Application of Feature Extraction based on Convolutional Neural Networks to Image Classification

Chang-Yu Pan (潘昶余), *Department of Applied Mathematics, National Sun Yat-sen University, Taiwan* **(dragon.105.me@gmail.com)**

Advisor: Prof. Tsung-Lin Lee (李宗錂)

Abstract

In the past, CNN trains the model with back-propagation. The model is lack of explanation and has large quantity of computation, so a CNN without back-propagation (FF-CNN) is proposed recently. The model replaces the convolution part with feature extraction method based on PCA. But PCA inputs the training data in a vector form. For images, it loses the information between different order so that the performance is limited.

This study proposed a classification model called Pixel-Anchored CNN (PA-CNN) which modifies the FF-CNN and replaces PCA stage with the High-Order Principal Component Analysis (HOPCA). It reduces quantity of computation and the loading of memory and the performance slightly increases.

Model Framework (PA-CNN)

Input: $\{(\mathcal{X}_1, y_1), (\mathcal{X}_2, y_2), \ldots, (\mathcal{X}_M, y_M)\}$: data ,
N_{conv}: number of convolution layer,
N_{fc}: number of FC layer,
$[k_1, k_2]$: kernel size.

Output: $\{\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_{N_{conv}}\}$: anchors and bias,
$\{\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_{N_{fc}}\}$: FC layer weight set.

For $l = 1,2, \ldots, N_{conv}$ **do**
  For $p = 1,2, \ldots,$ anchored pixel **do**
    Get anchors and bias by HOSaab.
  End
  Reshape feature map to tensor.
  Apply max-pooling.
End
For $l = 1,2, \ldots, N_{fc}$ **do**
  Create pseudo label by k-means clustering.
  Solve the least-square problem.
End

Problem description

We modify two parts of convolution in original CNN. One is the convolution step. It do the affine transformation.

$$y_k = \sum_{i=1}^{d} a_{k,i} x_i = \boldsymbol{a}_k^{\top} \boldsymbol{x} \quad \cdots\cdots (1)$$

Those filters in CNN are $\boldsymbol{a}_k^{\top}$ in Eq (1), we call them **anchor vectors**. We want to determine anchors directly rather than back-propagation. The other is activation function. Activation function in CNN do the job in resolving sign confusion problem.A feature extraction method called Saab can resolve these two problems. We propose a method called HOSaab by combining Saab and HOPCA and apply it on the PA-CNN framework.

HOSaab (High-Order Saab)

Let $\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M\}$ be a set of M tensor objects, where $\mathcal{X}_i \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. HOSaab first separate data to DC subspace S_{DC} and AC subspace S_{AC}. S_{DC} is spanned by $\mathbf{1}$ where $\mathbf{1}$ is the tensor with all elements equal to one . **Anchors selection**. We conduct HOPCA on the AC component. The anchor vectors are selected as the dominant singular vectors. HOPCA is to find the orthogonal projection set $\{\boldsymbol{U}^{(n)} \in \mathbb{R}^{I_n \times P_n} : P_n < I_n, n = 1,2, \ldots, N\}$ to maximize the total scatter $\Psi = \sum_{m=1}^{M} \|\mathcal{Y}_m - \bar{\mathcal{Y}}\|_2$ where $\bar{\mathcal{Y}}$ is the mean tensor and $\mathcal{Y}_m \in \mathbb{R}^{P_1 \times P_2 \times \cdots \times P_N}$ is defined by

$$\mathcal{Y}_m = \mathcal{X}_m \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times \cdots \times \boldsymbol{U}^{(N)}$$

Bias selection. Bias tensor is selected as the $\mathcal{B} = b\mathbf{1}$ where $b = \max_m \|\mathcal{X}_m\|_F$, which ensures that the tensor version affine transformation is non-negative. Namely,

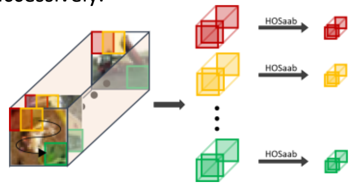
$$\mathcal{Y}_m = \mathcal{X}_m \times_1 \boldsymbol{U}^{(1)} \times_2 \boldsymbol{U}^{(2)} \times \cdots \times \boldsymbol{U}^{(N)} + \mathcal{B} \geq 0$$

Fully-connected (FC) Layer

Fully-connected layer in original CNN is treated as least-square problems with pseudo label. it can capture the diversity in the same class.

Pixel-Anchored CNN (PA-CNN)

The First part do the convolution on images in a specific kernel size to get patches. Then apply HOSaab on the patches obtained by the same pixels successively.

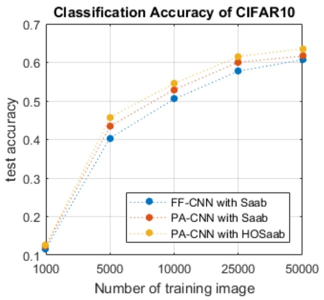


Second part, construct the least-square problem with the reduced features.

Results

CIFAR10

The CIFAR10 database contains 10 classes with 5000 images per class.



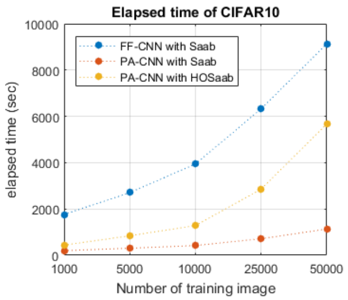


Table 1: classification accuracy of CIFAR10

train size	1000	5000	10000	25000	50000
FF-CNN with Saab	11.71	40.36	50.50	57.71	60.69
PA-CNN with Saab	12.35	43.43	52.92	60.05	61.63
PA-CNN with HOSaab	12.48	45.76	54.54	61.51	63.58

Table 2: elapsed time of CIFAR10

train size	1000	5000	10000	25000	50000
FF-CNN with Saab	1761.81	2706.42	3942.59	6322.54	9139.03
PA-CNN with Saab	198.56	303.33	419.46	715.15	1132.84
PA-CNN with HOSaab	436.39	842.83	1281.33	2856.55	5673.77

Conclusions

This study explore the classification accuracy and the computation time of models. The experiment shows that Pixel-anchored CNN reduce the computation time both on Saab and HOSaab. The accuracy of PA-CNN with HOSaab slightly higher than Saab. Explain that HOPCA captures more important information than PCA. Although HOSaab has more time-consuming than Saab, we can apply PA-CNN to reduce the training time instead of FF-CNN.

References

[1] C-C Jay Kuo, Min Zhang, Siyang Li, Jiali Duan, and Yueru Chen, Interpretable convolutional neural networks via feedforward design, Journal of Visual Communication and Image Representation, 60:346–359, 2019.

[2] C.-C. Jay Kuo and Yueru Chen, On data-driven Saak transform, Journal of Visual Communication and Image Representation, 50:237–246, 2018.

[3] Cheng-Ju Yang, Image classification via successive core tensor selection procedure, Master's thesis, National Sun Yat-sen University, 2018.

[4] H. Lu, K. N. Plataniotis and A. N. Venetsanopoulos, "MPCA: Multilinear Principal Component Analysis of Tensor Objects," in IEEE Transactions on Neural Networks, vol. 19, no. 1, pp. 18-39, Jan. 2008, doi: 10.1109/TNN.2007.901277.