COMP 546 HW 5

Kevin McCoy [kmm12]

colab notebook link

I did not collaborate with anyone.

1.0 Semantic Segmentation

a)

*** see colab notebook ***

b)

*** see colab notebook ***

c)

The "ego vehicle" class has the best IoU. Many classes have 0 IoU, which is most likely because they are not present in most images. For example, you can imagine a "caravan" is not a very popular class. This class imbalance probably also results in better performance for these classes, as the model has more data from these classes to train on.

Table 1: Labels Table with IoU

id	**name**	**category**	**Average IoU**
0	unlabeled	void	0.0
1	ego vehicle	void	0.871
2	rectification border	void	0.428
3	out of roi	void	0.802
4	static	void	0.003
5	dynamic	void	0.0
6	ground	void	0.0004
7	road	flat	0.870
8	sidewalk	flat	0.446
9	parking	flat	0.008
10	rail track	flat	0.0
11	building	construction	0.634
12	wall	construction	0.008
13	fence	construction	0.025
14	guard rail	construction	0.0
15	bridge	construction	0.00001
16	tunnel	construction	0.0
17	pole	object	0.047
18	polegroup	object	0.0
19	traffic light	object	0.010
20	traffic sign	object	0.102
21	vegetation	nature	0.711
22	terrain	nature	0.101
23	sky	sky	0.616
24	person	human	0.138
25	rider	human	0.0
26	car	vehicle	0.563
27	truck	vehicle	0.000001
28	bus	vehicle	0.0
29	caravan	vehicle	0.0
30	trailer	vehicle	0.0
31	train	vehicle	0.000015
32	motorcycle	vehicle	0.000061
33	bicycle	vehicle	0.086

d)

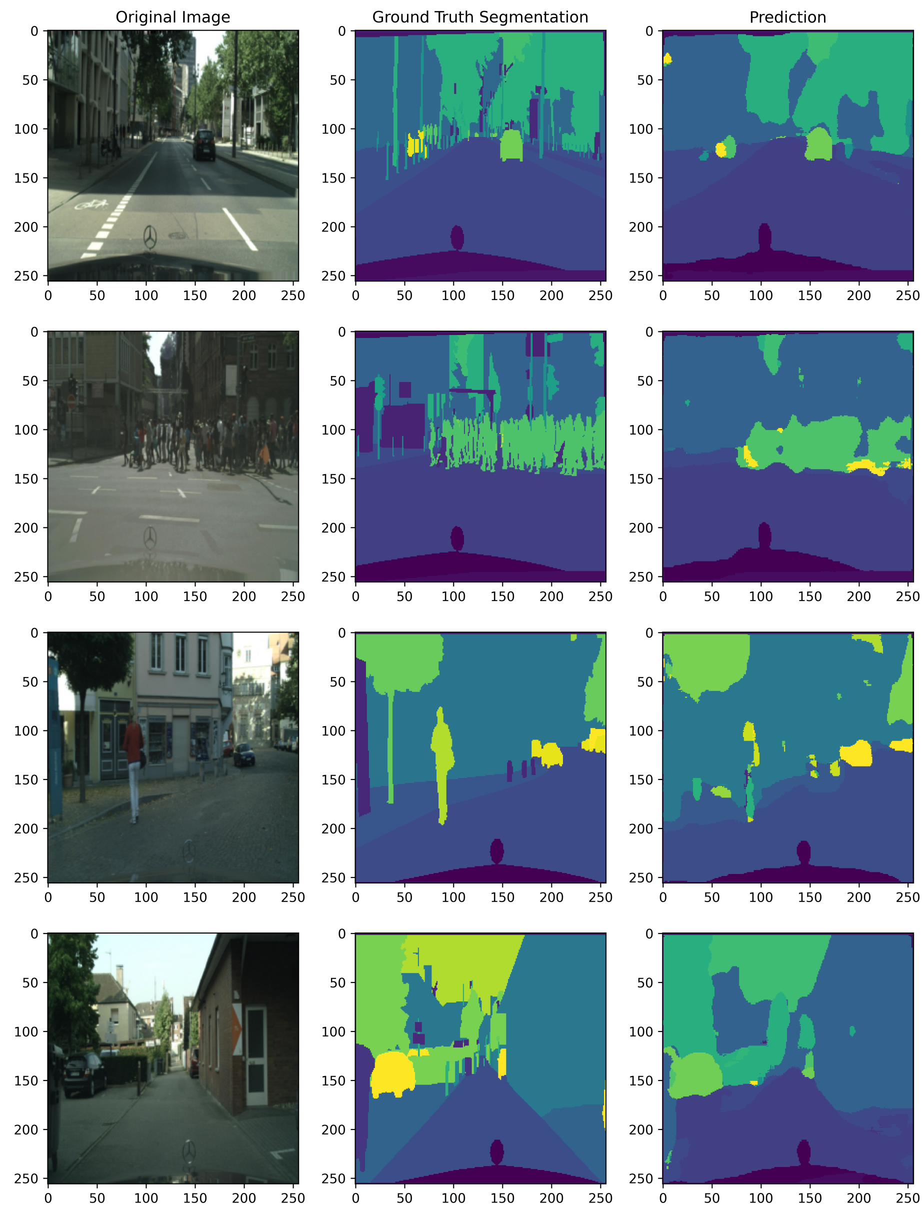

Figure 1: Original Images, Ground Truth, and Segmentation Predictions

e)

Using bilinear interpolation for the mask would not make sense as masks have discrete values for each respective class, and downsampling the mask should not result in a mixture of these discrete classes. Instead, it makes more sense to use nearest neighbors and just pick the closest class for that new pixel.

f)

Applying a horizontal flip augmentation to the image and mask using a random number generator ensures that the mask and image are either both not flipped or flipped. The model would not work well at all if the mask and image were opposite, as they would no longer correspond 1-1 to each other.

2.0 StyleGAN

a)

Figure 2: Interpolated Points Between Latent Vectors

Figure 3: Interpolated Points Between Style Vectors

The intermediate faces look quite realistic in both cases. However, they look more realistic when interpolating between points in style space. These images also make a nicer looking blend between the two endpoint faces. The 'middle' face in Figure 2 doesn't look quite like either endpoint face. These faces in Figure 2 are also much softer and fake looking.

b)

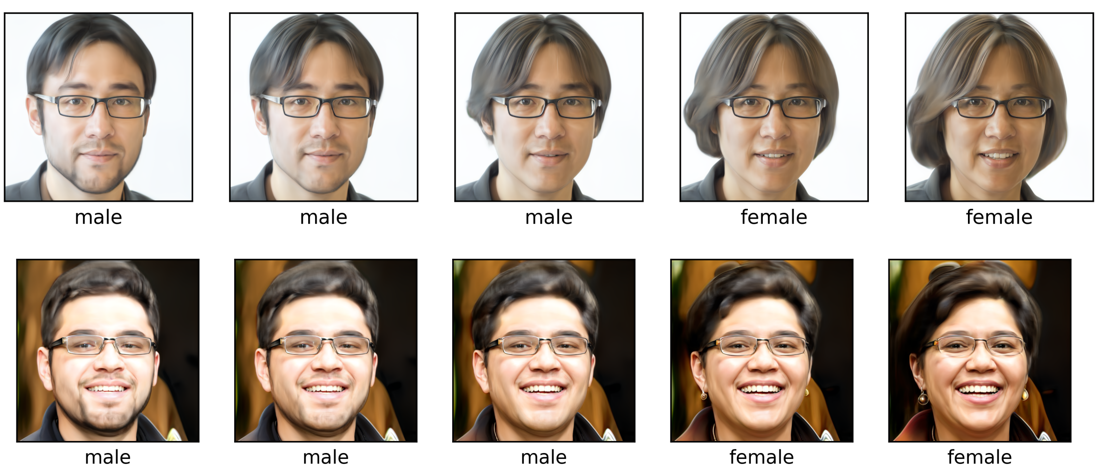

Figure 4: Two examples of latent space traversal. To the right travels along the direction of 'more female', and to the left travels along the direction of 'more male'.

Longer hair is usually found in the 'more female' pictures, and more facial hair is usually found in the 'more male' pictures. This makes sense, as women generally have longer hair and do not have facial hair. Females also have darker and fuller lips, as well as jewelry like earings.

3.0 Using CLIP for Zero-Shot Classification

a)

The overall accuracy using simple captions is 88.15%.

b)

The overall accuracy using the engineered captions is 88.76%.