Spring 2018	Report#6 - 02/11/2018	Tianxiao Zhang

I. Tasks achieved Last Week (***, **, *: order of priority)

- Project: Content-Adaptive Saak Transform

Purpose: Using content-adaptive method to further improve the accuracy of the MNIST test data.

Method: First, we should do the original Saak transform and get the class probability for each test images.

Then we compare the top 1 class probability of top 2 class probability and calculate the ratio of these two number. If the ratio is larger than the threshold, then we use the class of the maximum probability as the test label and the process is done. If the ratio is less than or equal to the threshold, then we should do the second stage Saak transform to further compare these two confusion classes.

Note: The time cost for running the codes should be controlled because what we want to do is beat CNN, not just running the codes without considering the time. If the time cost is long, the method is not that good.

Slogan: Beat CNN!

II. Feedback and Interaction

- ***Prof. Kuo's feedback***
    -
- ***Discussion***

III. Report

This week, I did further comparison of the confusion images. At first, I utilized the mean of the coefficients for each class to compare which coefficients were sufficient to separate the confusion classes. But it seemed not enough. Then a new method came out: K-L divergence.

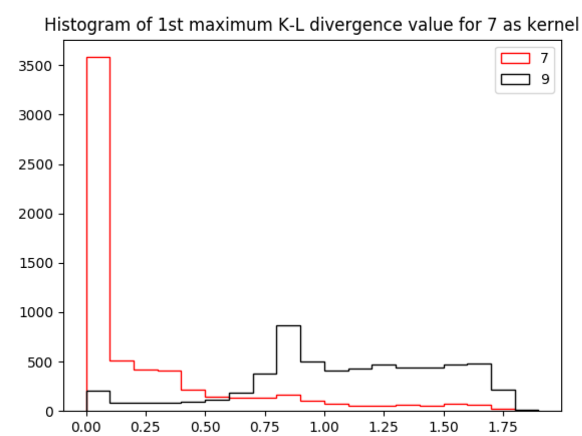
1. Using K-L divergence to separate confusion images

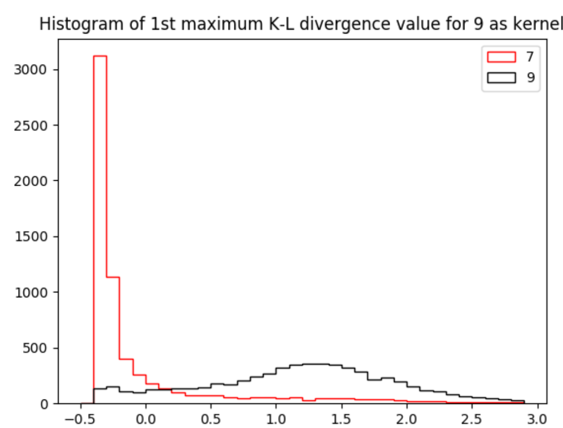
Suppose the density distributions of the coefficients for two confusion classes are P and Q.

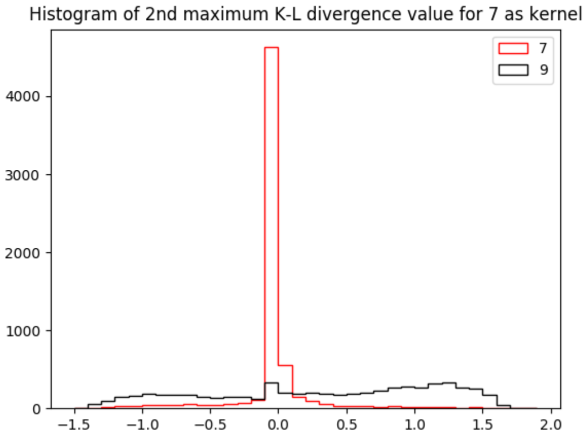
The equation of K-L divergence is:

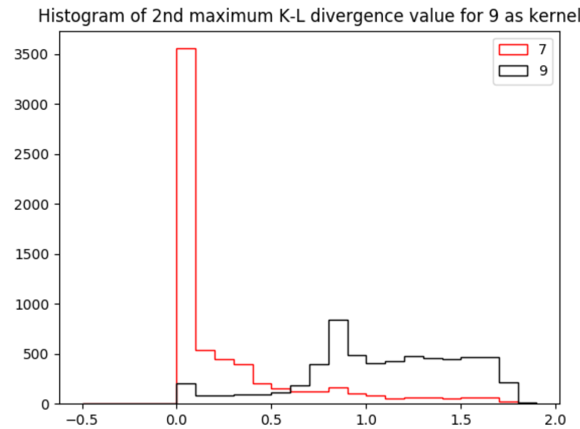
$$D_{KL}(P||Q) = \sum_i P(i) log \frac{P(i)}{Q(i)}$$

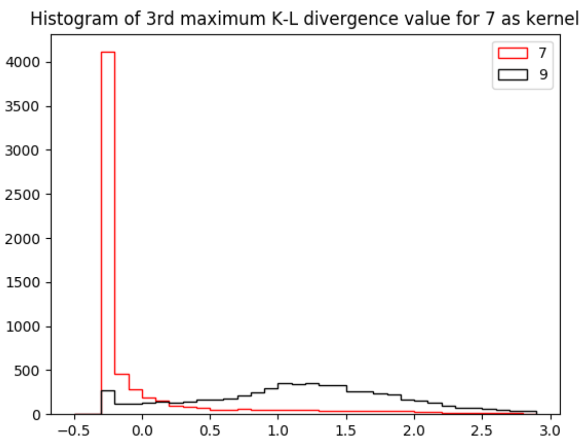
We can calculate the differences of the coefficients for two confusion classes. Suppose the confusion classes are “7” and “9”, we can easily get the distribution of the coefficients of top 10 maximum K-L divergence value.

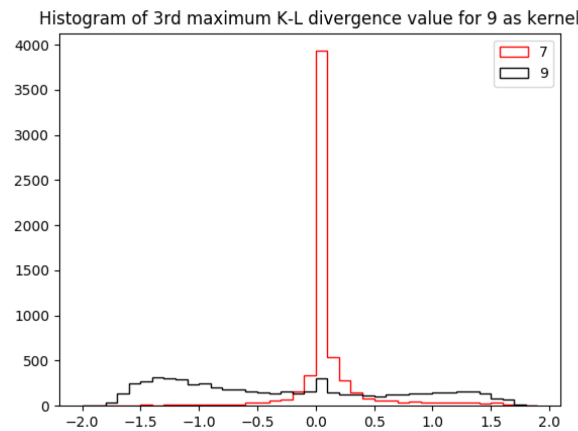


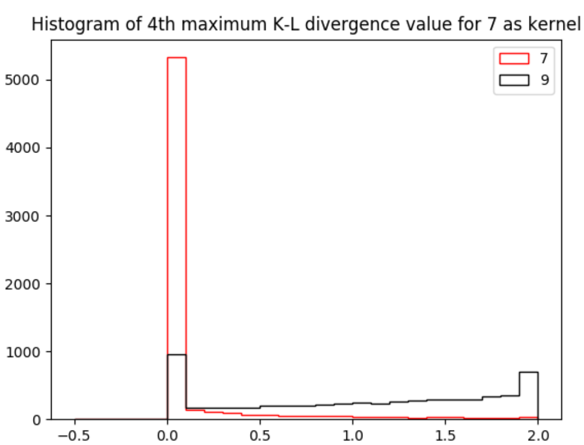


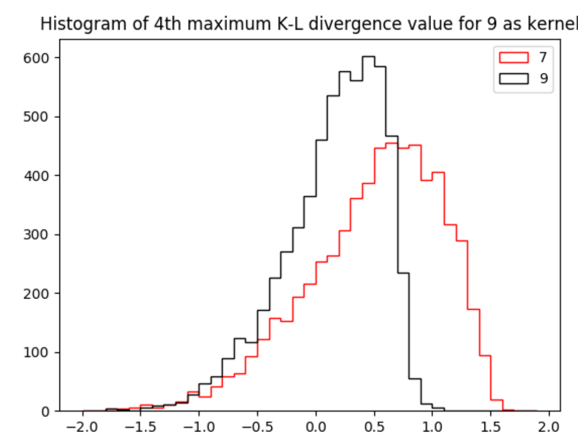

Histogram of 2nd maximum K-L divergence value for 7 as kernel
7
9

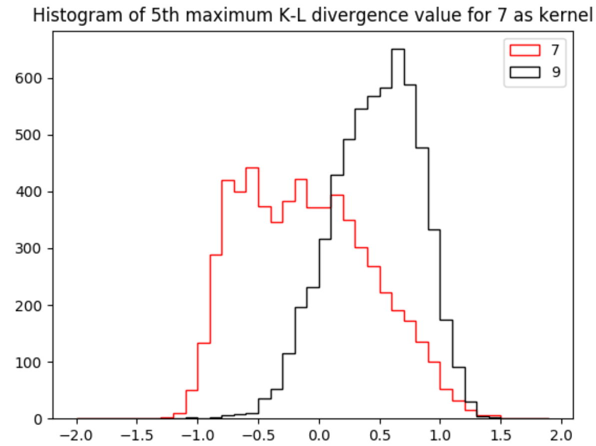

Histogram of 2nd maximum K-L divergence value for 9 as kernel
7
9

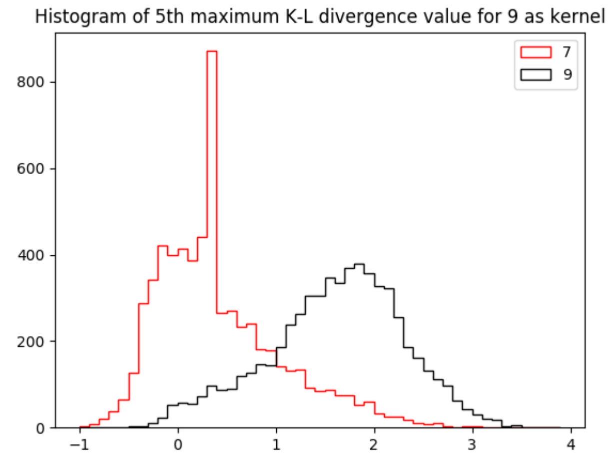

Histogram of 3rd maximum K-L divergence value for 7 as kernel
7
9

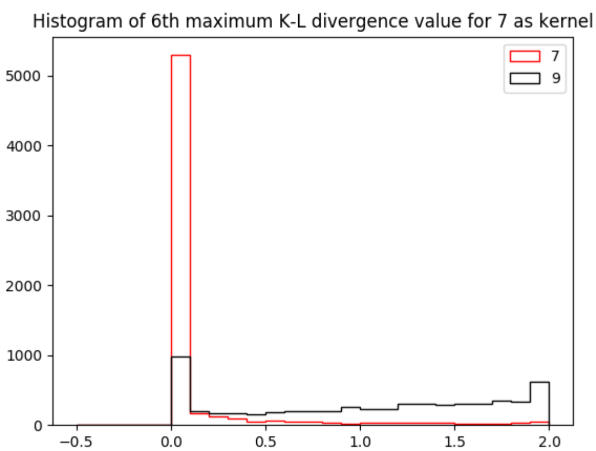

Histogram of 3rd maximum K-L divergence value for 9 as kernel
7
9

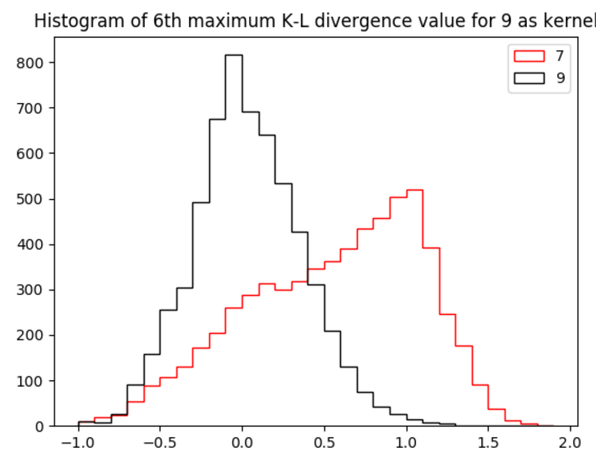

Histogram of 4th maximum K-L divergence value for 7 as kernel
7
9

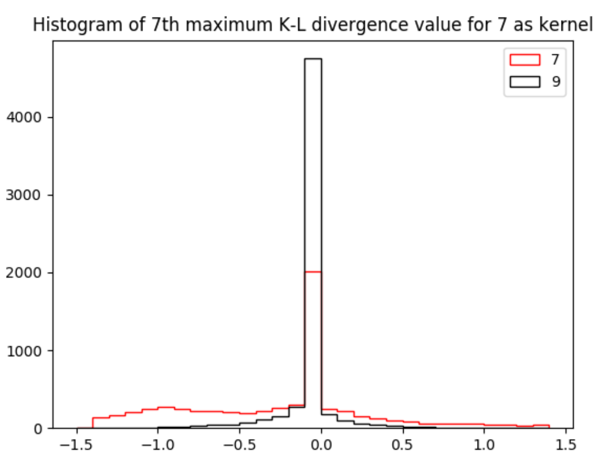

Histogram of 4th maximum K-L divergence value for 9 as kernel
7
9

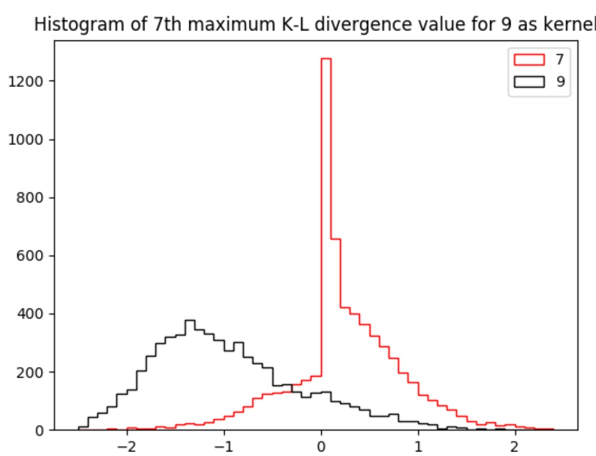

Histogram of 5th maximum K-L divergence value for 7 as kernel
7
9

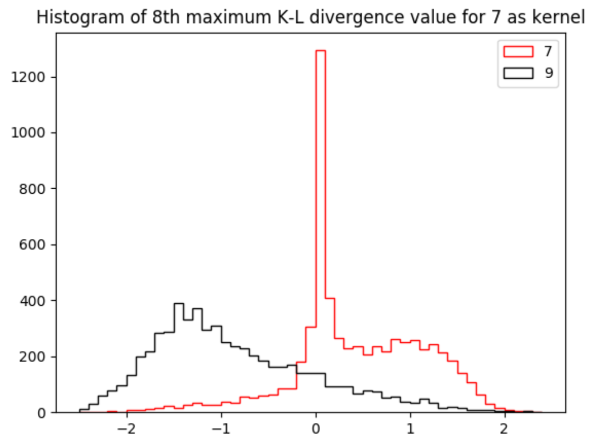

Histogram of 5th maximum K-L divergence value for 9 as kernel
7
9

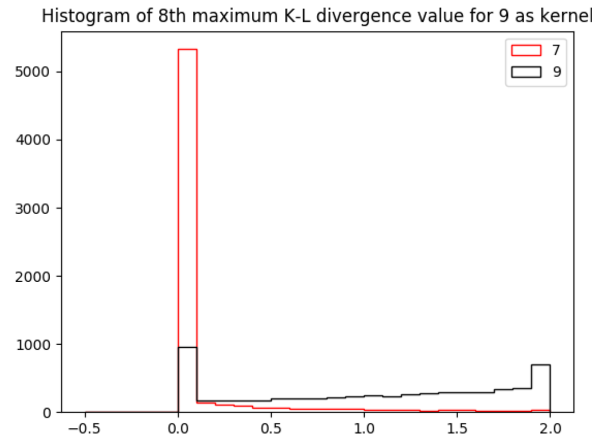

Histogram of 6th maximum K-L divergence value for 7 as kernel
7
9

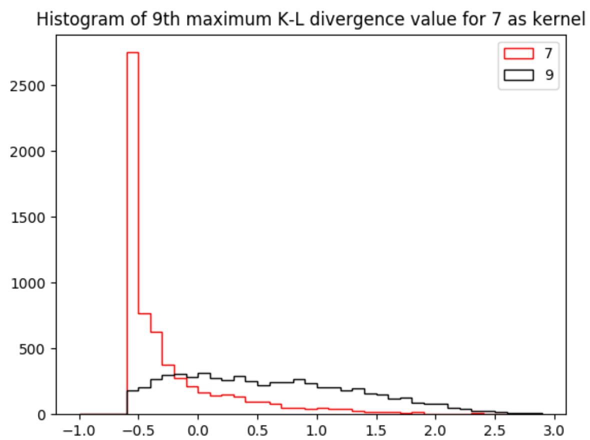

Histogram of 6th maximum K-L divergence value for 9 as kernel
7
9

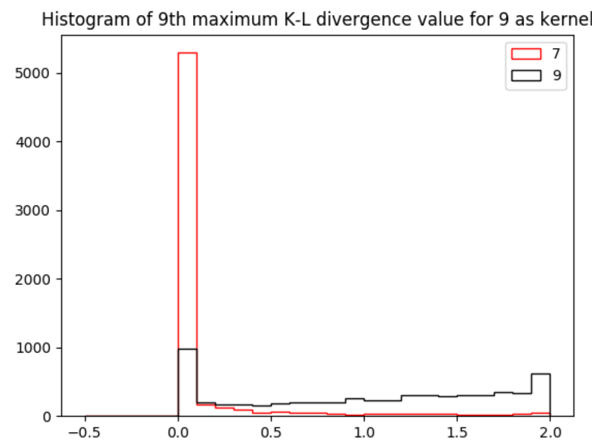

Histogram of 7th maximum K-L divergence value for 7 as kernel
7
9

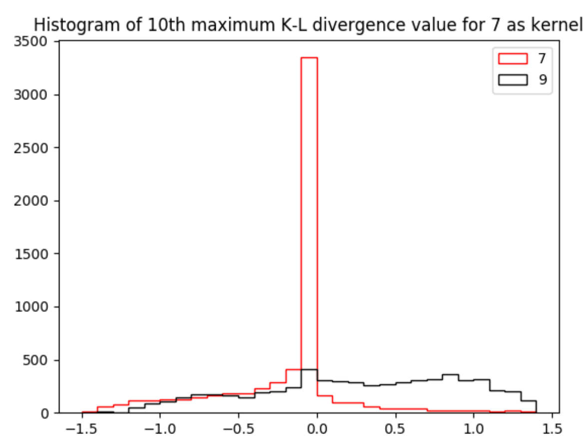

Histogram of 7th maximum K-L divergence value for 9 as kernel
7
9

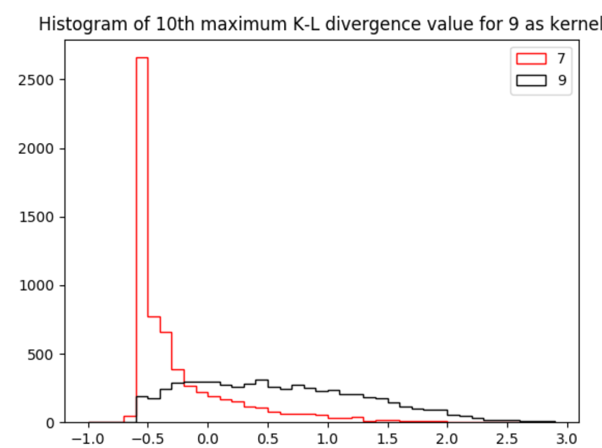

Histogram of 8th maximum K-L divergence value for 7 as kernel
7
9


Histogram of 8th maximum K-L divergence value for 9 as kernel
7
9


Histogram of 9th maximum K-L divergence value for 7 as kernel
7
9


Histogram of 9th maximum K-L divergence value for 9 as kernel
7
9


Histogram of 10th maximum K-L divergence value for 7 as kernel
7
9


Histogram of 10th maximum K-L divergence value for 9 as kernel
7
9

Even though the distribution is much better than that of using the mean of coefficients, the distribution of confusion classes still has much overlap. And the I can only improve 13 confusion test images at most under this method.

2. Using entropy to separate the confusion classes

An improvement of the method above is that we calculate the entropy of coefficients for each class and compare the entropy for the confusion classes to select some good ones to separate the confusion images. However, the results were almost the same as that using the method above.

3. Future Work

As we know, we only have 150 wrong labeled test images and maybe these images may have coefficients in these overlapping areas. If their coefficients do not in the overlapping areas, they may be right labeled in the first stage. For example, if "7" is wrong labeled as "9", then the coefficients for that specific "7" image may have most coefficients in the overlapping area, so that specific "7" image is wrong labeled as "9". Thus only by selecting good coefficients seems not enough because most of the coefficients are in the overlapping areas, even we select relatively good coefficients, these coefficients still cannot distinguish the two labels.

All in all, the reason for the wrong labeled image is that the coefficients generated by the specific image are mostly like the wrong class instead of its class, so just selecting relatively good coefficients seems not enough because the 150 images are tough to recognize, and some of them are unrecognizable through my eyes. If we can recognize them through selecting good coefficients, then they will not be labeled wrong in the original Saak transform.

In the future, we need to analysis the coefficients of the wrong labeled 150 images and further see what happened to these coefficients. If the coefficients of these wrong labeled images do not have similar properties to the label themselves, then we need to focus on how to generate coefficients which can distinguish these confusion classes.

References

[1] C.-C. Jay Kuo, “Understanding convolutional neural networks with a mathematical model,” the Journal of Visual Communications and Image Representation, Vol. 41, pp. 406-413, November 2016.
[2] C.-C. Jay Kuo, “The CNN as guided multi-layer RECOS transform,” the IEEE Signal Processing Magazine, Vol. 34, No. 3, pp. 81-89, May 2017.
[3] C.-C. Jay Kuo and Yueru Chen, “On data-driven Saak transform,” arXiv preprint arXiv: 1710.04176 (2017).

IV. Plan for the next week (***, **, *: order of priority)

- Analysis the wrong labeled images.

V. Milestone