
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, first I explored the coefficients with the maximum 10 K-L divergence values and the confusion test images and plotted them on the same graph to see if the selected coefficients of the confusion test images located in the overlapping areas of the two confusion classes. Then I explored the coefficients of the last stage of the Saak transform.

1. Investigating the coefficients of confusion test images

It seemed that using K-L divergence values can improve the performance of the Saak transform and the results showed that we can improve 13 test images and get the test accuracy of more than 0.986.

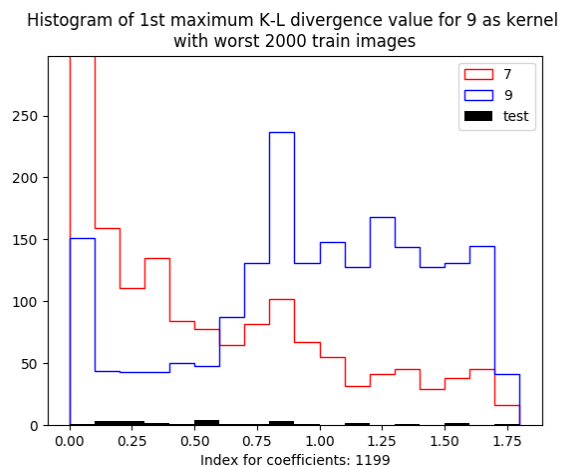
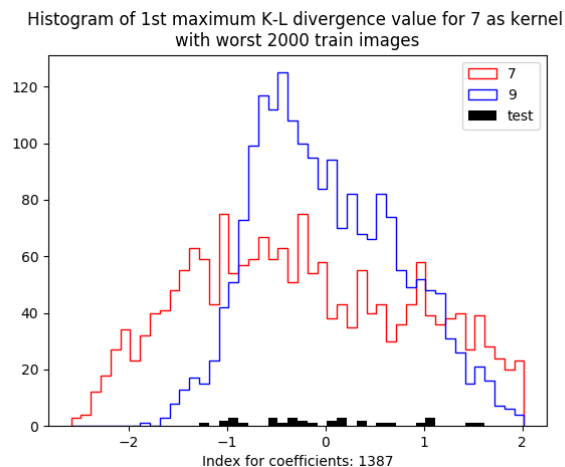
However, it seemed not enough because what we wanted is to get the test accuracy of more than 0.99 or even 0.995. Thus, the improvement we got seemed trivial. So we may have the question: why using K-L divergence values cannot improve the performance that well?

My proposal is that even though we selected the coefficients with the high K-L values, the selected coefficients of the confusion test images located in the overlapping areas of the two confusion classes, thus the performance was not that good enough.

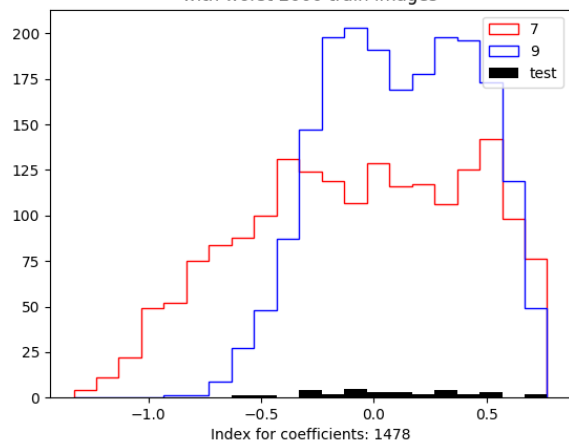
In this experiment, I utilized the highest 2000 probability for each class as the kernel to train, the lowest 2000 probability for each class as the kernel to train, the highest 1000 probability and the lowest 1000 probability for each class as the kernel to train, and all images for each class as the kernel to train.

I still used the “7” and “9” as the confusion classes and selected some confusion test images for “7” and “9” to do the experiment.

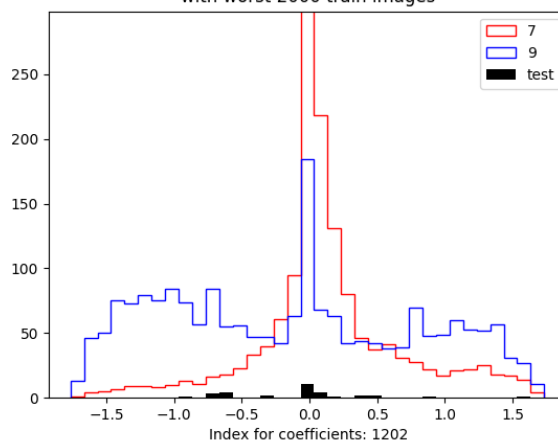
The lowest 2000 probability for each class as the kernel



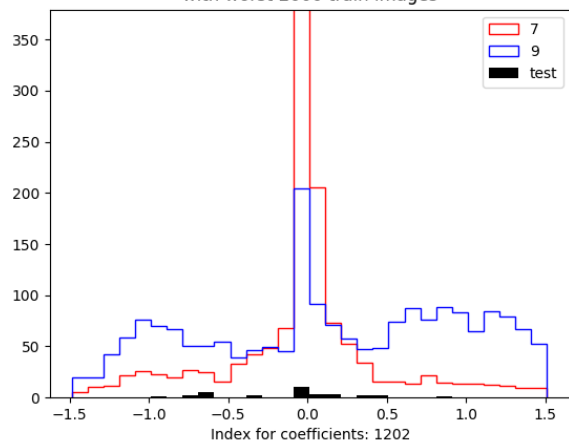
Histogram of 2nd maximum K-L divergence value for 7 as kernel with worst 2000 train images



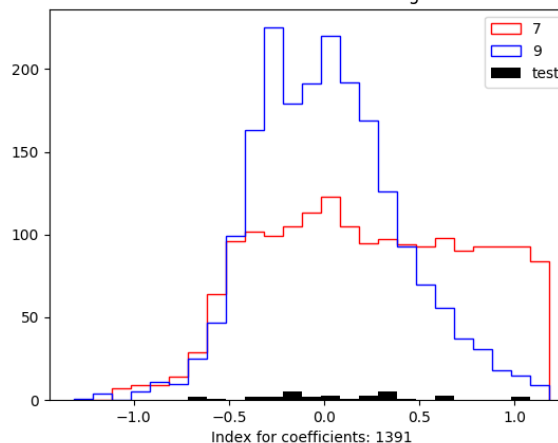
Histogram of 2nd maximum K-L divergence value for 9 as kernel with worst 2000 train images



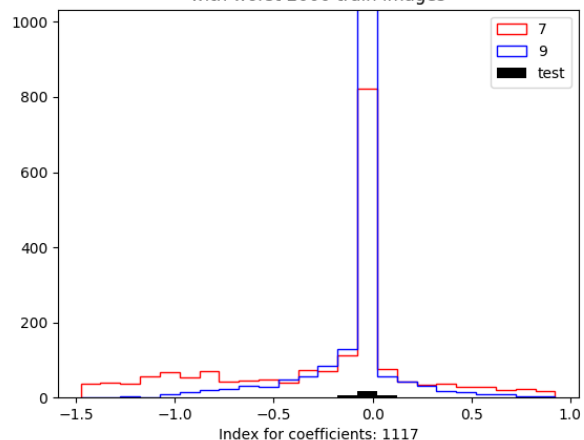
Histogram of 3rd maximum K-L divergence value for 7 as kernel with worst 2000 train images



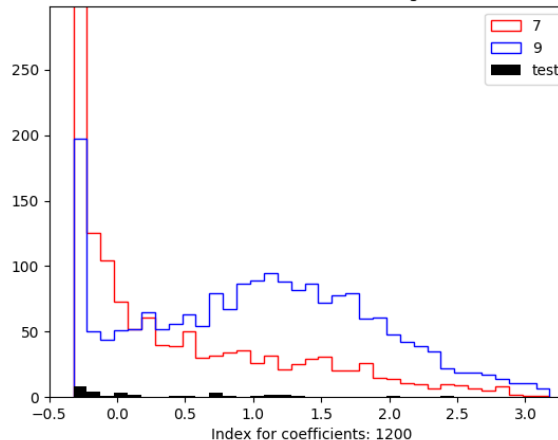
Histogram of 3rd maximum K-L divergence value for 9 as kernel with worst 2000 train images



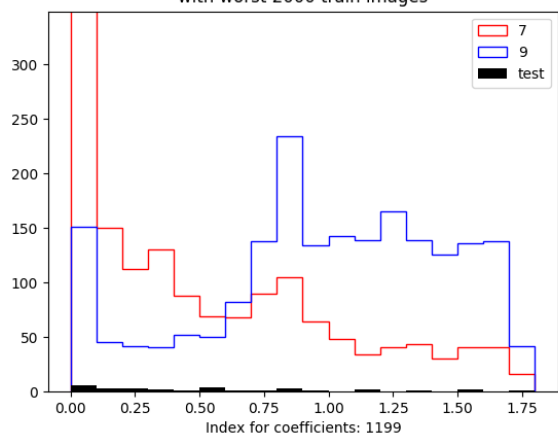
Histogram of 4th maximum K-L divergence value for 7 as kernel with worst 2000 train images



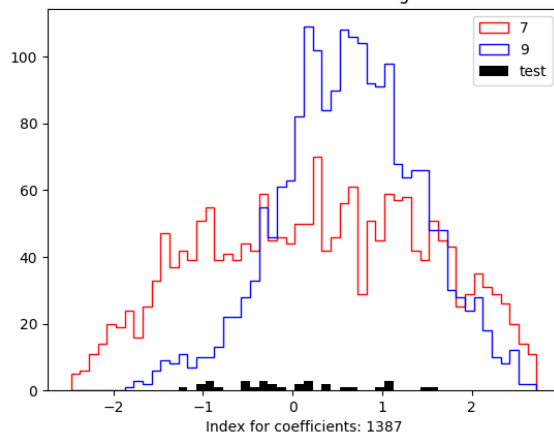
Histogram of 4th maximum K-L divergence value for 9 as kernel with worst 2000 train images



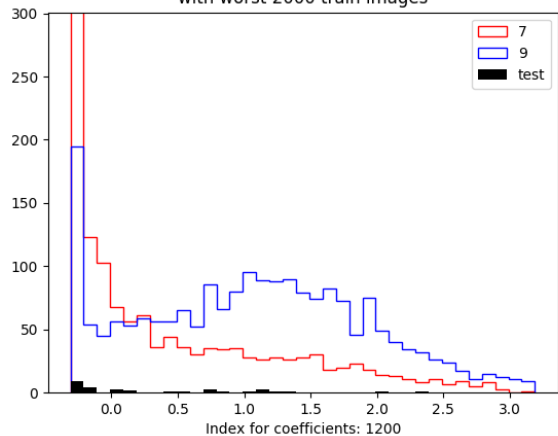
Histogram of 5th maximum K-L divergence value for 7 as kernel with worst 2000 train images



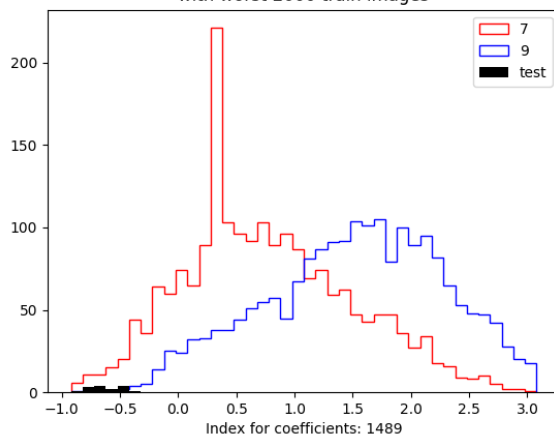
Histogram of 5th maximum K-L divergence value for 9 as kernel with worst 2000 train images



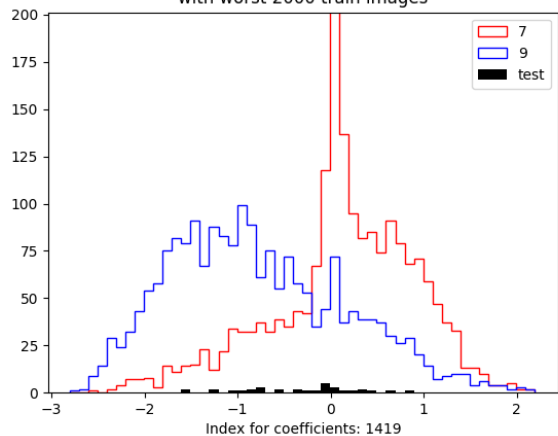
Histogram of 6th maximum K-L divergence value for 7 as kernel with worst 2000 train images



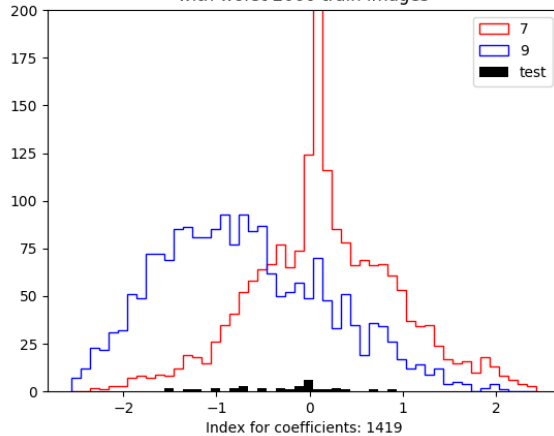
Histogram of 6th maximum K-L divergence value for 9 as kernel with worst 2000 train images



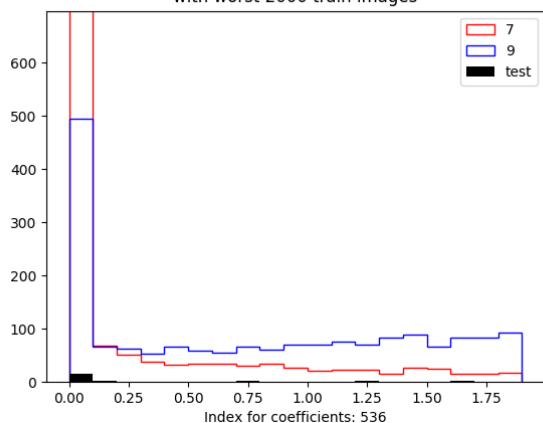
Histogram of 7th maximum K-L divergence value for 7 as kernel with worst 2000 train images



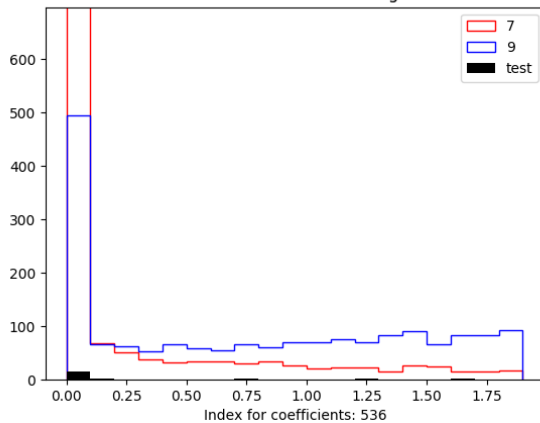
Histogram of 7th maximum K-L divergence value for 9 as kernel with worst 2000 train images



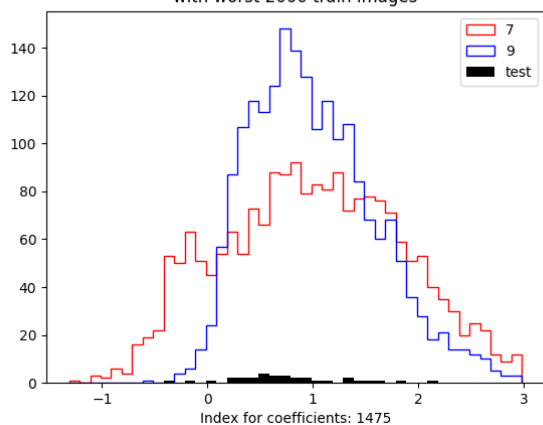
Histogram of 8th maximum K-L divergence value for 7 as kernel with worst 2000 train images



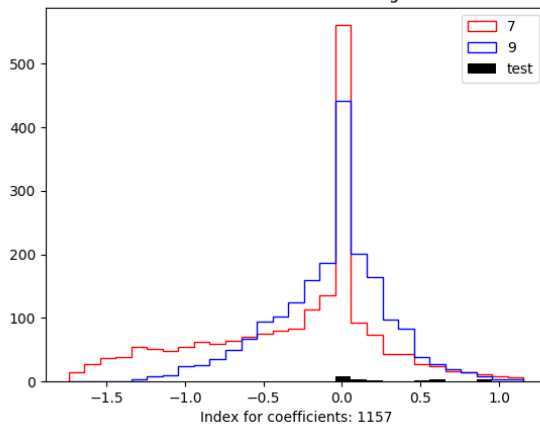
Histogram of 8th maximum K-L divergence value for 9 as kernel with worst 2000 train images



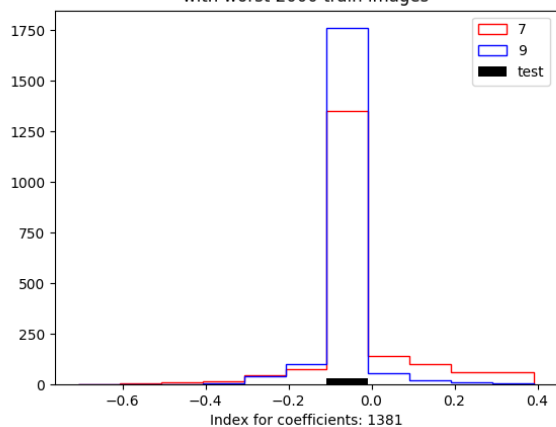
Histogram of 9th maximum K-L divergence value for 7 as kernel with worst 2000 train images



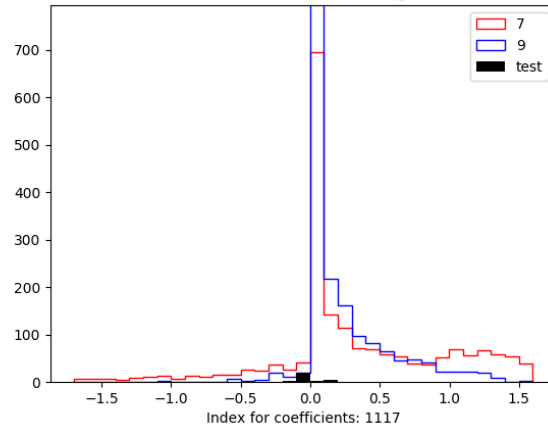
Histogram of 9th maximum K-L divergence value for 9 as kernel with worst 2000 train images



Histogram of 10th maximum K-L divergence value for 7 as kernel with worst 2000 train images

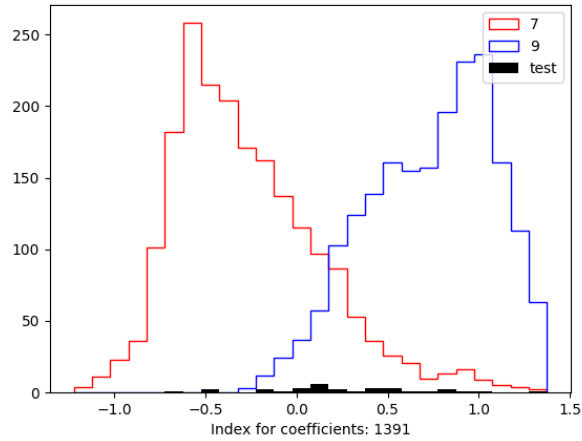


Histogram of 10th maximum K-L divergence value for 9 as kernel with worst 2000 train images

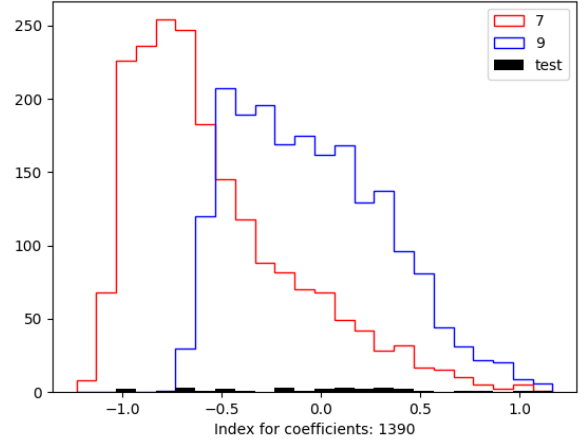


The highest 2000 probability for each class as the kernel

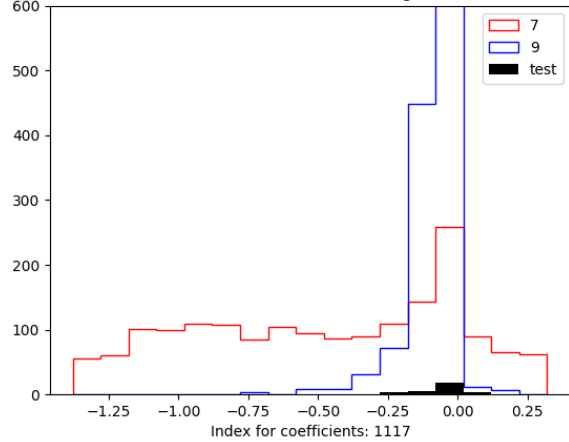
Histogram of 1st maximum K-L divergence value for 7 as kernel with best 2000 train images



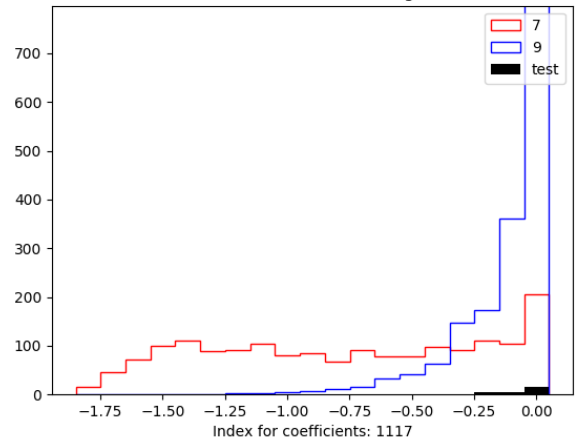
Histogram of 1st maximum K-L divergence value for 9 as kernel with best 2000 train images



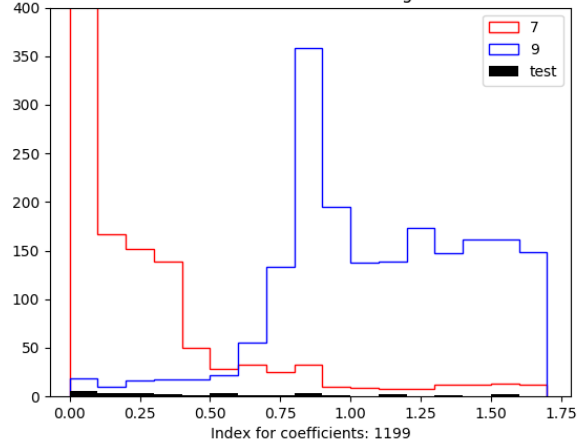
Histogram of 2nd maximum K-L divergence value for 7 as kernel with best 2000 train images



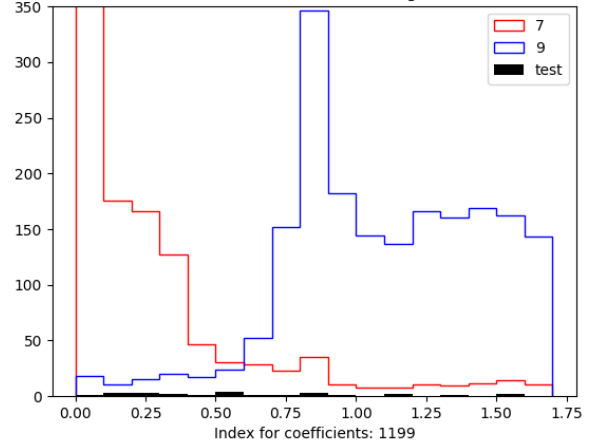
Histogram of 2nd maximum K-L divergence value for 9 as kernel with best 2000 train images



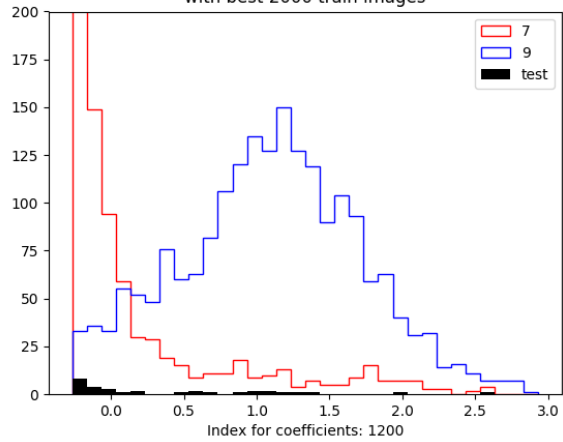
Histogram of 3rd maximum K-L divergence value for 7 as kernel with best 2000 train images



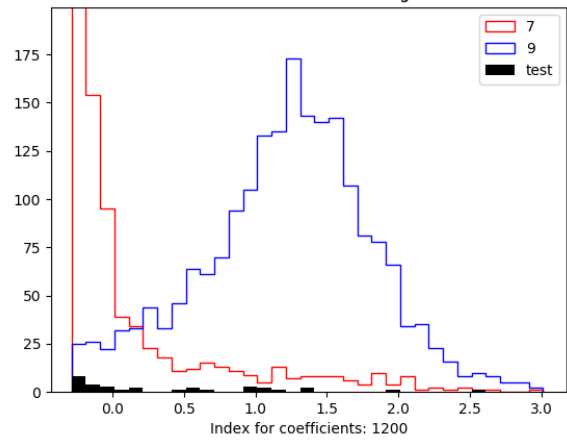
Histogram of 3rd maximum K-L divergence value for 9 as kernel with best 2000 train images



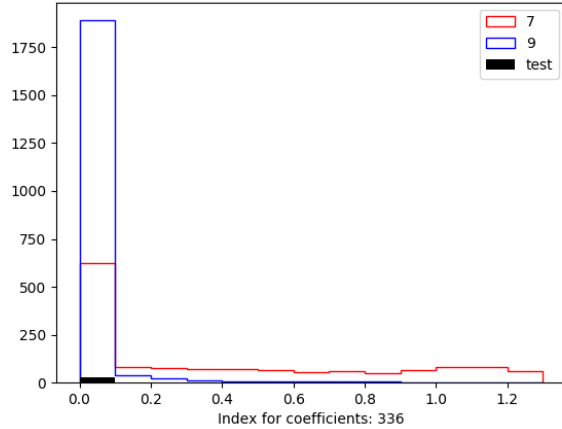
Histogram of 4th maximum K-L divergence value for 7 as kernel with best 2000 train images



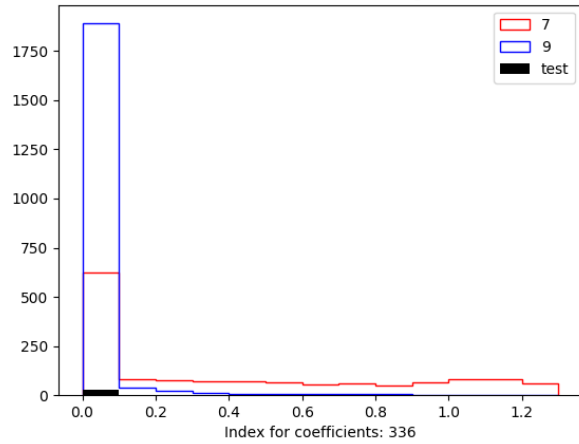
Histogram of 4th maximum K-L divergence value for 9 as kernel with best 2000 train images



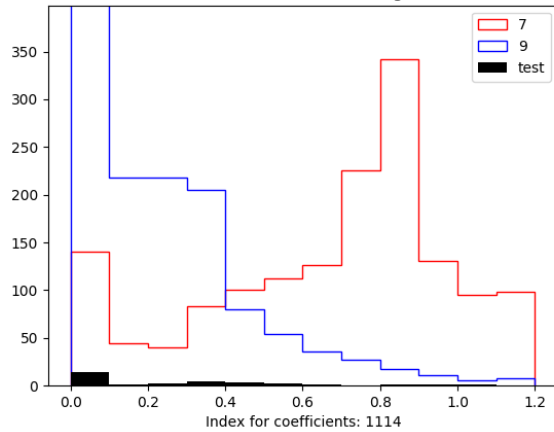
Histogram of 5th maximum K-L divergence value for 7 as kernel with best 2000 train images



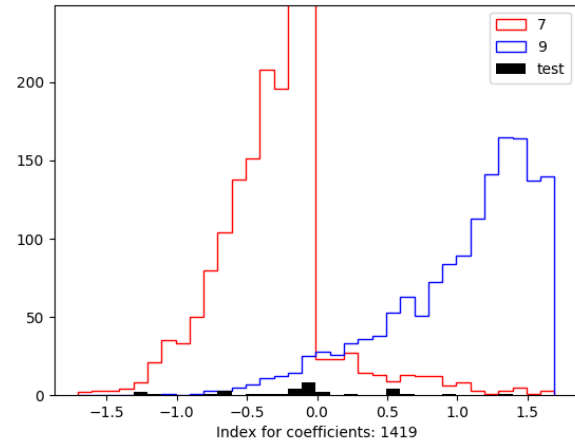
Histogram of 5th maximum K-L divergence value for 9 as kernel with best 2000 train images



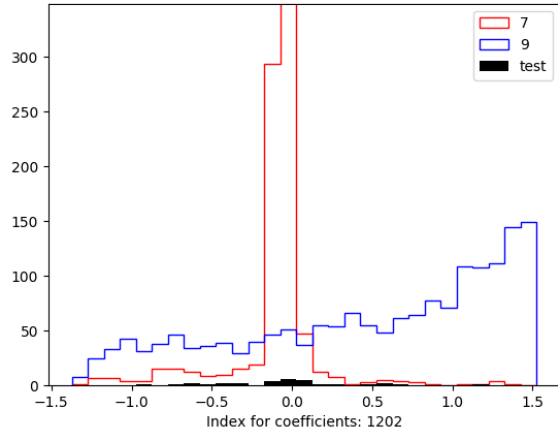
Histogram of 6th maximum K-L divergence value for 7 as kernel with best 2000 train images



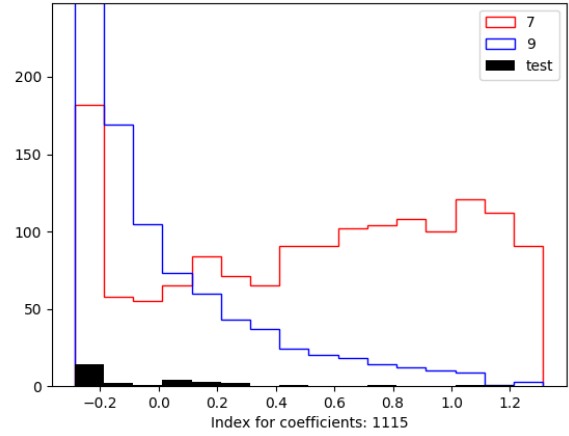
Histogram of 6th maximum K-L divergence value for 9 as kernel with best 2000 train images



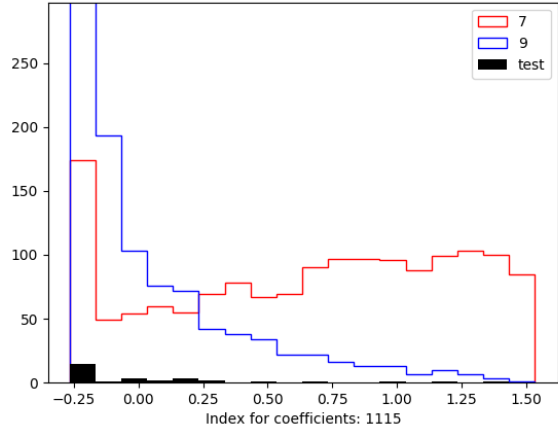
Histogram of 7th maximum K-L divergence value for 7 as kernel with best 2000 train images



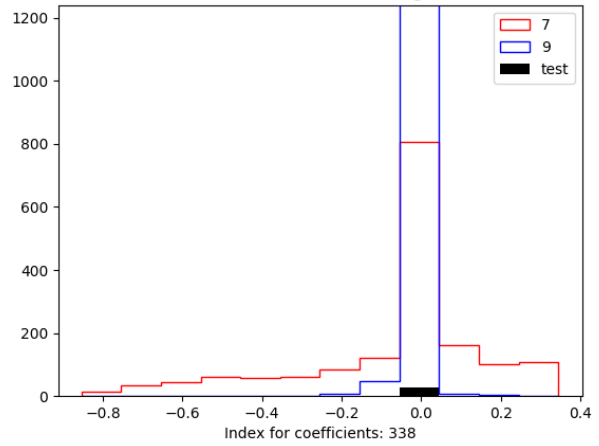
Histogram of 7th maximum K-L divergence value for 9 as kernel with best 2000 train images



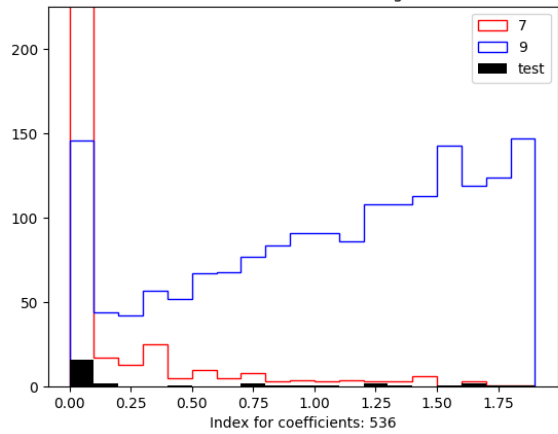
Histogram of 8th maximum K-L divergence value for 7 as kernel with best 2000 train images



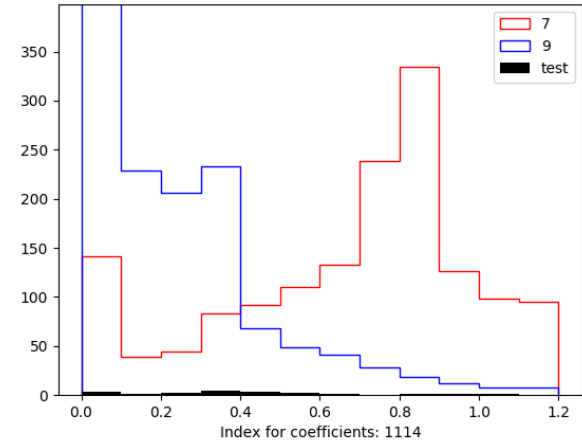
Histogram of 8th maximum K-L divergence value for 9 as kernel with best 2000 train images



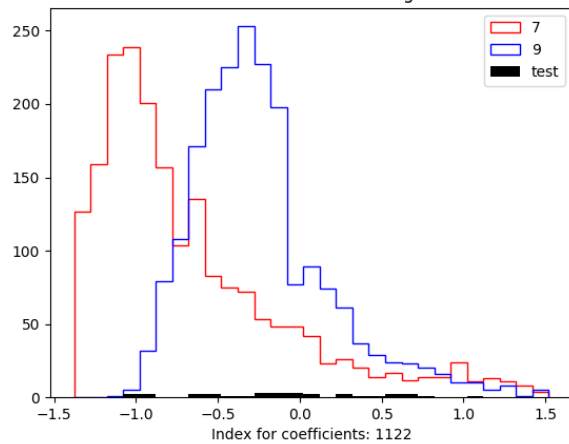
Histogram of 9th maximum K-L divergence value for 7 as kernel with best 2000 train images



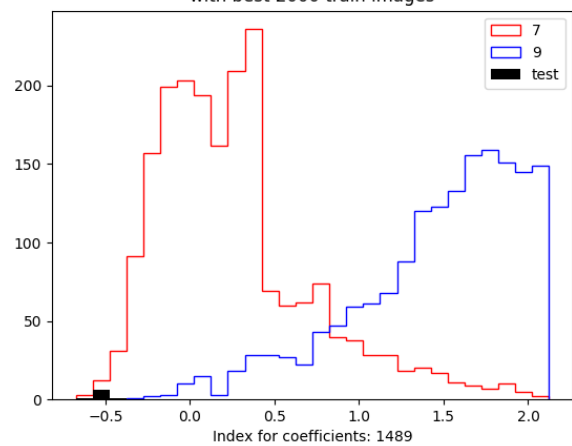
Histogram of 9th maximum K-L divergence value for 9 as kernel with best 2000 train images



Histogram of 10th maximum K-L divergence value for 7 as kernel with best 2000 train images

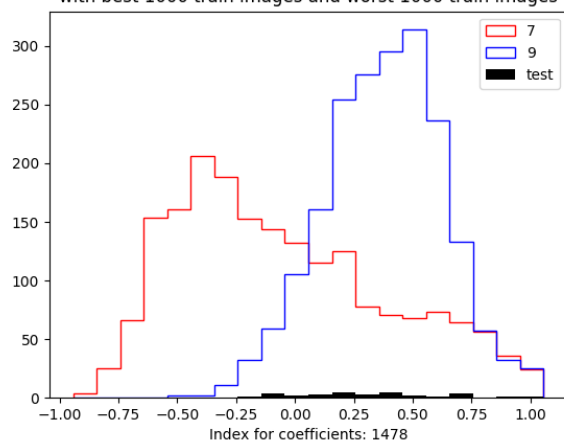


Histogram of 10th maximum K-L divergence value for 9 as kernel with best 2000 train images

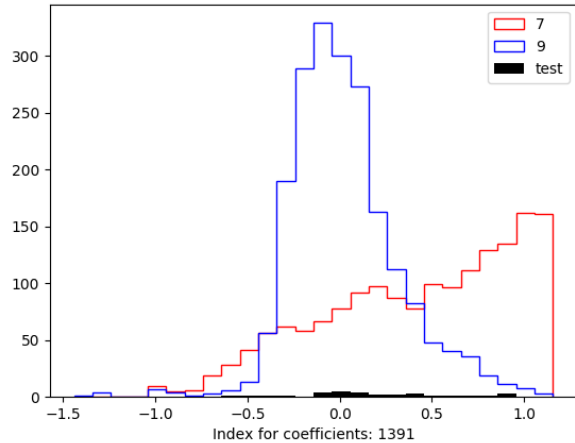


The lowest 1000 probability and the highest 1000 probability for each class as the kernel

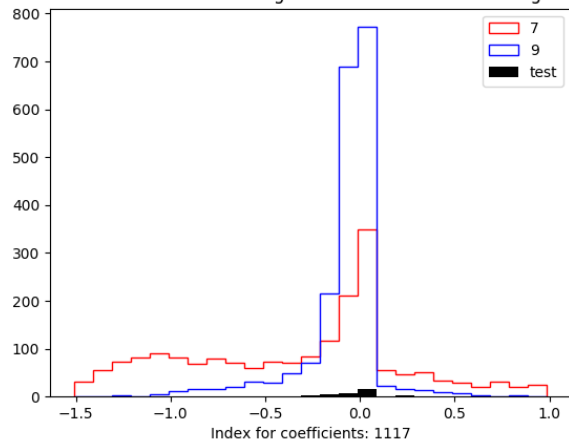
Histogram of 1st maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images



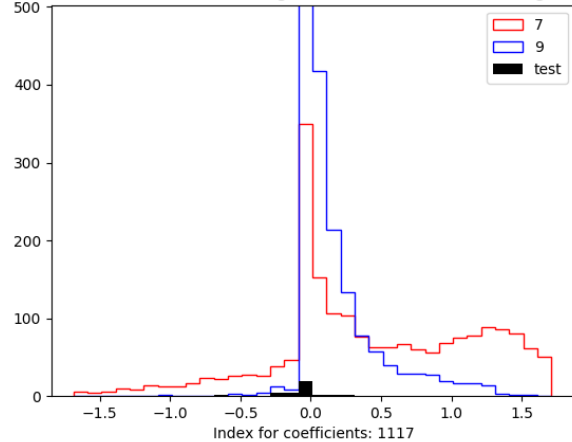
Histogram of 1st maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images



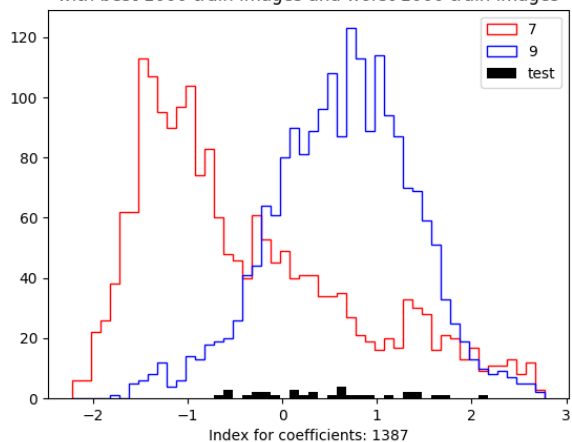
Histogram of 2nd maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images



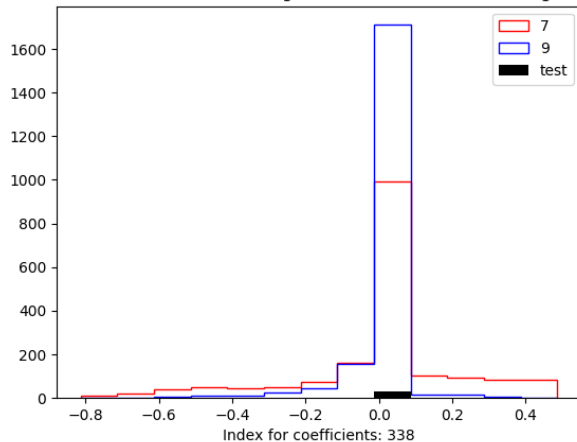
Histogram of 2nd maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images



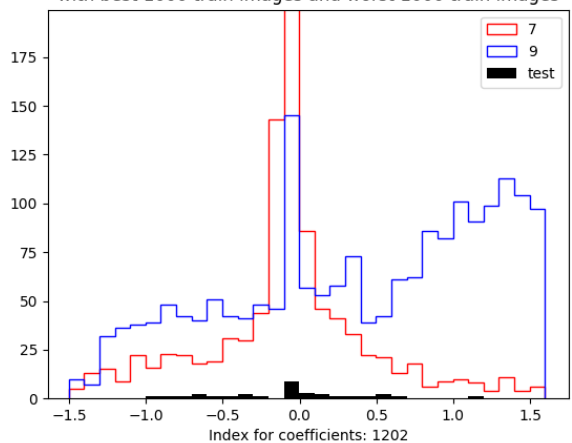
Histogram of 3rd maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images



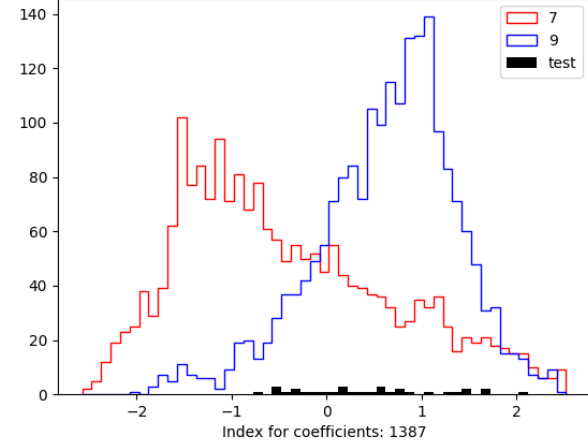
Histogram of 3rd maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images



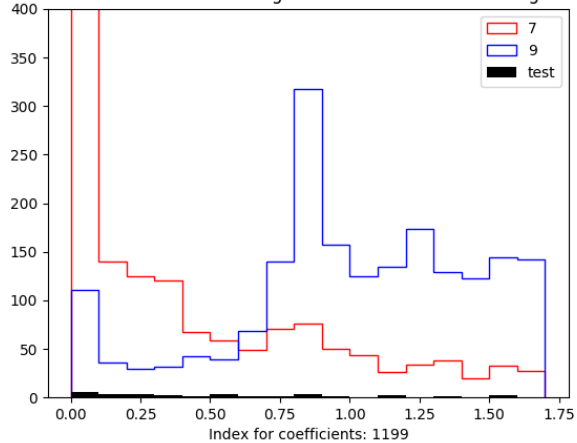
Histogram of 4th maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images



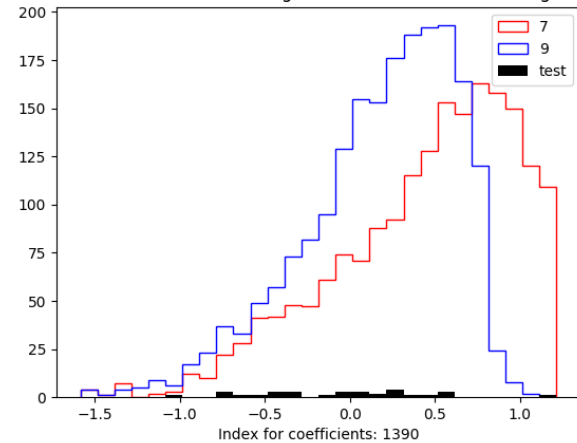
Histogram of 4th maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images



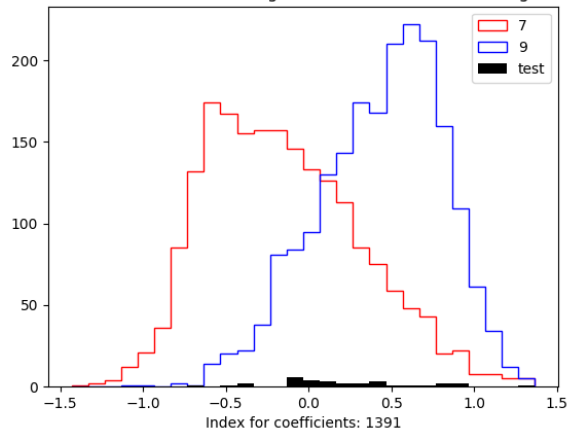
Histogram of 5th maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images



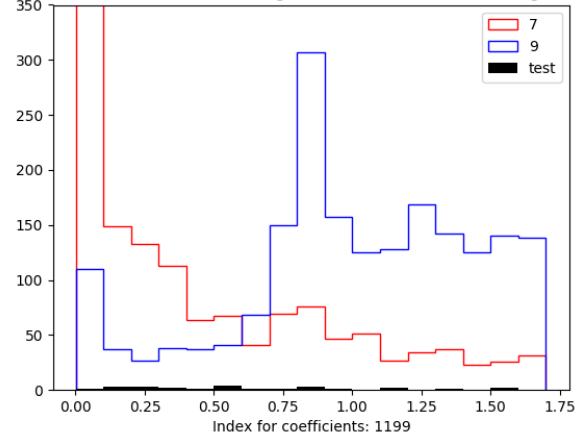
Histogram of 5th maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images



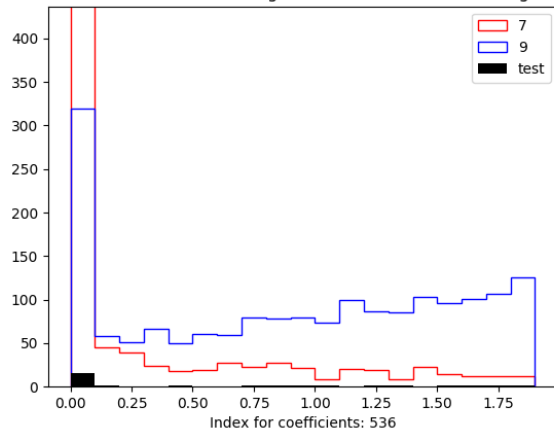
Histogram of 6th maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images



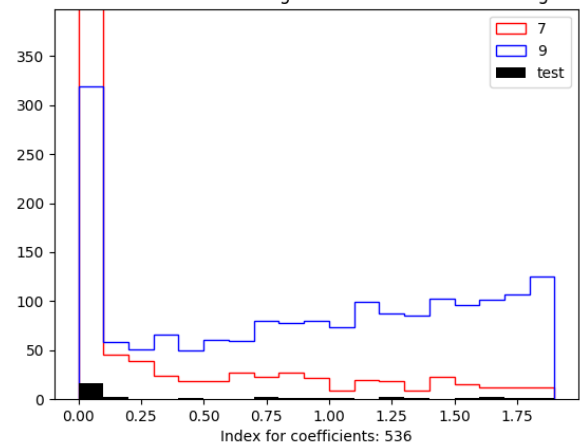
Histogram of 6th maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images



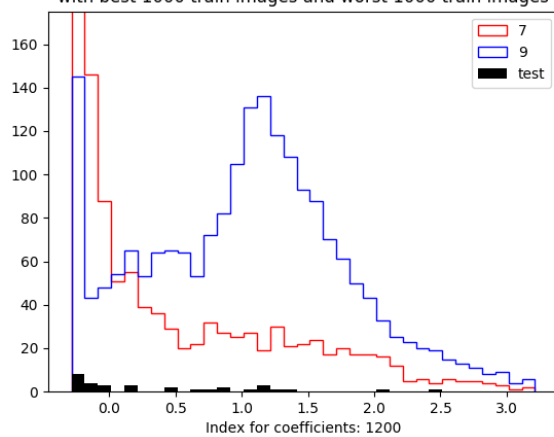
Histogram of 7th maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images



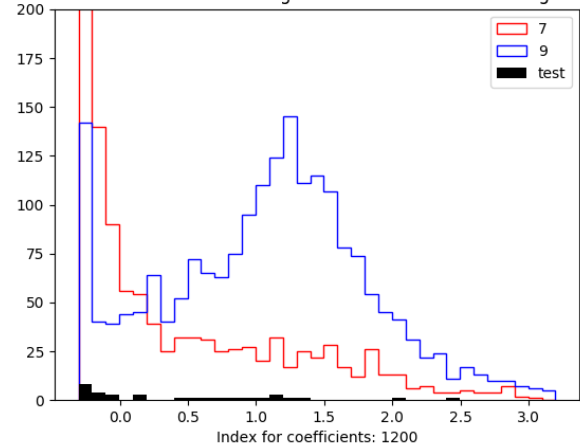
Histogram of 7th maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images



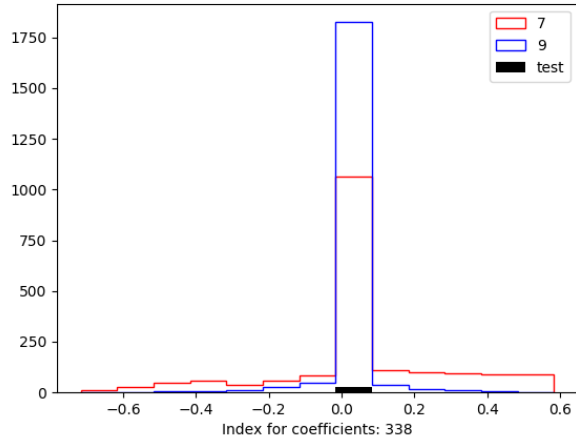
Histogram of 8th maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images



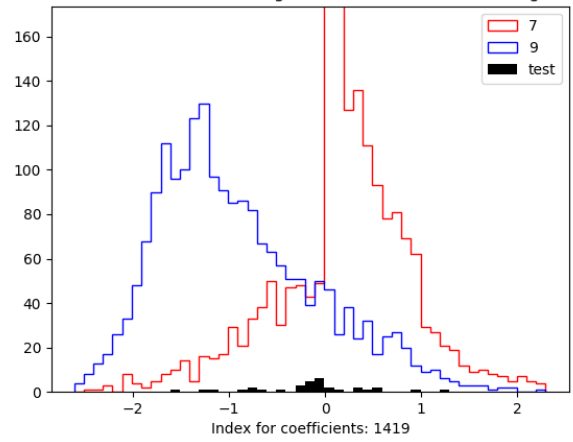
Histogram of 8th maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images



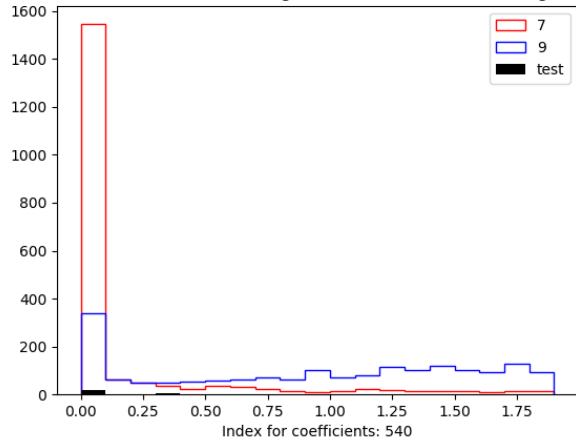
Histogram of 9th maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images



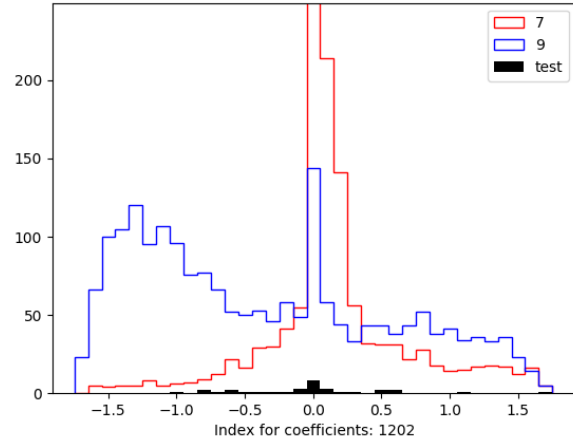
Histogram of 9th maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images



Histogram of 10th maximum K-L divergence value for 7 as kernel with best 1000 train images and worst 1000 train images

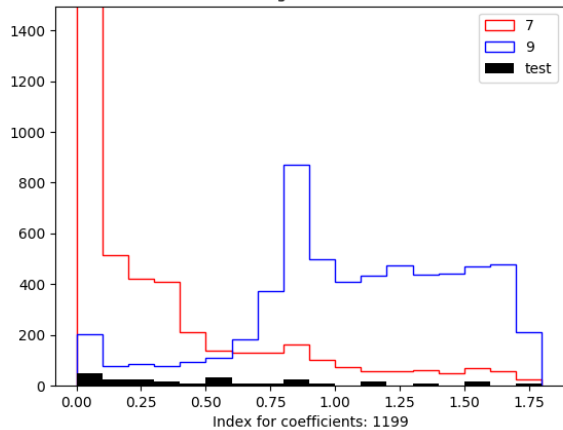


Histogram of 10th maximum K-L divergence value for 9 as kernel with best 1000 train images and worst 1000 train images

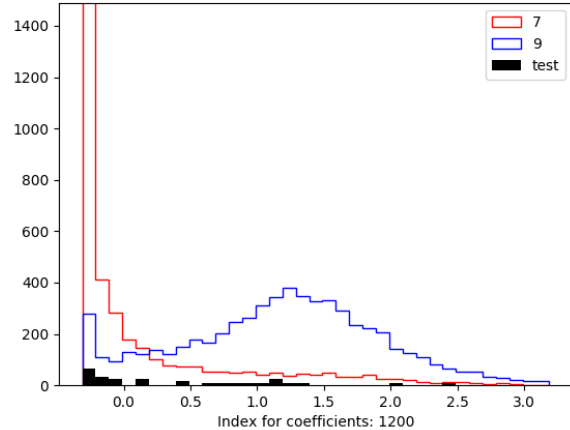


All images for each class as the kernel

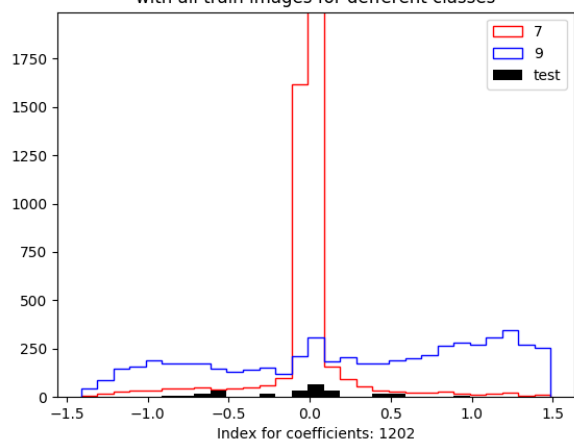
Histogram of 1st maximum K-L divergence value for 7 as kernel with all train images for defferent classes



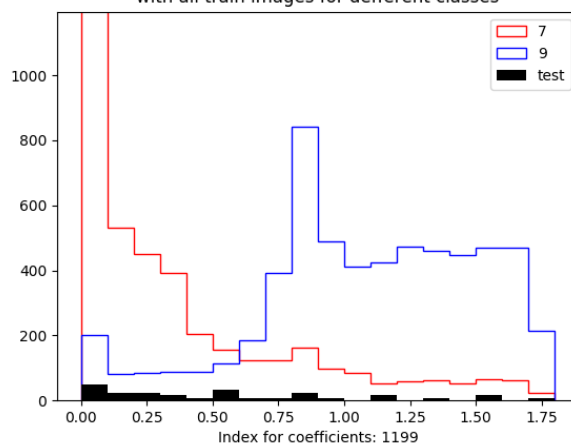
Histogram of 1st maximum K-L divergence value for 9 as kernel with all train images for defferent classes



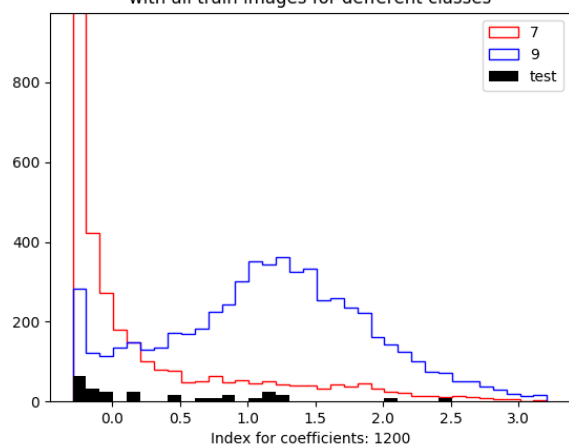
Histogram of 2nd maximum K-L divergence value for 7 as kernel with all train images for defferent classes



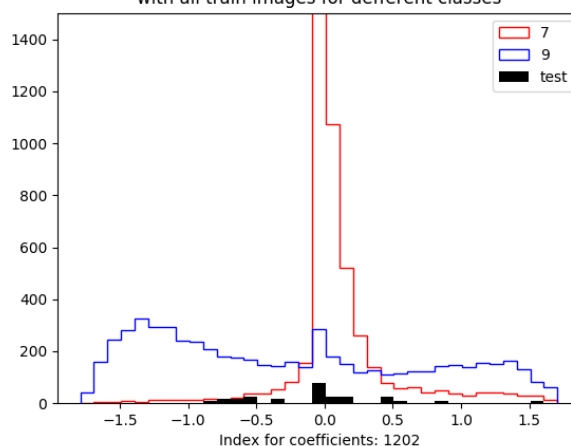
Histogram of 2nd maximum K-L divergence value for 9 as kernel with all train images for defferent classes



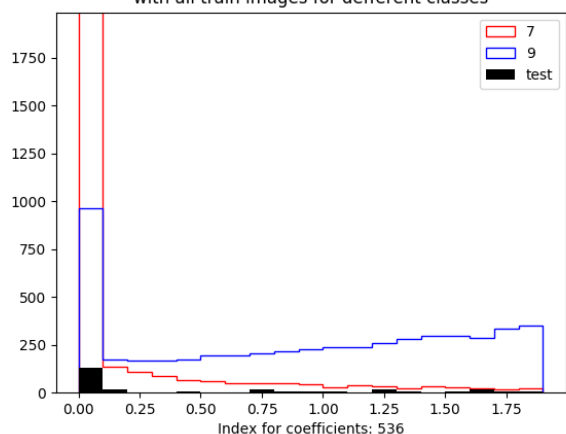
Histogram of 3rd maximum K-L divergence value for 7 as kernel with all train images for defferent classes



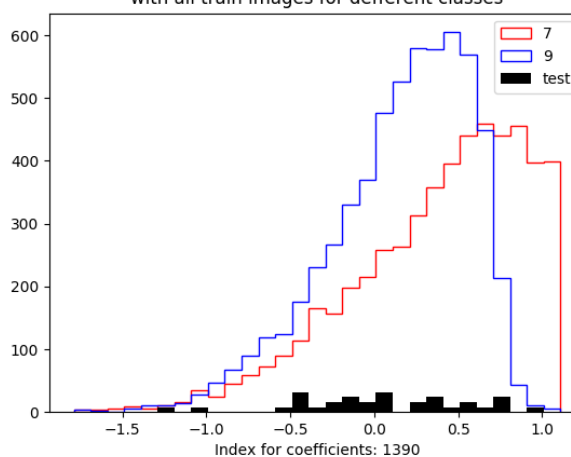
Histogram of 3rd maximum K-L divergence value for 9 as kernel with all train images for defferent classes



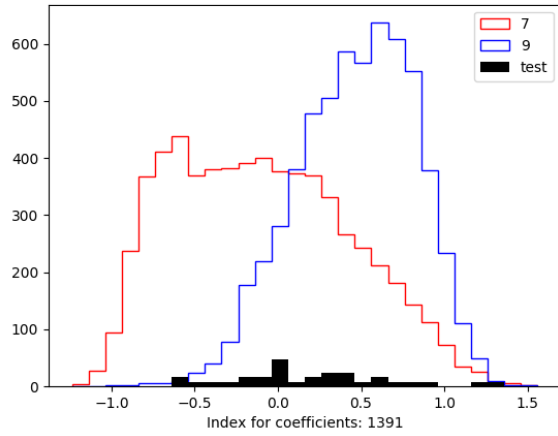
Histogram of 4th maximum K-L divergence value for 7 as kernel with all train images for defferent classes



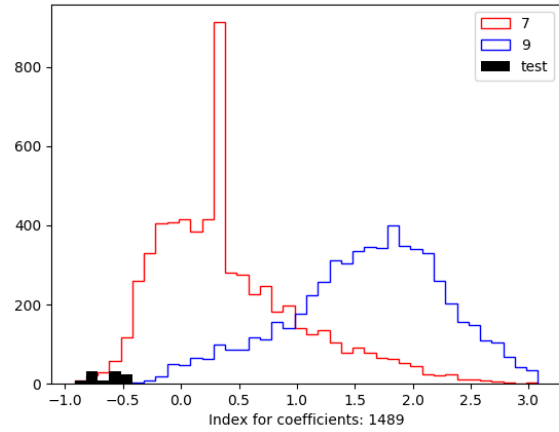
Histogram of 4th maximum K-L divergence value for 9 as kernel with all train images for defferent classes



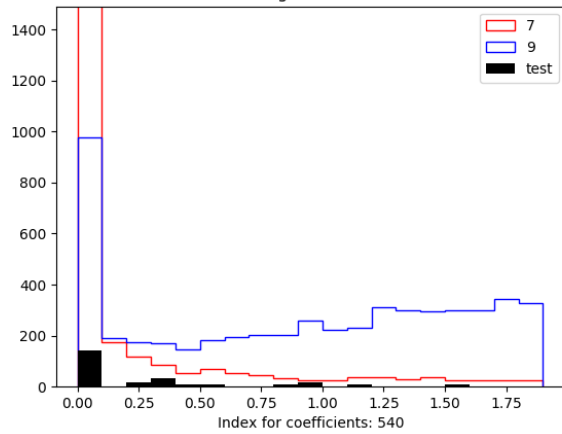
Histogram of 5th maximum K-L divergence value for 7 as kernel with all train images for defferent classes



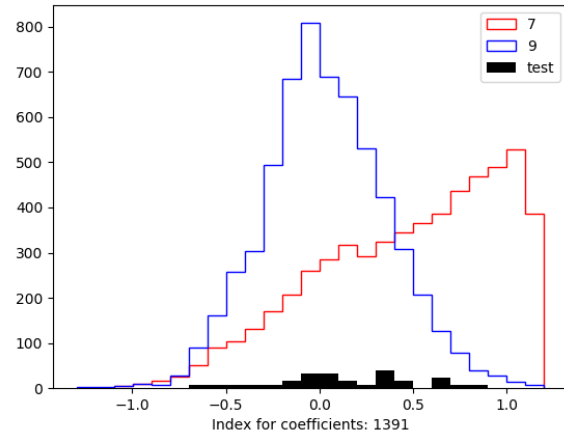
Histogram of 5th maximum K-L divergence value for 9 as kernel with all train images for defferent classes



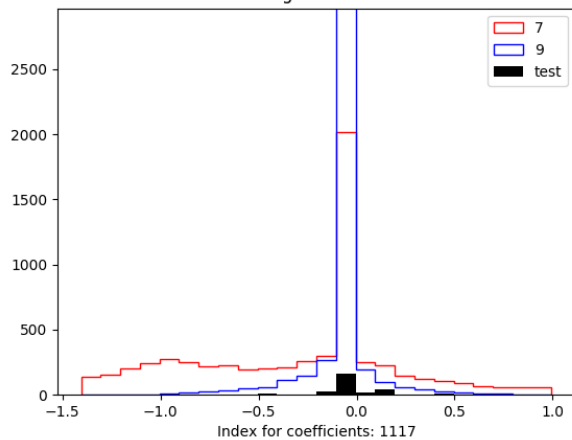
Histogram of 6th maximum K-L divergence value for 7 as kernel with all train images for defferent classes



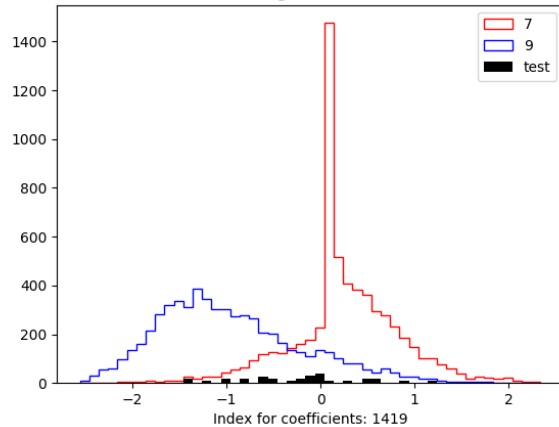
Histogram of 6th maximum K-L divergence value for 9 as kernel with all train images for defferent classes



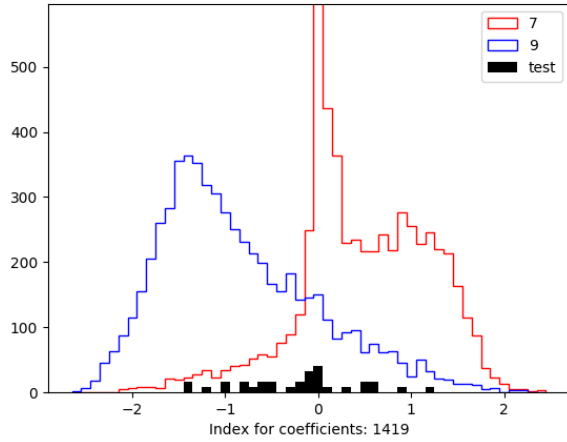
Histogram of 7th maximum K-L divergence value for 7 as kernel with all train images for defferent classes



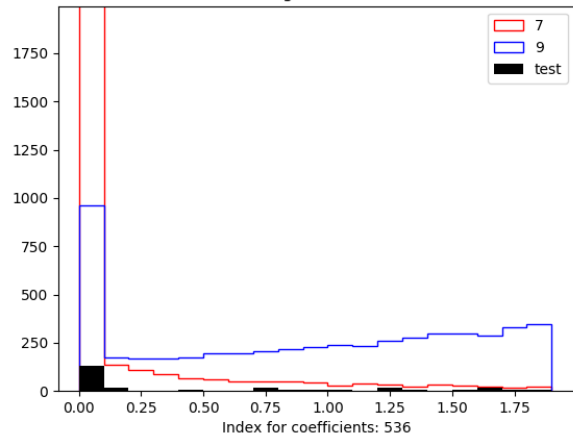
Histogram of 7th maximum K-L divergence value for 9 as kernel with all train images for defferent classes



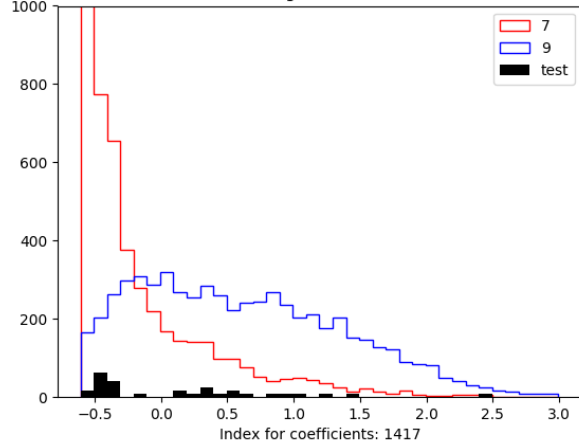
Histogram of 8th maximum K-L divergence value for 7 as kernel with all train images for defferent classes



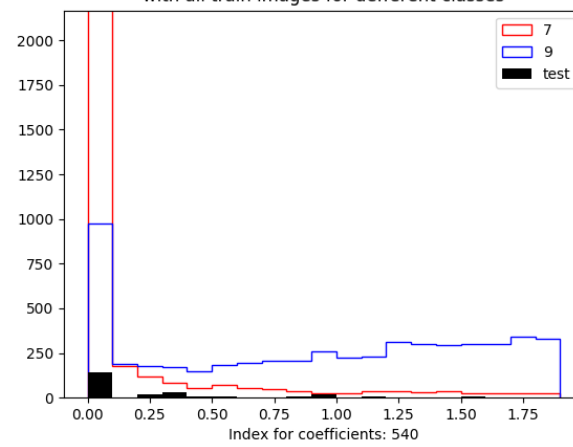
Histogram of 8th maximum K-L divergence value for 9 as kernel with all train images for defferent classes



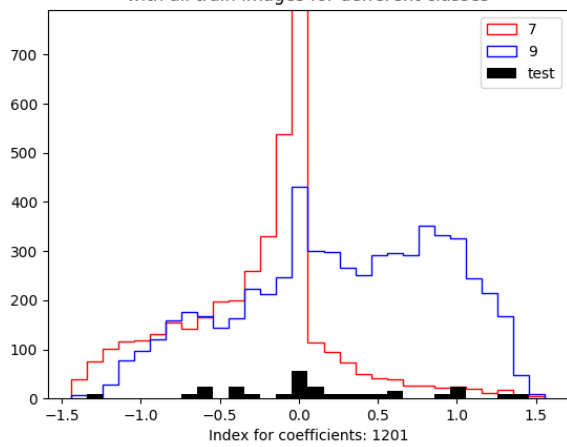
Histogram of 9th maximum K-L divergence value for 7 as kernel with all train images for defferent classes



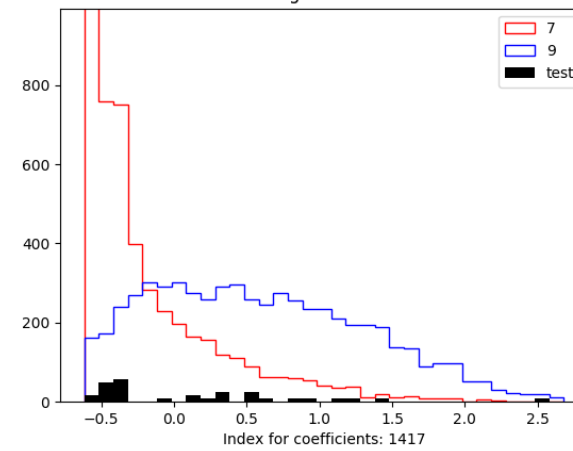
Histogram of 9th maximum K-L divergence value for 9 as kernel with all train images for defferent classes



Histogram of 10th maximum K-L divergence value for 7 as kernel with all train images for defferent classes



Histogram of 10th maximum K-L divergence value for 9 as kernel with all train images for defferent classes



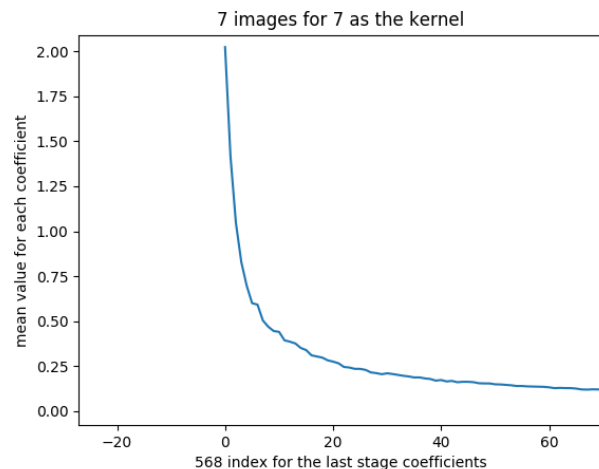
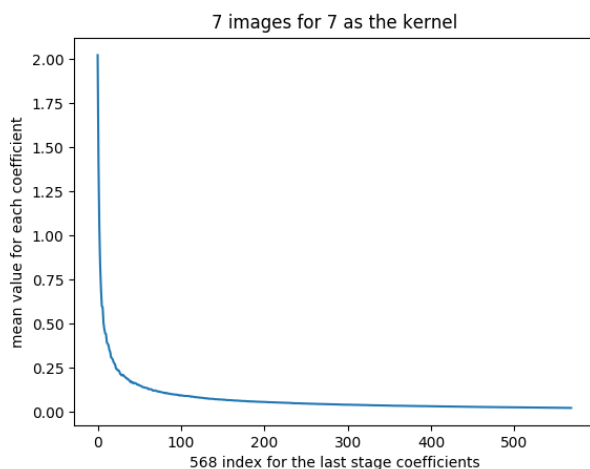
It seemed that the selected coefficients with top 10 K-L divergence values of the confusion test images mostly locate in the overlapping areas, that's why we cannot distinguish the confusion classes very well. By the way, on the bottom of each image shows the index of the coefficients, First stage: 1—1024, Second stage: 1025—1344, Third stage: 1345—1472, Fourth stage: 1473—1500, Fifth stage: 1501—1509.

2. Analysis of the coefficients of the last stage

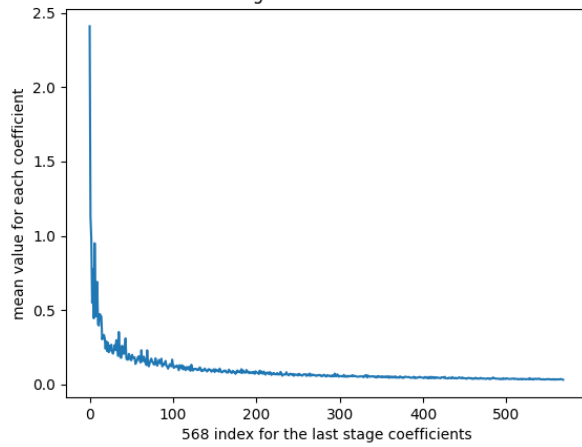
Then I utilized the entropy to distinguish the confusion classes, but I got almost the same results as using the K-L divergence values. So we need to calm down and analysis the coefficients stage by stage. First, we need to analysis the coefficients of the last stage to find out which filters we can get the high responses (absolute values).

In this experiment, I selected 3, 15, 51, 209, 568 principal components for each stage and for each image, we have 568 coefficients in the last stage. I still utilized “7” and “9” as the confusion classes. By the way, the left images are original images, and the right images are the left images zoomed in.

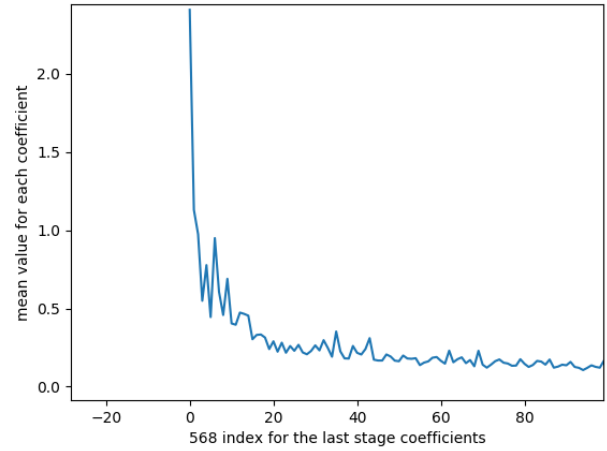
All images for each class as the kernel



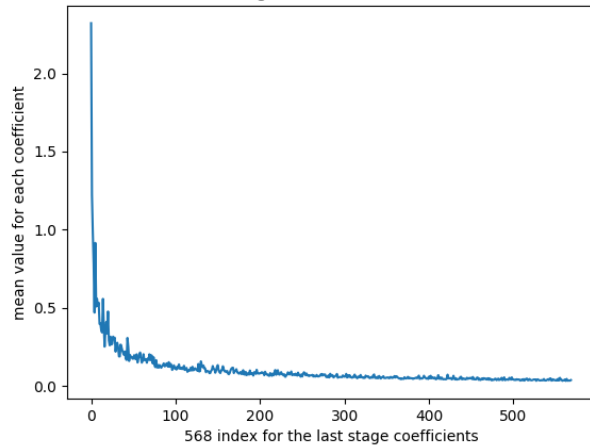
9 images for 7 as the kernel



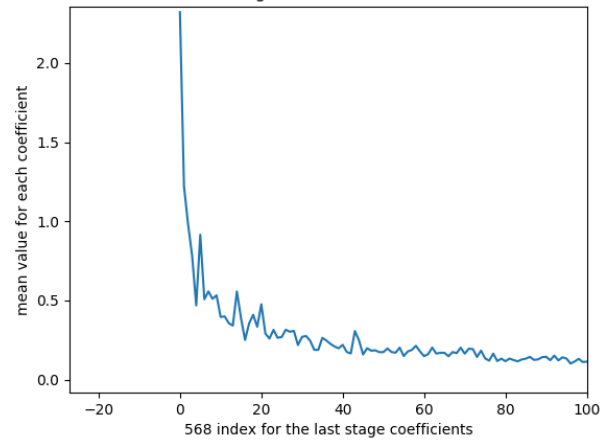
9 images for 7 as the kernel



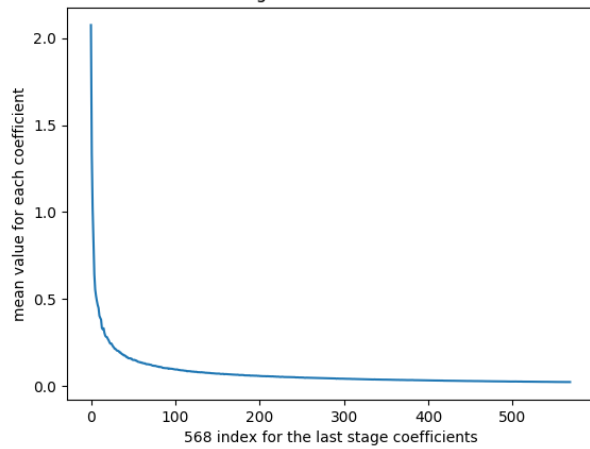
7 images for 9 as the kernel



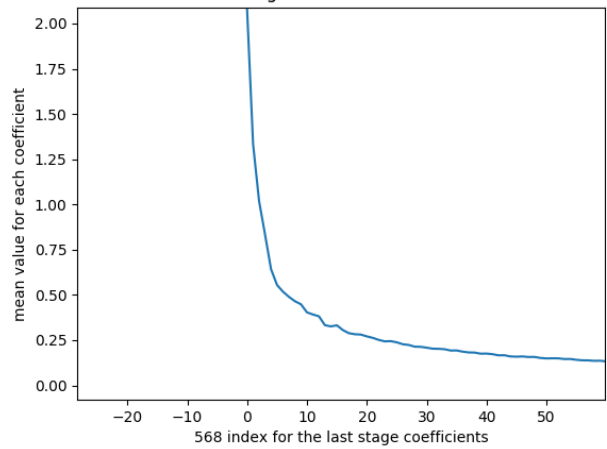
7 images for 9 as the kernel



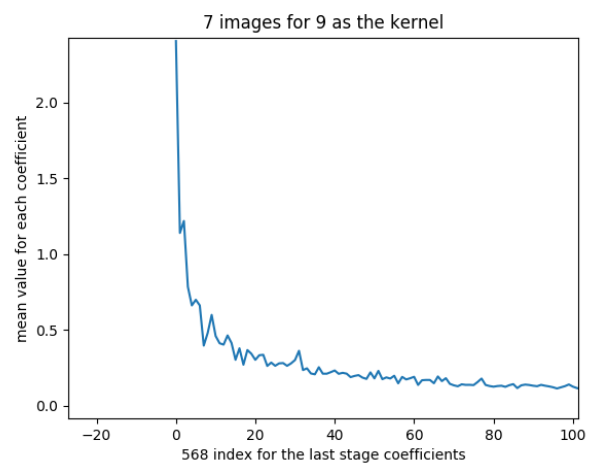
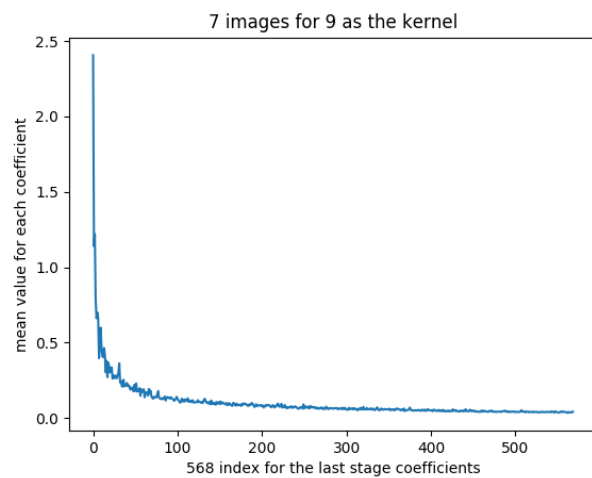
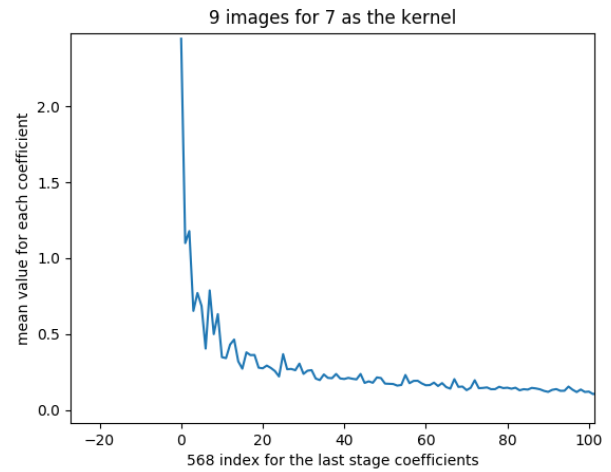
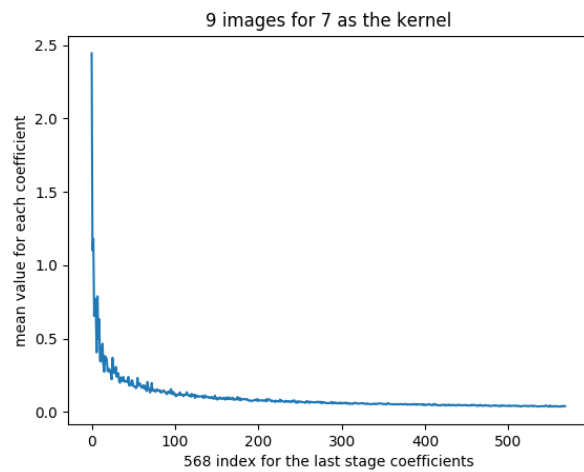
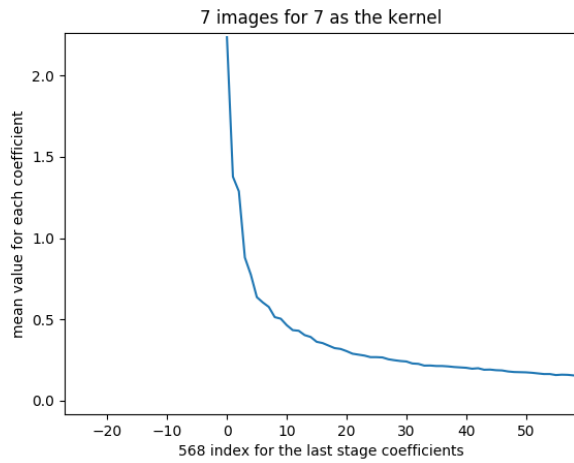
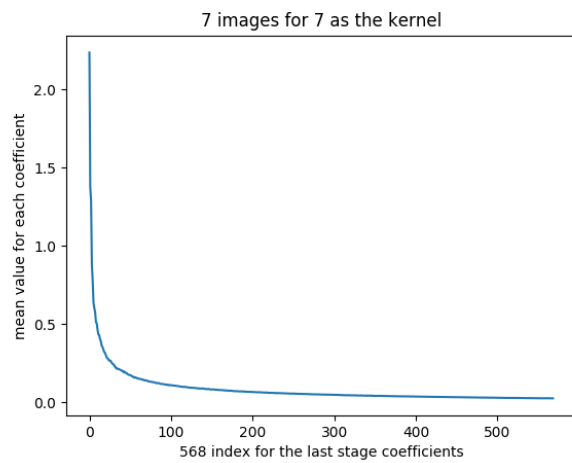
9 images for 9 as the kernel



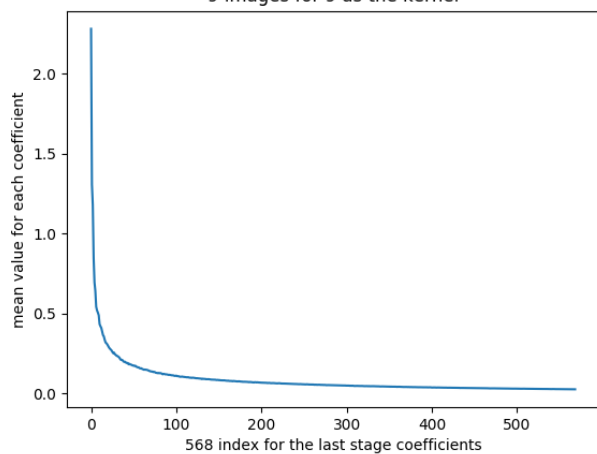
9 images for 9 as the kernel



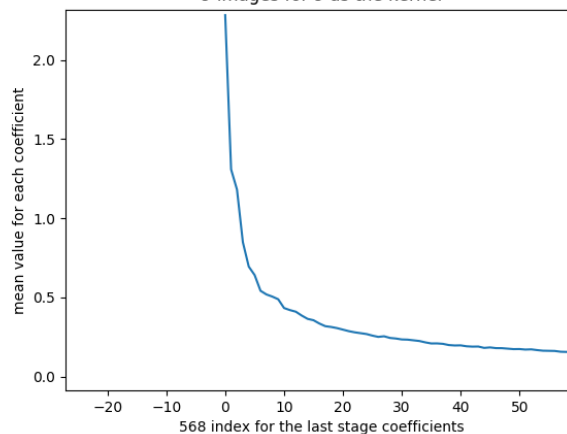
The lowest 2000 probability for each class as the kernel



9 images for 9 as the kernel

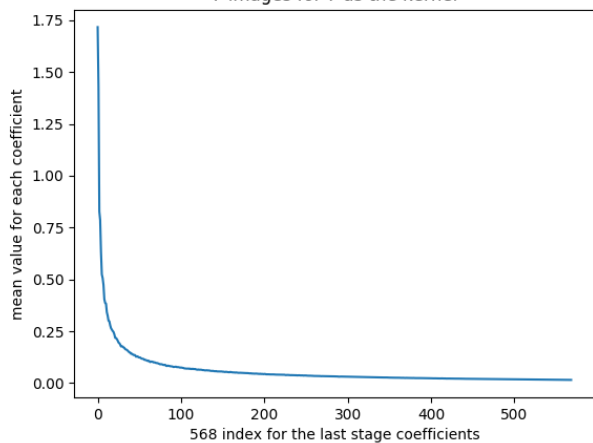


9 images for 9 as the kernel

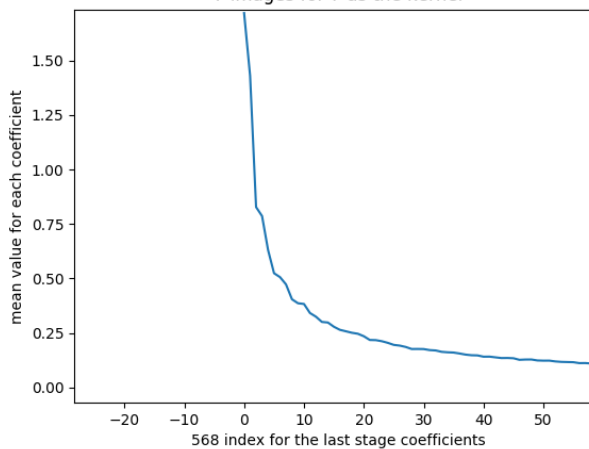


The highest 2000 probability for each class as the kernel

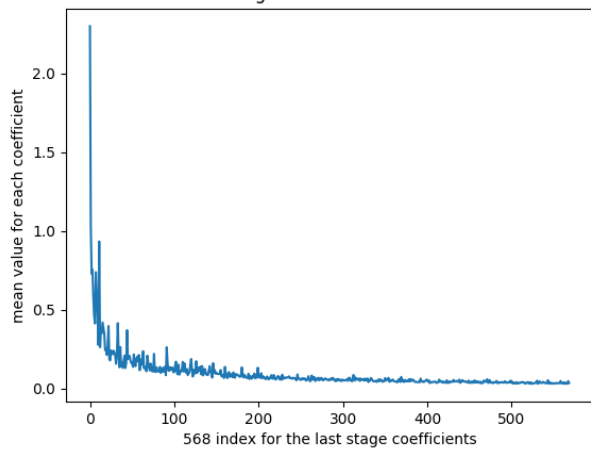
7 images for 7 as the kernel



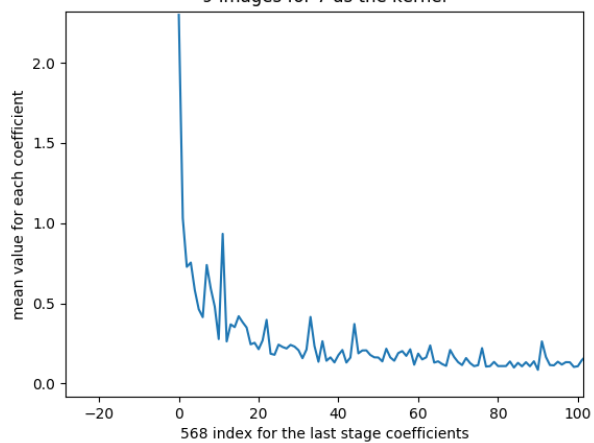
7 images for 7 as the kernel



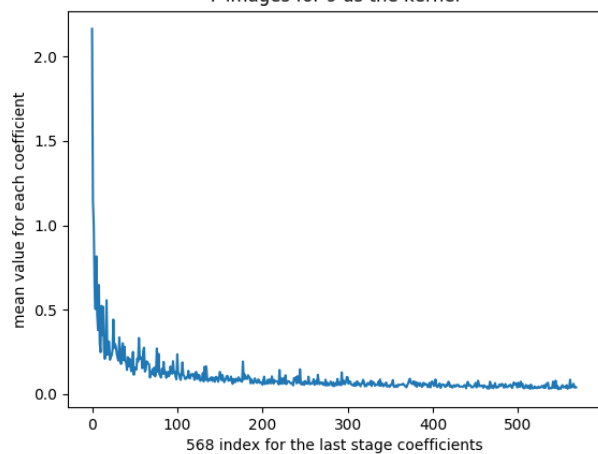
9 images for 7 as the kernel



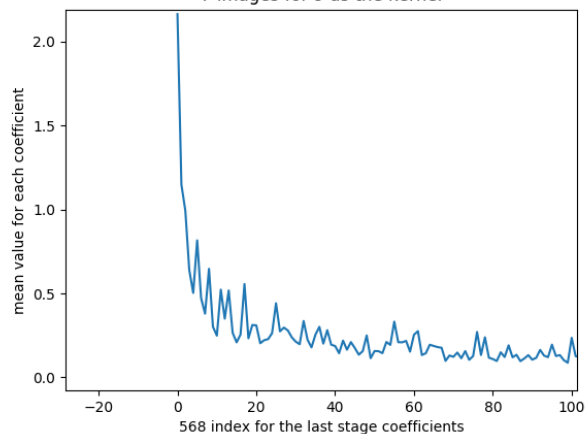
9 images for 7 as the kernel



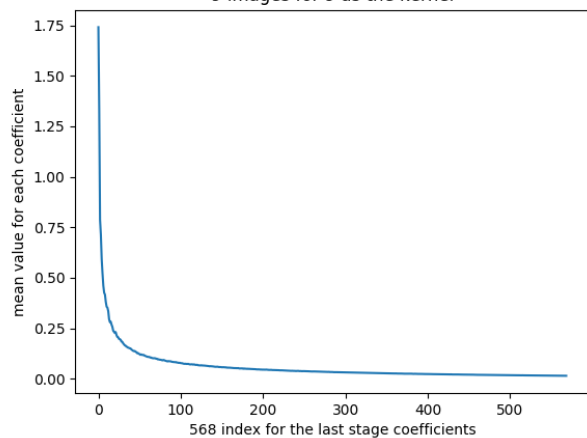
7 images for 9 as the kernel



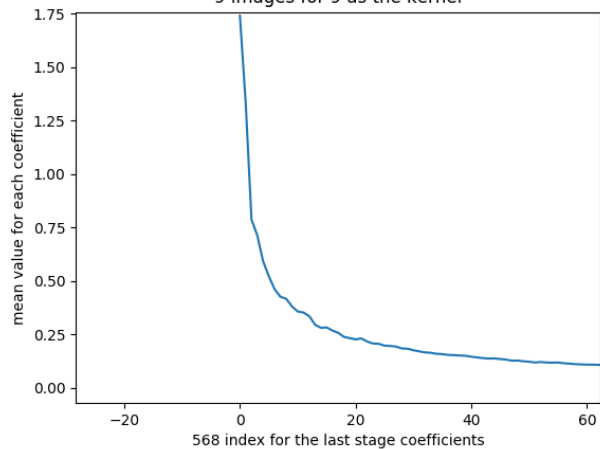
7 images for 9 as the kernel



9 images for 9 as the kernel

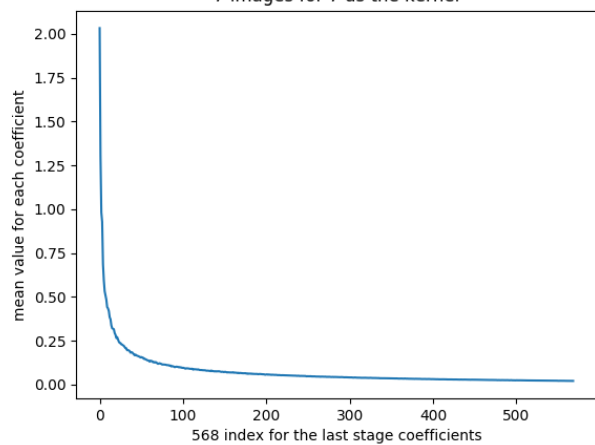


9 images for 9 as the kernel

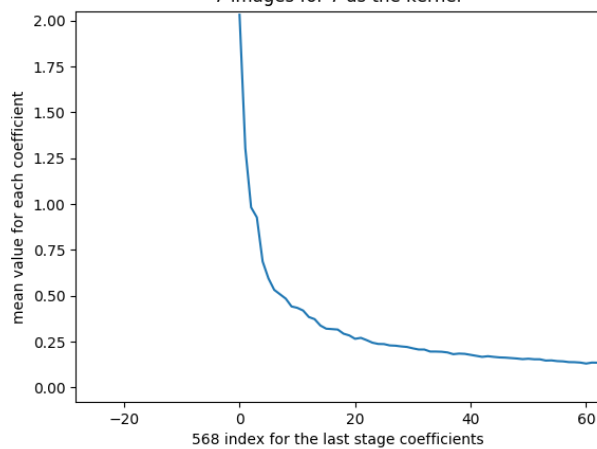


The lowest 1000 probability and the highest 1000 probability for each class as the kernel

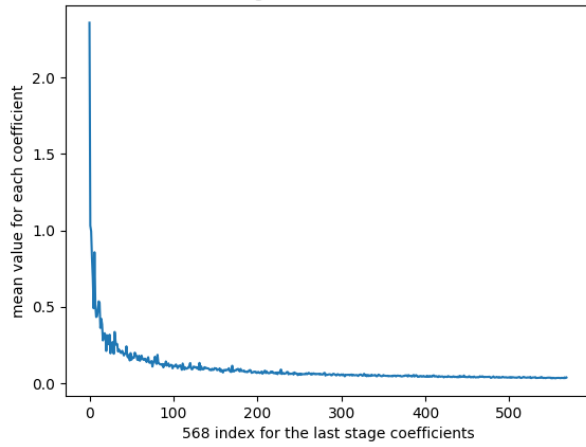
7 images for 7 as the kernel



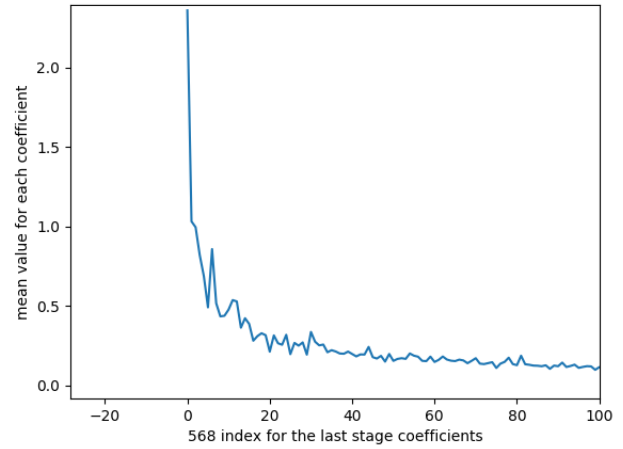
7 images for 7 as the kernel



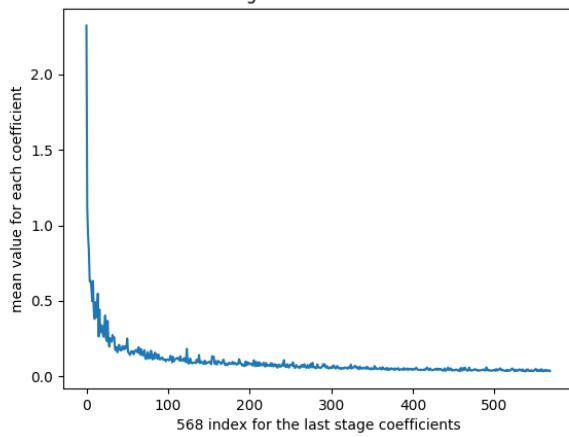
9 images for 7 as the kernel



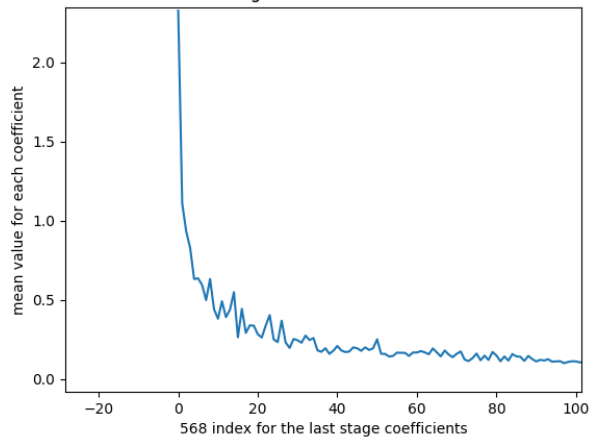
9 images for 7 as the kernel



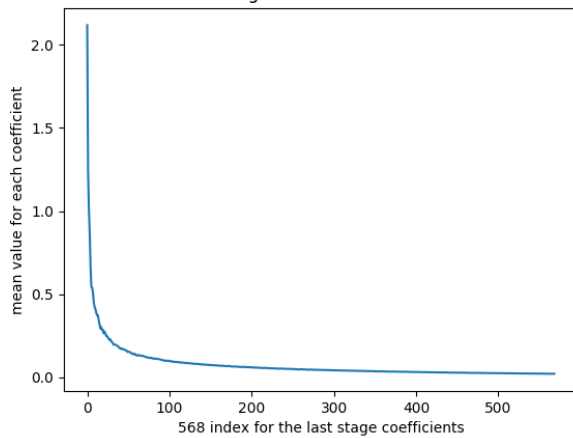
7 images for 9 as the kernel



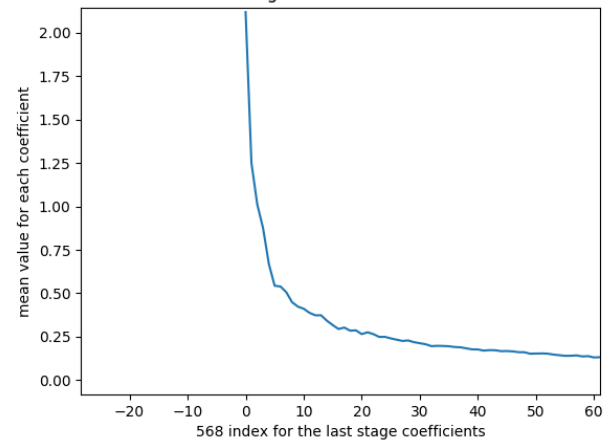
7 images for 9 as the kernel



9 images for 9 as the kernel



9 images for 9 as the kernel



We can see from above, the DC component has the strongest response, with the increasing of the index of the coefficients, the shape of the figure is decreasing, but it still has some maxima with the index increasing.

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 coefficients for all stages.

V. Milestone