BINOMIAL CLASSIFICATION BASED ON DLENE FEATURES IN SPARSE REPRESENTATION: APPLICATION IN KIDNEY DETECTION IN 3D ULTRASOUND

Mahdi Marsousi and Konstantinos N. Plataniotis

University of Toronto, Toronto, Ontario, Canada,
Email:{marsousi@comm.utoronto.ca}

ABSTRACT

Sparse representation-based classification (SRC) has been recently attracted a great interest among the signal processing society. SRC applies a discriminative representation using training samples to separate signals into their classes. In existing SRC methods, the dictionary size, which highly affects the performance, is manually set. Moreover, they are linear classifiers, and thus, they are not suitable for classifying nonlinear problems. In this paper, we propose a new classification method by cascading a dictionary learning and the neural network to take the advantages of both methods. We use dictionary learning with efficient number of elements (DLENE) to extract discriminative features. We also use the proposed binomial classifier to detect kidneys in 3D ultrasound images. A set of Caltech-101 images are used to compare the proposed method with the state-of-the-art. The proposed kidney detection is evaluated by a set of ultrasound volumes. The results confirm the superiority of our proposed method.

Index Terms— DLENE, Sparse Representation-Based Classification, Dictionary Learning, Kidney Detection

1. INTRODUCTION

Sparse representation-based on classification (SRC) was first proposed by Wright et al. [1], and then, rapidly expanded by many researchers. The idea behind SRC is that, using the sparse representation, samples of a particular class are more likely represented by the training samples of the corresponding class, rather than training samples from other classes. Compared to conventional classification methods, SRC better supports multi-subspace data structures [2]. For N_C number of classes, lets assume $y_i \in \Re^m$ is an input sample, and $D = [D_1, \cdots, D_j, \cdots, D_{N_C}]$, is a database of training samples in which $D_j = [y_{1,j}, \cdots, y_{N_j,j}] \in \Re^{m,N_j}$ is a subset of training samples belonging to the jth class. An input sample (patch) is sparsely represented as,

$$\min_{x_i} \|x_i\|_0 \; subject\, to \; \|y_i - Dx_i\|_2 < \epsilon, \tag{1}$$

where $x_i = [x_{i,1}, \cdots, x_{i,j}, \cdots, x_{i,N_j}]$ is a sparce vector of coefficients, and $x_{i,j}$ is a subvector of coefficients corresponding to the jth class. ϵ is the maximum reconstruction error and relates to the noise level. Then, the classification is performed simply by finding the minimum reconstruction error among $x_{i,j}$s as,

$$class\{y_i\} = \min_j \|y_i - D_j x_{i,j}\|_2. \tag{2}$$

Several methods have been proposed to enhance SRC performane, such as structured-SRC [2], Kernel-SRC [3] and SRC steered discriminative projection [4]. The main challenge of these approaches is that the dictionary requires large numbers of training samples to adequately span subspaces belonging to each class. This reduces the efficiency of representation, and increases the interference of noise in sparse classification. One immediate option to address this challenge is to use dictionary learning which reduces the number of required dictionary entries.

Since the introduction of K-SVD by Aharon et al. (2006) [5], learnt-based dictionaries have became very popular in sparse representation, since it improves the representation power with a fewer number of learnt dictionary atoms. Dictionary learning is generally formulated as,

$$\langle D, X \rangle = \min_{X,D} \sum_i \|y_i - Dx_i\|_2 \, subject\, to \; \|x_i\|_0 \leq \Gamma, \tag{3}$$

where X is the matrix of all sparse vectors, and Γ controls the sparsity level. K-SVD [5] solves the equation (3) simply and efficiently by adopting singular value decomposition, and thus, it has been widely used in different applications. Although, the dictionary learning problem of (3) is powerful for sparsely representing the entire dataset, it does not provide enough separability for different classes. Zhang and Le (2010) [6] has proposed the discriminative K-SVD (D-KSVD) method to learn dictionaries which are capable to distinguish patches of different classes, which is formulated as,

$$\begin{gathered}\langle D, W, X \rangle = \min_{X,W,D} \textstyle\sum_i \left\| \begin{pmatrix} y_i \\ \lambda h_i \end{pmatrix} - \begin{pmatrix} D \\ \lambda W \end{pmatrix} x_i \right\| \\ subject\, to \; \|x_i\|_0 \leq \Gamma, \end{gathered} \tag{4}$$

where each $h_i \in \Re^{N_C}$ has a non-zero element in jth row corresponding to the class of y_i, and λ is the regularization parameter. W is the weight matrix of the linear classifier,

$h_i = Wx_i$. The equation (4) implies that W and D are simultaneously trained while dictionary atoms are updated. This results in supporting both representation and class discrimination powers. The solution of (4) is similar to the K-SVD approach followed by an extra normalization step. The D-KSVD method has been followed by other researches, such as [7], [8] and [9]. As shown in [8], the number of dictionary atoms plays an important role in specificity and separability of the learnt-dictionary, though the existing methods do not consider an adaptive selection of a dictionary size to span sub-spaces of all classes. Moreover, all of them are linear classifiers in essence, and are not suitable for classifying non-linear problems.

1.1. Our contributions

In this paper, we propose a new binomial classification scheme for discriminating an object from its background by cascading a dictionary learning method and the neural network (NN) classifier [10]. The combination is used to take the advantages of: (i) supporting multi-subspace data structures by dictionary learning, and (ii) supporting non-linear classification by NN. We use the dictionary learning with efficient number of elements (DLENE) [11], which adaptively selects the number of dictionary atoms based on the structural complexity of the training dataset. DLENE is used to separately learn two dictionaries for background-class and object-class. Both dictionaries are learned for the same reconstruction error, while the object-class dictionary is learned with a higher sparsity level, compared with the background-class dictionary. This results in increasing the specificity of the object representation, and commonality of non-object and background representation. After sparsely representing training samples using the learnt dictionaries, the DLENE output including the sparsity level and indexes of major coefficients in sparse vectors are used as two sets of features for training the NN. The proposed binomial classification is used in a processing pipeline to automatically detect kidneys in 3D abdominal ultrasound images. The rest of the paper is organized as follows: in section (2), the proposed binomial classification, DLENE-NN, is represented in details; Also, its application in kidney detection is described; Section 3 provides experimental setup and results of the proposed method, and finally in section 4, conclusion, discussion, and limitations are presented.

2. THE PROPOSED METHOD

In this section, we first describe the DLENE method, and then, we apply it in our proposed classification to separate samples of two classes: background (bg), and object (obj). Finally, we apply the proposed method for kidney detection.

2.1. Using DLENE for Adaptive Dictionary Learning

DLENE [11] is an adaptive dictionary learning approach which automatically selects the number of atoms, given two parameters: (i) desired sparsity level, average number of non-zero coefficients ($ANNZC_{des}$), and (ii) desired reconstruction error, root-mean-square-error ($RMSE_{des}$). DLENE starts with an initial dictionary of two atoms, and then, automatically spreads the dictionary until the sufficient number of atoms are obtained by capturing all data structures to meet the desired parameters. DLENE is formulated as [11],

$$\begin{aligned}\min_{N_D,X,D}\ N_D\ \ subject\,to\ \textstyle\sum_i \|y_i - Dx_i\|_2^2 \leq N_j RMSE_{des}^2 \\ and\ sum_{k=1}^{N_j}\|x_k\|_0 \leq N_j ANNZC_{des},\end{aligned} \tag{5}$$

where N_D is the number of atoms in D. We apply DLENE to learn two dictionaries for two classes, D^{bg} and D^{obj}. We impose two constraints on the dictionary learning problem to provide discriminative learnt-based dictionaries as,

1. Each class is trained with its own training samples;
2. D^{bg} and D^{obj} are learned with equal $RMSE_{des}$, and different sparsity levels, $ANNZC_{des}^{obj} < ANNZC_{des}^{bg}$.

The second constraint makes D^{obj} to be specific to the object-class training samples by having a small $ANNZC_{des}^{obj}$, while D^{bg} is trained with a high $ANNZC_{des}^{obj}$ to represent a wide range of non-object data structures in more condensed format.

2.2. DLENE-NN: Binomial Classification

2.2.1. *Training*

After learning dictionaries, D^{bg} and D^{obj}, two types of discriminative features are extracted from sparse vectors, x_i:

1. Sparsity level: $f_i^{SL} = \|x_i\|_0$;
2. indexes of maximum coefficients in sparse vectors, $f_i^{ind} = [ind_1, \cdots, ind_{N_k}]$.

To find f_i^{ind}, we first sort non-zero coefficients in x_i in the ascent order. Then, we select the first N_K indexes. Since D^{obj} is learned to represent an obj-sample with a few non-zero coefficients, and D^{bg} is learned to represent a bg-sample with more non-zero coefficients, the sparsity level, f_i^{SL}, is a discriminative feature for separating samples of the two classes. For each sample, x_i we have a feature vector, $F_i = [f_i^{SL}, f_{i,1}^{ind}, \cdots, f_{i,N_k}^{ind}]^T \in \Re^{N_k+1}$. By collecting all feature vectors for the bg and obj classes, two feature matrices are obtained, F^{bg} and F^{obj}. These features along with their corresponding labels, L^{bg} and L^{obj} are then used to train a NN with N_{HL} number of hidden layers.

Algorithm 1: Pseudo-algorithm of Proposed Method

input : train:$\{y_i\}^{bg}$,$\{y_i\}^{obj}$, eval: y_i, $ANNZC^{bg}_{des}$, $ANNZC^{obj}_{des}$, $RMSE_{des}$, N_{HL},N_K

begin

Training Step:

1. **learning dictionaries**: $D^{bg} = DLENE(ANNZC^{bg}_{des}, RMSE_{des}, \{y_i\}^{bg})$; $D^{obj} = DLENE(ANNZC^{obj}_{des}, RMSE_{des}, \{y_i\}^{obj})$;
2. **Sparse coding**: $X^{bg} = BatchOMP([D^{bg}, D^{obj}], \{y_i\}^{bg})$; $X^{obj} = BatchOMP([D^{bg}, D^{obj}], \{y_i\}^{obj})$;
3. **Extracting features from training samples**: $f^{SL}_i = \|x_i\|_0$, and $f^{ind}_i = [ind_1, \cdots, ind_{N_k}]$ $\forall i$;
4. **Creating feature matrices**: F^{bg} and F^{obj};
5. **Training NN**: $Net = NNTrain([F^{bg}, F^{obj}], [L^{bg}, L^{obj}], N_{HL})$;

Classifying Step:

1. **Sparse Coding**: $x_{in} = BatchOMP([D^{bg}, D^{obj}], y_{in})$
2. **Extracting features**: $f^{SL}_{in} = \|x_i\|_0$, and $f^{ind}_{in} = [ind_1, \cdots, ind_{N_k}]$;
3. **Creating feature vector**: F_{in};
4. **Classify with NN**: $class\{y_{in}\} = NNClassify(Net, F_{in})$

2.2.2. *Classification*

For classifying incoming samples into bg and obj classes, we cascade sparse coding (using the learned dictionaries) and the trained NN. For each incoming sample, y_{in}, we calculate its sparse vector, x_{in}, using the batch OMP approach [12] with $D = [D^{bg}, D^{obj}]$. It is important to preserve the order of dictionaries and their atoms in both training and Classifying stages, since the indexes in sparse vectors are used as the features. Then, the feature vector F_{in} is generated, and the sample is classified using the trained NN. The algorithm of the proposed method is represented in Algorithm 1.

2.3. Application in Kidney Detection and Segmentation

Automated kidney diagnosis in 3-D ultrasound has a vital significance in abdominal trauma detection. It has been shown that an internal bleeding can be detected as a free fluid, which appears as a dark region in an ultrasound [13], in a region between the left kidney and left lung [14]. Thus, kidney detection is a necessary step for trauma detection, and automated kidney detection contributes to automated trauma detection, which promotes emergency trauma diagnosis. The kidney has a unique structure in 3-D ultrasound among all other internal organs, which makes the kidney to be distinguishable. But, the automated detection of kidney is very challenging: (1) low contrast profile, inhomogeneous intensity profile and speckle noise result in low quality ultrasound images; (2) There exist gaps among the kidney boundary; and (3) the kidney might be partially visible due to shadows caused by stones or misaligned probe [15]. Due to these challenges, kidney detection and segmentation using 3-D ultrasound has been only investigated by a few researches [16] [17]. Here, we extend our previous kidney detection method [17] by applying the proposed DLENE-NN. The block diagram of the proposed kidney detection is shown in Fig 1.

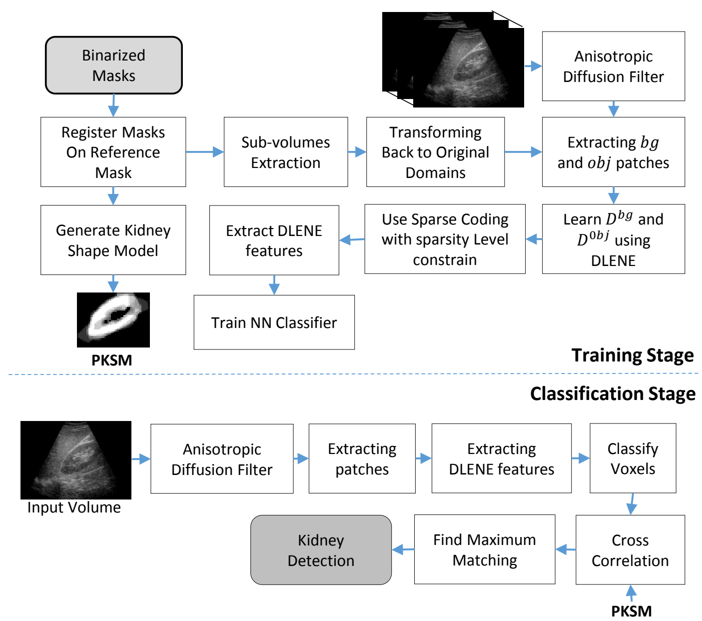

Fig. 1. Displaying block diagram of the kidney detection.

We collect a set of training volumes with their ground-truth data (segmented kidneys as binarized masks). To reduce speckle noise of each training volume, the anisotropic diffusion filter is applied [18]. The training volumes and their binarized masks are used for two purposes: (1) to train dictionaries and the NN, and (2) to generate a probabilistic kidney shape model. Since the kidney shape has a high level of variability, we split the kidney shape into 18 sub-volumes to reduce the complexity of each sub-volume. The sub-volumes are extracted by evenly splitting the shape into 3, 3 and 2 divisions along x-, y- and z- axes, respectively. To achieve this, we select a training volume as a reference, and register all other training volumes on the reference volume, based on the affine transformation with manually specified landmarks. Then, the registered volumes are divided on the same lines into 18 sub-volumes, and finally, the divided sub-volumes of each training volume is transformed back into its original space. Now, for each sub-volume, we use the corresponding binarized mask to extract patches which are placed on the kidney, and use the patches to learn a dictionary for the corresponding sub-volume using the DLENE method. The obj-class dictionary is formed by combining the learnt-dictionaries of all the sub-volumes. Then, we extract non-kidney patches from all training volumes, and use the patches to learn the bg-class dictionary. Having the registration transforms of the training volumes on the reference volume, we generate a probabilistic kidney shape model by finding the average of registered masks [17], and we call it, the probabilistic kidney shape model ($PKSM$).

In the kidney detection stage, an input volume, V_{in}, is first enhanced by the anisotropic diffusion filter. Then, patches are extracted from the enhanced volume, and are classified with the trained DLENE-NN classifier. The classification result, C_{in}shows the probability of a voxel in the input volume to belong to obj-class. Then, the maximum matching is obtained by calculating the maximum spatial cross correlation of $PKSM$ and C_{in}.

3. EXPERIMENTS AND RESULTS

We developed two sets of experiments, aiming to evaluate the performance of the proposed DLENE-NN: (1) comparing DLENE-NN with D-KSVD [6], and (2) evaluating the presented kidney detection method and compare with our previously presented method [17].

For the first experiment, we used the Caltech-101 database [19] (30 volumes of car-side category), and manually generated masks to specify cars from backgrounds. The objective of the first experiment is to classify each pixel into obj (car) and bg (non-car) classes. We trained our DLENE-NN using 6 arbitrarily selected images, and evaluated the methods with the rest of images. The patch sizes for both DLENE-NN and D-KSVD are set to 21×21 $pixel^2$. In both methods, patches are extracting with 90% of pixel overlapping. For learning dictionaries with DLENE, parameters are set to $RMSE_{des} = 0.045$, $ANNZC_{des}^{obj} = 5$ and $ANNZC_{des}^{obj} = 12$. We have empirically selected $N_{HL} = 30$. For the training D-KSVD method, we set the sparsity level $\Gamma = 16$, and the dictionary contains 1024 atoms. We applied the Dice's coefficient to calculate the detection accuracy. For D-KSVD and DLENE-NN, the classification results are $dice = 0.2584 \pm 0.0698$ and $dice = 0.3385 \pm 0.0948$. Classification of a sample image from the Caltech-101 database is shown in Fig 2. According to the obtained results, the DLENE-NN is performing better than D-KSVD on average.

For the second experiment, we have utilized a database of 28 3-D ultrasound images, in which 14 volumes are correct Morison's pouch views with kidneys, and 14 volumes are randomly selected without kidneys. For the volumes with kidneys, ground truth data of their kidneys as binarized masks are manually generated. 4 binarized masks are used to generate PKSM, and the rest are used for evaluating our kidney detection. The detection accuracy is computed using $\frac{(\#true\, positive\, detections + \#true\, negative\, detections)}{(\#total\, number\, of\, detections)}$, where

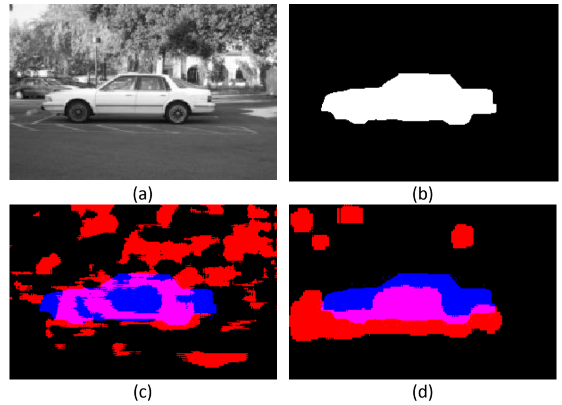

Fig. 2. Classifying pixels belonging to a car from the caltech-101 database [19]. (a) original image, (b) binarized mask, (c) D-KSVD, and (d) DLENE-NN.

the detection accuracy for the previous approach [17] and DLENE-NN are obtained as $\%92.86$ and $\%96.43$, respectively. The class separability index, $\frac{N_{C1}(\mu_{C1}-\mu)^2 + N_{C2}(\mu_{C2}-\mu)^2}{\sigma_{C1}^2 + \sigma_{C2}^2}$, of the DLENE-NN based Kidney detection and our previous method [17] are 30.75 and 27.31, respectively. The DLENE-NN based kidney detection provides higher accuracy, however its computational cost is also relatively higher than [17].

4. CONCLUSION

We have proposed a new sparse representation-based classification by cascading an adaptive size dictionary learning, DLENE, with the neural network. The discriminativity of the sparse representation comes from two characteristics of the proposed method: (i) learning a separated dictionary for each of the classes, and (ii) setting a low ANNZC for obj-class and a high ANNZC for bg-class. By combining DLENE and NN, we take the advantage of both SRC methods and NN classifier: (i) Our method is able to span multiple sup-spaces using sparse representation-based features; (ii) Since we use the NN classifier, the proposed classification method is able to classify non-linear problems. The proposed method is applied to detect kidneys in 3D abdominal ultrasound images. As a limitation of the proposed method, it does not support multi-classes problems, which can be addressed in our future research.

5. REFERENCES

[1] J. Wright, AY. Yang, A Ganesh, S.S. Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, Feb 2009.

[2] E. Elhamifar and R. Vidal. Robust classification using structured sparse representation. In *2011 IEEE Con-*

ference on Computer Vision and Pattern Recognition (CVPR), pages 1873–1879, June 2011.

[3] Li Zhang, Wei-Da Zhou, Pei-Chann Chang, Jing Liu, Zhe Yan, Ting Wang, and Fan-Zhang Li. Kernel sparse representation-based classifier. *IEEE Transactions on Signal Processing*, 60(4):1684–1695, April 2012.

[4] Jian Yang, Delin Chu, Lei Zhang, Yong Xu, and Jingyu Yang. Sparse representation classifier steered discriminative projection with applications to face recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 24(7):1023–1035, July 2013.

[5] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

[6] Qiang Zhang and Baoxin Li. Discriminative k-svd for dictionary learning in face recognition. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2691–2698, June 2010.

[7] Shu Kong and Donghui Wang. A dictionary learning approach for classification: Separating the particularity and the commonality. 7572:186–199, 2012.

[8] Zhuolin Jiang, Zhe Lin, and L.S. Davis. Label consistent k-svd: Learning a discriminative dictionary for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2651–2664, Nov 2013.

[9] Ning Zhou and Jianping Fan. Jointly learning visually correlated dictionaries for large-scale visual recognition applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(4):715–730, April 2014.

[10] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[11] M. Marsousi, K. Abhari, P. Babyn, and J. Alirezaie. An adaptive approach to learn overcomplete dictionaries with efficient numbers of elements. *IEEE Transactions on Signal Processing*, 62(12):3272–3283, June 2014.

[12] Ron Rubinstein, Michael Zibulevsky, and Michael Elad. Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit. Technical report, 2008.

[13] K. Matsushima and H.L. Frankel. Beyond focused assessment with sonography for trauma: ultrasound creep in the trauma resuscitation area and beyond. *Curr Opin Crit Care*, 17(6):606–612, 2011.

[14] D. V. Feliciano B. D. Thomas B. R. Boulanger F. E. Davis R. E. Falcone J. A. Schmidt G. S. Rozycki, M. G. Ochsner. Early detection of hemoperitoneum by ultrasound examination of the right upper quadrant: A multicenter study. *The Journal of Trauma: Injury, Infection, and Critical Care*, 45(5):878–883, 1998.

[15] A. Belaid, D. Boukerroui, Y. Maingourd, and J. F Lerallut. Implicit active contours for ultrasound images segmentation driven by phase information and local maximum likelihood. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 630–635, 2011.

[16] Marcos Martin-Fernandez and Carlos Alberola-Lopez. An approach for contour detection of human kidneys from ultrasound images using markov random fields and active contours. *Medical image analysis*, 9(1):1–23, 2005.

[17] Mahdi Marsousi, Konstantinos N. Plataniotis, and Stergios Stergiopoulos. Shape-based kidney detection and segmentation in three-dimensional abdominal ultrasound images. In *2014 IEEE Conference on engineering in medicine and biology (EMBC-2014), (accepted)*, 2014.

[18] Yongjian Yu and S.T. Acton. Speckle reducing anisotropic diffusion. *Image Processing, IEEE Transactions on*, 11(11):1260–1270, 2002.

[19] Li Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04. Conference on*, pages 178–178, June 2004.