Analysis Dictionary Learning: An Efficient and Discriminative Solution

NC STATE UNIVERSITY

Wen Tang, Ashkan Panahi, Hamid Krim, Liyi Dai[1]
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC, USA
[1]Raytheon Integrated Defense Systems, Tewksbury, MA, USA

Abstract

Discriminative Dictionary Learning (DL) methods have been widely advocated for image classification problems. To further sharpen their discriminative capabilities, most state-of- the-art DL methods have additional constraints included in the learning stages. These various constraints, however, lead to additional computational complexity. We hence propose an efficient Discriminative Convolutional Analysis Dictionary Learning (DCADL) method, as a lower cost Discriminative DL framework, to both characterize the image structures and refine the interclass structure representations.

Introduction

In the past decade, sparse representation has been widely and successfully applied to address a variety of image processing and computer vision problems. One well known approach is the Synthesis Dictionary Learning (SDL), which recovers the signal by learning a dictionary with corresponding coefficients. To overcome the shortcomings of classical patch-based sparse representation and better translation invariance, convolutional filters were also introduced in SDL for signal and image processing applications. Due to this success in image processing, SDL has also been explored in image inference problems, such as image classification, by augmenting with some supervised learning constraints. Besides SDL, Analysis Dictionary Learning (ADL) has recently been of interest on account of its fast encoding and stability attributes. Inspired by the SDL methodology in image classification, ADL has also been adapted to the supervised learning problems by promoting discriminative sparse representations. In all above methods, both the structure of images and the structure between different classes play important roles in the classification task. Such structures increase the accuracy, but they also require a substantial amount of computation and time for training and testing. It is hence desirable to forego this potentially costly structure-promoting regularization and to instead embed the discriminating characteristics of ADL methods in the dictionary formulation itself.

To this end, we introduce a convolutional mapping within the ADL framework, and embed its resulting feature resolution using its translation invariant structure. We thus propose the Discriminative Convolutional ADL (DCADL) method, which amounts to jointly learning a convolutional ADL and a linear classifier to ensure the capability of characterizing structures among individual images and across classes, while taking advantage of fast ADL encoding. To reduce the excessive training time, we propose a novel algorithmic technique which transforms convolution to a low-cost matrix multiplication. This turns DCADL into an efficiently solvable conventional discriminative ADL framework.

Discriminative Dictionary Learning

Let X denote a training data matrix of C classes and Ω be an associated dictionary, Discriminative DL methods generally belong to the following optimization framework,

$$\arg\min_{\Omega,U} f(\Omega,U,X)+\lambda\|U\|_p+\Phi_s(\Omega,U,Y)+\Phi_G(\Omega,U,Y,W) \quad (1)$$

with $f(\Omega,U,X)=\frac{1}{2}\|\Omega X-U\|_F^2$ or $f(\Omega,U,X)=\frac{1}{2}\|X-\Omega U\|_F^2$, $Y\in R^{C\times n}$ represents the labels of the training data, where $Y_{ij}=1$ if and only if image j belongs to class i, with some structure-promoting constraint function $\Phi_S(\Omega,U,Y)$, and a general classification objective functional $\Phi_G(\Omega,U,Y,W)$.

Later, we simplify the Discriminative DL framework in Eq. (1) to the following by replacing $\Phi_S(\Omega,U,Y)$ by matrix reshaping operators:

$$\begin{aligned}&\arg\min_{\Omega,U} f(\Omega,U,X)+\lambda\|\widehat{U}\|_p+\Phi_G(\Omega,\widetilde{U},Y,W)\\ &s.t.\ \widehat{U}=RS_1(U);\quad \widetilde{U}=RS_2(U),\end{aligned} \quad (2)$$

where RS_1, RS_2 are some matrix reshaping.

Discriminative Convolutional Analysis Dictionary Learning

This intuitive DCADL framework is defined as follows

$$\arg\min_{\Omega,U}\sum_{j=1}^{n}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|\omega_i * x_j-u_j^i\right\|_2^2+\lambda_1\left\|u_j^i\right\|_1\right)+\frac{\lambda_2}{2}\left\|Y-W\widetilde{U}\right\|_F^2 \quad (3)$$

$$s.t.\ \|\omega_i\|_2^2\le 1;\ \forall i=1,\dots,m,\ \widetilde{U}=\begin{bmatrix}u_1^1 & \cdots & u_n^1\\ \vdots & \ddots & \vdots\\ u_1^m & \cdots & u_n^m\end{bmatrix},$$

where $*$ is convolutional operator, $\omega_i^T\in R^{s^2}$ is the ith atom (row) of size $s\times s$ in the analysis dictionary Ω, $x_j\in R^r$ is the jth image, and $u_i^j\in R^p$ is the ith response map of the jth image corresponding to the convolution of the ith atom.

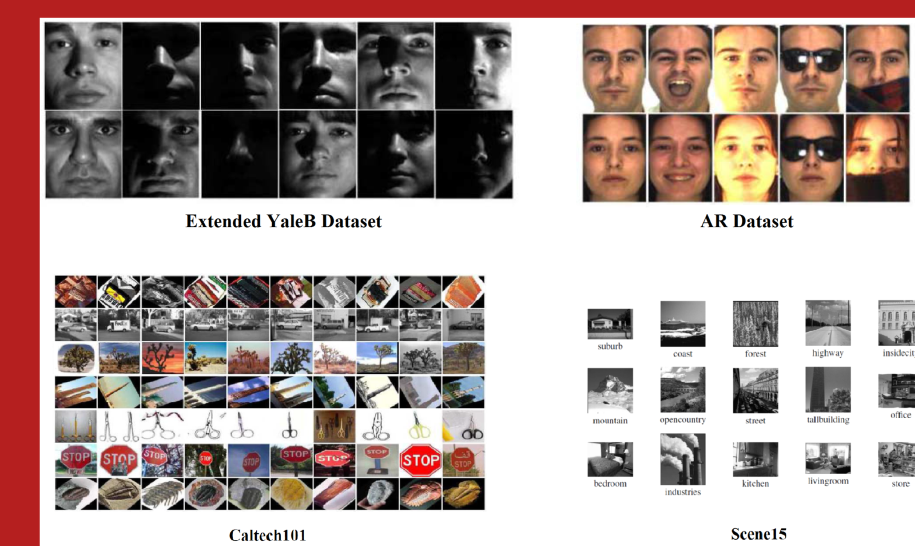
Segment an image x_i into p patches $[x_{i_1},\dots,x_{i_p}]$ with $s\times s$ pixels, being of the same size as the atom, and let $\bar{X}=\left[x_{1_1},\dots,x_{1_p},\dots,x_{n_1},\dots,x_{n_p}\right]\in R^{s^2\times np}$ and $\bar{U}=\begin{bmatrix}u_{1_1}^1 & \cdots & u_{1_p}^1 & \cdots & u_{n_1}^1 & \cdots & u_{n_p}^1\\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots\\ u_{1_1}^m & \cdots & u_{1_p}^m & \cdots & u_{n_1}^m & \cdots & u_{n_p}^m\end{bmatrix}\in R^{m\times np}$. The problem in Eq. (3) can then be rewritten in the same form as in Eq. (2):

$$\begin{gathered}\arg\min_{\Omega,U}\frac{1}{2}\|\Omega\bar{X}-\bar{U}\|_F^2+\lambda_1\|\widehat{U}\|_1+\frac{\lambda_2}{2}\left\|Y-W\widetilde{U}\right\|_F^2\\ s.t.\ \|\omega_i\|_2^2\le 1;\ \forall i=1,\dots,m,\end{gathered}$$

$$\widehat{U}=\begin{bmatrix}u_{1_1}^1 u_{1_1}^2 & \cdots & u_{n_1}^m\\ \vdots\quad\vdots & \ddots & \vdots\\ u_{1_p}^1 u_{1_p}^2 & \cdots & u_{n_p}^m\end{bmatrix}\ \widetilde{U}=\begin{bmatrix}u_1^1 & \cdots & u_n^1\\ \vdots & \ddots & \vdots\\ u_1^m & \cdots & u_n^m\end{bmatrix}.$$

Experiments and Results

Four widely used visual classification datasets have been applied to evaluate our proposed DCADL.



To evaluate our proposed DCADL, we carry out a comparative study with the following methods: The first one is ADL+SVM[1], which serves as a baseline. LC-KSVD [2] is a state-of-the-art SDL. Then SADL[3] and DADL[4] are up-to-date ADL approaches. The last method, DPL[5] is a hybrid technique of SDL and ADL.

Table 1. Classification Results on Extended YaleB Dataset

Methods(#atoms)	Accuracy(%)	Training Time(s)	Testing Time(s)
ADL+SVM(1216) [1]	88.91 ± 0.73	274.16	6.78×10^{-4}
LC-KSVD(570) [2]	94.74 ± 0.47	183.55	1.36×10^{-3}
LC-KSVD(1216) [2]	66.05 ± 2.35	244.77	1.23×10^{-3}
SADL(1216) [3]	97.58 ± 0.39	257.31	$\mathbf{1.53 \times 10^{-5}}$
DADL(2031) [4]	98.33 ± 0.28	6.40	2.19×10^{-4}
DPL(1216) [5]	98.01 ± 0.45	20.25	2.09×10^{-4}
HDL-2 (-) [6]	98.50	-	-
PCANet-1 (-) [7]	97.77	-	-
DCADL(50)	$\mathbf{99.57 \pm 0.08}$	**3.82**	1.93×10^{-5}

For Extended YaleB: Our proposed DCADL method achieves the highest classification accuracy with the shortest training time and an extremely fast testing time, while securing an at least 1% greater accuracy relative to others'. In the second part of Table 1, a 2- layer hierarchical dictionary learning approach[6] and a 1- layer convolutional network[7] are also included for comparison. Both of these two methods also worked on raw pixels of images.

Table 2. Classification Results on AR Dataset

Methods(#atoms)	Accuracy(%)	Training Time(s)	Testing Time(s)
ADL+SVM(2000) [1]	85.35 ± 2.34	1301.97	9.05×10^{-3}
LC-KSVD(500) [2]	91.97 ± 1.09	275.18	3.93×10^{-4}
LC-KSVD(2000) [2]	67.70 ± 5.14	253.55	2.31×10^{-3}
SADL(2000) [3]	98.55 ± 0.33	69.93	2.88×10^{-5}
DADL(2211) [4]	$\mathbf{99.20 \pm 0.28}$	**10.42**	4.26×10^{-4}
DPL(2000) [5]	99.03 ± 0.32	24.03	8.45×10^{-5}
CNN-3 (-) [8]	96.50	-	-
DCADL(50)	98.93 ± 0.43	14.52	$\mathbf{2.78 \times 10^{-5}}$

For AR dataset: The accuracy of our proposed DCADL is barely lower than DADL and DPL, but it is still much higher than other methods with a very quick training and testing time. It is even better than the performance of a 3-layer Convolutional Network[8], which also worked on the raw pixel of the AR dataset.

Table 3. Classification Results on Caltech101 Dataset

Methods(#atoms)	Accuracy(%)	Training Time(s)	Testing Time(s)
ADL+SVM(3060) [1]	66.75 ± 1.08	1943.47	1.33×10^{-2}
LC-KSVD(3060) [2]	73.67 ± 0.93 (73.6[2])	2144.90	2.49×10^{-3}
SADL(3060) [3]	$\mathbf{74.17 \pm 0.49}$	1406.68	4.76×10^{-5}
DADL(3061) [4]	71.77 ± 0.44 (**74.6** [4])	26.29	7.90×10^{-4}
DPL(3060) [5]	71.64 ± 0.50 (73.9 [5])	64.33	3.79×10^{-4}
DCADL(152)	$\mathbf{74.17 \pm 0.42}$	**17.55**	$\mathbf{2.52 \times 10^{-5}}$

Table 4. Classification Results on Scene15 Dataset

Methods(#atoms)	Accuracy(%)	Training Time(s)	Testing Time(s)
ADL+SVM(1500) [1]	80.55 ± 3.20	494.41	1.73×10^{-4}
LC-KSVD(1500) [2]	$\mathbf{99.21 \pm 0.18}$ (92.9[2])	390.22	1.81×10^{-3}
SADL(1500) [3]	98.40 ± 0.21 (-)	219.80	2.41×10^{-5}
DADL(3001) [4]	97.81 ± 0.27 (98.3 [4])	15.00	4.62×10^{-4}
DPL(1500) [5]	98.35 ± 0.17 (97.7 [4])	8.83	5.67×10^{-5}
DCADL(50)	$\mathbf{98.41 \pm 0.26}$	**2.59**	$\mathbf{1.00 \times 10^{-5}}$

For Caltech 101: DCADL achieves the highest performance again in our experiments, achieving the fastest training and testing time. Though its accuracy is slightly lower than the reported one in DADL, DCADL is at least 1.5 times faster than DADL in training and testing time.

For Scene 15: Our accuracy is barely lower than LC-KSVD, but is still higher than all other methods and the reported performance in LC-KSVD. In addition, compared with all other methods, DCADL still registers a much greater training and testing time gain.

Conclusion

We proposed an efficient discriminative convolutional ADL method for classification tasks. Our DCADL consists of learning a convolutional ADL together with a universal linear classifier. We further transformed the optimization framework of DCADL to a more efficient discriminative DL framework by eliminating structural constraint costs, while preserving the discriminative power. Our extensive numerical studies show the DCADL exhibits its highly competitive accuracies with significant efficiency.

Acknowledgements

This research work was generously supported in part by the U.S. Army Research Office under agreement W911NF-16-2-0005.

Reference

[1] Sumit Shekhar, Vishal M Patel, and Rama Chellappa, "Analy- sis sparse coding models for image-based classification," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5207–5211.
[2] Zhuolin Jiang, Zhe Lin, and Larry S Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelli- gence*, vol. 35, no. 11, pp. 2651–2664, 2013.
[3] Wen Tang, Ashkan Panahi, Hamid Krim, and Liyi Dai, "Anal- ysis dictionary learning based classification: Structure for ro- bustness," ," in *Acoustics, Speech and Signal Processing (ICASSP), 2018 IEEE International Conference on*. IEEE, 2018.
[4] Jun Guo, Yanqing Guo, Xiangwei Kong, Man Zhang, and Ran He, "Discriminative analysis dictionary learning," in Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[5] ShuhangGu,LeiZhang,WangmengZuo,andXiangchuFeng, "Projective dictionary pair learning for pattern classification," in *Advances in neural information processing systems*, 2014, pp. 793–801.
[6] ShahinMahdizadehaghdam,LiyiDai,HamidKrim,ErikSkau, and Han Wang, "Image classification: A hierarchical dic- tionary learning approach," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2597–2601.
[7] Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma, "Pcanet: A simple deep learning baseline for image classification?," *IEEE Transactions on Image Pro- cessing*, vol. 24, no. 12, pp. 5017–5032, 2015.
[8] AR Syafeeza, M Khalil-Hani, SS Liew, and R Bakhteri, "Con- volutional neural network for face recognition with pose and illumination variation," *International Journal of Engineering & Technology*, vol. 6, no. 1, pp. 0975–4024, 2014.