## Proposal of Example Perplexity and Class Label Cleaning

QIU Yaowen, 20784389, [yqiuau@connect.ust.hk](mailto:yqiuau@connect.ust.hk)

1. **Project Description**

For Deep Neural Networks (DNNs), some images are easier to classify than others. The term perplexity is used to describe the extent of difficulty of classifying. In previous work, C-perplexity and X-perplexity are proposed to measure the image perplexity quantitatively. In brief, for a specific example **x**, C-perplexity means the average uncertainty of the population of classifiers when classifying **x**, X-perplexity means the fraction of the classifiers that misclassifies **x**. For a class, the C/X-perplexity is defined to be the average of the C/X-perplexity over all the examples belong to the class. The paper believes that attention confusion and class confusion are two main factors that make an example difficult to classify; visual similarity and class co-occurrence are another two factors that make class perplexity high.

An intuitive idea is to control the level of perplexity of training examples, for example, filter those examples with perplexity larger than a certain threshold. Training examples with relatively lower complexity should lead to better classification result. Therefore, the project will investigate on the effect of reducing the complexity of the training examples by comparing the classification performance (e.g., precision, recall) on validation examples. Furthermore, the project should try to analyze the experiment data and explain the experimental result.

1. **Final Report, Demo, and Presentation**

The final report should contain following parts:

* The motivation/background of the project
* Literature review or concept explanation
* Methodology
* Experiment
* Dataset (Definition of training dataset and validation dataset)
* Models (VGG, ResNet, Inception, etc.)
* Parameters (Loss function, epochs, batch size, evaluations metrics, etc.)
* Experimental Result & Analysis of data
* Conclusion

The demo should contain the whole procedure of the project with clear structure. The Presentation should demonstrate the whole project in 10-15 minutes to the supervisor.

1. **Workload and Workload and Grading Criteria**

This project is regarded as a standard 3-credit project. Final grade depends on final project report, demo, and presentation. The contribution to the final grade should 40%, 30%, and 30%, respectively.