Jaiden Hausler

January 26, 2022

ECGR 4106

Homework 0 Report

GITHUB LINK: https://github.com/jnhausler/ECGR4106/tree/main/Homework%200

Problem 1

For problem 1, we were instructed to pass five random images to ResNet 101 and analyze the outcomes.

To set up for problem 1 I imported the pre-trained models from torchvision in section [1]. Then I created an instance of the CNN ResNet 101 in section [3]. Before processing the images with ResNet 101, we must "clean" them up to match the format that ResNet trained on. The preprocessing function was created in section [5] to resize, crop around the center, transform it to a tensor, and normalize the RGB components.

Image 1

Then, I imported a picture of a cat with its eyes closed in section [6]. The image of the cat is shown below.

Figure 1 – Cat with eyes closed

In section [8], I used the preprocessing function to normalize the image for ResNet 101. The import tensor was reshaped, cropped, and normalized for the expectations of the ResNet 101 network in section [9].

The network was put into eval mode in section [10] to do inference. In section [11] the trained model was run on the image in Figure 1. This produces 1000 labels which are loaded in section [12]. To find the highest prediction, the max function in PyTorch was used to load the max value and the indices where the maximum value occurred in section [13]. Using that index, the label was loaded from the 1000 label list and the percentage of confidence was calculated in section [14]

The output of figure 1 shown in section [14] was "tabby, tabby cat" with a confidence percentage of 69.1064%.

The top five predictions were shown in section [15] using the top five indices. A table of the top 5 predictions for figure 1 is shown below

Table 1 – Top 5 predictions of figure 1

Rank	Label	Confidence Percent [%]
1	tabby, tabby cat	69.1064
2	tiger cat	14.8119
3	Egyptian cat	11.0667
4	lynx, catamount	4.2454
5	grey fox, gray fox, Urocyon cinereoargenteus	0.1243

The top prediction was correct, and you can see that the confidence drops from 1 to 2. This is expected from CNN ResNet 101 when a correct prediction is made. ResNet was trained on numerous cats so it makes sense it would correctly predict a close-up image of one. Prediction 2 through 4 are also types of cats, which means the network had roughly a 99.2304% confidence that the picture was a cat. The image of the tabby cat was very close up and clear which could be why the network was confident it was a cat. The last guess of a gray fox could be because the cat in the image is gray and animal-like.

Image 2

The next image used was another picture of a kitten with its eyes opened. The image is shown in figure 2 below.

Figure 2 – Kitten with eyes open

The image in figure 2 was imported in section [18]. Then the same process was repeated for figure 2 in sections [19-23]. The output of the top 5 predictions for Figure 2 is shown in section [23] and in table 2 below.

Table 2 – Top 5 predictions of figure 2

Rank	Label	Confidence Percent [%]
1	tabby, tabby cat	72.0715
2	Egyptian cat	14.0447
3	tiger cat	7.8706
4	lynx, catamount	4.9407
5	plastic bag	0.0701

The top prediction for this image was correct with confidence of 72.0715%. The top prediction shows 5 times higher confidence which was expected from this network. Both figure 1 and figure 2 show the same type of cat, but the second image had higher confidence. My assumption to this is because figure 2 has a cat with eyes open which could make it clearer as to what kind of animal it is. Also, figure 1 has human hands in the photo around the cat's face while figure 2 does not. Predictions 2-4 are also types of cats. The 5th prediction was a plastic bag, which I assume comes from the gray color of the cat.

Image 3

Image 3 is shown in figure 3 and was imported in section [24] of a green smoothie.

Figure 3 – Green Smoothie

The same process was repeated for figure 3 in sections [25-30]. The output of the top 5 predictions for Figure 3 is shown in section [30] and in table 3 below.

Table 3 – Top 5 predictions of figure 3

Rank	Label	Confidence Percent [%]
1	eggnog	37.4073
2	cucumber, cuke	14.0840
3	guacamole	7.7756
4	lemon	6.9965
5	plate	6.5336

The top prediction was incorrect with a confidence of 37.4073%. You'll notice that the confidence is about half of the previous image. This is expected because the top prediction was incorrect. The prediction being incorrect is most likely because the ResNet label list does not actually include smoothies. My assumption for the top prediction being eggnog is because it was the closest to being a thick drink. Predictions 2 and 3 seem to come from the color of the drink. The fourth prediction could come from the lime at the bottom of the figure. The plate could come from the background color of white, making it look like the fruits and smoothie are on a plate. These confidence levels are significantly lower than the previous images. This picture has a lot of fruits in the background which can interfere with the main object being a smoothie.

Image 4

Image 4 is shown in figure 4 and was imported in section [31].

Figure 4 – Succulent plants

The same process was repeated for figure 4 in sections [31-37]. The output of the top 5 predictions for Figure 4 is shown in section [37] and in table 4 below.

Table 4 – Top 5 predictions of figure 4

Rank	Label	Confidence Percent [%]
1	artichoke, globe artichoke	28.2051
2	hen-of-the-woods, hen of the woods, Polyporus frondosus, Grifola frondosa	23.5403
3	coil, spiral, volute, whorl, helix	16.2332
4	earthstar	6.1243
5	confectionery, confectionary, candy store	4.0311

The top prediction was incorrect with confidence of 28.2051%. Lower confidence than images 1 and 2 is expected since the prediction was wrong. There is no succulent label in the ResNet therefore it could not correctly label this picture. The top prediction came from the color and shape of the succulents. The second prediction is a type of brown plant that looks similar in shape to the succulents, therefore it has closer confidence to the top prediction. The remaining predictions seem to come from the shape of the succulents. Overall, the top 5 predictions show total confidence of 78.134% which is about 21% lower than images 1 and 2 which were correctly predicted.

Image 5

Image 5 is shown in figure 5 and was imported in section [38].

Figure 5 – Rottweiler

The same process was repeated for figure 5 in sections [39-44]. The output of the top 5 predictions for figure 5 is shown in section [44] and in table 5 below.

Rank	Label	Confidence Percent [%]
1	Rottweiler	99.4157
2	black-and-tan coonhound	0.2149
3	Doberman, Doberman pinscher	0.1154
4	Appenzeller	0.0477
5	miniature pinscher	0.0345

The top prediction was correct with a confidence of 99.4157%. This is the highest of the five images and the second prediction drops to below 1%. This is expected when the network correctly classifies an image. A correct prediction could be because the ResNet was trained on dogs and their breeds. Also, this image shows a close-up of the face of a rottweiler, therefore it might be easier to analyze. The other four predictions are all dog breeds with either similar colors or build.

Problem 2

For problem 2, we were instructed to pass 5 images of horses and use the ResNetGenerator network to convert them into pictures of zebras by adding stripes.

Sections [45-47] include code to import the ResNetGenerator and run the pre-trained horse2zebra dataset. Sections [48-49] include putting the network into eval mode and defining the preprocessing function.

Image 1

The first image is shown in figure 6 and was imported in section [50].

Figure 6 – First Horse Image

The image was preprocessed in section [51] and the variable was shaped in this section as well. The image was run through the generator in section [52]. The tensor output was converted back to an image in section [53]. The image output is shown below in figure 7 and in section [53].

Figure 7 – First Output Zebra

The first thing noticed is that the size of the image has changed. This is because the preprocessing is done on the original image to make it the needed size, as well as the changes made to turn the tensor output into an image. This image converted from a horse to a zebra well. This is because it is one single horse, and it shows the side of the horse making it more clear to the generator.

Image 2

The second image is shown in figure 8 and was imported in section [54].

Figure 8 – Second Horse Image

The second image was processed the same way in sections [55-57]. The output is shown in figure 9 and in section [57].

Figure 9 – Second Output Zebra

The generation for the second image is less clear than the first image. This is because the image contains two horses of two different colors, and they overlap each other making the horses less clear. The light tan horse at the front seems to be generated into a zebra better than the darker horse in the back.

Image 3

The third image is shown in figure 10 and was imported in section [58].

Figure 10 – Third Horse Image

The third image was processed the same way in sections [59-61]. The output is shown in figure 11 and in section [61].

Figure 11 – Third Output Zebra

The generation for this image successfully changed one horse of a light tan color but did not change the horse on the left. This might be because the horse on the left is a mixture of white and brown and this could have made the horse not recognizable to the model if horses of this pattern were not used in the model.

Image 4

The fourth image is shown in figure 12 and was imported in section [62].

Figure 12 – Fourth Horse Image

The fourth image was processed the same way in sections [63-65]. The output is shown in figure 13 below and in section [65].

Figure 13 – Fourth Output Zebra

This image did not convert from a horse to a zebra very well. It had some sections that show stripes, but it is faint. This could be because the horse is very dark, and it seems as though the generator does better on medium brown horses.

Image 5

The fifth image is shown in figure 14 and was imported in section [66].

Figure 14 – Fifth Horse Image

The fifth image was processed in section [67-69]. The output is shown in figure 15 below and in section [69].

Figure 15 – Fifth Output Zebra

The original image was a group of horses. The generator did well in converting the two medium brown horses. It did not as well on the dark and light tan horses. This may be that the pre-trained model was shown more medium brown horses therefore it does better at converting them. You can also see the third horse does not have many stripes on its neck and face. This could be because it is a group of horses that overlap each other, and the generator may have had a hard time deciding the outline of each horse.

Overall, the horse2zebra generator seemed to do better on medium brown pictures of horses. It also did slightly better on images of medium brown horses that did not have overlapping horses.

PROBLEM 3

In sections [70-71], the computational complexity and number of parameters were computed for ResNet 101 and ResNetGenerator. ResNet 101 had a computational complexity of 7.85 GMac and 44.55 M parameters. ResNetGenerator had a computational complexity of 56.89 GMac and 11.38 M parameters.

It makes sense for ResNet 101 to have less computational complexity because it is only taking in an image and outputting strings. ResNetGenerator is taking in an image but also creating a tensor that has enough information to output a changed image.

ResNet 101 has more parameters because it compares 1000 labels to the entire image and gives a percent for each categorization. ResNetGenerator was pre-trained just to search for a horse.

PROBLEM 4

The objective of this problem is to use the MobileNet model instead of ResNet 101 for problem 1. I loaded the model in section [72].

Image 1

In sections [73-76], I used the MobileNet to predict the image shown in Figure 1. The output of the top five predictions is shown in table 6 below and in section [76].

Table 6 – Top 5 Predictions for Figure 1 using MobileNet

Rank	Label	Confidence Percent [%]
1	tabby, tabby cat	68.0913
2	Eqyptian cat	25.6895
3	tiger cat	5.0017
4	lynx, catamount	0.8038
5	snow leopard, ounce, Panthera uncia	0.0801

The top prediction was correct with confidence of 68.0913%. This is lower confidence than using ResNet. However, all five predictions were a type of cat and in ResNet 101 only four predictions were cats.

Image 2

In sections [77-80], I used the MobileNet to predict the image shown in Figure 2. The output of the top five predictions is shown in table 7 below and in section [80].

Table 7 – Top 5 Predictions for Figure 2 using MobileNet

Rank	Label	Confidence Percent [%]
1	tabby, tabby cat	79.2443
2	Egyptian cat	9.9444
3	tiger cat	8.8594
4	lynx, catamount	1.3299
5	tiger, Panthera tigris	0.1587

The top prediction was correct with confidence of 79.2443%. This is higher confidence than using ResNet. Also, all five predictions were a type of cat whereas ResNet 101 only predicted four types of cats.

Image 3

In sections [81-84], I used the MobileNet to predict the image shown in Figure 3. The output of the top five predictions is shown in table 8 and in section [84].

Table 8 – Top 5 Predictions for Figure 3 using MobileNet

Rank	Label	Confidence Percent [%]
1	eggnog	78.7878
2	head cabbage	4.7354
3	lemon	2.0321
4	guacamole	1.8342
5	strawberry	1.8029

The top prediction was incorrect with a confidence of 78.7878%. This is higher confidence than using ResNet 101. Although there is no label for the smoothie in figure 3, MobileNet has shown to be more confident in a wrong answer. Both had the same top prediction and two other matching predictions.

Image 4

In sections [85-88], I used the MobileNet to predict the image shown in Figure 4. The output of the top five predictions is shown in table 9 and in section [88].

Table 9 – Top 5 Predictions for Figure 4 using MobileNet

Rank	Label	Confidence Percent [%]
1	artichoke, globe artichoke	67.4080
2	coil, spiral, volute, helix	14.5463
3	custard apple	5.4111
4	confectionery, confectionary, candy store	3.8310
5	grocery store, grocery, food market, market	1.4279

The top prediction was incorrect with confidence of 67.4080%. This is almost double the confidence of ResNet 101. Although there is no label for the succulent in figure 4, MobileNet has shown again to be more confident in a wrong answer. Both had the same top prediction and two other matching predictions.

Image 5

In sections [89-92], I used the MobileNet to predict the image shown in Figure 5. The output of the top five predictions is shown in table 10 and in section [92].

Table 10 – Top 5 Predictions for Figure 5 using MobileNet

Rank	Label	Confidence Percent [%]
1	Rottweiler	99.4157
2	black-and-tan coonhound	0.2149
3	Greater Swiss Mountain dog	0.1154
4	Doberman, Doberman pinscher	0.0477
5	EntleBucher	0.0345

The top prediction was correct with confidence of 99.2630%. This is slightly lower confidence than using ResNet 101. MobileNet and ResNet 101 were both similar with having all five predictions being dogs.

Computational Complexity and Number of Parameters

In section [93], the computation complexity of 0.32 GMac and 3.5 M parameters was found for the MobileNet network. This is significantly less computational complexity and parameters than ResNet 101. Overall, MobileNet is less accurate than ResNet because it is more confident in wrong answers. Still made the same number of correct answers as ResNet 101. MobileNet is great if you need less power or a faster network. ResNet 101 is better if you need more accuracy.