Fall 2023
Intro to Deep Learning and Pattern Recognition for Computer Vision 18794

Homework 3 Semantic Segmentation

Due: Tuesday, Nov. 14, 2023 11:59 pm

Submitting your work: For all problems on this homework we will be using **Gradescope.** To receive full credit, you need to submit both the PDF solution and a zip file containing your code solution. Please be super clear, neat, and complete when you write your solutions. **If the TAs cannot follow your work, you will not receive full credit for that problem.**

Collaboration Policy: Homework will be done individually: each student must hand in their own answers. It is acceptable for students to collaborate in figuring out answers and helping each other solve the problems. We will be assuming that, as participants in a graduate course, you will be taking the responsibility to make sure you personally understand the solution to any work arising from such collaboration. You must also indicate on each homework with whom you collaborated. **This homework has 100 points.**

In this assignment, we'll be building our own DeepLab network, a framework designed for high resolution, precise image segmentation, and using it to predict a categorical label for every single pixel in an image.

This type of image classification is called semantic image segmentation. It's similar to object detection in that both ask the question: "What objects are in this image and where in the image are those objects located?," but where object detection labels objects with bounding boxes that may include pixels that aren't part of the object, semantic image segmentation allows you to predict a precise mask for each object in the image by labeling each pixel in the image with its corresponding class.

To train a segmentation network, you will need an annotated dataset where a training pair contains an RGB image and the annotated segmentation map. Each pixel is labeled by a categorical number similar to the classification.

To finish this assignment, you need to submit a zip file containing

both the finished code and a report.

Problem 1: Prepare Data Pipeline (20pts)

Finish data pipeline code for training and evaluation.

1. (5pts) In *datasets/voc.py*, complete the *VOCSegmentation* class. Specifically, please finish the *__getitem()__* method and *decode_target* method.

2. (5pts) How many categories are there in the dataset? In the training set, do you think this is a class-balanced set? And what can be the potential challenges while training a model on this dataset?

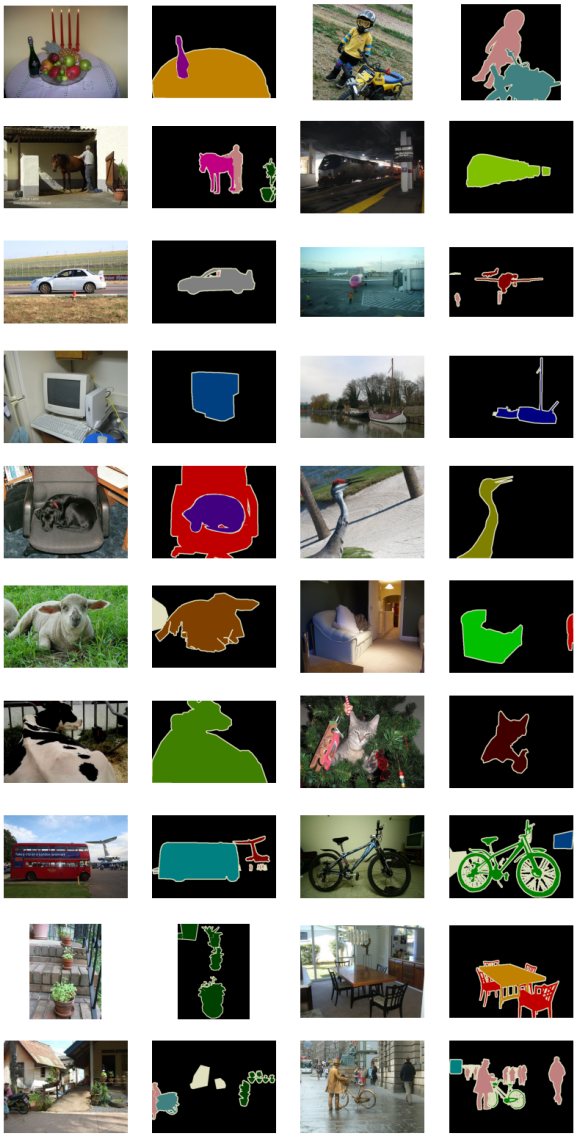
1. There are 20 categories with a background category, so total number is 21 categories
2. No, in the datasets, some categories have more instances than others.
3. If this isn't class-balanced, it may happens that model perform better in ones with larger datasets category, but performs badly in ones with less datasets. Besides, there can be chanllenges if the locations of objects aren't precise enough. Also, the real shape of object can vary a lot, not all of the shapes can be well included.

3. (5pts) Semantic Segmentation utilizes Accuracy and Mean IoU as evaluation metrics. Can you give an example to show why mIoU is a better metric than accuracy?

accurancy: correctly prediction/total predition
mIoU: mean(intersection/union)
Example: In automation driving, we want to distinguish between roads and non-roads area. Since it's dangerous to treat non-road as roads and there are more roads than non-roads, we want to give more penality to that roads are treated as non-roads. Using mIoU makes this possible for it uses intersection and union and penalize misclassifications based on the significance of each class.

4. (5pts) Pick any 20 training images that covers all available categories. Plot the ground truth annotations side-by-side with the training images and show them in your report. (You can organize this as a large picture with 10 columns x 4 rows)

visualization of 20 images
chuangm: 2023-11-13



Problem 2: Build Segmentation Network (30 pts)

Please read the original DeepLabV3 and DeepLabV3+ paper and implement the networks.

1. (15pts) In network/_deeplab.py, complete ASPPConv, ASPPPooling and ASPP.

2. (15pts) Complete DeepLabHead & DeepLabHeadPlus. What are their differences? Please write your understanding in the report.

The first difference is that the Plus uses both out and the low_level features which are from the backbone. But DeepLabHead not contain that and don't do the conv-net for both.
Besides, plus conains low_level_channels as input for processing and make model works better for containing additional useful inforamtion and more complex net-work.
DeepLabHeadPlus considers the low level part of backbone and can imporve the performance.

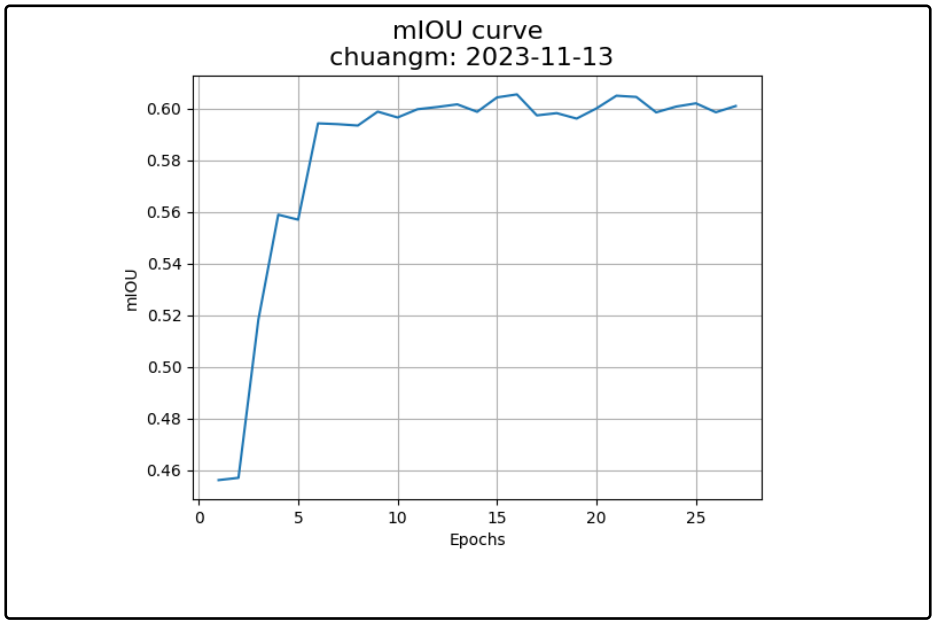
Problem 3: Training and Evaluation (35 pts)

In "main.py" file, please complete the training loop and train the networks.

1. (5pts) Build the optimizer. Since we will be using an ImageNet pretrained ResNet, we want to scale down the learning rate on the *__backbone__* component. Set up an optimizer such that the learning rate of the *__backbone__* is 0.1x the main learning rate.

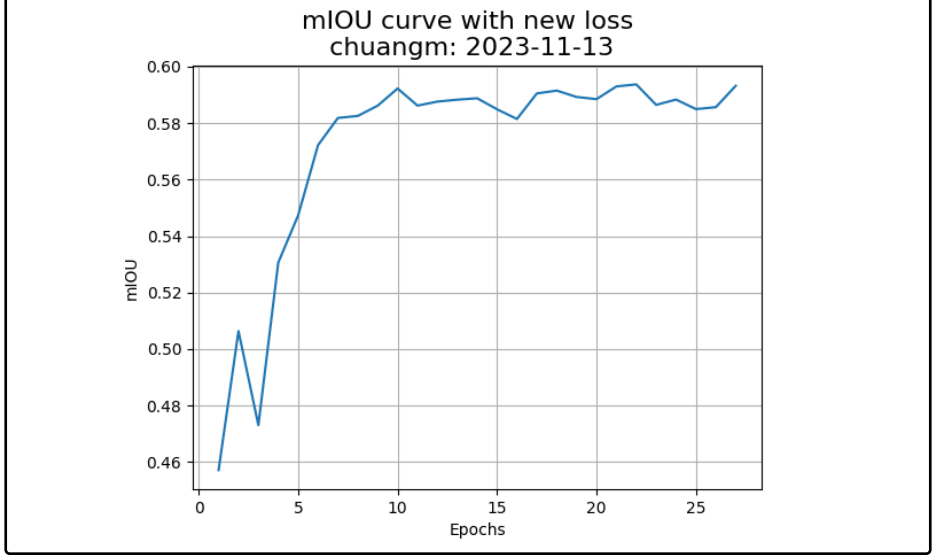
2. (5pts) Build the learning rate scheduler. We will be using a step learning rate scheduler that reduce the learning rate by 0.9x every 1,000 iterations.

3. (10pts) Train the DeepLabV3+ with resnet50 backbone (imagenet pretrained) on PascalVOC train split for 5k iteration, with *CrossEntropyLoss*. Plot the evaluation mIOU with an interval of 1 epoch in your report. (You should be able to train the model with a batch size of 8. In our test run, GPU memory usage is 6861MB.)



4. (10pts) Based on what you have learned in the dataset, train the same network using the same settings, but in this time, with a different loss function. Write your loss function in *utils/loss.py*. What loss function do you propose to use and why? Plot the evaluation mIOU with an interval of 1 epoch in your report. (Hint: What do you learn from the Detection assignment when it comes to class-imbalanced training?)

Based on what I learned from detection assignment of class-imbalanced. I have tried many lossed includeing diceloss, exponential cross entropy loss... I finally choose exponential cross entropy loss, for the exponential part are often used in dealing with class imbalance and give more weight less frequent classes.



5. (5pts) For both models, report the best performances and pick 5 images

in problem 1, show the ground-truth vs predictions side-by-side. There should be 4 columns, | RGB Image | Ground Truth annotation | Model 1 prediction | Model 2 prediction|

Model1: Best result:

```
Training state restored from checkpoints/latest_deeplabv3plus_resnet50_VOC_os16.pth
best score and itr: 4941, 0.6055021500756477
Model restored from checkpoints/latest_deeplabv3plus_resnet50_VOC_os16.pth
```

It happens in this epoch:

```
 Epoch 16, Itrs 2920/5000, Loss=0.134392
 Model saved as checkpoints/latest_deeplabv3plus_resnet50_VOC_os16.pth
 validation...
 1449it [01:29, 16.27it/s]

 Overall Acc: 0.885850
 Mean Acc: 0.830480
 FreqW Acc: 0.814458
 Mean IoU: 0.605502

 new best mIOU:  0.6055021500756477
 Model saved as checkpoints/best_deeplabv3plus_resnet50_VOC_os16.pth
 Epoch 17, Itrs 2930/5000, Loss=0.147954
```

And this is the last epoch:

```
Epoch 27, Itrs 4940/5000, Loss=0.137646
Model saved as checkpoints/latest_deeplabv3plus_resnet50_VOC_os16.pth
validation...
1449it [01:42, 14.11it/s]

Overall Acc: 0.882715
Mean Acc: 0.834167
FreqW Acc: 0.810483
Mean IoU: 0.601003

Epoch 28, Itrs 4950/5000, Loss=0.137955
```

Model2: Best result:

```
Device: cuda
Dataset: VOC, Train set: 1464, Val set: 1449
Training state restored from checkpoints/best_deeplabv3plus_resnet50_VOC_os17.pth
best score with itr: 0.5937002663362606, 4026
Model restored from checkpoints/best_deeplabv3plus_resnet50_VOC_os17.pth
```

It happens in this epoch:

```
 Model saved as checkpoints/latest_deeplabv3plus_resnet50_VOC_os17.pth
 validation...
 1449it [01:25, 16.87it/s]

 Overall Acc: 0.877684
 Mean Acc: 0.835104
 FreqW Acc: 0.804323
 Mean IoU: 0.593700

 new best mIOU:  0.5937002663362606
 Model saved as checkpoints/best_deeplabv3plus_resnet50_VOC_os17.pth
 Epoch 23, Itrs 4030/5000, Loss=0.112314
 Epoch 23, Itrs 4040/5000, Loss=0.126264
```

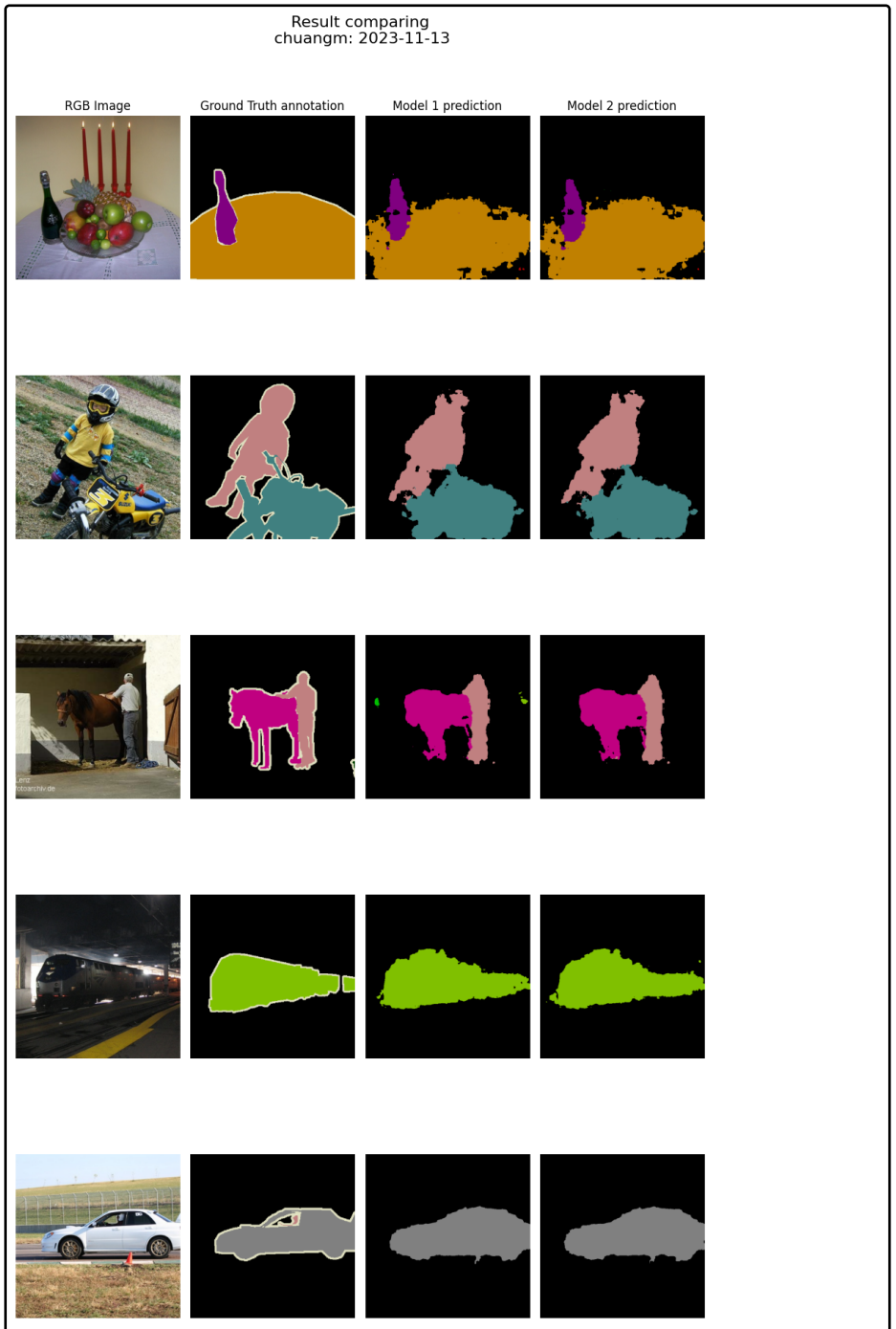
And this is the last epoch:

```
Model saved as checkpoints/latest_deeplabv3plus_resnet50_VOC_os17.pth
validation...
1449it [01:38, 14.70it/s]

Overall Acc: 0.877290
Mean Acc: 0.836523
FreqW Acc: 0.804294
Mean IoU: 0.593212

Epoch 28, Itrs 4950/5000, Loss=0.136432
Epoch 28, Itrs 4960/5000, Loss=0.145185
```


Please see comapring result in next page.


Result comparing
chuangm: 2023-11-13
RGB Image
Ground Truth annotation
Model 1 prediction
Model 2 prediction

Problem 4: Segment-Anything-Model (15pts)

Segment-Anything-Modelis recently proposed by Meta AI Research that produces complete high-quality object masks. Please follow the [installation] instruction, and download the pretrained model checkpoint. You can also modify the provided [Google Colab script] as well to save some AWS credits.

1. Run the model on the same 5 images. What do you think are the differences between semantic segmentation and SAM?

I think semantic segmentation more focuses on classification in pixels level and identify which image pixels belong to an object. It needs the number of classes and name of classes. Also, it needs manually annotated objects to train. While SAM is a generalized approach and no longer need to collect segmentation data and fine-tune a model for use. Also, SAM is trained on very large number datasets of masks.

Result of same 5 images:

see next page for rest ones

please see next page
see next page for rest ones

source image

segmented image

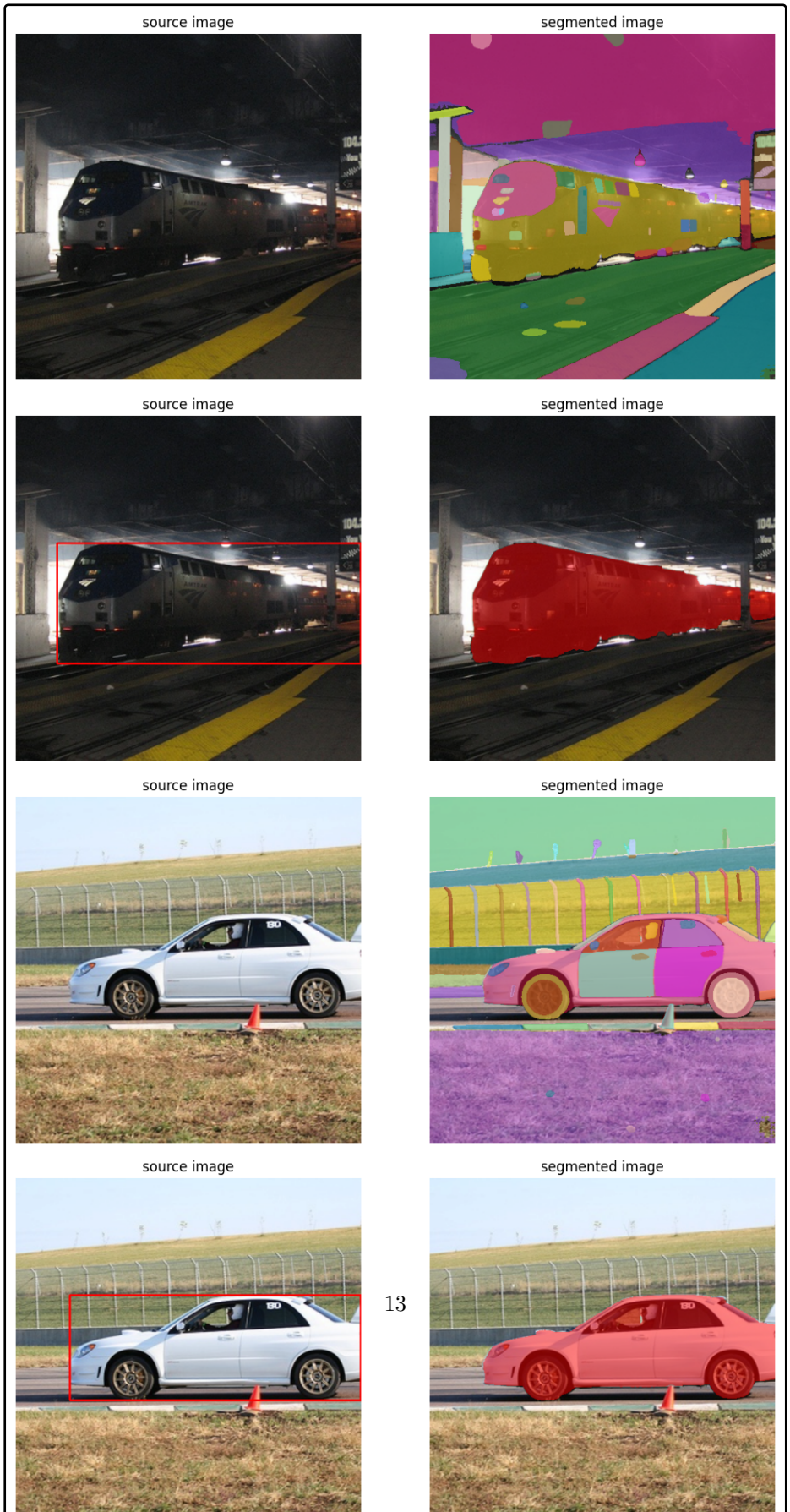
source image

segmented image

source image

segmented image

please see next page
see next page for rest ones


source image
segmented image
source image
segmented image
source image
segmented image
source image
segmented image