Histopathology Image Categorization with Discriminative Dimension Reduction of Fisher Vectors

Anonymous submission

Paper ID 10

Abstract. In this paper, we present a histopathology image categorization method based on Fisher vector descriptors. While Fisher vector has been broadly successful for general computer vision and recently applied to microscopy image analysis, its feature dimension is very high and this could affect the classification performance especially when there is small amount of training images available. To address this issue, we design a dimension reduction algorithm in a discriminative learning model with similarity and representation constraints. In addition, to obtain the image-level Fisher vectors, we incorporate two types of local descriptors based on the standard texture feature and unsupervised feature learning. We use three publicly available datasets for experiments. Our evaluation shows that our overall approach achieves consistent performance improvement over existing approaches, our proposed discriminative dimension reduction algorithm outperforms the common dimension reduction techniques, and different local descriptors have varying effects on different datasets.

1 Introduction

Tissue examination using histopathology images is regularly performed in the clinical routine for cancer diagnosis and treatment. Computerized histopathology image classification supports automated categorization of cancer status (benign or malignant) or subtypes, where manual analysis can be subjective and error-prone due to the complex visual characteristics of histopathology images. The majority of existing studies in this area have focused on image feature representation. For example, custom feature descriptors have been designed based on structures of cells or regions [1–3]. Other approaches propose to use automated feature learning techniques such as convolutional sparse coding [4], dictionary learning [5], and deconvolution network [6].

As a less studied feature representation for biomedical imaging, Fisher vector (FV) [7] has recently been applied to microscopy image analysis [6, 3]. FV is a feature encoding algorithm that aggregates a dense set of locally extracted features, such as dense scale-invariant feature transform (DSIFT) descriptors, based on the Gaussian mixture model (GMM) to form a high-dimensional image-level

descriptor. Classification of FV descriptors is then performed using the linear-kernel support vector machine (SVM). This approach has shown excellent performance in many general imaging applications such as face recognition, texture classification, and object detection.

On the other hand, due to the high dimensionality, the classification performance using FV descriptors could be affected if there are insufficient training data to represent the complex visual characteristics of the problem domain. This issue would be especially important for histopathology image studies, due to the cost of preparing training data. While it could be intuitive to apply standard dimension reduction techniques such as principal component analysis (PCA), it has been shown that with FV descriptors, discriminative dimension reduction is more effective for face recognition [8]. However, we are not aware of existing studies on the design of suitable dimension reduction algorithms for FV descriptors of histopathology images.

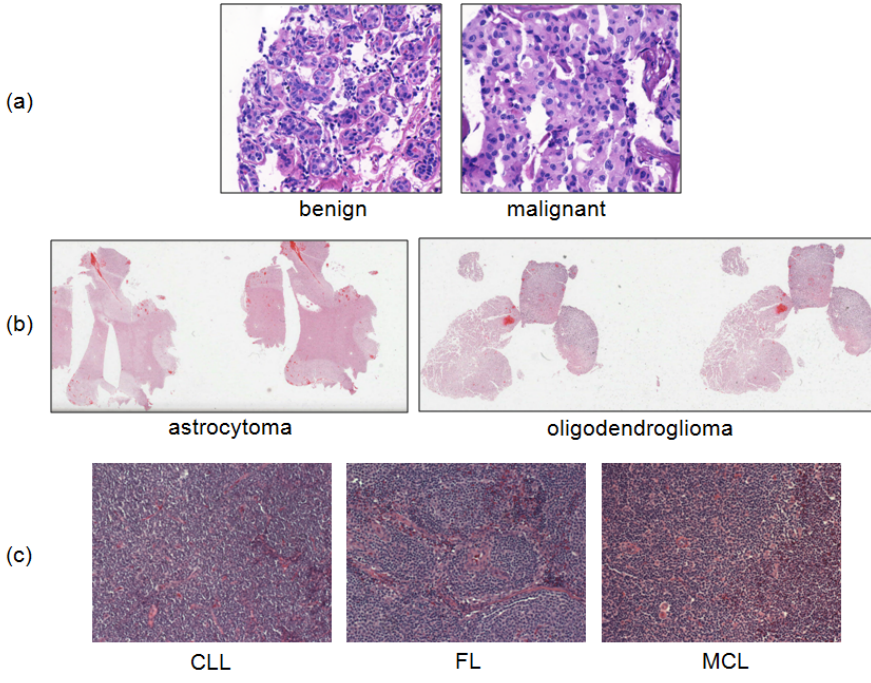


Fig. 1. Example images of three datasets: (a) UCSB breast cancer dataset, (b) MICCAI 2015 CBTC challenge training set, and (c) IICBU 2008 malignant lymphoma dataset. Each image represents one image category.

In this work, we propose an automated method to categorize histopathology images. Our algorithm contributions are two-fold. First, we design two FV

descriptors to represent the histopathology images, based on the local DSIFT features and patch-level features that are learned using a deep belief network (DBN) [9] model, respectively. Second, to address the issue of small training data, we design a discriminative dimension reduction method to reduce the feature dimension and enhance the discriminative power of FV descriptors. Our design is inspired by [8], but we find direct application of this method reduces the classification performance on some datasets. We thus further improve the model [8] by devising a classification score-based training set selection technique and introducing an additional representation constraint into the optimization objective. For evaluation, three public histopathology image datasets (Fig. 1) are used. We demonstrate better categorization performance compared to other feature representation and dimension reduction techniques.

2 Methods

2.1 Fisher Vector Descriptors

FV is analogous to the bag-of-words (BOW) encoding that it encodes a dense set of local descriptors into an image-level descriptor. The encoding works by first generating a GMM from all local descriptors. Based on the soft assignments of local descriptors to the Gaussian modes, the first and second-order difference vectors between the local descriptors and each of the Gaussian centers are computed. The FV descriptor is then the concatenation of all difference vectors. Assume that the local descriptor is d dimensional and k Gaussian modes are used. The resultant FV descriptor is $2dk$ dimensional.

The local descriptors can be any patch-level features. In this study, we use two types of local descriptors. 1) Following the standard FV computation, DSIFT descriptors are extracted at multiple scales with spatial bins of 4, 6, 8, 10 and 12 pixels and sampled every two pixels. 2) Unsupervised feature learning using DBN is performed on half-overlapping patches of 8×8 pixels. The network comprises two layers with each layer producing 64 features. The patch-level local descriptor is the concatenation of features from the two layers. The use of unsupervised feature learning is inspired by existing work [6, 10], which shows that unsupervised feature learning is highly effective and can be more representative than supervised feature learning for microscopy images. We have also evaluated DBN with other numbers of layers and nodes, using only the features from the last layer, or generating local descriptors at multiple scales. The proposed structure is found to provide the best results.

For each type of local descriptor, we set a common feature dimension of 128 to make the different local descriptors directly comparable. PCA is then used to reduce the local descriptor dimension to 64, and based on which, GMM of 64 modes is generated. The image-level FV descriptor is consistently $d = 2 \times 64 \times 64$ dimensional for each of the two types of local descriptors.

2.2 Discriminative Dimension Reduction

We design a discriminative learning-based dimension reduction algorithm, which helps to enhance the discriminative power of the FV descriptor. Formally, given a training set of n images $\{I_i : i = 1, \ldots, n\}$, the FV descriptor (regardless of the local descriptor type) of image I_i is denoted as f_i with category label y_i. The objective is to learn a linear projection matrix $W \in \mathbb{R}^{h \times d}$ with $h \ll d$ indicating the reduced dimension. The descriptor f_i is then transformed to Wf_i, which is expected to meet the similarity and representation constraints. The dimension-reduced descriptors are finally used to train a linear-kernel SVM to categorize the histopathology images. The overall method flow is illustrated in Fig. 2.

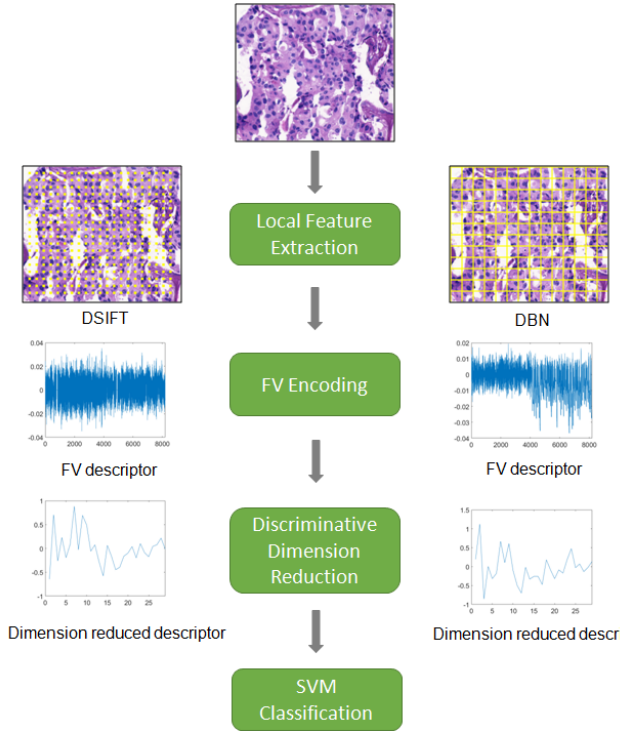


Fig. 2. Illustration of our method design. To categorize an image, the local DSIFT or patch-wise DBN features are first extracted then encoded as FV descriptors. These descriptors are then dimension reduced using our discriminative dimension reduction method, and finally classified using linear-kernel SVM. Unsupervised learning is involved in DBN feature extraction and FV encoding, and supervised learning is used in discriminative dimension reduction and SVM classification.

Similarity constraint. This constraint specifies that after dimension reduction, descriptors of the same class should be similar and descriptors of different

classes should be dissimilar. As a result, such descriptors would be more easily classified compared to the original FV descriptors. We formulate this constraint based on the Euclidean distance between pairs of descriptors:

$$\|Wf_i - Wf_j\|_2^2 < b, \ \forall y_i = y_j; \quad \|Wf_i - Wf_j\|_2^2 > b, \ \forall y_i \neq y_j; \tag{1}$$

where i and j index the n training data, and b is a learned threshold. By imposing a margin of at least one, the constraint is rewritten as:

$$\theta_{ij}(b - \|Wf_i - Wf_j\|_2^2) > 1 \tag{2}$$

where θ_{ij} is 1 if $y_i = y_j$, and -1 otherwise. This is equivalent to the following learning objective:

$$\underset{W,b}{\operatorname{argmin}} \sum_{i,j} \max[1 - \theta_{ij}(b - (f_i - f_j)^T W^T W (f_i - f_j)), 0] \tag{3}$$

An iterative optimization process can be used to update W and b by incorporating one pair of descriptors at each iteration [8].

An important issue of this learning process is the selection of descriptor pairs f_i and f_j. If all possible pairs of training data are enumerated, there would be a large number of descriptor pairs imposing contradicting optimization goals hence affecting the effectiveness of the learned matrix. To overcome this issue, we design a classification score-based training data selection approach. Specifically, our idea is that for each f_i, we select positive f_j (i.e. $\theta_{ij} = 1$) that has high classification score of class y_j, and negative f_j (i.e. $\theta_{ij} = -1$) that has low classification score of class y_j. The positive f_j represents the good data that f_i should move towards, and the negative f_j represents the hard examples that f_i should move further from. The classification score is the probability estimate from a linear-kernel SVM trained using all the training data. The scores of all training data from one class are then sorted and the score at the pth percentile is chosen as the threshold. The positive (or negative) f_j having a score above (or below) this threshold is selected to form a training pair with f_i. In addition, we further reduce the number of descriptor pairs by restricting that only images with classification scores lower than the threshold are used as f_i, so that training will focus on hard examples.

Representation constraint. While the similarity constraint measures the pairwise similarity between descriptors, the representation constraint evaluates the representativeness of the descriptor by the overall class space. In the lower dimensional space, we expect the descriptor f_i to be well represented by the correct class only. Specifically, assume that there are n_j descriptors from class j, and the descriptor set is denoted as $R_j \in \mathbb{R}^{d \times n_j}$ (note for simplicity j is now used to index the classes). The representation of f_i by class j, defined as $\mu_{i,j}$, is computed as the sparse reconstruction from R_j, $\mu_{i,j} = R_j c_{i,j}$. Here $c_{i,j} \in \mathbb{R}^{n_j}$ is a weight vector with q nonzero elements and is derived analytically using locality-constrained linear coding (LLC) [11]:

$$\min_{c_{i,j}} \|f_i - R_j c_{i,j}\|^2 + \lambda \|v_{i,j} \odot c_{i,j}\|^2 \quad s.t. \mathbf{1}^T c_{i,j} = 1, \ \|c_{i,j}\|_0 = q \tag{4}$$

where $v_{i,j} \in \mathbb{R}^{n_j}$ contains the Euclidean distances between f_i and each descriptor in R_j, and the constant $\lambda = 0.01$. Note that f_i is excluded from R_j if $y_i = j$.

The representation constraint is then defined in a similar construct to the similarity constraint, but replacing the distance between pairs of descriptors by distance between the descriptor and its reconstructions. Formally, this constraint is formulated as:

$$\theta_{ij}(b - \|Wf_i - W\mu_{i,j}\|_2^2) > 1 \tag{5}$$

where $\theta_{ij} = 1$ if $y_i = j$, and otherwise $\theta_{ij} = -1$. The learning objective is thus:

$$\underset{W,b}{\operatorname{argmin}} \sum_{i,j} \max[1 - \theta_{ij}(b - (f_i - \mu_{i,j})^T W^T W(f_i - \mu_{i,j})), 0] \tag{6}$$

Note that we can set multiple q values $q = \{q_1, q_2, \ldots\}$, so that for f_i and each class j, we obtain multiple representations $\{\mu_{i,j} : q = q_1, q_2, \ldots\}$ and expect each of the representation to satisfy Eq. (5). In this way, we increase the number of training pairs for the representation constraint.

Algorithm 1: Discriminative Dimension Reduction

Data: Training data $\{f_1, y_1\}, \ldots, \{f_n, y_n\}$.
Result: Linear projection matrix W.
Train a linear kernel SVM on the training data;
Based on the classification scores of the training data, create a set of training pairs $\{(f_i, f_j), i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\}\}$;
Initialize W using PCA on the training data;
Initialize b as the threshold that can satisfy $\theta_{ij}(b - \|Wf_i - Wf_j\|_2^2) > 1$ for most training pairs;
repeat
  for *each training pair* (f_i, f_j) **do**
    if $\theta_{ij}(b - \|Wf_i - Wf_j\|_2^2) \leq 1$ **then**
      $\Delta_{ij} = (f_i - f_j)(f_i - f_j)^T$;
      $W_{t+1} = W_t - \gamma\theta_{ij}W_t\Delta_{ij}, \;\; b = b + \alpha\theta_{ij}$;
    end
  end
  for *each training data* f_i*, class* j*, and parameter* q **do**
    $\min_{c_{i,j}} \|f_i - R_j c_{i,j}\|^2 + \lambda\|v_{i,j} \odot c_{i,j}\|^2 \;\; s.t. \mathbf{1}^T c_{i,j} = 1, \; \|c_{i,j}\|_0 = q$;
    $\mu_{i,j} = R_j c_{i,j}$;
    if $\theta_{ij}(b - \|Wf_i - W\mu_{i,j}\|_2^2) \leq 1$ **then**
      $\Delta_{ij} = (f_i - \mu_{i,j})(f_i - \mu_{i,j})^T$;
      $W_{t+1} = W_t - \gamma\theta_{ij}W_t\Delta_{ij}, \;\; b = b + \alpha\theta_{ij}$;
    end
  end
until *convergence or maximum number of iterations is reached*;

Combined optimization. W is first initialized using PCA on the training data. Then, W and b are learned using a stochastic sub-gradient method combin-

ing Eqs. (3) and (6). First, at each iteration t, the algorithm takes a descriptor pair f_i and f_j. If Eq. (2) is not met, W and b are updated by:

$$W_{t+1} = W_t - \gamma\theta_{ij}W_t\Delta_{ij}, \quad b = b + \alpha\theta_{ij} \tag{7}$$

where $\Delta_{ij} = (f_i - f_j)(f_i - f_j)^T$, γ and α are constant learning rates and default to 0.25 and 1. After the selected descriptor pairs are enumerated, the training pair f_i and $\mu_{i,j}$ is used at each iteration to update W and b using Eq. (7) with $\Delta_{ij} = (f_i - \mu_{i,j})(f_i - \mu_{i,j})^T$. The iteration continues until convergence or maximum number of iterations is reached. The overall discriminative dimension reduction method is summarized in Algorithm 1.

With the derived linear projection matrix W, a descriptor f_i is dimension reduced to Wf_i. SVM training and classification are then performed on these lower dimensional descriptors to obtain the image category. Our empirical analysis shows that a linear-kernel SVM is more effective than the polynomial and radial basis function (RBF) kernels.

2.3 Datasets and Implementation

We use three public histopathology image datasets: 1) UCSB breast cancer dataset of 58 hematoxylin and eosin (H&E) stained tissue microarray (TMA) images, including 32 benign (B) and 26 malignant (M) cases; 2) MICCAI 2015 CBTC challenge training set, containing 32 whole-slide images of brain tumors with 40x apparent magnification, out of which 16 are astrocytoma (A) and 16 are oligodendroglioma (O); and 3) IICBU 2008 malignant lymphoma dataset of 374 H&E stained image sections from brightfield microscopy, including 113 chronic lymphocytic leukemia (CLL), 139 follicular lymphoma (FL), and 122 mantle cell lymphoma (MCL) cases. The first two datasets represent problems with small numbers of training data, and the third dataset demonstrates multi-class categorization. Example images are shown in Fig. 1.

For all datasets, we employ four-fold cross validation with three parts of the data for training and one part for testing, following the same protocol used in existing studies for the UCSB breast cancer dataset [2, 12]. For the first two datasets, we set h (the reduced feature dimension) to half of the number of images and p (the pth percentile of classification score) to 80. For the third dataset, due to the relatively large number of images, h is set to a quarter of the number of images, and p is set to 60. For the representation constraint, we set q to 3 and 5 for all datasets. These parameter settings are determined using 10-fold cross validation within the training set.

3 Experimental Results

We first evaluate the effect of different local descriptors (DSIFT, DBN) and feature encoding algorithms (BOW, FV), and our proposed method. Table 1 shows that in general FV encoding outperformed the BOW encoding by a large extent.

DBN-FV is particularly effective for the CBTC dataset while DSIFT-FV is more descriptive for the other two datasets. With our proposed discriminative dimension reduction (DDR), we achieve the best performance for all three datasets. DBN-FV-DDR is more effective for the CBTC dataset while DSIFT-FV-DDR is more effective for the IICBU dataset. Overall, the results suggest that FV can provide more discriminative power than BOW, but the choice of local descriptors (DSIFT or DBN) needs to be evaluated for different datasets.

Table 1. Classification accuracy (%) comparing feature representations.

	UCSB	CBTC	IICBU
DSIFT-BOW	47.9	53.1	73.8
DBN-BOW	44.5	43.8	70.9
DSIFT-FV	87.8	50.0	90.9
DBN-FV	86.2	65.6	90.1
DSIFT-FV-DDR (proposed)	**89.7**	56.3	**93.3**
DBN-FV-DDR (proposed)	**89.7**	**75.0**	91.2

Fig. 3 shows the results using the original FV descriptor (no dimension reduction), our proposed discriminative dimension reduction (DDR) algorithm, and the other popular dimension reduction techniques: PCA (unsupervised), linear discriminative analysis (LDA) and generalized discriminative analysis (GDA) [13] (supervised). It can be seen that PCA is useful for the CBTC dataset only. While GDA with DSIFT local descriptors performs the best for the IICBU dataset, the performance using GDA fluctuates largely for different datasets and local descriptors. Our DDR method provides more consistent advantage.

The receiver operating characteristic (ROC) curves using our methods are shown in Fig. 3 as well. For the UCSB dataset, although the DSIFT-FV-DDR and DBN-FV-DDR methods produced the same classification accuracy (89.7%) and AUC (0.93), differences exist in the ROC curves. It can be seen that to obtain 100% true positive rate, DBN-FV-DDR outputs a lower false positive rate; and with 0% false positive rate, DBN-FV-DDR gives a higher true positive rate, when compared to SIFT-FV-DDR. Similar performance differences are observed for the other two datasets as well. This indicates the overall advantage of using DBN as the local descriptors over DSIFT.

Table 2 lists the AUCs of our methods and those reported in the existing studies [2, 12, 14] for the UCSB dataset. The existing methods are based on multiple instance learning (MIL) models and a combination of texture and morphological features. Our method can be considered as related to MIL that only image-level label is available; however, in our method, local descriptors are combined using FV encoding and classification is performed at the image-level only. The results show that while the advanced MIL method with joint clustering and classification [14] is more effective, our method performs better than or comparably to

the other approaches [2, 12]. We suggest that our method can be a viable and conceptually simpler alternative to the MIL models.

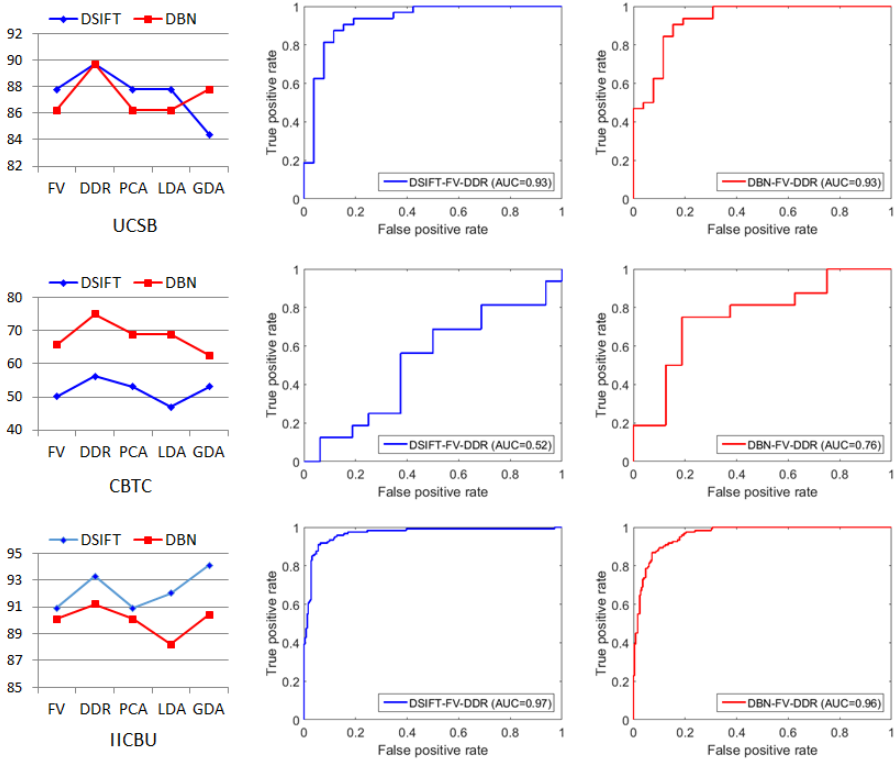


Fig. 3. Classification results comparing dimension reduction approaches when DSIFT or DBN is used as the local descriptor. The left column shows the classification accuracies (%). The middle and right columns show the ROC curves and AUCs using our approaches DSIFT-FV-DDR and DBN-FV-DDR. The ROC curves of the compared techniques are not shown, so that the curves of our methods can be seen clearly.

Table 2. AUCs on the UCSB breast cancer dataset compared to existing studies.

DSIFT-FV-DDR	DBN-FV-DDR	[2]	[12]	[14]
0.93	0.93	0.92	0.93	0.95

Table 3 lists the classification accuracies of methods and those reported in the literature [15, 16] for the IICBU dataset. The approach [15] is standard benchmark for the IICBU dataset, and performs image classification by extracting

texture and shape features and then using a variation of kNN classification. The more recent approach [16] provides another open platform for the microscopy image classification task with a different set of features and SVM classification. Our results show large improvement over these approaches. We note that for the CBTC dataset, to the best of our knowledge, there is no available benchmark for performance comparison.

Table 3. Classification accuracies (%) on the IICBU 2008 lymphoma dataset compared to existing studies.

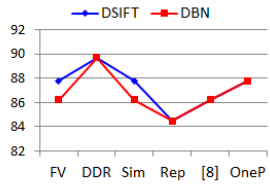
DSIFT-FV-DDR	DBN-FV-DDR	[15]	[16]
93.3	91.2	85	70.9

The effects of individual components in our DDR method are evaluated by comparing the following approaches: 1) Sim: only the similarity constraint is used for discriminative dimension reduction; 2) Rep: only the representation constraint is used; and 3) OneP: a single p value ($p = 3$) is used when constructing the representation constraint. We also compare with the original discriminative dimension reduction algorithm [8], which is equivalent to using only the similarity constraint but without the classification score-based training data selection, i.e. all possible positive and negative pairs are used in training. As shown in Fig. 4, when DSIFT is used as the local descriptor, Sim provides some performance improvement over FV; but when DBN is used as the local descriptor, Rep is more effective than Sim. Combining Sim and Rep (i.e. DDR) gives the best classification. The performance difference between Sim and [8] illustrates the usefulness of having training data selection. It can be seen that Sim outperforms [8] in most cases except when DBN is used as the local descriptor on the CBTC dataset. We suggest that this is because the CBTC dataset contains a small number of images, and including all possible training data is helpful to better exploit the feature space characteristics. The overall advantage of our DDR method over the compared approaches demonstrates that it is essential to have both the similarity and representation constraints in the optimization objective, and to use only a subset of descriptor pairs for training.

The confusion matrices of classification using our DSIFT-FV-DDR and DBN-FV-DDR methods are shown in Fig. 4. For the CBTC dataset, with DBN-FV-DDR, there is a tendency that more images are classified as oligodendroglioma than astrocytoma, hence resulting in a high recall (81%) but low precision rate (72%) of oligodendroglioma. For the IICBU dataset, DSIFT-FV-DDR generates more balanced results among the various classes compared to DBN-FV-DDR. Compared to using FV (without dimension reduction), the main difference is that more MCL cases are correctly identified.

We would also like to mention that our discriminative dimension reduction method is quite efficient. With Matlab implementation on a PC, the training

process of DDR needs about 3.8 s on the UCSB dataset, 1.8 s on the CBTC dataset, and 36.4 s on the IICBU dataset. The different sizes of datasets affect the size and complexity of training data, and subsequently affect the number of iterations and time required for optimization.

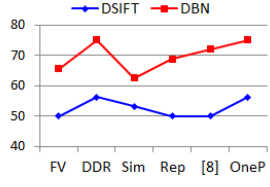


	B	M
B	0.94	0.06
M	0.15	0.85

DSIFT-FV-DDR

	B	M
B	0.94	0.06
M	0.15	0.85

DBN-FV-DDR

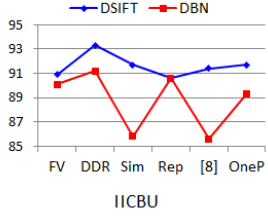


	A	O
A	0.56	0.44
O	0.44	0.56

DSIFT-FV-DDR

	A	O
A	0.69	0.31
O	0.19	0.81

DBN-FV-DDR



	CLL	FL	MCL
CLL	0.93	0.01	0.06
FL	0.01	0.93	0.06
MCL	0.05	0.04	0.91

DSIFT-FV-DDR

	CLL	FL	MCL
CLL	0.87	0.04	0.09
FL	0.02	0.95	0.03
MCL	0.05	0.04	0.91

DBN-FV-DDR

Fig. 4. Classification accuracy (%) comparing our DDR method with several variations (Sim, Rep, and OneP) when DSIFT or DBN is used as the local descriptor. Results using FV (without dimension reduction) or [8] are also included. Confusion matrices of classification using our methods are shown as well.

4 Conclusions

We present a histopathology image categorization method in this paper. Our method comprises two major components: FV encoding of local descriptors that are computed using DSIFT or based on unsupervised learning with DBN; and discriminative dimension reduction of FV descriptors with similarity and representation constraints. Our method is evaluated on three public datasets of breast cancer, brain tumor and malignant lymphoma images, and we show better performance in comparison with some existing approaches and other feature representation and dimension reduction techniques.

References

1. Sparks, R., Madabhushi, A.: Explicit shape descriptors: novel morphologic features for histopathology classification. Medical Image Analysis **17**(1) (2013) 997–1009
2. Kandemir, M., Zhang, C., Hamprecht, F.A.: Empowering multiple instance histopathology cancer diagnosis by cell graphs. In:MICCAI 2014, Part II, LNCS 8674, (2014) 228–235
3. Xu, X., Lin, F., Ng, C., Leong, K.P.: Adaptive co-occurrence differential texton space for HEp-2 cell classification. In:MICCAI 2015, Part III, LNCS 9351, (2015) 260–267
4. Zhou, Y., Chang, H., Barner, K., Spellman, P., Parvin, B.: Classification of histology sections via multispectral convolutional sparse coding. In: CVPR, (2014) 3081–3088
5. Vu, T.H., Mousavi, H.S., Monga, V., Rao, G., Rao, A.: Histopathological image classification using discriminative feature-oriented dictionary learning. IEEE Trans. Med. Imag. (2015) 1–13
6. BenTaieb, A., Li-Chang, H., Huntsman, D., Hamarneh, G.: Automatic diagnosis of ovarian carcinomas via sparse multiresolution tissue representation. In:MICCAI 2015, Part I, LNCS 9349, (2015) 629–636
7. Perronnin, F., Sanchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In:ECCV 2010, Part IV, LNCS 6314, (2010) 143–156
8. Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A.: Fisher vector faces in the wild. In: BMVC, (2013) 1–12
9. Keyvanrad, M.A., Homayounpour, M.M.: A brief survey on deep belief networks and introducing a new object oriented toolbox (DeeBNet). arXiv:1408.3264, (2014)
10. Otalora, S., Cruz-Roa, A., Arevalo, J., Atzori, M., Madabhushi, A., Judkins, A.R., Gonzalez, F., Muller, H., Depeursinge, A.: Combining unsupervised feature learning and riesz wavelets for histopathology image representation: application to identifying anaplastic medulloblastoma. In:MICCAI 2015, Part I, LNCS 9349, (2015) 581–588
11. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: CVPR, (2010) 3360–3367
12. Li, W., Zhang, J., McKenna, S.J.: Multiple instance cancer detection by boosting regularised trees. In:MICCAI 2015, Part I, LNCS 9349, (2015) 645–652
13. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. Neural Computation **12**(10) (2000) 2385–2404
14. Sikka, K., Giri, R., Bartlett, M.: Joint clustering and classification for multiple instance learning. In: BMVC, (2015) 1–12
15. Shamir, L., Orlov, N., Eckley, D.M., Macura, T.J., Johnston, J., Goldberg, I.G.: Wndchrm - an open source utility for biological image analysis. Source Code Biol. Med. **3**(1) (2008) 13
16. Zhou, J., Lamichhane, S., Sterne, G., Ye, B., Peng, H.: BIOCAT: a pattern recognition platform for customizable biological image classification and annotation. BMC Bioinformatics **14** (2013) 291