Assignment 3 Training a Scene Segmentation Network on an Automotive Dataset

January 28, 2019

Objective

The objective is to gain experience with the architecture of modern convolutional neural networks (CNN) for perception tasks in automated driving.

For this purpose we will use a lab and infrastructure sponsored to us by Nvidia Deep Learning Institute.

By working though this lab, you will learn about state of the art convolutional neural networks, MobileNets, which target embedded domains. You will see how a pre-trained CNN can be reused for a new task, specifically image segmentation, and how a complete network architecture is designed and sized to fit the task and the target execution environment. You will also experience working with an automotive dataset for image segmentation (Cityscapes), and the process of training and evaluating a segmentation network.

This lab builds on the concepts you have learned in class and the previous assignments, but it also includes additional concepts and reading. In contrast to the previous assignments, working with a segmentation net requires a GPGPU to speed up the training and inference. Thus, you will use an Amazon cloud environment, with a server including a GPGPU and all the necessary software to work through the lab.

Resources and Instructions

The lab instructions, background material, and the questions and code to be executed are contained in a Jupyter notebook, which you will access after login into your lab account at NVIDIA Deep Learning Institute.

To access online notebook and the cloud session, go to this page:
https://courses.nvidia.com/dli-event
and enter this event code: L-AV-02-WA
Notes: If the enrollment does not work the first time, try again. Once you're enrolled, select the course from the top menu, and then scroll to the bottom and press the start button. Once you hit start, make yourself a cup of coffee… starting the Jupyter session may take 10-15 minutes.

You'll have to create an account. Read the instructions below **before** starting your session.

The Jupyter notebook has multiple sections with background material, questions (also note the additional questions in the Marking section below), and code cells to be executed. You will spend the majority of the time needed to complete this lab on reading and understanding the notebook. Assuming that you are familiar with the material covered in class, the notebook text is self-contained; however, you may find it helpful to consult some of the referenced materials, especially the MobileNets paper.

Before you access the account and start working on the Jupyter notebook, I strongly recommend that you work through the PDF export of this notebook offline. The PDF export is available on UW learn. **The online Jupyter notebook session is timed for 2 hours and all your work will be deleted after the time is up.** Working though the PDF export and answering all the questions listed there (except the few that need running the code) may take 6-10 hours, depending on how deeply you work through the material, and how familiar you may already be with some of the concepts. I suggest that you prepare a separate file with the answers to the questions and specific code updates for the later online session, so that you can save time when you access the online Jupyter notebook during the limited online session.

Once you have answered the questions offline based on the PDF export, you will be able to execute the code portions of the notebook by accessing your account at NVIDIA Deep Learning Institute. At that point, you will edit the text cells of the Jupyter notebook to answer the questions; ideally, by just pasting your answers from the previously prepared file. You will also execute the code cells and ensure that the output is correct. In the last part of the lab, you will be accessing the DIGITS environment (linked from the Jupyter notebook), to load the dataset, setup the network, and execute and monitor the training. **When loading the dataset, name it <user id>_7.** Finally, you will run the network on the test data, which will report the overall test score. With good offline preparation, the online portion will take about 1 hour, but should not exceed 2 hours.

Notes:

1. Section 6.2 – After running data import, you need to refresh the page in your browser to see the import results and verify that your dataset DB contains training and validation data.
2. Section 7.2 – If you get an error "Resource exhausted: OOM when allocating tensor", you've run out of GPU memory. Try a smaller depth multiplier or batch size; 0.25 and 4 should work.

Deliverables

Submit a zip file of a directory containing the following deliverables:

1. HTML export of the final Jupyter notebook including your answers to the questions (see below for details on the questions) and the correct outputs from executing the code cells;
2. PDF printout of the same final Jupyter notebook (use the browser print function);
3. Screenshot (png file) of the tensorflow board that plots the loss and IoU over iterations after the training is done; the board is part of the DIGITS interface showing the training process; **make sure that your screenshot shows the dataset name, which will use your your user id**;
4. Export of the trained model from the DIGITS environment (tar.gz).

Please follow the naming convention of your zip file: **a3_<user_id>.zip**

Marking

This assignment has a completion mark. You will be awarded full points for submitting the required files, and 0 otherwise.

Due Date

11:59pm on February 4, 2019

There will be no deadline extensions.

Policies

Collaboration

You can discuss the problem with peers, but you must design and implement your own solution independently.

Use of online resources

You may consult any online resources to get ideas or to troubleshoot, but you must develop your own code.