

# CS5242 August 2023, Assignment 3

Due week 11, 31 October 2023, 1800hours

## 1 Task

Implement the code and run experiments for this paper: WEAKLY SUPERVISED CLUSTERING BY EXPLOITING UNIQUE CLASS COUNT, Oner et al <https://openreview.net/forum?id=B1xIj3VYvr>

## 2 Requirements

### 2.1 Reproduce Results on CIFAR10 (7 Marks)

Use full CIFAR10 dataset 40,000 for training, 10,000 for validation, to generate testing results for JS divergence, UCC and clustering as in Table 1 (on Page 7) of the paper. The splitted dataset is provided in *splitted\_cifar10\_dataset.npz*.

- Step 1: Create the dataloader for CIFAR10, and explain your choices for augmentation, normalization, and etc. (1 mark)
- Step 2: Implement autoencoder  $\theta_{\text{feature}} + \theta_{\text{decoder}}$ . (1 mark)
- Step 3: Implement KDE module. (0.5 mark)
- Step 4: Implement distribution regression module  $\theta_{\text{drn}}$ . (0.5 mark)
- Step 5: Realize the losses used in the paper. (1 mark)
- Step 6: Set and explain your optimization strategy (optimizer, learning rate, etc). (1 mark)
- Step 7: Organize your above implementations to realize training, validation and test processes. (1 mark)
- Step 8: Evaluate your model's JS divergence, ucc accuracy and clustering accuracy. (1 mark)

### 2.2 Reporting results and Analysis (5 Marks)

Plot all your metrics (e.g. training and validation loss and many more) during training for analysis, and explain them. (marks given will depend on clarity of explanation, criticality of analysis)

### 2.3 Improvement (3 Marks)

Devise a way to improve this algorithm and explain clearly what are your innovations. (1 mark: clear motivation and introduction of the improved method; 1 mark: improved accuracy; 1 mark: novelty)

### 2.4 Report format

Limit is 3 pages. In your report, it should include at least four sections: introduction, your improved method, implementation details (including architecture, explanation of your choices, etc), experiment results (including reproduced and improved results, analysis on metrics, etc). **Reports exceeding page length limit will not be graded!**

### 3 Submission

Make a zip file with the following folder structure:

```
YourNUSNETID_YourName
├── report.pdf
└── code/
    ├── main.ipynb
    └── ..
```

- main.ipynb is your code file, which should keep the cell outputs.
- Your main.ipynb should be reproducible.
- report.pdf is your report file.
- The authors have open-sourced their code at <http://bit.ly/uniqueclasscount>. You can check their code to better understand the algorithm, but you should reimplement the algorithm yourself.
- DO NOT copy the code from the internet, e.g. GitHub.
- You can post your questions on the Canvas/Discussions/Assignment3.
- Name the zip file as YourNUSNETID\_YourName.zip, and upload it to LumiNUS before deadline.