Discriminative Learned Dictionaries for Local Image Analysis

-Julien Mairal

-Francis Bach

-Jean Ponce

-Guillermo Sapiro

-Andrew Zisserman

- Motivation
- Approach
- Dictionary learning for reconstruction
    A. Learning reconstruction dictionaries
    B. A reconstruction approach to discrimination
- Learning discriminative dictionaries
    A. Discriminative model
    B. Optimization procedure
    C. Data Pre-processing
- Experimental results and applications
    A. Texture segmentation of the Brodatz dataset
    B. Learning discriminant image patches
    C. Weakly-supervised feature selection
- Conclusions and future directions

Motivation

Sparse signal models are broadly used in improving upon state-of-art results in signal, image, and video restoration. So, the objective of this paper is to extend the use of sparse techniques for local image discrimination tasks, proposing an energy formulation with both sparse reconstruction and class discrimination components. So, the main objective is to use the mixture of discrimination and sparse techniques for better local image analysis.

Approach

- learning of multiple dictionaries which are simultaneously reconstructive and discriminative, and the use of the reconstruction error of these dictionaries on image patches to derive a pixel-wise classification.
- This approach can be broken down in 2-steps.
  1. Redundant non-parametric dictionaries learning, in contrast with the common use of predefined feature and dictionaries. Than, the sparse local representations are learned with explicit discriminative goals.
  2. After introducing the discriminative framework, optimization is used for better performance of the model.

Dictionary learning for reconstruction

- **Learning reconstructive dictionaries**

here K-SVD is used for sparse reconstruction.

Input: $\mathbf{D} \in \mathbb{R}^{n\times k}$ (dictionary from the previous iteration); M vectors $\mathbf{x}_l \in \mathbb{R}^n$ (input data); $\alpha \in \mathbb{R}^{k\times M}$ (coefficients from the previous iteration).
Output: $\mathbf{D}$ and α.
Loop: For $j = 1 \ldots k$, update $\mathbf{d}$, j-th column of $\mathbf{D}$,
• Select the set of patches that use $\mathbf{d}$:

$$\omega \leftarrow \{l \in 1, \ldots, M \text{ s.t. } \alpha_l[j] \neq 0\}. \tag{5}$$

• For each patch $l \in \omega$, compute the residual of the decomposition of $\mathbf{x}_l$: $\mathbf{r}_l = \mathbf{x}_l - \mathbf{D}\alpha_l$.
• Compute a new atom $\mathbf{d}' \in \mathbb{R}^n$ and the associated coefficients $\beta \in \mathbb{R}^{|\omega|}$ that minimize the residual error on the selected set ω, using a truncated SVD:

$$\min_{||\mathbf{d}'||_2=1, \beta\in\mathbb{R}^{|\omega|}} \sum_{l\in\omega} ||\mathbf{r}_l + \alpha_l[j]\mathbf{d} - \beta_l\mathbf{d}'||_2^2. \tag{6}$$

• Update $\mathbf{D}$ using the new atom $\mathbf{d}'$, and replace the scalars $\alpha_l[j] \neq 0$ from Eq. (5) in α using β.

MOD can also be used for sparse reconstruction, but K-SVD converges fast.

Dictionary learning for reconstruction

- **A reconstructive approach to discrimination**

  N=no. of training set, S_i = no. of training patches in i^{th} training set

  D_i = Dictionary for i^{th} training set.

we learn dictionary for each class. Approximating each patch using a constant sparsity L and N different dictionaries provides N different residual errors, which can than be used as classification features. So, the naïve way to estimating the class i_o for each patch x ,

$$\hat{i}_0 = \arg\min_{i=1\ldots N} \mathcal{R}^\star(\mathrm{x}, \mathrm{D}_i)$$

Learning discriminative dictionaries

- **Discriminative model**

This model also uses the residual error as a discriminant. Based on the fact that a dictionary D_i associated to its class S_i should be good at reconstructing this class and same time bad at reconstructing other classes. So, we introduces a discriminative term as discriminative cost function.

$$\mathcal{C}_i^{\lambda}(y_1, y_2, \ldots, y_N) \equiv \log\left(\sum_{j=1}^{N} e^{-\lambda(y_j - y_i)}\right)$$

$\lambda > 0$, increasing the value of λ provides higher relative penalty cost for each misclassified patch. In our case,

$$\min_{\{\mathbf{D}_j\}_{j=1}^N} \sum_{\substack{i=1\ldots N \\ l \in S_i}} \mathcal{C}_i^{\lambda}\left(\{\mathcal{R}^\star(\mathbf{x}_l, \mathbf{D}_j)\}_{j=1}^N\right) + \lambda\gamma\mathcal{R}^\star(\mathbf{x}_l, \mathbf{D}_i)$$

$\gamma \geq 0$ controls the trade-off between reconstruction and discrimination.

Learning discriminative dictionaries

Optimization Procedure

Initializes the dictionaries D_i in R^{nxk} for i=1…..N, using few iterations of K-SVD on each class separately(reconstruction approach). Then the main loop at iteration k is composed of

- Sparse coding:- D_i remain fix, we calculate $\alpha_{li} = \alpha^*(x_l,D_i)$ for each patch l and dictionary i. here our main purpose is to compute α^* . Use of this step to make dictionary constructive.
- Dictionary update:- this step will make dictionary more discriminative. once we want to compute new D_i , this is done by updating each single atom of dictionary sequentially. We will also let our α change. For dictionary update we can either use MOD or K-SVD.
- Choice of parameters λ and γ:-

choose a varying λ by choosing it in ascending series.

choose a varying γ by choosing it in descending series. Choose γ value large enough to insure the stability of framework, but as small as possible to enforce the discriminative power.

Learning discriminative dictionaries

Optimization Procedure

- K-SVD for update stage

K-SVD converges faster than MOD and also gives better results. The main idea we use is the local linear approximation of the cost function. So, our algorithm can be written as-

here d' is the eigenvector associated to the largest eigen value of

$$B = \sum_{p=1}^{N} \sum_{l \in S_p \cap \omega} w_l (r_l + \alpha_{li}[j] d)(r_l + \alpha_{li}[j] d)^T$$

Input: N dictionaries $\mathbf{D}_i \in \mathbb{R}^{n \times k}$; M vectors $\mathbf{x}_l \in \mathbb{R}^n$ (input data); S_i (classification of the input data); $\boldsymbol{\alpha} \in \mathbb{R}^{k \times MN}$ (coefficients from the previous iteration).
Output: $\mathbf{D}$ and $\boldsymbol{\alpha}$.
Loop: For $i = 1 \dots N$, for $j = 1 \dots k$, update $\mathbf{d}$, the j-th column of $\mathbf{D}_i$:
- Select the set of patches that uses $\mathbf{d}$:

$$\omega \leftarrow \{l \in 1 \dots M | \boldsymbol{\alpha}_{li}[j] \neq 0\}. \qquad (14)$$

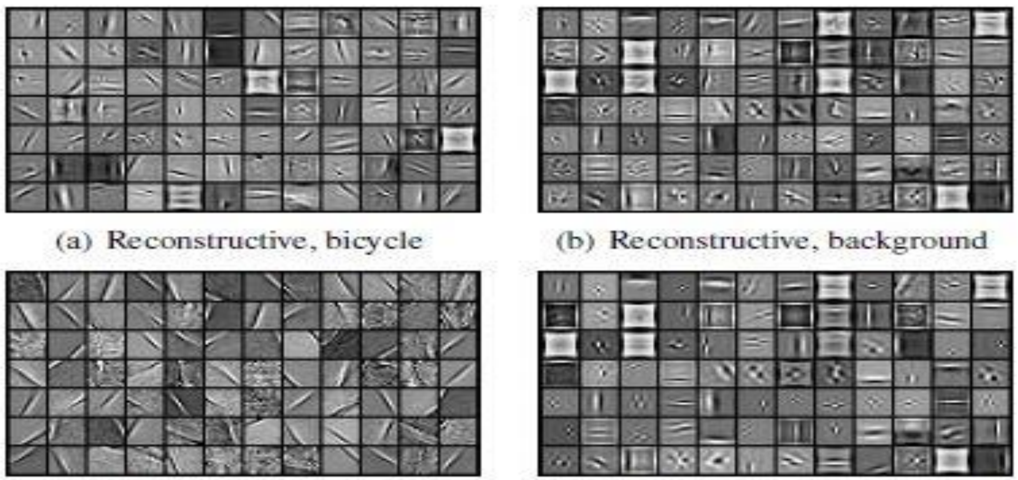
- For each patch l in ω, compute the residual of the decomposition of $\mathbf{x}_l$: $\mathbf{r}_l = \mathbf{x}_l - \mathbf{D}_i \boldsymbol{\alpha}_{li}$.
- Compute the same weights w_l as in Equation (12), for all $p = 1 \dots N$, for all l in $S_p \cap \omega$.
- Compute a new atom $\mathbf{d}' \in \mathbb{R}^n$ and the associated coefficients $\boldsymbol{\beta} \in \mathbb{R}^{|\omega|}$ that minimize the residual error on the selected set ω, using Eq. (13)

$$\min_{\substack{||\mathbf{d}'||_2=1 \\ \boldsymbol{\beta} \in \mathbb{R}^{|\omega|}}} \sum_{\substack{p=1 \dots N \\ l \in S_p \cap \omega}} w_l ||\mathbf{r}_l + \boldsymbol{\alpha}_{li}[j]\mathbf{d} - \boldsymbol{\beta}_l \mathbf{d}'||_2^2. \qquad (15)$$

- Update $\mathbf{D}_i$ and $\boldsymbol{\alpha}$ using the new atom $\mathbf{d}'$, and replace the scalars $\boldsymbol{\alpha}_{li}[j] \neq 0$ from Eq. (14) in $\boldsymbol{\alpha}$ using $\boldsymbol{\beta}$.

Learning discriminative dictionaries

Ex of dictionary learned from a bicycle image.



(a) Reconstructive, bicycle (b) Reconstructive, background

(c) Discriminative, bicycle (d) Discriminative, background

Figure 6. *Parts of the dictionaries, learned on the class 'bicycle' from the Pascal VOC06 dataset. The left part has been learned on bounding boxes containing a bicycle, the right part on background regions. The resulting dictionaries from the two approaches, reconstructive and discriminative, are presented. Visually, the dictionaries produced by the discriminative approach are less similar to each other than with the reconstructive one.*

Learning discriminative dictionaries

- **Data pre-processing**

since this framework is designed for local discrimination. Depending on the particular application on proposed algorithm, different operations can be applied to the input data.

Ex.- Gaussian Mask- give more weight to the center of the patches.

Laplacian filter- helpful in discrimination task.

Depending on the data extra information can be added.

Ex.- using color patches for the K-SVD by concatenating R,G,B information from a patch into single vectors.

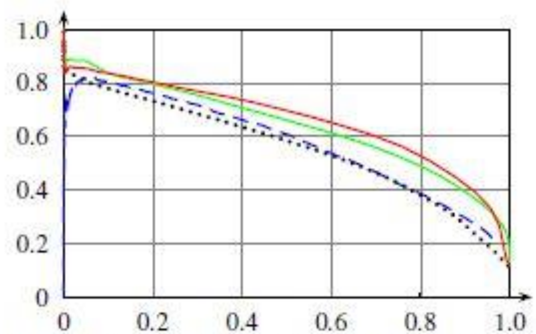


Figure 7. *Precision-recall curve obtained by our framework for the bikes, without pruning of the training dataset (green, continuous), and after 5 pruning iterations (red, continuous), compared with the one from [25] (blue, dashed) and [36] (black, dotted).*

Experimental Results and Applications

Texture segmentation of Brodatz database

this dataset provides a set of patchwork Images composed of texture from diffe- rent classes, and a training sample for Each class. Patches are size of 144(12x12)

Dictionary size=128, sparsity=4.

No. of iterations=30.

Gaussian mask of s.d.=4, and laplacian are Applied for prefiltering.

♯	[28]	[17]	[34]	[16]	R1	R2	D1	D2
1	7.2	6.7	5.5	3.37	2.22	1.69	1.89	**1.61**
2	18.9	14.3	**7.3**	16.05	24.66	36.5	16.38	16.42
3	20.6	10.2	13.2	13.03	10.20	5.49	9.11	**4.15**
4	16.8	9.1	5.6	6.62	6.66	4.60	3.79	**3.67**
5	17.2	8.0	10.5	8.15	5.26	**4.32**	5.10	4.58
6	34.7	15.3	17.1	18.66	16.88	15.50	12.91	**9.04**
7	41.7	20.7	17.2	21.67	19.32	21.89	11.44	**8.80**
8	32.3	18.1	18.9	21.96	13.27	11.80	14.77	**2.24**
9	27.8	21.4	21.4	9.61	18.85	21.88	10.12	**2.04**
10	0.7	0.4	NA	0.36	0.35	**0.17**	0.20	**0.17**
11	**0.2**	0.8	NA	1.33	0.58	0.73	0.41	0.60
12	2.5	5.3	NA	1.14	1.36	**0.37**	1.97	0.78
Av.	18.4	10.9	NA	10.16	9.97	10.41	7.34	**4.50**

Table 1. *Error rates for the segmentation/classification task for the Brodatz dataset. The proposed framework is compared with a number of reported state-of-the-art results [16, 17, 34] and the best results reported in [28]. R1 and R2 denote the reconstructive approach, while D1 and D2 stand for the discriminative one. A Gaussian regularization has been used for R1 and D1, a graph-cut-based one for R2 and D2. The best results for each image are in bold.*

Experimental Results and Applications

Texture segmentation of Brodatz database

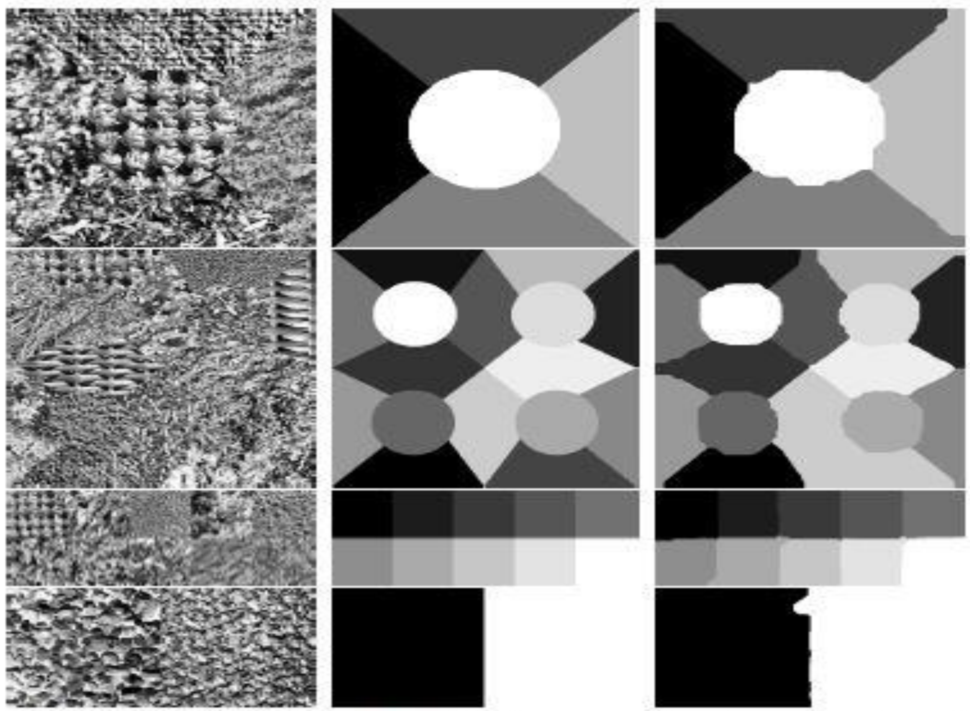


Figure 4. *Subset of the Brodatz dataset with various number of classes: From top to bottom, images 4, 7, 9 and 12. The ground-truth segmentation is displayed in the middle and the resulting segmentation on the right side with a graph-cut regularization. Note that the segmentation is in general very precise but fails at separating two classes on image 7.*

Experimental Results and Applications

Learning discriminant image patches

Two set of patches

S_1 = 200000 12x12 patches extracted from bounding boxes that contain one object of class A. S_2 = 200000 12x12 background patches from an image that contain one object of class A. so, there is overlap between S_1 and S_2 .

Learning of discriminative dictionary:-

Sparsity coefficient =4, no. of iterations=15, k=128.

Result:- after 15 iterations, it pruned the set S_1 by keeping the 90% best classified patches.

Conclusion:- The learned key-patches focus on parts of the object that stand locally discriminative compared to the background.

Experimental Results and Applications

Weakly-supervised feature selection

Without using any ground truth information or bounding box during the training.

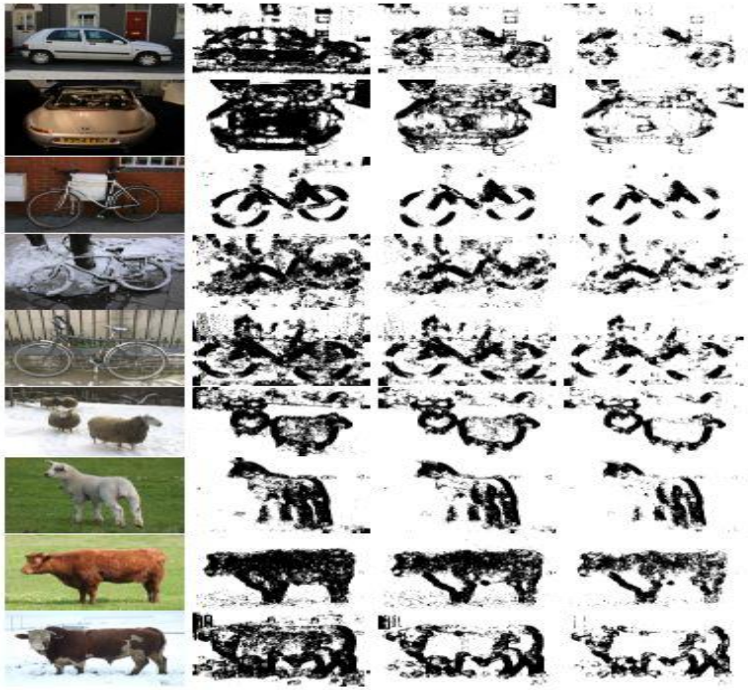


Figure 5. *Learning of key-patches from the Pascal VOC06 dataset. Column 1 presents the test image. Columns 2,3,4 present the raw pixelwise classification results obtained respectively at iterations 20,25 and 30 of the procedure, during the pruning of the dataset. Interestingly, the vertical and horizontal edges of the bicycles are not considered as locally discriminative in an urban environment.*

Conclusions and future directions

- Optimization of the cost function, leading to the learning of over-complete and non-parametric dictionaries that are explicitly optimized to be both representative and discriminative.
- This algorithm has significant improvement over previously published methods. Applied to more general image databases, mainly of natural images.
- It also allow to learn some key-patches of objects and to perform local discrimination.
- For, future work, this could be embedded into a graph-cut-based segmentation framework, which should take into account both the local classification and more global image characteristics.

References

Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro and Andrew Zisserman. Discriminative learned dictionaries for local image analysis. IEEE Trans. 05aug2008.

Questions?

-Thank you