

MIT CSAIL
6.8300/1 Advances in Computer Vision
Spring 2024

Problem Set 4

Posted: Tuesday, Mar 12, 2024

Due: Tuesday 11:59 pm, Mar 19, 2024

Gradescope Instructions: This week, (1) convert your completed .ipynb notebook to .pdf and submit. (2) Also submit a .ipynb submission to a second portal.

Note that colab no longer allows downloading as html or pdf, so we have provided a helper notebook to automate that process. **Make sure your conversion to PDF includes your correct answers**, since sometimes the conversion cuts off answers. If you want to write comments, add markdown text boxes with prominent headers. Please note that 6.8301 students will not receive credit for problems which are only for 6.8300.

Attention: Unreported work gets a zero: please run all cells first.

Attention: No .zip. Only one file needed per submission portal.

Late Submission Policy: If your problem set is submitted within a week of the original deadline, you will receive partial credit. Such submissions will be penalized by a multiplicative coefficient that decreases from 1 to 0.5, daily stepwise.

In this problem set, you will be experimenting with neural networks using PyTorch. To reduce training time, we recommend using GPU acceleration. Colab comes with free GPU support. On Colab, select GPU as your runtime type as follows:

Runtime → Change runtime type → Hardware accelerator → GPU → Save.

If you exceed your GPU usage limit on Colab, don't fret. You can also complete your problem set using a regular CPU in a reasonable amount of time.

Problem 0 *Notebook Submission, .ipynb version* (1 point, required)

Problem 1 *Filter Visualization* (20 points)

Convolutional kernels in different layers are used to extract information from the input image at different levels. In this problem, we will focus on visualizing the filters of the ResNet network pretrained on the Places365 dataset. You must download the model and the pretrained weights first.

(a) Implement `normalize_tensor` to normalize the input kernel for better visualization.

(b) Visualize filters for another convolutional layer in ResNet (If what you see is unshapely it's probably not right. Also, see course notes, Sec. 24.2.1, to distinguish layers from color channels.)

Problem 2 *Internal Activation Visualization* (6.8301: 20 points, 6.8300: 50 points)

In this problem, we will visualize the activations of the internal units in the model. We will chop the fully connected layers of ResNet and visualize the activations of the last convolutional layer into different locations of the input image.

(a) Create a version of the ResNet model without the last two layers by completing `remove_last_layers` to expose the last convolutional layers in the `generate_featuremap_unit` function.

(b) As you can see, unit 300 activates the mountain part of the image. Find other units that detect 1) the sky and 2) buildings in the image.

(c) **[6.8300 only]** Each unit contributes differently to the final prediction. The contribution weight for each unit is determined by the weights of the fully connected layer (last layer in ResNet). If we deactivate top-5 units (force the kernel to be zeros) that have the maximum weights for the top 1 category of some input image, the prediction for that category will drop dramatically. The code to find the index of those units is given, try to deactivate those units and compare the difference between the original and new top 5 predictions. **Show the feature map of each of the 5 deactivated units.**

(d) **[6.8300 only]** In (c), we deactivate top-5 units and observe dramatic changes in predictions. Try to deactivate fewer units and. **Report the lowest number of dropped units required to change the top prediction class.**

Problem 3 *Class Activation Map (CAM)* (30 points)

So far, we know the output of the last convolutional layer activate different parts of the input image. In this problem, we will explore how to visualize which parts of the image are responsible for the final decision. Use the chopped ResNet from problem 2 as the model.

(a) Running the chopped ResNet will produce a tensor with the size of $(1 \times 2048 \times 8 \times 8)$. 1 is the batch size (b), 2048 is the number of channels (c) and 8 & 8 is the height (h) and width (w) of the kernel. Convert the output tensor to a new tensor with the size of $(hw \times c)$.

(b) Feed the new tensor from (a) to the fully connected layer of ResNet to compute the weighted average. You will get a output tensor with the size of $(hw \times 365)$, where 365 is the number of classes in the Places365 Dataset.

(c) Show the feature maps (combined with input image) of the top 5 predicted categories and explain the relationship between the feature activation and the corresponding category.

Problem 4 *Network Training* (6.8301: 30 points, 6.8300: 50 points)

The goal of this problem is to train a small convolutional neural network to classify images of clothing items from the FashionMNIST dataset. You'll first fill in critical components of a simple PyTorch training pipeline, evaluate the model on the test set, and explore the impact of specific design choices and hyperparameters on the model's performance.

(a) Read and execute the notebook cells until you reach the training loop section. Here you will complete the function `train` that trains a network for 1 epoch by filling in the missing code snippets.

(b) Implement the `get_prediction` function which returns the index of the predict class given an image and a network and is called during the evaluation routine. Also implement the accuracy computation in the (evaluate) function

(c) Here, we bring all of the pieces together to train the network. Instantiate an optimizer to update the weights of the network during training. Run training and validation for 10 epochs and report your final validation accuracy.

(d) **[6.8300 only]** We want you to get a feel for the impact of specific design choices on the performance of the network. Experiment with two or more of the following hyperparameters / techniques:

- Data augmentation
- Weight initialization
- Number of layers, or number of layer features
- Type of optimizer
- Learning rate and/or schedule
- Regularization

For the techniques you choose, plot the top-1 accuracy of your modified network against the top-1 accuracy of the original network for both the training and validation sets. Try several different hyperparameter values! For example, if you choose to modify the learning rate, you can plot a chart of learning rate vs. top-1 accuracy. Briefly describe the techniques you tried, and suggest an explanation for your results. **Full credit will only be given if at least 90% validation accuracy is achieved** (*i.e.*, the accuracy must be above 89.999999%).

Results: at least two plots, greater than 90% accuracy network.