Final Project (40 points)

Due on: 11 pm. 24th, November, 2019

Late policy: late submission will **not** be marked (no matter in what reason)!!!

# Marking regulation.

1. Implement our algorithms in Keras, Tensorflow, or Matlab environments. If codes are implemented in another environment, no points will be given.
2. Students will be required to make their group to complete the final project. Each group could have up to **two** students. However, **Thesis-based masters students should complete the final project individually.**
3. **Thesis masters students** could use any machine learning algorithms for their own research topics. For these students, research report along with the codes, at least draft version of the research report, should be submitted to D2L.
4. There will be 10 points for the project presentation (15 min for a group). The group which makes a qualified presentation and accepted project report will receive the full 10 points. The final project report should include source codes, method descriptions, as well as the experimental results.
5. The presentation time is set in the last three classes. (November of 25th, November of 27th, and December of 2nd).
6. GAs and instructor will evaluate all the codes. If the project codes fail to be repeated, 30 points will be lost.
7. There are 30 points for the project results. The detailed marking rule is as below.

* Each group has to select and complete one of the three tasks. For each dataset. The top performance of the group will receive the full 100 percentage points; the second top performance of the group will receive 90 percentage points; the third top performance of the group will receive 85 percentage points; the fourth top performance of the group will receive 80 percentage points; the fifth top performance of the group will receive 75 percentage points; other remaining groups will **maximal** receive 60 percentage points.
* All the experimental results should be obtained by at least 3 runs. The experimental results must include average testing accuray.
* The requirement of how to organize training samples and testing samples is mentioned in the following subsection. **Students must follow the requirement to construct the training and testing set. If the training set and testing set are not constructed based on the requirement, the 30 points will be lost.**

1. If the group cannot complete all the tests for the given dataset, points will be cut accordingly.
2. Each group should give Gas and instructor their group information by **20th of November** so we can arrange the order of the presentation.
3. The codes may upload to Moss(https://theory.stanford.edu/~aiken/moss/) to check the similarity. If the similarity rate is above 70 percent, the related final projects will be marked as zero. The instructor will also report it to the Departmental Chair for a further penalty.

# Description

The final project has three classification tasks(problems). Students should complete one of the three tasks(problems).

1. Object-centric Image recognition task (Datasets of Task 1)
2. Scene-centric Image recognition task (Datasets of Task 2)
3. EEG brain signal dataset for emotion prediction (Datasets of Task 3)

# Required Databases

* Datasets of Task 1: CIFAR10, Caltech101, Caltech256, CIFAR100
* Datasets of Task 2: SUN397
* Datasets of Task 3: EEG Brain DE Features (https://github.com/PerforMance308/EEG\_Dataset)

# How to construct the training and testing dataset (**Very important**)

1. For CIFAR10/100, the training set, and testing set were organized.
2. For Caltech101/256, 30 image per category will be selected to construct the training dataset. The remaining images of the datasets will be construct as a testing dataset.
3. For SUN397 datasets, please randomly split the datasets into a training dataset and a testing dataset, each with 50 images per class.
4. For DE features from SEED dataset, the detailed steps are shown in <https://github.com/PerforMance308/EEG_Dataset>

# Where to download these datasets

1. For CIFAR10/100, and Caltech101/256, the download link is already provided in D2L.
2. Please found the link (<https://vision.princeton.edu/projects/2010/SUN/>) to download SUN397 dataset.

# How to obtain top-1 accuracy/total training time for each given dataset.

1. Any classification methods are allowed to use to obtain the average top-1 accuracy such as deep convolutional neural networks, RBF network, Extreme Learning Machine, Support Vector Machine, etc.
2. Any feature extraction methods are allowed to use to boost the performance such as PCA, autoencoder, deep believe networks, etc.
3. Any data argumentation algorithms also allowed to use to further boost the performance.
4. The training time should be included all the training time provided by the used classification methods, feature extraction methods, and data argumentation algorithms.

# How to obtain average top-1 accuracy for each given task.

For example, group A selects object-centric Image recognition task. In this case, they obtained four average testing accuracies: 88 percent for CIFAR10, 65 percent for CIFAR100, 85 percent for Caltech256 and 91 percent for Caltech101. Hence the average top-1 accuracy of object-centric Image recognition task is (88+65+85+91)/4=82.25.

# Useful Resource

1. Places365 pretrained DCNN models: <https://github.com/CSAILVision/places365>
2. DenseNet pretrained models: <https://github.com/liuzhuang13/DenseNet>
3. Deep Learning in Matlab: <https://www.mathworks.com/solutions/deep-learning.html>
4. Pretrained DCNN model in MATLAB: <https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html>