accurate.

Problem

Problem

- Semantic segmentation on synthetic images

Problem

- Semantic segmentation on synthetic images
  - StyleGAN
  - Extend to more

Problem

- Semantic segmentation on synthetic images
  - StyleGAN
  - Extend to more
- Not a new problem
  - DatasetGAN, Label-4-Free, Segment-in-Style, furryGAN

Our contribution

Previous methods

- Use a masking branch
- Take generator intermediate activation as input

This is problematic because

- Every new generator -> re-train masking branch

Our contribution

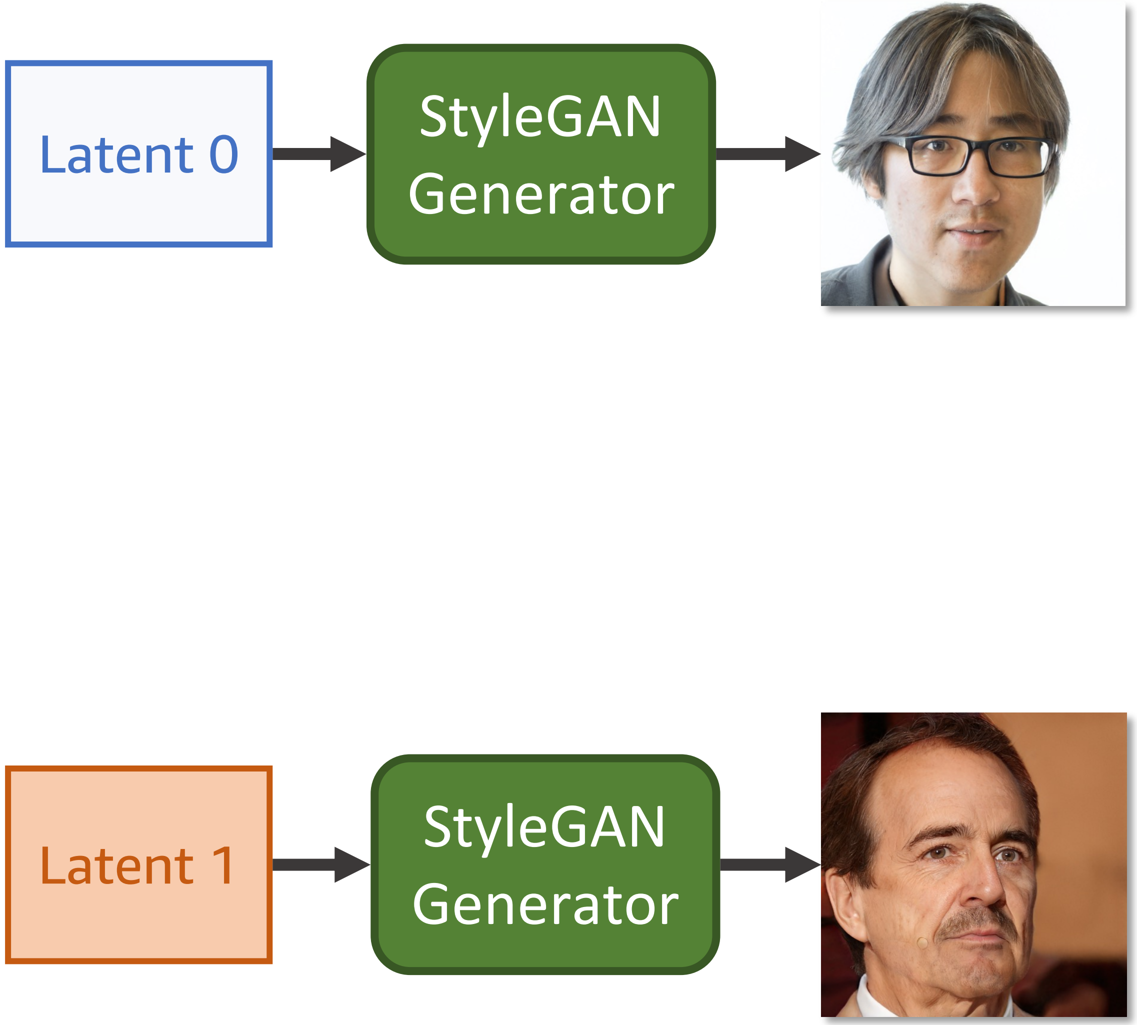
We proposed a simple, novel segmentation method that

1. Does not need to train a new network
2. Unsupervised
3. Highly accurate

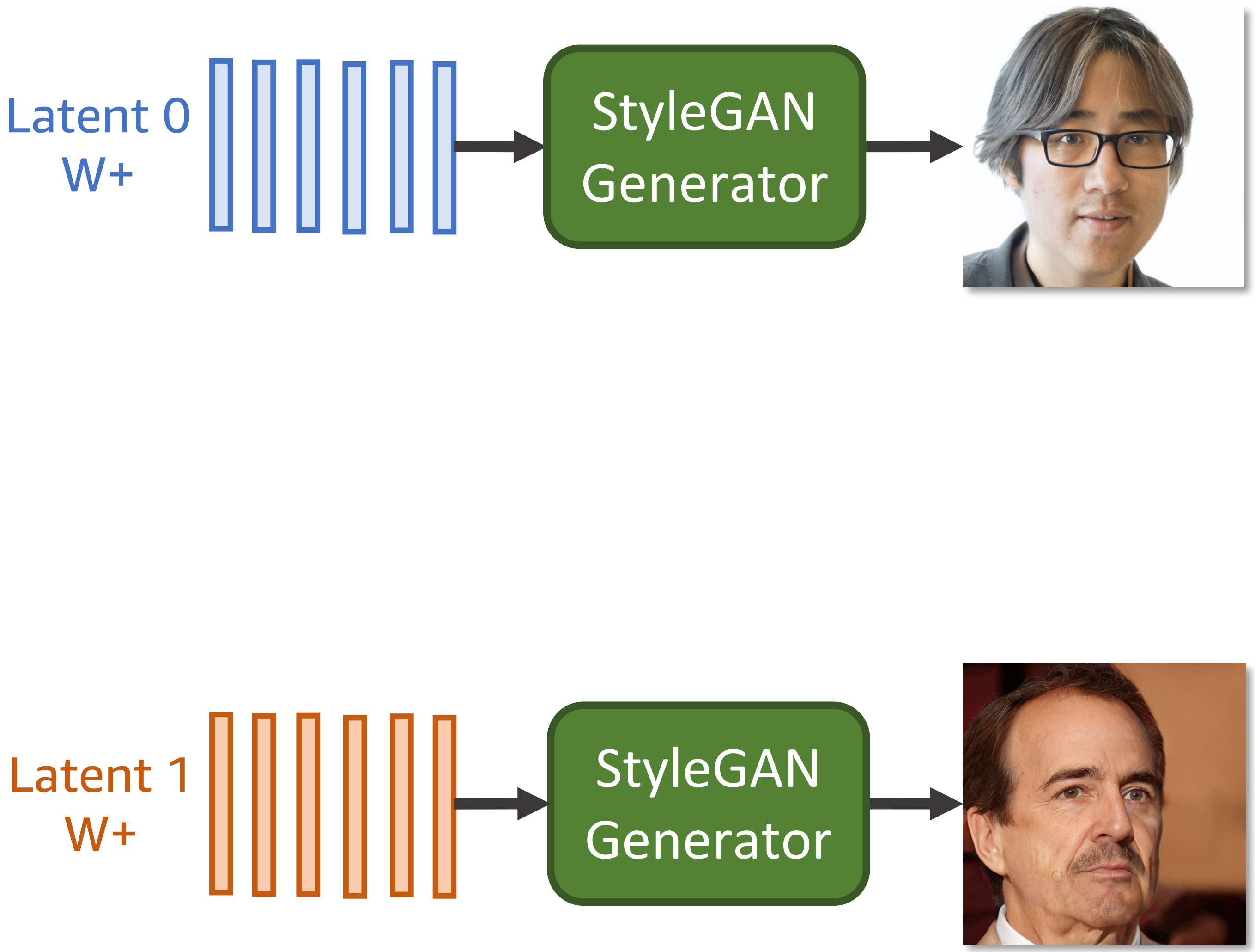
A key observation that, in StyleGAN2

Pixels that belong to the same semantics class, change their color together across different style mixings.

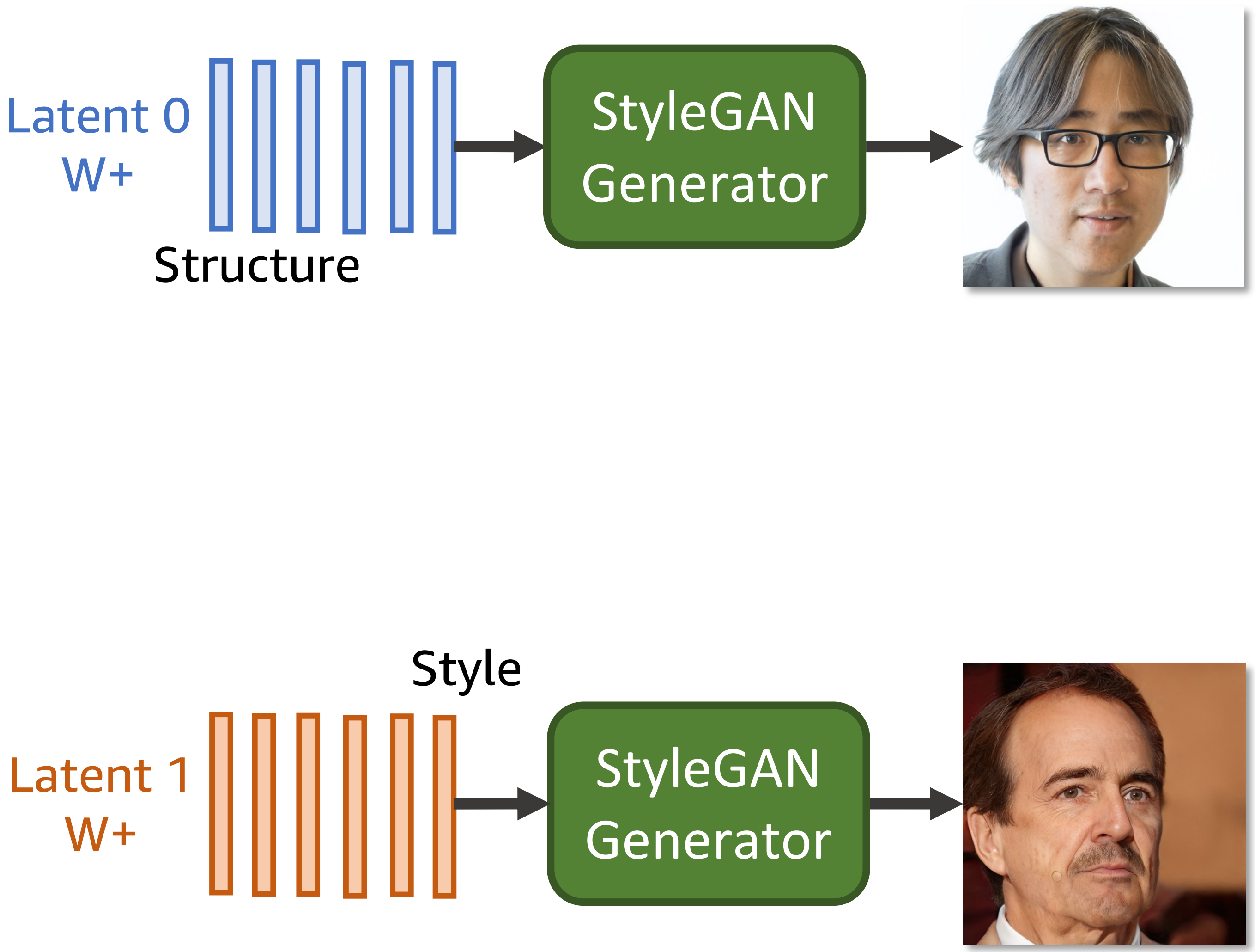
Background – Style Mixing in StyleGAN2



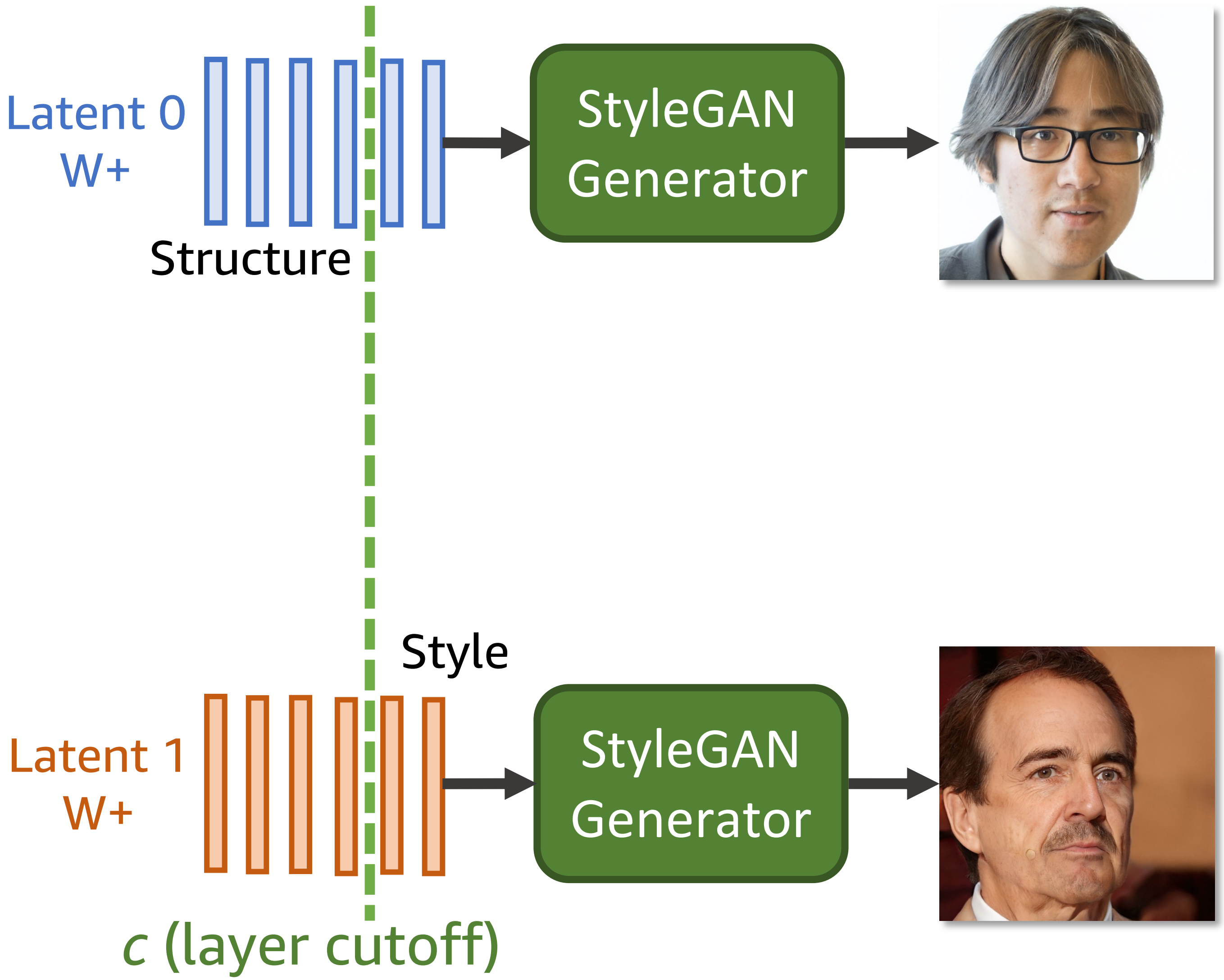
Background – Style Mixing in StyleGAN2



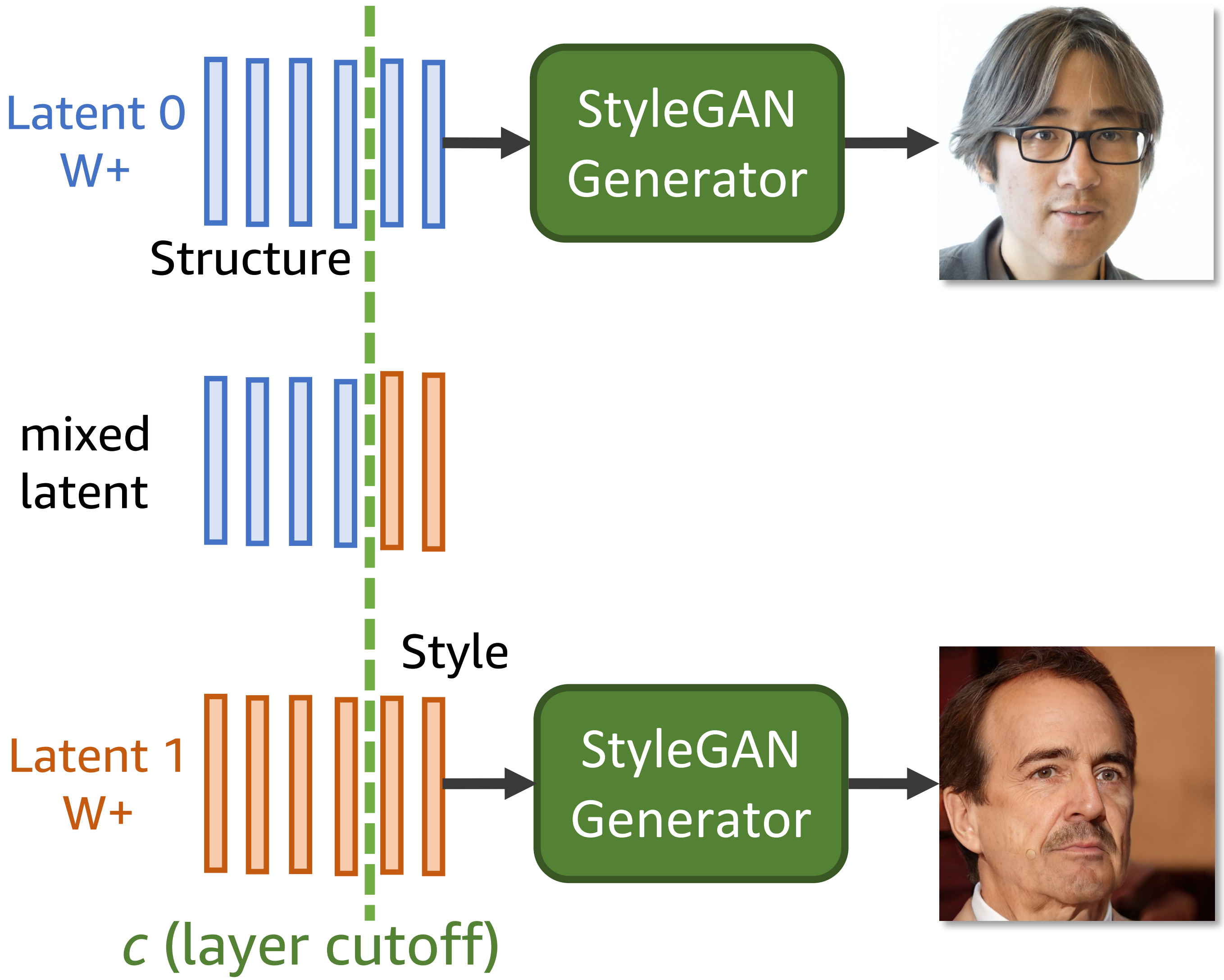
Background – Style Mixing in StyleGAN2



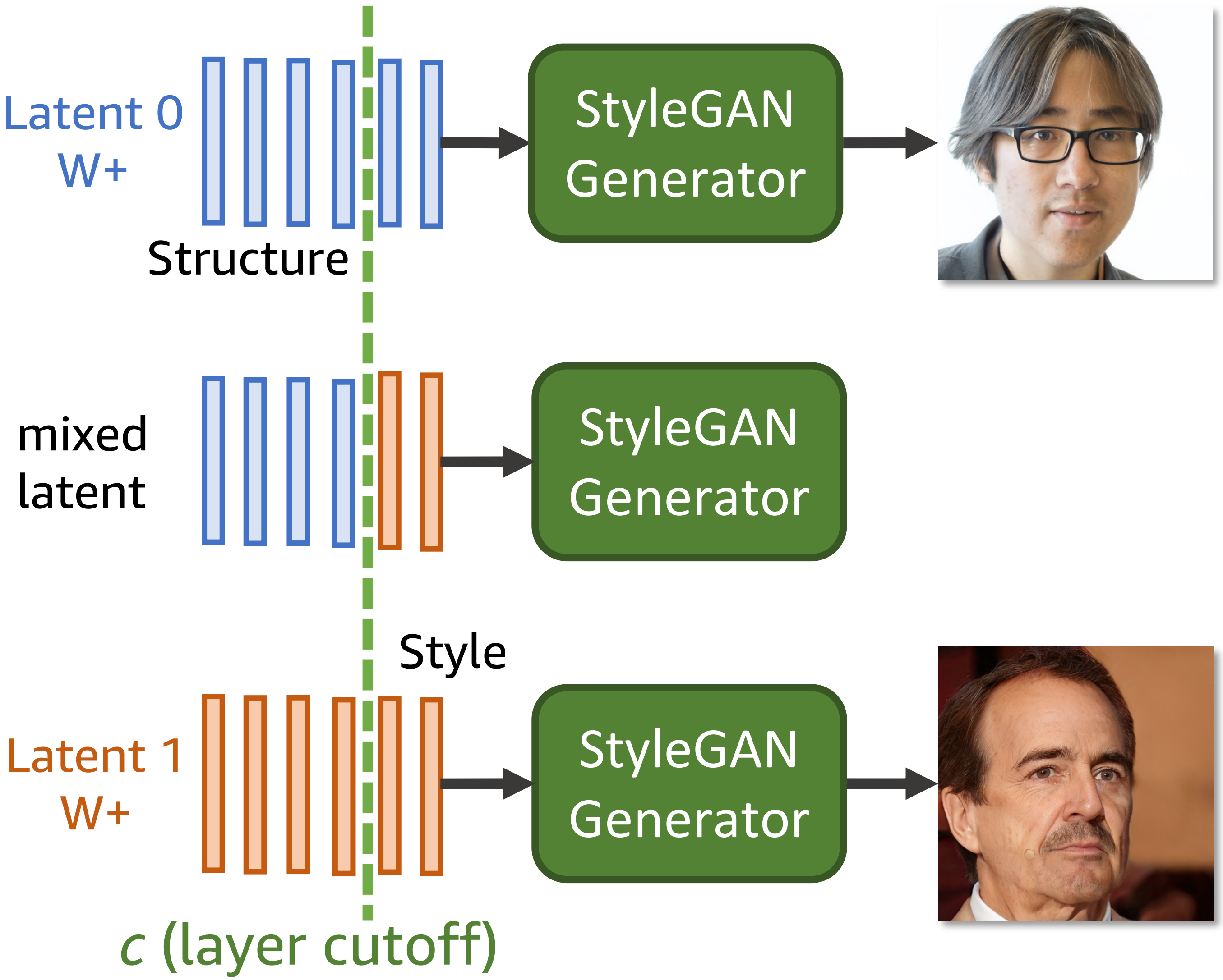
Background – Style Mixing in StyleGAN2



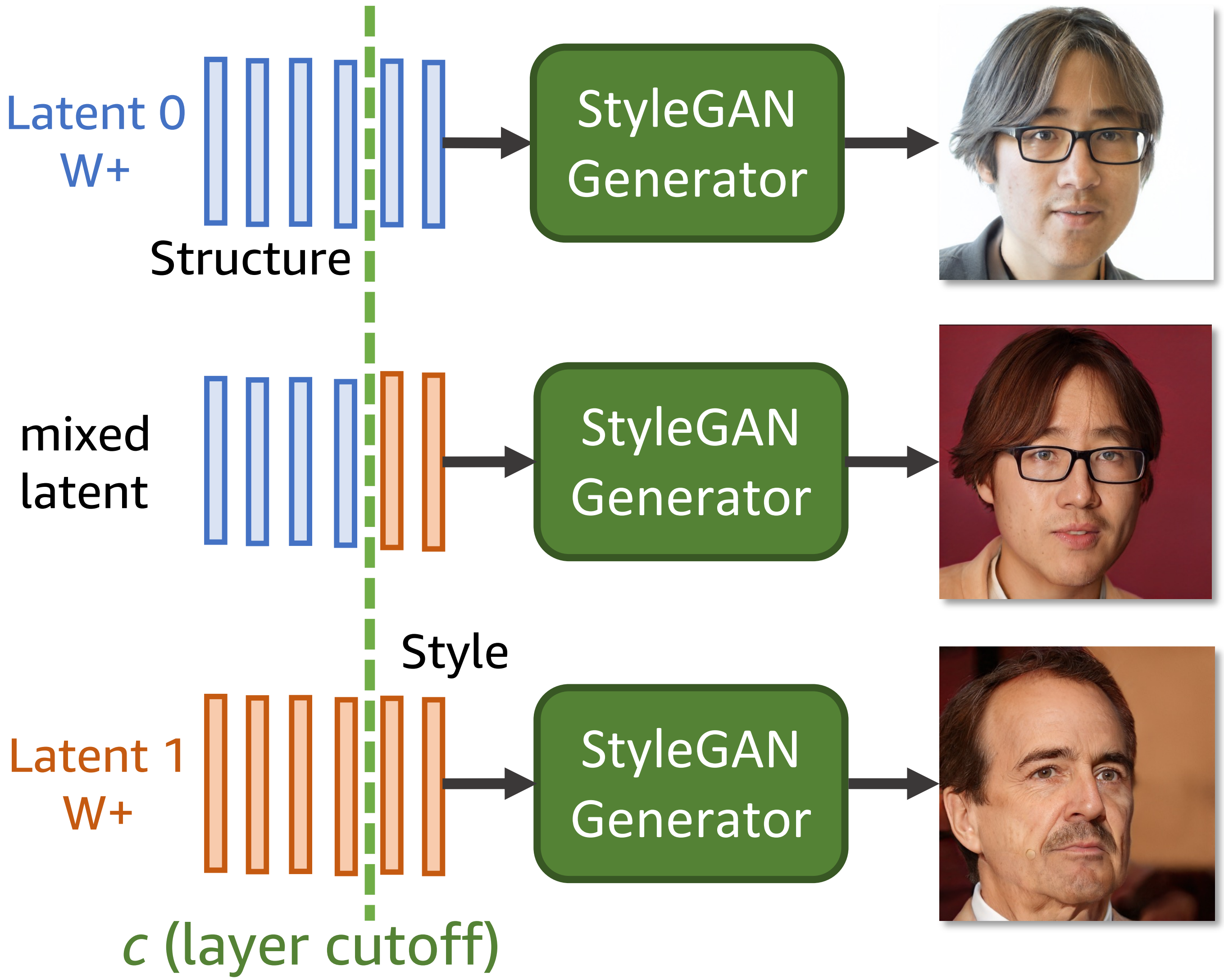
Background – Style Mixing in StyleGAN2



Background – Style Mixing in StyleGAN2



Background – Style Mixing in StyleGAN2



Key observation

latent 1

latent 2

latent 3

latent 4

In StyleGAN2, pixels that belong to the same semantics class, change their color together across different style mixings.

Key observation

latent 1

latent 2

latent 3

latent 4

latent

pixel

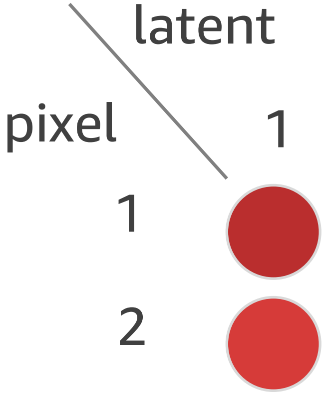
Key observation

latent 1

latent 2

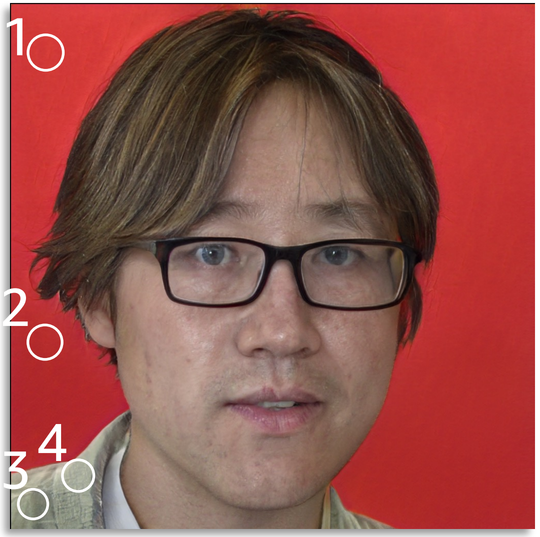
latent 3

latent 4



Key observation

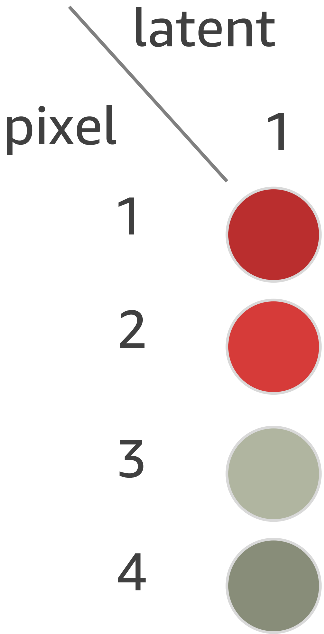
latent 1



latent 2

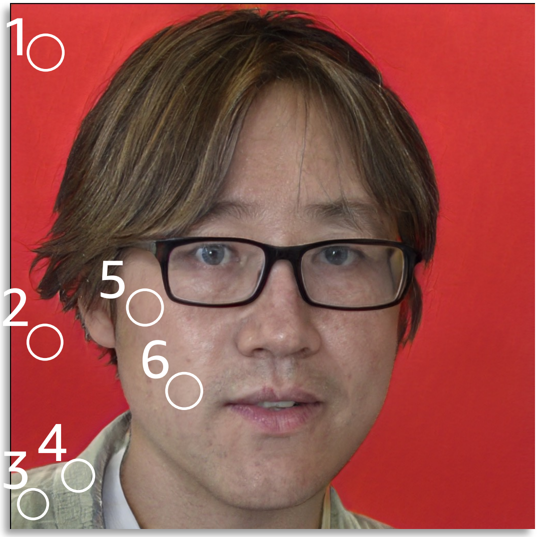
latent 3

latent 4



Key observation

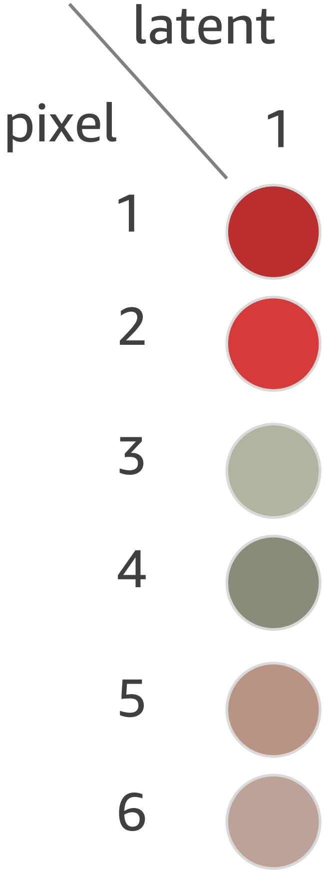
latent 1



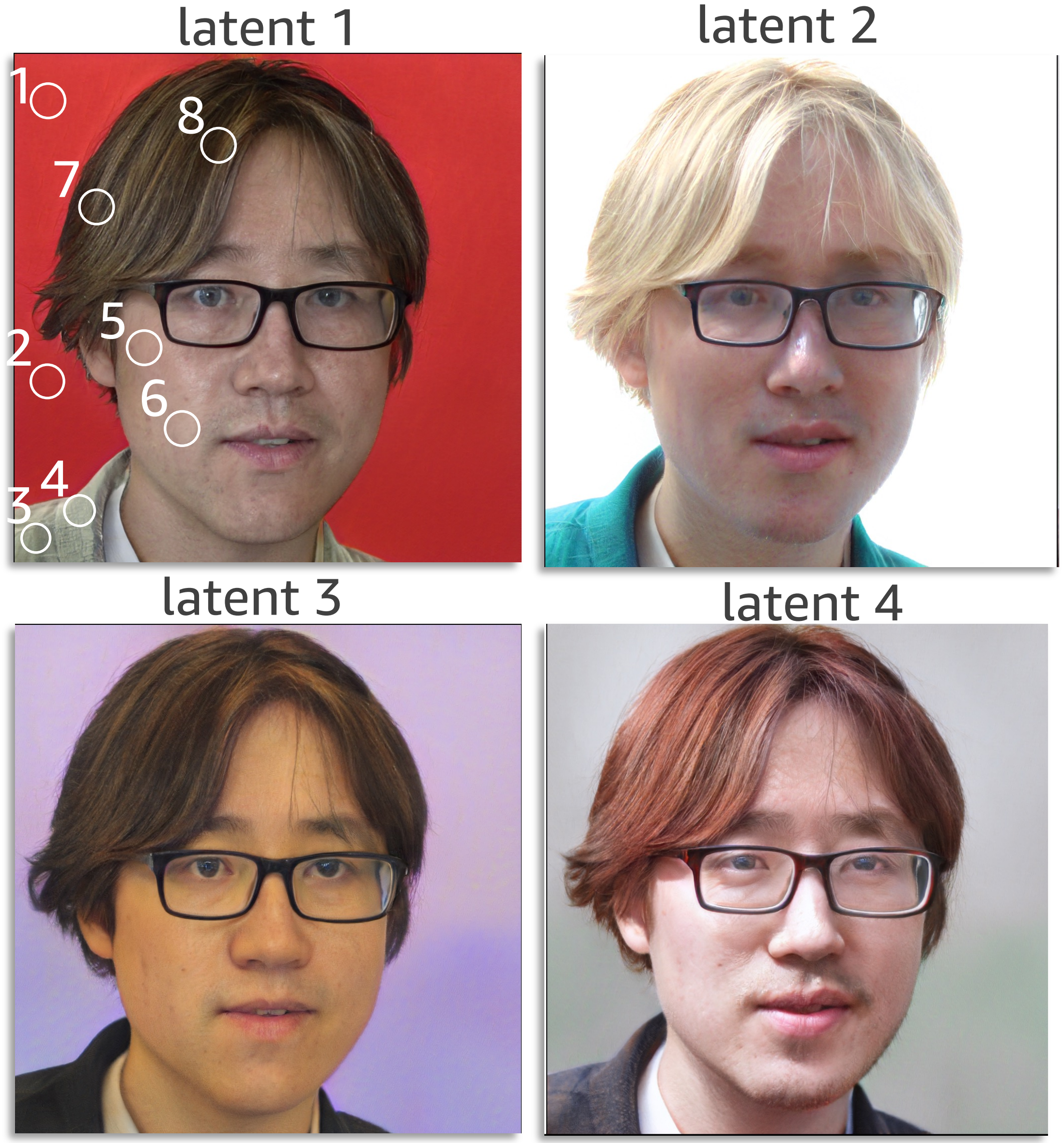
latent 2

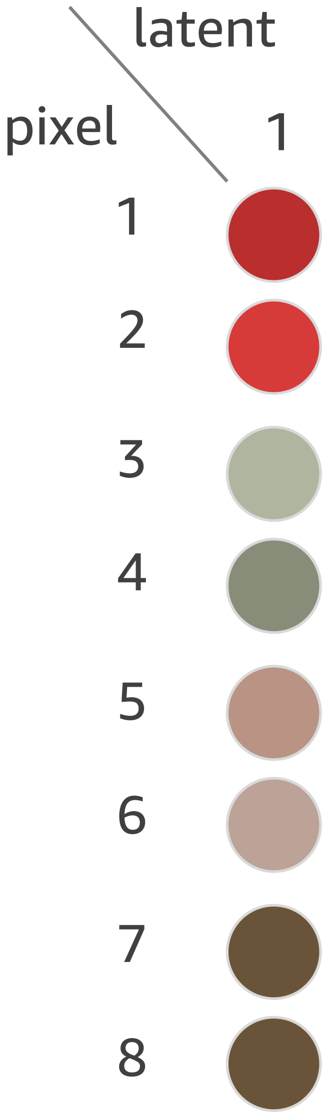
latent 3

latent 4

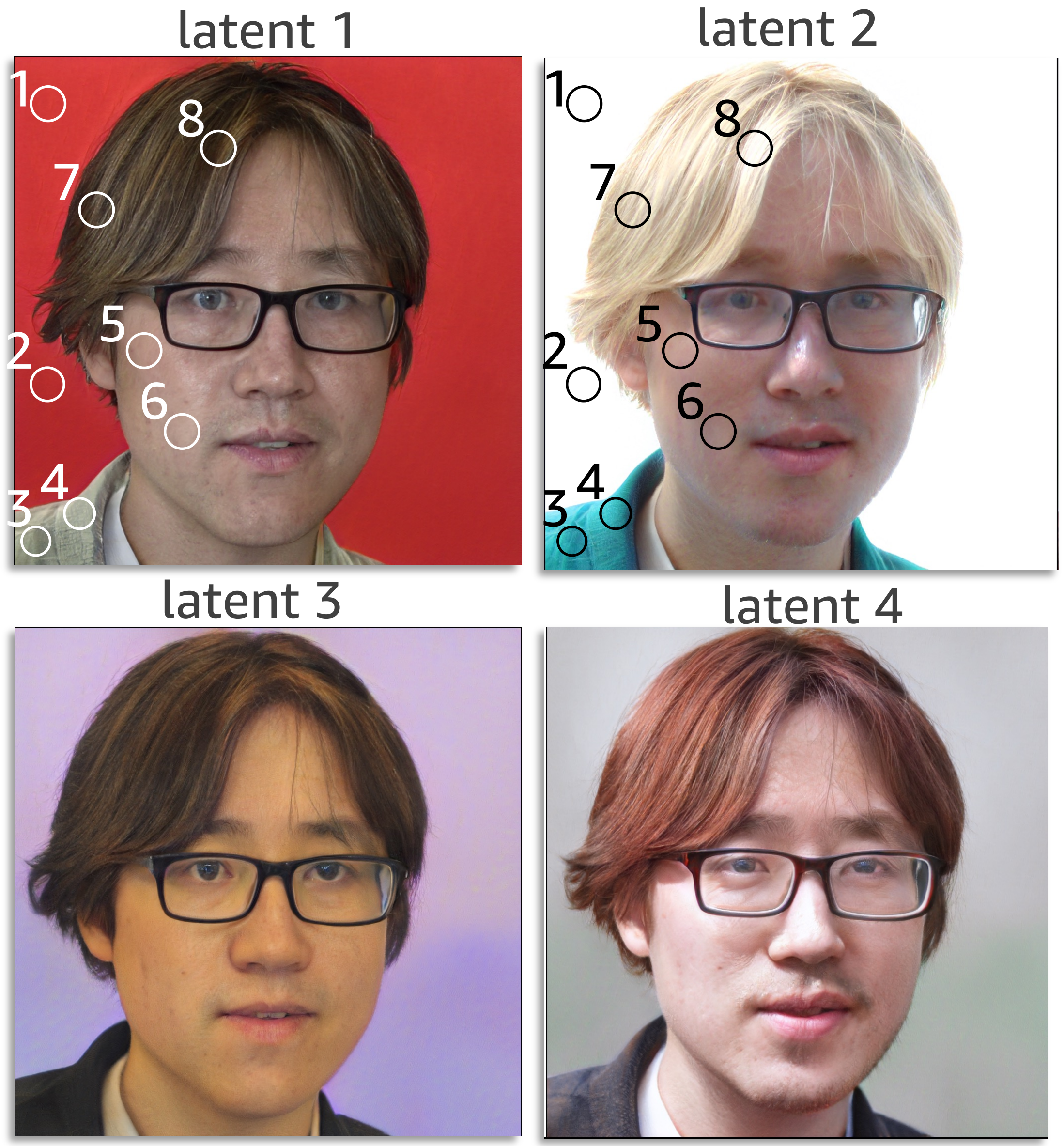


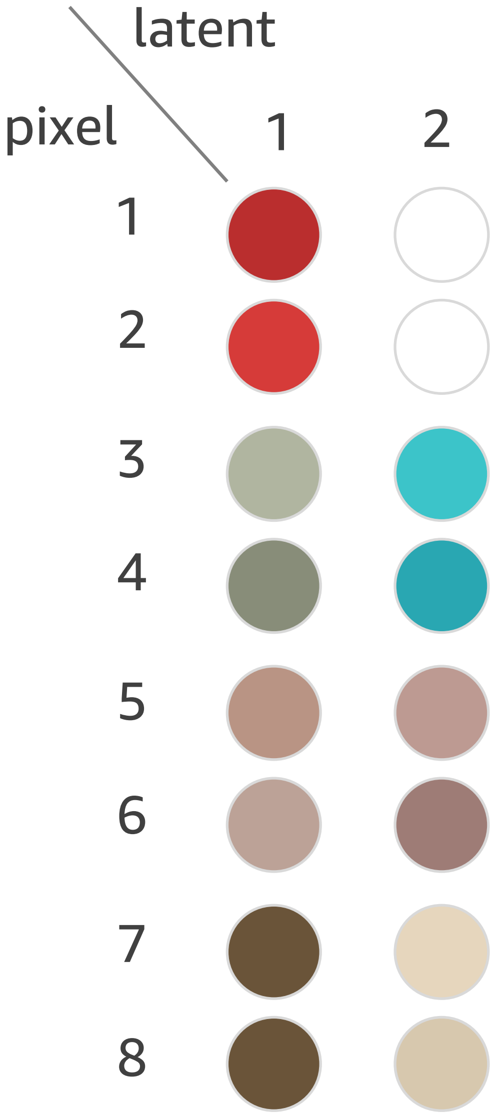
Key observation



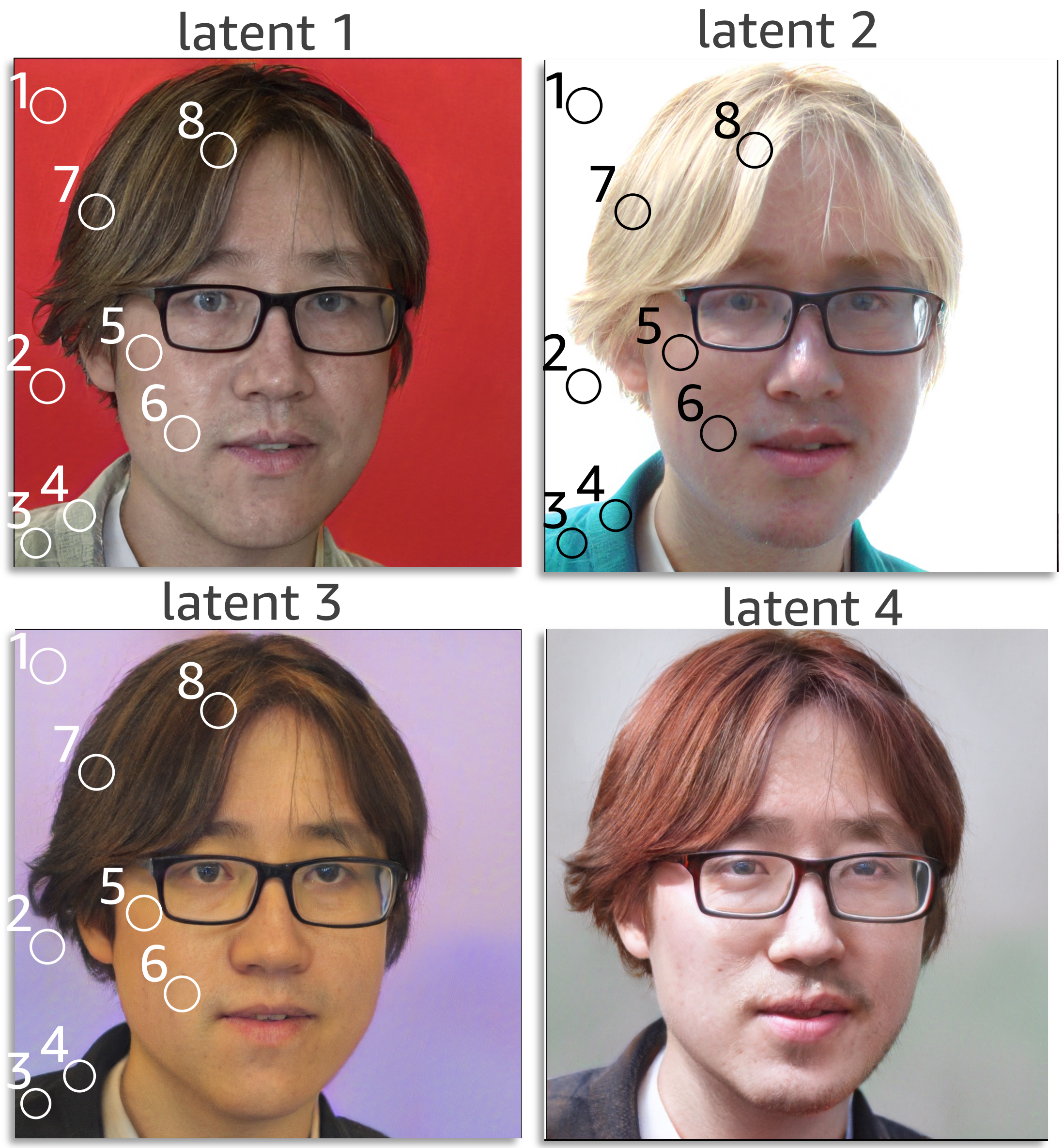


Key observation

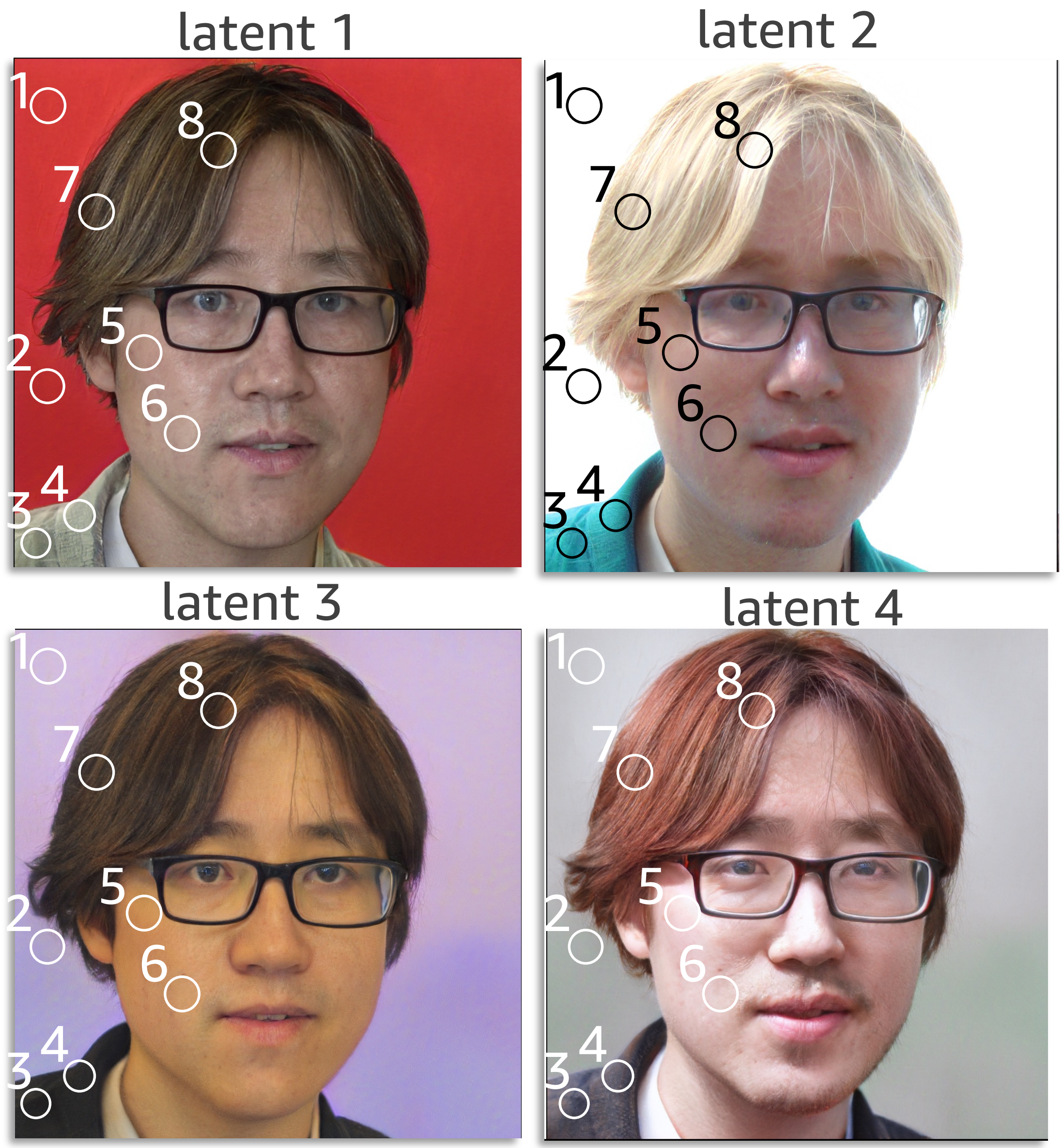


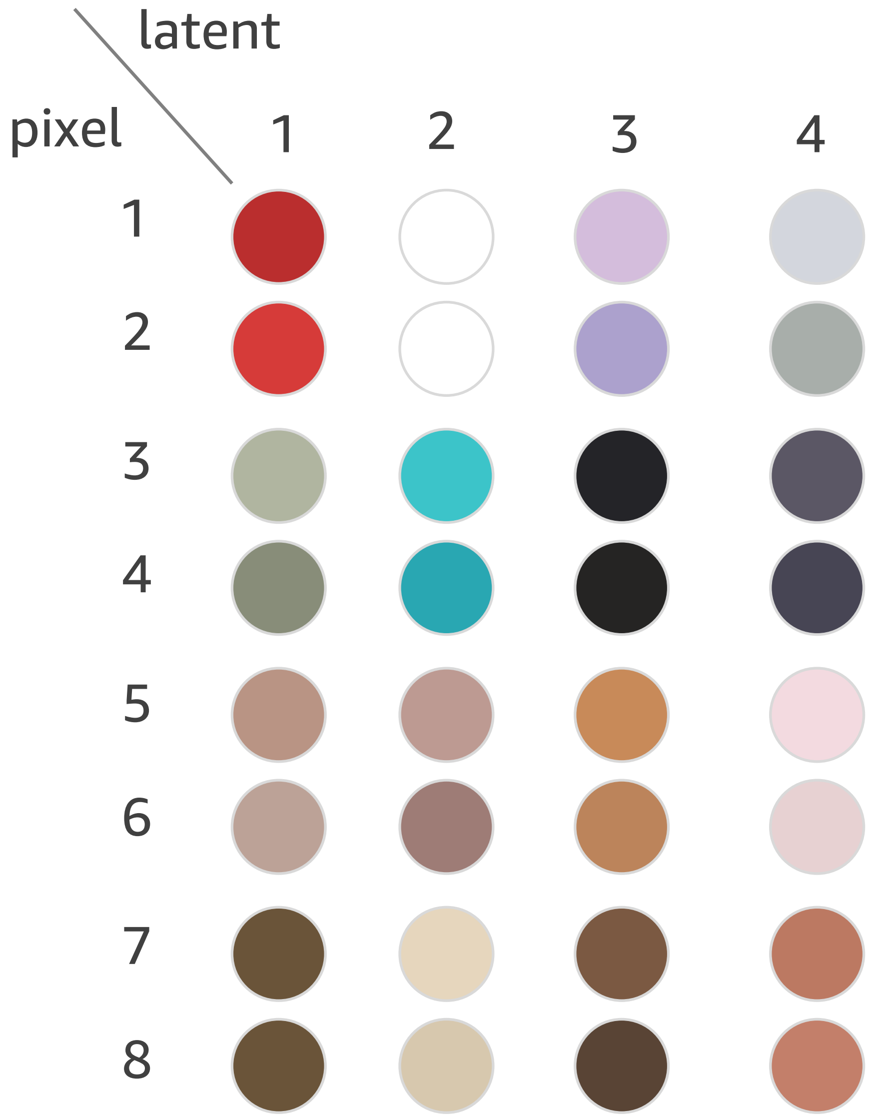


Key observation

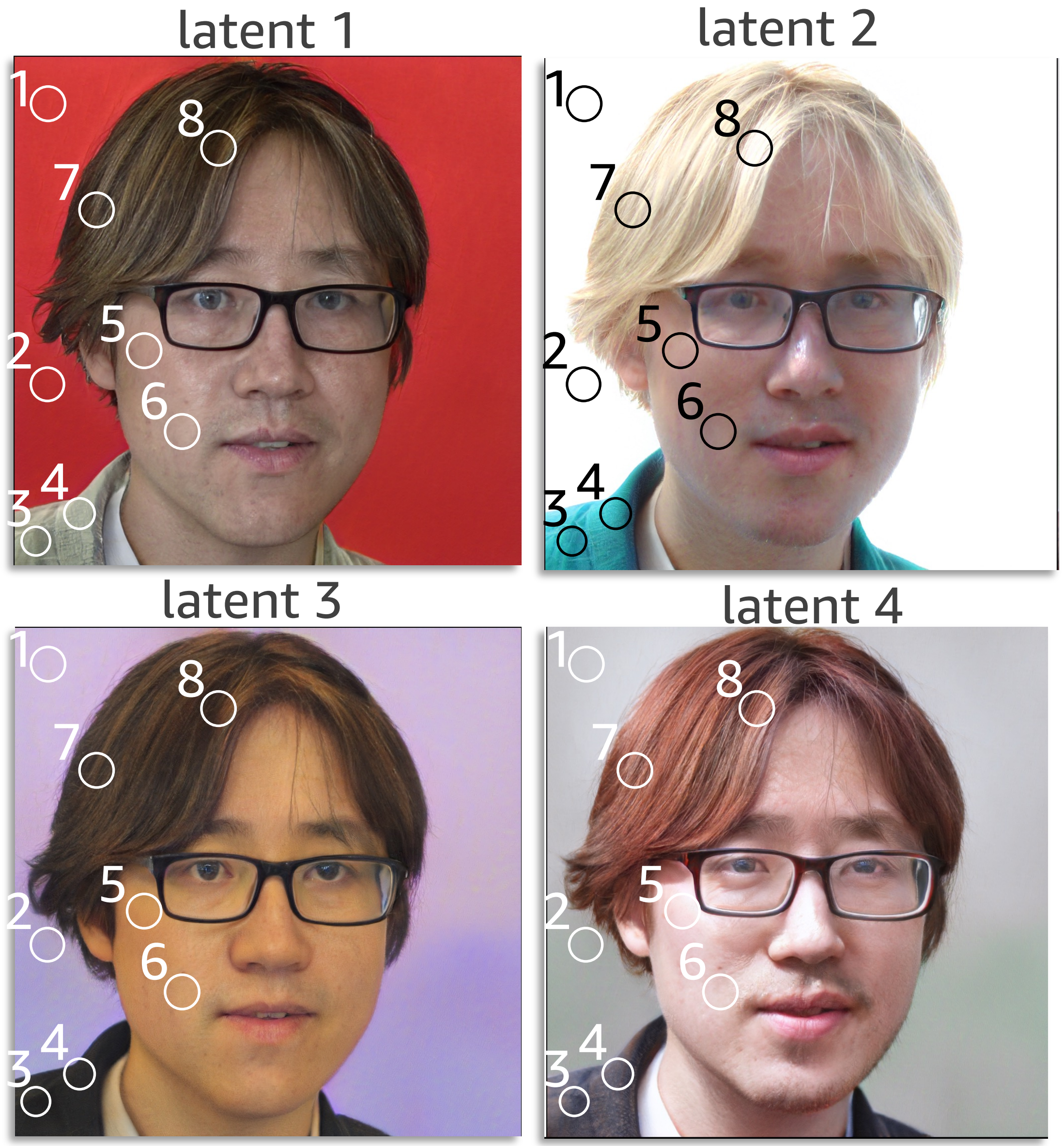


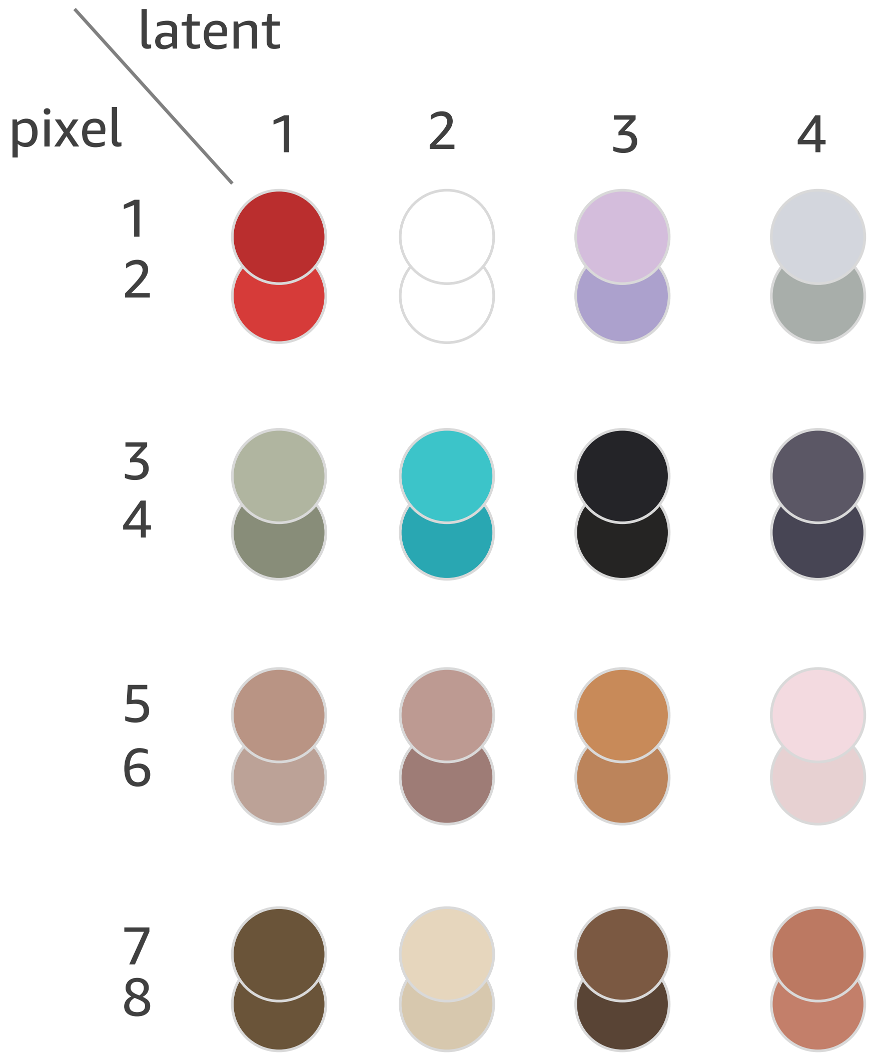
Key observation



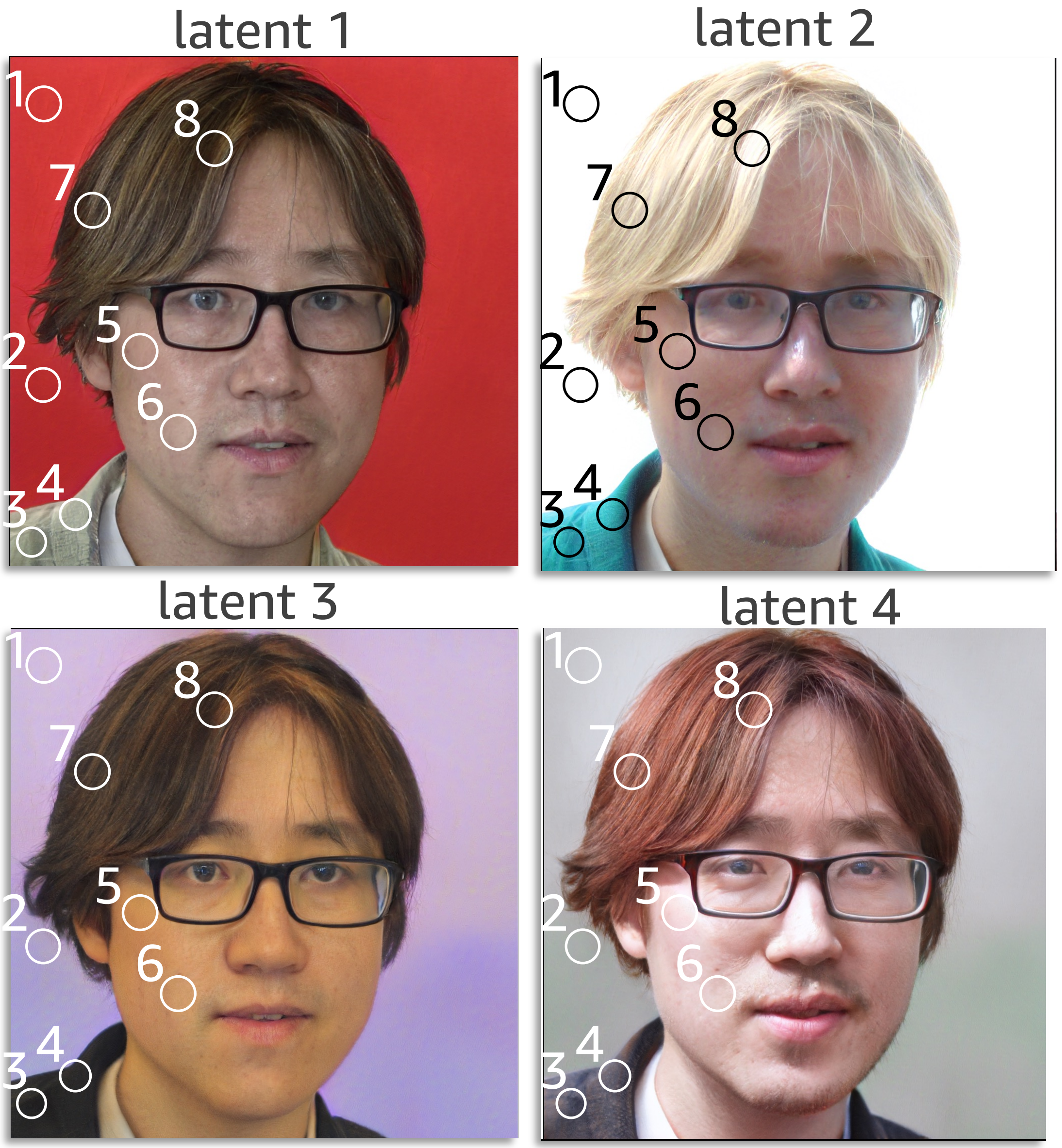


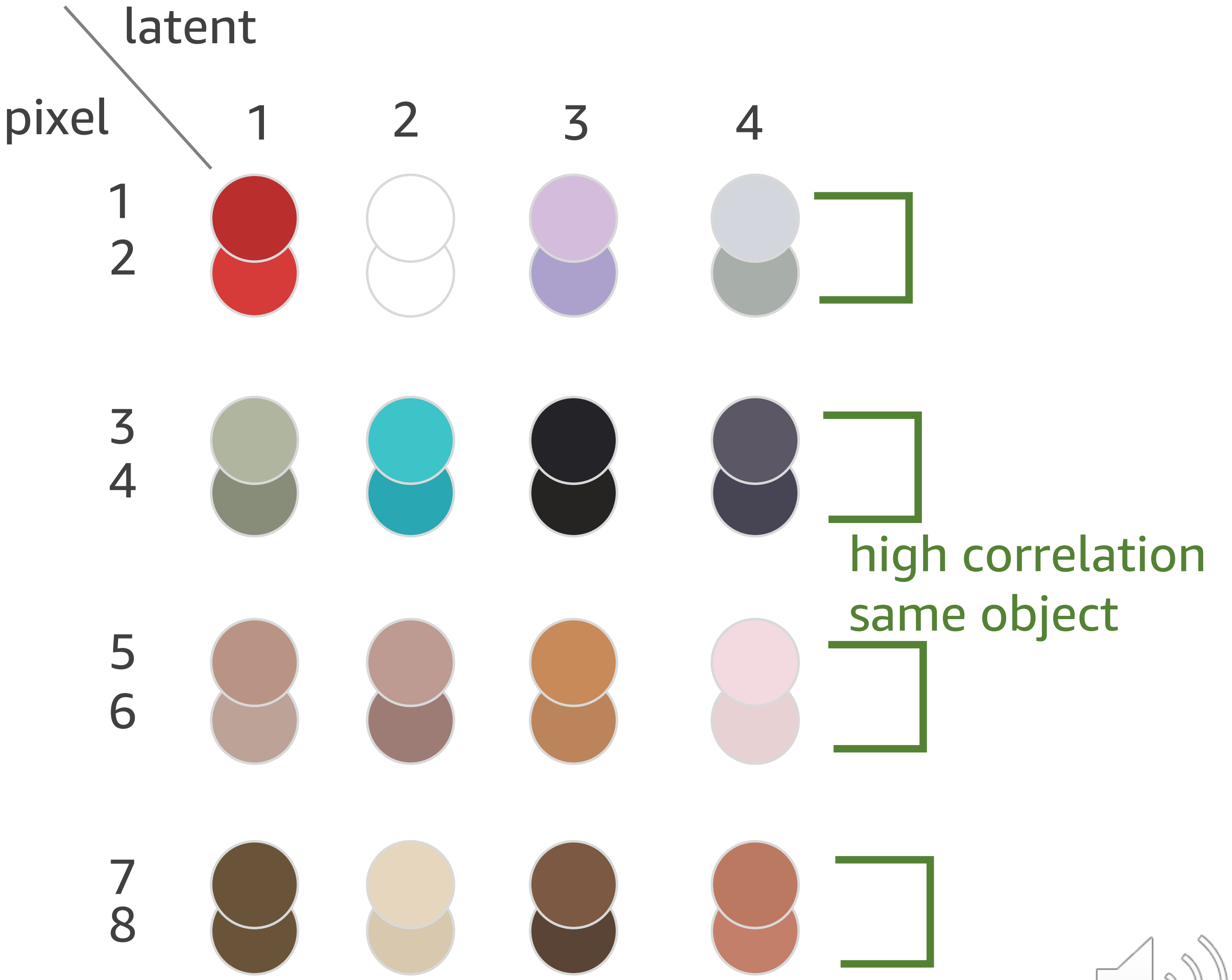
Key observation



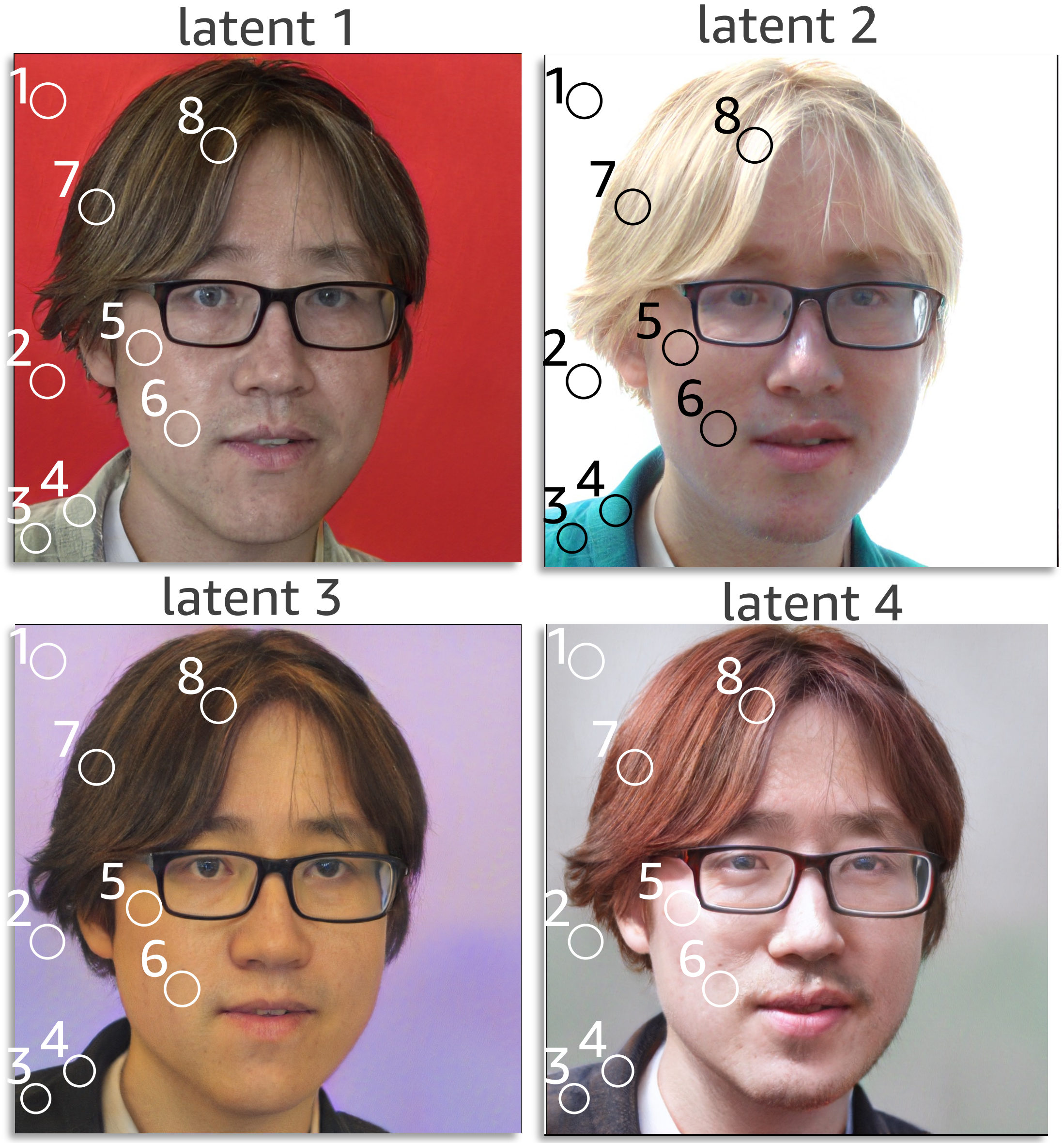


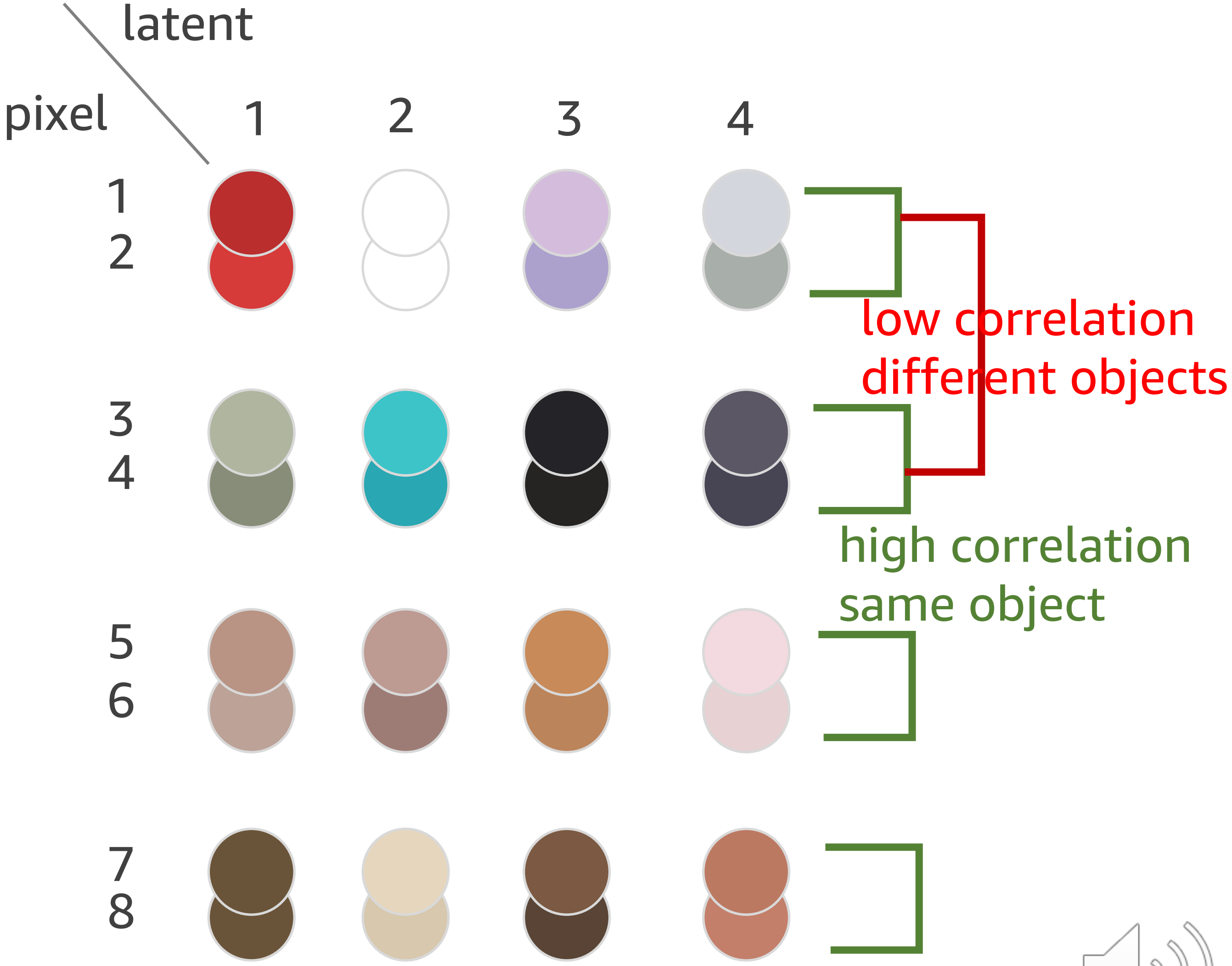
Key observation



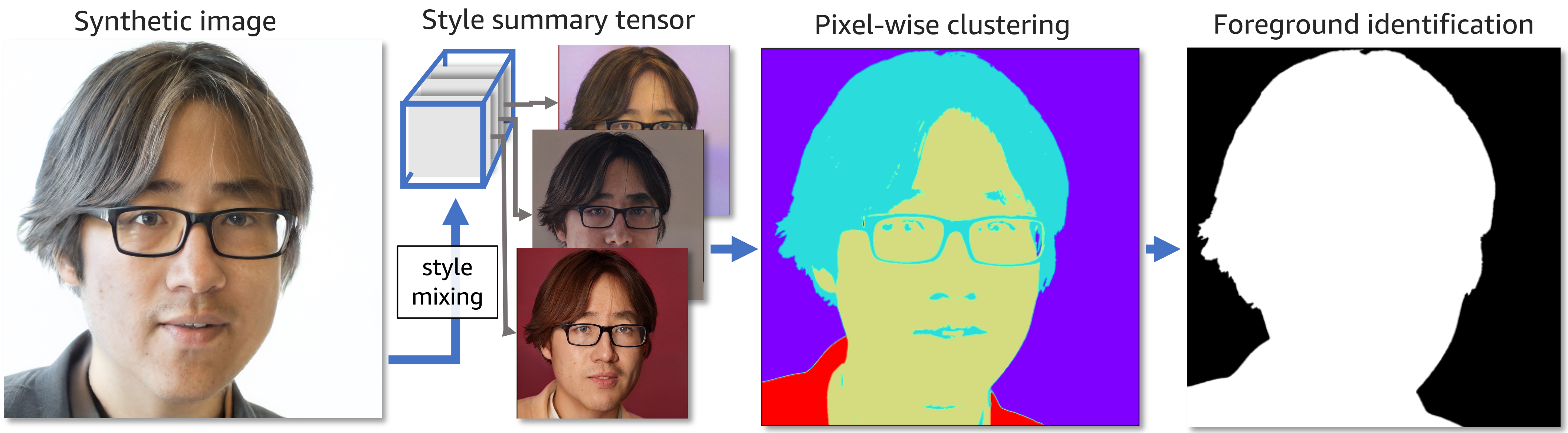


Key observation





Our method



- For a given input image, our algorithm has the following steps
  1. Generate N style variants
  2. Concatenate style variants, get NxCxHxW tensor
  3. Cluster HxW pixels into k clusters using flattened N-C dimensions
  4. Perform foreground identification for fg/bg task
- Our method can be extended to object/instance segmentation with the help of a detector. Simply do the detection first, then run through steps 1-4.

Experiment / Results

- Experiment on
  - Fg/bg segmentation directly on synthetic images

				FFHQ		CelabAHQ-Mask	
Methods	Training data	supervision	additional network	IOU (fg/bg)	mIOU	IOU (fg/bg)	mIOU
DatasetGAN [28]	16	✓	✓	0.83/0.73	0.78	0.87/0.73	0.80
L4F [1]	10k	×	✓	0.92/0.85	**0.88**	0.92/0.80	0.86
SiS [21]	50+15k	×	✓	0.89/0.77	0.83	0.92/0.81	**0.87**
Ours	0	×	×	0.87/0.73	0.80	0.91/0.81	0.86

Table 1. Image segmentation performance on FFHQ (*i.e.*, on synthetic data) and CelebA-Mask-HQ (*i.e.*, on real data). IOU (fg/bg) is the IOU for foreground/background segmentation. mIOU is the average between the IOU (fg) and IOU (bg).

Experiment / Results

- Experiment on
  - Fg/bg segmentation directly on synthetic images
  - Instance segmentation on synthetic images

	LSUN-Horse		DeepRoom-livingroom			
Methods	IOU (horse-fg/bg)	mIOU	IOU (sofa-fg/bg)	mIOU	IOU (table-fg/bg)	mIOU
L4F [1]	0.51/0.73	0.62	×	×	×	×
SiS [21]	0.44/0.78	0.61	×	×	×	×
Ours	0.64/0.89	**0.77**	**0.88/0.97**	**0.93**	**0.14/0.96**	**0.55**

Table 2. Semantic segmentation performance on LSUN-horses, and DeepRoom-livingroom datasets, all with synthetic images and DeepLabV3 as psuedo ground-truth. ×: method not easily extendable to segment the target class.

Experiment / Results

- Experiment on
  - Fg/bg segmentation directly on synthetic images
  - Instance segmentation on synthetic images
  - Generate synthetic segmentation data for downstream training

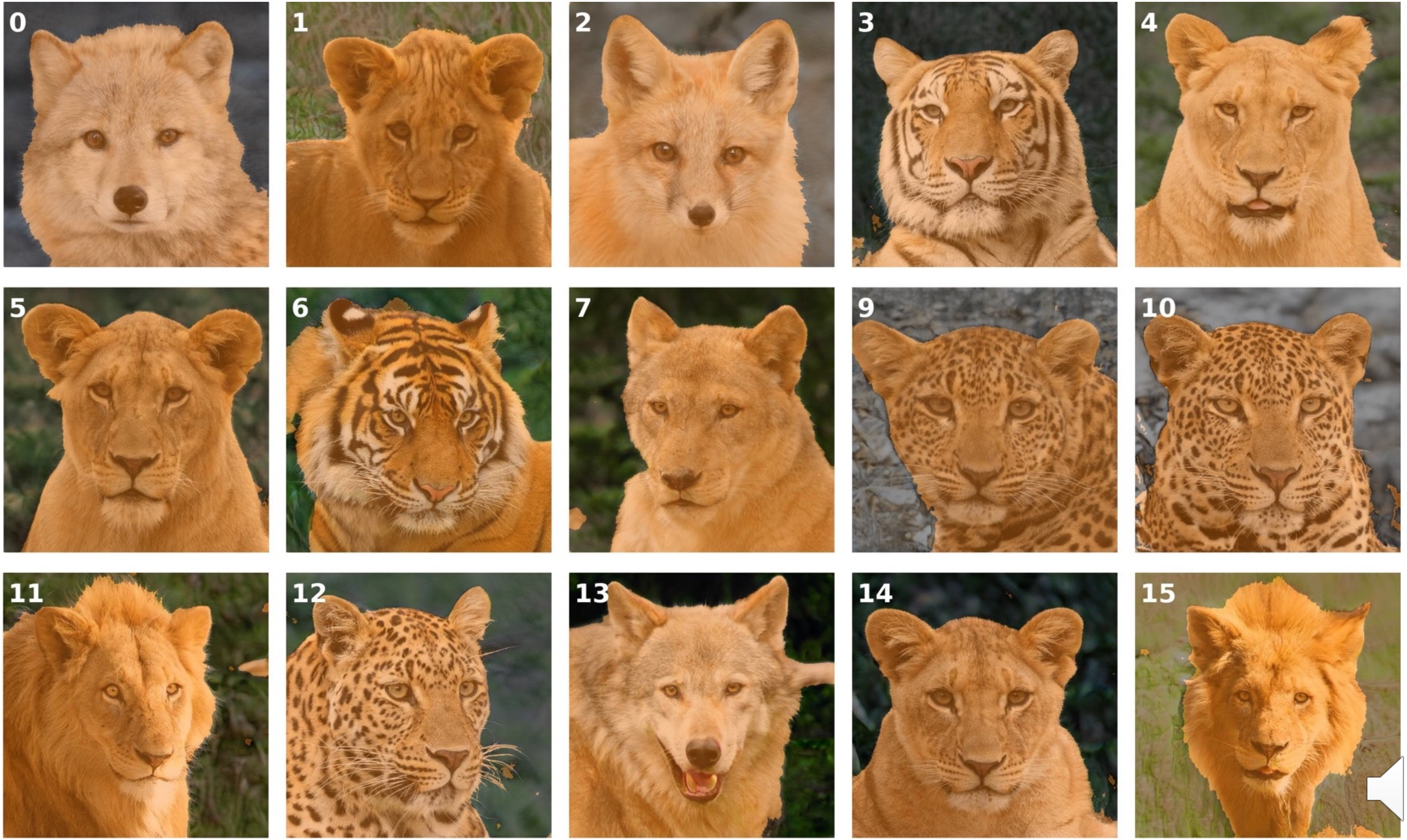
		IOU		mIOU	Trimap IOU		Trimap mIOU
Methods	# manual gt	fg	bg	fg/bg	fg	bg	fg/bg
U-net [23]	1000	0.95	0.87	**0.91**	0.53	0.45	**0.49**
w/ DatasetGAN [28]	16	0.90	0.79	0.84	0.43	0.39	0.41
w/ L4F [1]	0	0.92	0.82	0.87	0.43	0.38	0.41
w/ SiS [21]	0	0.92	0.80	0.86	0.45	0.33	0.39
w/ Ours	0	0.92	0.82	**0.87**	0.42	0.43	**0.42**

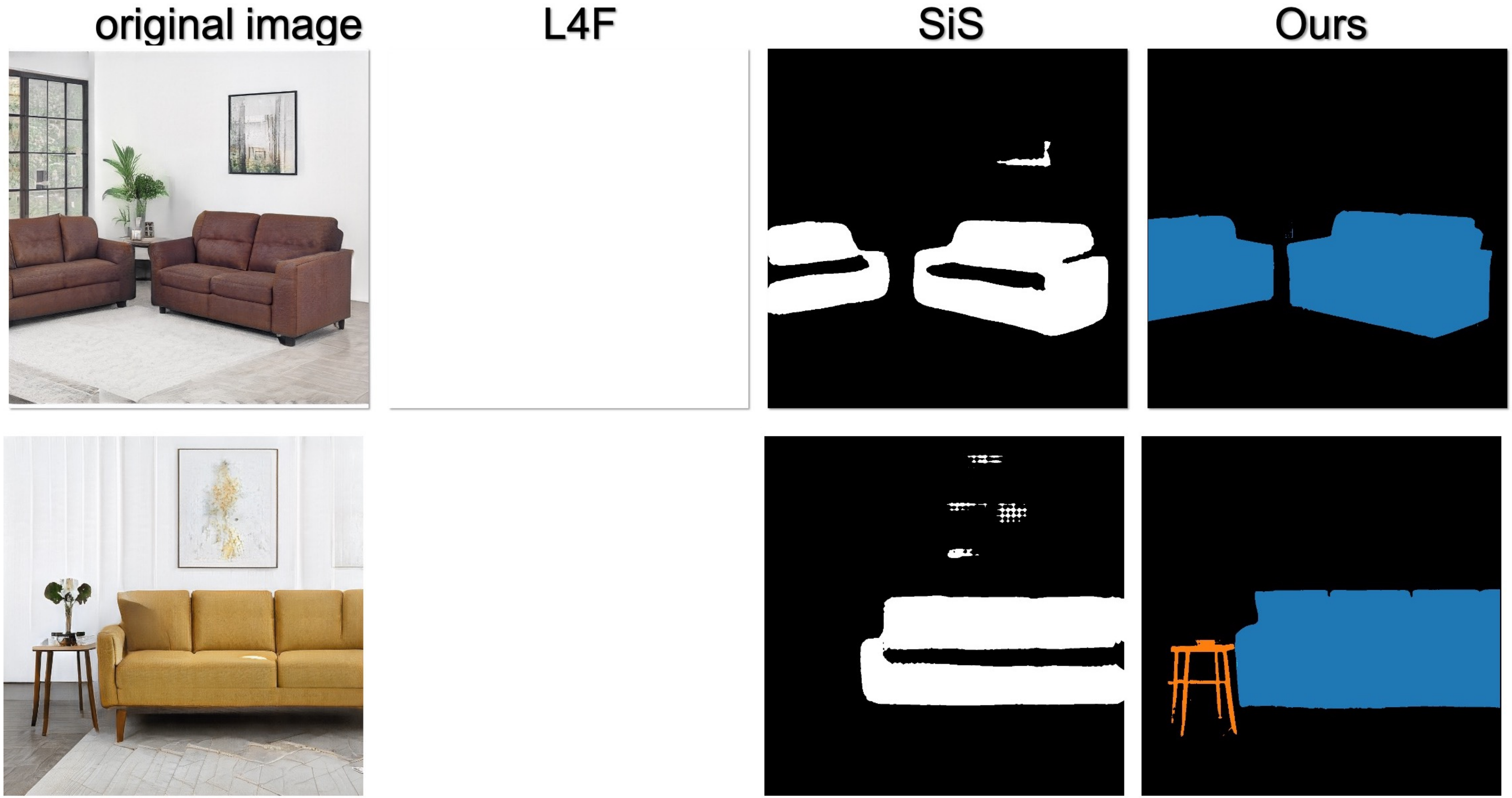
Table 3. Using synthetic data as training data for image segmentation. Trained on images generated from FFHQ model, test on CelebA-Mask-HQ (real data). The supervised segmentation method is DeepLabV3. All synthetic data performances are trained from scratch using synthetic data only. Trimap width is 3 pixels.

More qualitative results


0
1
2
3
4
5
6
8
9
11
12
13
14
15
16


0
1
2
3
4
5
6
7
8
9
10
11
13
14
15


0
1
2
3
4
5
6
7
9
10
11
12
13
14
15


original image
L4F
SiS
Ours

amazon

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Thanks! Please check out our paper

Network-free, unsupervised semantic segmentation with synthetic images

@THU-PM-286