International Institute of Information Technology, Hyderabad
Spring 2024 CS7.505: Computer Vision
Assignment 2: Image Classification
20 February 2024

Instructions:

- The goal of this assignment is to get familiar with image classification.
- You should upload the assignment as a jupyter notebook with appropriate cells (markdown and code) containing (1) code that you wrote, (2) keep relevant outputs, and (3) your report and observations (markdown cells). The file should be uploaded in the courses portal.
- We recommend Python.
- Include the assignment number, your name, and roll number in the first cell of the notebook submission.
- **Make sure that the assignment that you submit is your own work. Any breach of this rule could result in serious actions including an F grade in the course.**
- The experiments and writing it all together can take time. Start your work early and do not wait till the deadline.

Submission: Any time before *8th Mar 2024, 23:59 IST*

1 Assignment

This assignment requires you to implement image recognition methods. Please understand and use relevant libraries. You are expected to solve both questions.

Data preparation and rules

Please use the images of the MNIST hand-written digits recognition dataset. You may use `torchvision.datasets` library to obtain the images and splits. You should have 60,000 training images and 10,000 test images. Use test images only to evaluate your model performance.

Q1: SIFT-BoVW-SVM [4 points]

1. [2 points] Implement the SIFT detector and descriptor. Compute cluster centers for the Bag-of-Visual-Words approach. Represent the images as histograms (of visual words) and train a linear SVM model for 10-way classification.

   Note 1: You may want to use libraries such as cv2 (OpenCV) and sklearn (Sci-kit learn) for doing this question. `https://scikit-learn.org/stable/modules/svm.html#multi-class-classification` may be useful for the SVM.

   Note 2: Seed random numbers for reproducibility (running the notebook again should give you the same results!).

2. [1 point] Keeping everything else constant, plot how classification accuracy changes as you sweep across 6 different values for the number of clusters. Please decide what numbers are meaningful for this question. Explain the trends in classification accuracy that you observe.

   Note 1: It is recommended to try hyperparameters in logarithmic steps such as 2x or 3x multiples. An example of 2x multiples is: 1, 2, 5, 10, 20, ... An example of 3x multiples is: 1, 3, 10, 30, 100, ...

3. [1 point] Show the results for 6 different hyperparameter settings. You may play with the SIFT detector or descriptor and the linear SVM. Keep the number of clusters constant based on the answer to the previous question. Explain the trends in classification accuracy that you observe.

Q2: CNNs and Transformers [6 points]

1. [2.5 points] Set up a modular codebase for training a CNN (LeNet) on the task of handwritten digit recognition. You should have clear functional separation between the data (dataset and dataloader), model (nn.Module), and trainer (train/test epoch loops). Implement logging: using Weights & Biases is highly recommended, alternatively, create your own plots using other plotting libraries. Log the training and evaluation losses and accuracies at every epoch, show the plots for at least one training and evaluation run.

   Note 1: Seed random numbers for reproducibility (running the notebook again should give you the same results!).

2. [1 point] Show the results for 6 different settings of hyperparameters. You may want to change the batch size, learning rate, and optimizer. Explain the trends in classification accuracy that you observe. Which hyperparameters are most important?

3. [0.5 points] Compare the best performing CNN (from above) against the SIFT-BoVW-SVM approach. Explain the differences.

4. [0.5 points] How does the performance change if you double the number of convolutional layers?

5. [0.5 points] How does the performance change as you increase the number of training samples: [0.6K, 1.8K, 6K, 18K, 60K]? Explain the trends in classification accuracy that you observe.

   Note 1: Make sure that all classes are represented equally within different subsets of the training sets.

6. [1 point] Replace the CNN model with a 2 layer TransformerEncoder. Using a ViT style prediction scheme, evaluate classification accuracy when training with 6K and 60K images. How do the results compare against CNNs? Explain the trends.

2 Submission

We recommend that you submit a report as a single jupyter notebook with relevant cell outputs as mentioned at the top. In case you are not using Python, please share code (with instructions on how to execute) and a separate pdf report.

Submit the file in the courses / moodle portal before the deadline: **8th Mar 2024 23:59 IST**. The moodle portal may show a different date due to the grace period, do not get confused.

The report/notebook should contain:

- A description of the problem, algorithms, results and comparison of the calibration methods based on the experiments you performed.
- Code for calibration that you wrote.
- Images of the inputs and outputs for the different algorithms and cameras.
- Challenges you faced and learnings from the experiments.

Remember, you are expected to write the complete code for the assignment yourselves. DO NOT COPY ANY PART FROM ANY SOURCE not limited to your friends, seniors or the internet.