

Neural Networks and Deep Learning: Lab Assignment 4

Summary: In this assignment, you will demonstrate you understand how to train, evaluate, and analyze neural networks for a multimodal problem (i.e., language and vision).

Vision and Language task: You will develop a model that, when given a visual question, predicts if it is answerable.

1. Design and conduct your experiment (Code)

- Use the VizWiz-VQA dataset with its pre-defined train/validation/test split to develop models. You can find the dataset and the paper describing the dataset at the following [link](#). Alternatively, you can refer to the TA's programming tutorial for how to access publicly posted versions of the train, validation, and test data.
- Train at least three models using neural network architectures and optimization schemes of your design. The one requirement is that each trained model leverages information about both the input image and the input question as predictive cues. For the three variants, you are welcome to use the same architecture and optimization scheme with three different sets of hyperparameters. For full credit, you need to train each model with at least 1,500 visual questions and test on at least 600 validation examples.
- Report the performance of each trained model on the validation set using average precision (covered in the TA's programming tutorial and object detection lecture).
- Using the top-performing approach found on the validation set, create your final model that you will use on the test data. You can either use the model that you already trained from the previous step or you can retrain with more data (e.g., on all of the training and validation data). When training your final model, create a plot that shows the loss curves on both the training and validation splits (i.e., the plot should show two curves indicating the computed loss with respect to the number of epochs used during training).
- With your final trained model, compute prediction results on the following 1,000 examples in the test split: i.e.,
VizWiz_test_00000000.jpg – VizWiz_test_00000999.jpg. All prediction results must be stored in a csv file with the predictions listed in the first 1,000 rows in column A. Note, for submitting your results in a csv file, your result file **MUST** be named "results.csv" and be included directly in the ZIP file you submit for your assignment (NOT in a folder inside the ZIP). This is to enable automated analysis of your predictions. To generate your results.csv file, you can use the code provided in class. Verify your results.csv file contains 1,000 rows of data without any header. The CSV values should specify prediction confidence scores, such that each confidence score provided by a prediction model indicates if the visual question is 'answerable' using values in the range $[0,1]$.

- Extra credit: Your prediction results will be included as part of a competition between all students in the course. The first place winner will receive 10 extra points and the second place winner will receive 5 extra points. Performance will be judged using the average precision evaluation metric.
2. Describe your methods, results, and analysis on the validation and test sets. Feel free to refer to [examples](#) shared on Canvas to revisit expectations. (Report)
- Validation set
 - Describe the methods you tested on the validation set such that the reader could reproduce your experiments. This should include a discussion of the dataset (e.g., source? number of examples?), what neural network architectures and hyperparameters were used to train all the models, and what type of hardware was used during training.
 - Describe your results on the validation split.
 - Discuss your analysis of general trends that emerge from your results. Your discussion should be approximately 3-5 paragraphs long. To format each paragraph, please first identify one general trend observed from the results and then offer insights/speculations into **why** you think the trend/results may occur (regardless of whether you deem the results good or bad).
 - Test set
 - Indicate which method was used on the test data and describe what, if anything, changed when using it on the test set such that the reader could reproduce your set-up (e.g., trained on additional data).
 - Visualize the loss curves for your final prediction system.
 - Discuss what you can infer by looking at the loss curves.

How to Submit Lab Assignment 4: Please submit a zip file named with your first and last name; i.e., `firstname_lastname.zip`. It must include two contributions. First, include a csv file with your prediction results, named “results.csv”. The second item to include in the zip file is a PDF file that merges two self-contained, separate contributions. First, it should include a report describing all methods, results, and analysis (i.e., portions indicated by “Report”) **as the first part of the PDF file**. Second, it should include the source code of your implementation **as the second part of the PDF file** (i.e., portions indicated by “Code”).¹ We will only review the code in detail when the report is not a self-contained document, in order to provide partial credit.

¹We require submitting the code as a PDF to avoid many issues that we have observed in the past with being able to access submitted code. These issues have arisen, in part, because we make no programming language requirements. Issues also have arisen from students not providing read permissions for links to their files; e.g., for Google Colab.