Title

Stacked Predictive Sparse Decomposition for Classification of Histology Sections

Journal

International Journal of Computer Vision, 113(1)

ISSN

0920-5691

Authors

Chang, Hang
Zhou, Yin
Borowsky, Alexander
et al.

Publication Date

2015-05-01

DOI

10.1007/s11263-014-0790-9

Peer reviewed

HHS Public Access

Author manuscript

Int J Comput Vis. Author manuscript; available in PMC 2016 October 05.

Published in final edited form as:

Int J Comput Vis. 2015 May ; 113(1): 3–18. doi:10.1007/s11263-014-0790-9.

Stacked Predictive Sparse Decomposition for Classification of Histology Sections

Hang Chang,
Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Department of Electrical and Computer Engineering, University of California, Riverside, USA

Yin Zhou,
Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Alexander Borowsky,
Center for Comparative Medicine, UC Davis, Davis, CA, USA

Kenneth Barner,
ECE Department, University of Delaware, Newark, DE, USA

Paul Spellman, and
Center for Spatial Systems Biomedicine, Oregon Health Sciences University, Portland, ON, USA

Bahram Parvin
Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Department of Electrical and Computer Engineering, University of California, Riverside, USA

Department of Biomedical Engineering, University of Nevada, Reno, Nevada, USA

Hang Chang: hchang@lbl.gov; Yin Zhou: yinzhou@lbl.gov; Alexander Borowsky: adborowsky@ucdavis.edu; Kenneth Barner: barner@udel.edu; Paul Spellman: spellmap@ohsu.edu; Bahram Parvin: b_parvin@lbl.gov

Abstract

Image-based classification of histology sections, in terms of distinct components (e.g., tumor, stroma, normal), provides a series of indices for histology composition (e.g., the percentage of each distinct components in histology sections), and enables the study of nuclear properties within each component. Furthermore, the study of these indices, constructed from each whole slide image in a large cohort, has the potential to provide predictive models of clinical outcome. For example, correlations can be established between the constructed indices and the patients' survival

Correspondence to: Hang Chang, hchang@lbl.gov; Bahram Parvin, b_parvin@lbl.gov.

Hang Chang and Yin Zhou are Co-First Authors.
Hang Chang and Bahram Parvin are Co-Corresponding Authors.

information at cohort level, which is a fundamental step towards personalized medicine. However, performance of the existing techniques is hindered as a result of large technical variations (e.g., variations of color/textures in tissue images due to non-standard experimental protocols) and biological heterogeneities (e.g., cell type, cell state) that are always present in a large cohort. We propose a system that automatically learns a series of dictionary elements for representing the underlying spatial distribution using stacked predictive sparse decomposition. The learned representation is then fed into the spatial pyramid matching framework with a linear support vector machine classifier. The system has been evaluated for classification of distinct histological components for two cohorts of tumor types. Throughput has been increased by using of graphical processing unit (GPU), and evaluation indicates a superior performance results, compared with previous research.

Keywords

Tissue histology; Classification; Sparse coding; Unsupervised feature learning

1 Introduction

Tumor histology provides detailed insight into cellular morphology, organization, and heterogeneity. For example, histology sections can be used to identify mitotic cells, cellular aneuploidy, and autoimmune responses. More importantly, if tumor morphology and architecture can be quantified in a large cohort, it will provide the basis for predictive models in a similar way that genomic techniques have identified predictive molecular subtypes. Genome analysis techniques (e.g., microarray analysis) have the advantages of standardized tools for data analysis and pathway enrichment, which enables hypothesis generation for the underlying mechanism. On the other hand, histological signatures are hard to compute because phenotypic signatures are not standardized and advanced methods for image analysis remain at a deficit. Image analysis is further complicated by technical variations as a result of sample preparation (e.g., fixation, staining) and biological heterogeneity, where the latter can originate both within a whole slide image (WSI) and between WSIs of different patients with the same aberrant signature. These variations are manifested in terms of wide swings in the texture, color, and intensity of the histology sections at multiple scales. If morphometric indices can be successfully computed then they can be linked with clinical and genomic data for building predictive models Chang et al. (2013c).

One of the main technical barriers for processing a large collection of histological data is the diverse phenotypic signature, which is shared by the object recognition problem in computer vision, i.e., the number of different ways that a *sitting chair* can be manufactured. In this case of histology sections, the diversity in the aberrant signatures is rendered by technical variations (e.g., fixation, staining, uncalibrated imaging system), and biological heterogeneity as no two patients have the same signature. Biological heterogeneity can be as a result of the (i) protein macromolecules that are being secreted into the immediate environment; thus, altering color composition, (ii) tumor being at a specific temporal states

at the time of sample collection. The latter has a tremendous impact on the tumor composition that ranges from cellular density, cell death, and immune response.

In this paper, we aim to classify components of each histology section in terms of distinct phenotypes (e.g., tumor, stroma, necrosis), which is disease specific. We show that, compared with human engineered features, unsupervised feature learning is more tolerant to batch effect (e.g., technical variations associated with sample preparation) and can learn pertinent features without user intervention. The key concept is that stacked predictive sparse decomposition (PSD) Kavukcuoglu et al. (2008) can elucidate a superior representation that captures intrinsic phenotypic signature. When this representation is coupled with spatial pyramid matching (SPM) Lazebnik et al. (2006), which utilizes sparse tissue morphometric signatures at various locations and scales, an improved classification performance is realized.

Organization of this paper is as follows: Sect. 2 reviews related works. Section 3 describes the details of our proposed approach. Section 4 elaborates the details of our experimental setup, followed by a detailed discussion (as in Sect. 5) on the experimental results. Lastly, Sect. 6 concludes the paper.

2 Related Work

Several outstanding reviews for the analysis of histology sections can be found in Demir and Yener (2009); Gurcan et al. (2009). From our perspective, four distinct works have defined the trends in tissue histology analysis: (i) one group of researchers proposed nuclear segmentation, classification, and organization for tumor grading and/or the prediction of tumor recurrence Axelrod et al. (2008); Datar et al. (2008); Basavanhally et al. (2009); Doyle et al. (2011); Chang et al. (2013c). It is worth to mention that deep max-pooling convolutional neural networks has been used to detect mitotic cells in in breast histology images Ciresan et al. (2013) with a significant success. (ii) A second group of researchers focused on patch level analysis (e.g., small regions) Bhagavatula et al. (2010); Kong et al. (2010); Han et al. (2011); Cruz-Roa et al. (2013) for tumor representation, among which, Cruz-Roa et al. (2013) utilized deep learning for the detection of basal-cell carcinoma. (iii) A third group focused on block-level analysis to distinguish different states of tissue development using cell-graph representation Acar et al. (2012); Bilgin et al. (2012). (iv) Finally, a fourth group has suggested detection and representation of the auto-immune response as a prognostic tool for cancer Fatakdawala et al. (2010).

The major challenge for tissue classification is the large amounts of technical variations and biological heterogeneities preserved in large scale dataset Kothari et al. (2012), which shares the same threads with object recognition and big data in computer vision. To overcome this problem, recent studies have focused on either fine tuning human engineered features Bhagavatula et al. (2010); Kong et al. (2010); Kothari et al. (2012), or applying automatic feature learning Huang et al. (2011) for robust representation.

In the context of computer vision research on image categorization, the traditional bag of features (BoF) model has been widely studied and improved through different variations

Bosch et al. (2008); Boiman et al. (2008); Elad and Aharon (2006); Moosmann et al. (2006); Lazebnik et al. (2006), among which SPM Lazebnik et al. (2006) has clearly become the major component of the state-of-art systems Everingham et al. (2012) for its effectiveness in practice.

To a large degree, applications of deep learning for classification of histology sections have been driven by advances in machine learning and computer vision literature. The evolution of our research in patch level analysis has been SIFT-like feature extraction followed by an evaluation of several kernel-based classification policies Han et al. (2011); independent subspace analysis that utilizes unsupervised learning without the constraint of being able to reconstruct the original signal Le et al. (2012); a single layer predictive sparse coding with support vector machine (SVM) classifier Nayak et al. (2013); and more recently, coupling of either prior knowledge Chang et al. (2013a) or unsupervised feature learning Chang et al. (2013b); Zhou et al. (2014), with spatial pyramid matching. The current research builds on these results to render an unsupervised feature learning approach with superior performance.

In summary, motivated by the fact that (i) pathologists often use “context” to assess the disease state, (ii) SPM partially captures context Lazebnik et al. (2006); Kavukcuoglu et al. (2008), and (iii) unsupervised feature learning is preferable to capture the variance in large cohorts, we have, for the first time, combined hierarchical unsupervised feature learning with SPM framework for the classification of histology sections, which enables (1) automatic discovery of intrinsic patterns from large histology tissue cohort via hierarchical unsupervised feature learning; and (2) effective representation of tissue morphometric context through spatial pyramid feature pooling. As a result, our approach leads to superior performance for tissue classification on large scale cohorts, across different tumor types.

3 Approach

The Proposed approach (PSD^{n}SPM) utilizes predictive sparse decomposition (PSD) Kavukcuoglu et al. (2008) as a building block for the purpose of constructing hierarchical learning framework, which is suggested to be able to capture higher-level dependencies of input variables, thereby improving the ability of the system to capture underlying regularities in the data Ranzato et al. (2008). Unlike many unsupervised feature learning algorithms Lee et al. (2006, 2007); Poultney et al. (2006); Yu et al. (2009), the feed-forward feature inference of PSD is very efficient, as it involves only element-wise nonlinearity and matrix multiplication. For classification, the predicted sparse features are used in a similar fashion as SIFT features in the traditional framework of SPM, as shown in Fig. 1.

3.1 Unsupervised Feature Learning

Given $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{m \times N}$ as a set of vectorized image patches, we formulate the PSD optimization problem in three different ways:

3.1.1 PSD with Linear Regressor (LR-PSD)—The formulation of PSD with linear regressor is as follows,

$$\min_{\mathbf{B},\boldsymbol{Z},\mathbf{W}} \|\mathbf{X}-\mathbf{B}\boldsymbol{Z}\|_F^2+\lambda\|\boldsymbol{Z}\|_1+\|\boldsymbol{Z}-\mathbf{W}\mathbf{X}\|_F^2$$

$$\text{s.t.}\|\mathbf{b}_i\|_2^2=1,\forall i=1,\ldots,h \quad (1)$$

where $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_h] \in \mathbb{R}^{m\times h}$ is a set of the dictionary elements; $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N] \in \mathbb{R}^{h\times N}$ is the sparse feature matrix; $\mathbf{W} \in \mathbb{R}^{h\times m}$ is the auto-encoder; and λ is a regularization constant. Joint minimization of Eq. (1) with respect to the triple $\langle \mathbf{B}, \mathbf{Z}, \mathbf{W} \rangle$, enforces the inference of the linear regressor **WX** to be similar to the optimal sparse codes, **Z**, which can reconstruct **X** over **B** Kavukcuoglu et al. (2008).

As shown below, optimization of Eq. (1) is iterative, where the algorithm terminates when either the objective function is below a preset threshold or the maximum number of iterations has been reached.

1. Randomly initialize **B**, and **W**.
2. Fixing **B**, and **W**, minimize Eq. (1) with respect to **Z**, where **Z** can be either solved as a ℓ_1-minimization problem Lee et al. (2006) or equivalently solved by greedy algorithms, e.g., Orthogonal Matching Pursuit (OMP) Tropp and Gilbert (2007).
3. Fixing **Z**, update **B** and **W**, respectively, using the stochastic gradient descent algorithm.
4. Repeat [2]–[3] until stopping condition is satisfied.

3.1.2 PSD with Nonlinear Regressor (NR-PSD)—The formulation of PSD with nonlinear regressor is as follows,

$$\min_{\mathbf{B},\boldsymbol{Z},\mathbf{G},\mathbf{W}} \|\mathbf{X}-\mathbf{B}\boldsymbol{Z}\|_F^2+\lambda\|\boldsymbol{Z}\|_1+\|\boldsymbol{Z}-\mathbf{G}\sigma(\mathbf{W}\mathbf{X})\|_F^2$$

$$\text{s.t.}\|\mathbf{b}_i\|_2^2=1,\forall i=1,\ldots,h \quad (2)$$

where $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_h] \in \mathbb{R}^{m\times h}$ is a set of the dictionary elements; $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N] \in \mathbb{R}^{h\times N}$ is the sparse feature matrix; $\mathbf{W} \in \mathbb{R}^{h\times m}$ is the auto-encoder; $\mathbf{G} = \text{diag}(g_1, \ldots, g_h) \in \mathbb{R}^{h\times h}$ is a scaling matrix with diag being an operator aligning vector, $[g_1, \ldots, g_h]$, along the diagonal; $\sigma(\cdot)$ is the element-wise sigmoid function; and λ is a regularization constant. Joint minimization of Eq. (2) with respect to the quadruple $\langle \mathbf{B}, \mathbf{Z}, \mathbf{G}, \mathbf{W} \rangle$, enforces the inference of the nonlinear regressor $\mathbf{G}\sigma(\mathbf{WX})$ to be similar to the optimal sparse codes, **Z**, which can reconstruct **X** over **B** Kavukcuoglu et al. (2008).

As shown below, optimization of Eq. (2) is iterative, where the algorithm terminates when either the objective function is below a preset threshold or the maximum number of iterations has been reached.

1. Randomly initialize **B, W**, and **G**.
2. Fixing **B, W** and **G**, minimize Eq. (2) with respect to **Z**, where **Z** can be either solved as a ℓ_1-minimization problem Lee et al. (2006) or equivalently solved by greedy algorithms, e.g., Orthogonal Matching Pursuit (OMP) Tropp and Gilbert (2007).
3. Fixing **B, W** and **Z**, solve for **G**, which is a simple least-square problem with analytic solution.
4. Fixing **Z** and **G**, update **B** and**W**, respectively, using the stochastic gradient descent algorithm.
5. Repeat [2]–[4] until stopping condition is satisfied.

Figure 2 illustrates 1,024 dictionary elements computed from GBM and KIRC datasets, respectively, which capture both color and texture information from the data and is generally difficult to realize using hand-engineered features.

3.1.3 PSD with Saliency (Salient-PSD)—After training, both LR-PSD and NR-PSD extract features (sparse codes) on regularly-spaced patches over the input image. A recent study Wu et al. (2013) shows that saliency may help improve the classification performance. In tissue histology, nuclear structure is considered with the highest saliency, and, as a result, the Salient-PSD is designed to extract features on patches aligned at segmented nuclear centers Chang et al. (2013c). An example of tissue image saliency mask is shown in Fig. 3.

3.1.4 Speeding up the Unsupervised Feature Learning—In large-scale feature learning problems, involving ~ 10^5 image patches, it is computationally intensive to evaluate the sum-gradient over the entire training set. However, both stochastic gradient descent algorithm and graphical processing unit (GPU) parallel computing can provide a significant increase in speed. The former approximates the true gradient of the objective function by the gradient evaluated over mini-batches, and the latter further accelerates the process (up to 5×) with our Matlab implementation based on an Nvidia GTX 580 graphics card.

3.2 Spatial Pyramid Matching (SPM)

The codebook, $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K] \in \mathbb{R}^{h \times K}$, consisting of K sparse tissue morphometric types, is constructed by solving the following optimization problem:

$$\min_{\mathbf{D},\mathbf{C}} \sum_{i=1}^{N} \|\mathbf{z}_i - \mathbf{D}\mathbf{c}_i\|^2$$

$$\text{s.t.} \operatorname{card}(\mathbf{c}_i) = 1, \|\mathbf{c}_i\|_1 = 1, \mathbf{c}_i \succeq 0, \forall i \quad (3)$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N] \in \mathbb{R}^{K \times N}$ is the code matrix assigning each $\mathbf{z}_i$ to its closest sparse tissue morphometric type in $\mathbf{D}$, $card(\mathbf{c}_i)$ is a cardinality constraint enforcing only one nonzero element in $\mathbf{c}_i$, and $\mathbf{c}_i \geq 0$ is a non-negative constraint on all vector elements. Equation (3) is optimized by alternating between the two variables, *i.e.*, minimizing one while keeping the other fixed. After training, $\mathbf{D}$ is fixed and the query signal set, $\mathbf{Z}$, is encoded by solving Eq. (3) with respect to $\mathbf{C}$ only.

The next step is to construct a spatial histogram for SPM Lazebnik et al. (2006). By repeatedly subdividing an image, histograms of different sparse tissue morphometric types over the resulting subregions are computed. The spatial histogram, H, is then formed by concatenating the appropriately weighted histograms of sparse tissue morphometric types at all resolutions, *i.e.*,

$$H_0 = H_0^0$$

$$H_l = (H_l^1, \ldots, H_l^{4^l}), 1 \leq l \leq 1$$

$$H = \left(\frac{1}{2^L} H_0, \frac{1}{2^L} H_1, \ldots, \frac{1}{2^{L-l+1}} H_l, \ldots, \frac{1}{2} H_L \right) \quad (4)$$

where $(\cdot)$ denotes the vector concatenation operator, $l \in \{0, \ldots, L\}$ is the resolution level of the image pyramid, and H_l represents the concatenation of histograms for all image subregions at pyramid level l (Fig. 4).

Finally, a χ^2 SVM was transferred into a linear SVM based on a homogeneous kernel map Vedaldi and Zisserman (2012). In practice, the intersection kernel and χ^2 kernel have been found to be the most suitable for histogram representations Yang et al. (2009). Thus, a homogenous kernel map is applied to approximate the χ^2 kernels, which enables the efficiency by adopting learning methods for linear kernels, e.g., linear SVM. For more details about the homogeneous kernel map, please refer to Vedaldi and Zisserman (2012).

4 Experiments and Results

In this section, we provide details of the experimental design that includes data from tumor histopathology. The tumor data includes curated sets of glioblastoma multiforme (GBM) and kidney clear cell carcinoma (KIRC) from The Cancer Genome Atlas (TCGA), which are publicly available from both the NIH repository and our website.[1] The methods and the detailed configuration involved in the evaluation are listed as follows,

1. NR-PSD^{n}SPM: The nonlinear kernel SPM that uses spatial-pyramid histograms of sparse tissue morphometric types. In this implementation,

[1] http://vision.lbl.gov

a. $n = 1, 2$;

b. The nonlinear regressor ($\mathbf{Z} = \mathbf{G}\sigma(\mathbf{WX})$) was trained for the inference of $\mathbf{Z}$;

c. The image patch size was fixed to be 20 × 20 and the number of dictionary elements in the top layer was fixed to be 1,024. We adopted the SPAMS optimization toolbox Mairal et al. (2010) for efficient implementation of OMP to compute the sparse code, $\mathbf{Z}$, with sparsity prior set to 30;

d. The PSD features were extracted on regularly-spaced patches over the input image, with fixed step-size (20 pixels).

e. Standard K-means clustering was used for the construction of the dictionary;

f. The level of pyramid was fixed to be 3; and

g. The homogeneous kernel map was applied, followed by the linear SVM for classification.

2. LR-PSD[1]SPM Chang et al. (2013b): The nonlinear kernel SPM that uses spatial-pyramid histograms of sparse tissue morphometric types. In this implementation,

a. The linear regressor ($\mathbf{Z} = \mathbf{WX}$) was trained for the inference of $\mathbf{Z}$;

b. For consistency, the image patch size and the number of dictionary elements was fixed at 20 × 20 and 1,024, respectively. The sparsity constraint was set at 0.3 for best performance following cross validation.

c. The PSD features were extracted on regularly-spaced patches over the input image, with fixed step-size (20 pixels).

d. Standard K-means clustering was used for the construction of the dictionary;

e. The level of pyramid was fixed to be 3;

f. The homogeneous kernel map was applied, followed by linear SVM for classification.

3. Salient-LR-PSD[1]SPM: The nonlinear kernel SPM that uses spatial-pyramid histograms of sparse tissue morphometric types extracted at nuclear centers. In this implementation,

a. The linear regressor (**Z** = **WX**) was trained for the inference of **Z**;

b. For consistency, the image patch size and the number of dictionary elements was fixed at 20 × 20 and 1,024, respectively. The sparsity constraint was set at 0.3 for best performance following cross validation.

c. The PSD features were extracted on patches centered at segmented nuclear centers Chang et al. (2013c) over the input image.

d. Standard K-means clustering was used for the construction of the dictionary;

e. The level of pyramid was fixed to be 3;

f. The homogeneous kernel map was applied, followed by linear SVM for classification.

4. LR-PSD Chang et al. (2013b): The sparse tissue morphometric features with max-pooling strategy, and RBF kernel. In the implementation,

a. The linear regressor (**Z** = **WX**) was trained for the inference of **Z**;

b. For consistency, the image patch size and the number of dictionary elements was fixed at 20 × 20 and 1,024, respectively. The sparsity constraint was set at 0.3 for best performance following cross validation.

c. The PSD features were extracted on regularly-spaced patches over the input image, with fixed step-size (20 pixels).

d. Max-pooling strategy was used for sparse feature summarization.

e. nonlinear SVM with RBF kernel was used for classification.

5. ScSPM Yang et al. (2009): The linear SPM that utilizes linear kernel on spatial-pyramid pooling of SIFT sparse codes. In this implementation,

a. The dense SIFT features was extracted on 16 × 16 patches sampled from each image on a grid with step-size 8 pixels;

b. The sparse constraint parameter λ was fixed to be 0.15, which was determined empirically to achieve the best performance;

c. The level of pyramid was fixed to be 3;

d. Linear SVM was used for classification.

6. KSPM Lazebnik et al. (2006): The nonlinear kernel SPM that uses spatial-pyramid histograms of SIFT features; In the implementation,
   a. The dense SIFT features was extracted on 16 × 16 patches sampled from each image on a grid with step-size 8 pixels;
   b. Standard K-means clustering was used for the construction of the dictionary;
   c. The level of pyramid was fixed to be 3;
   d. The homogeneous kernel map was applied, followed by linear SVM for classification.
7. CTSPM: The nonlinear kernel SPM that uses spatial-pyramid histograms of color and texture features; In this implementation,
   a. Color features were extracted from the RGB color space;
   b. Texture features were extracted via steerable filters Young and Lesperance (2001) with 4 directions $(\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\})$ and 5 scales ($\sigma \in \{1, 2, 3, 4, 5\}$) from the grayscale image;
   c. The feature vector was constructed by concatenating texture and mean color on 20 × 20 patches, empirically, to achieve the best performance;
   d. Standard K-means clustering was used for the construction of the dictionary;
   e. The level of pyramid was fixed to be 3;
   f. The homogeneous kernel map was applied, followed by linear SVM for classification.

All experimental processes were repeated 10 times with randomly selected training and testing images. The final results were reported as the mean and standard deviation of the classification rates on the following two distinct datasets, which included vastly different tumor types:

1. GBM Dataset. The GBM dataset contains 3 classes: tumor, necrosis, and transition to necrosis, which were curated from WSIs scanned with a 20× objective (0.502 micron/pixel). Examples can be found in Fig. 5. The number of images per category are 628, 428 and 324, respectively. Most images are 1, 000 × 1, 000 pixels. In this experiment, we trained on 40, 80 and 160 images per category and tested on the rest, using three different dictionary sizes: 256, 512 and 1,024. Detailed comparisons are shown in Table 1.
2. KIRC Dataset. The KIRC dataset contains 3 classes: tumor, normal, and stromal, which were curated from WSIs scanned with a 40× objective

(0.252 micron/pixel). Examples can be found in Fig. 6. The number of images per category are 568, 796 and 784, respectively. Most images are 1, 000 × 1, 000 pixels. In this experiment, we trained on 70, 140 and 280 images per category and tested on the rest, using three different dictionary sizes: 256, 512 and 1,024. Detailed comparisons are shown in Table 2.

5 Discussion

5.1 Does Unsupervised Feature Leaning Provide an Improvement over Human Engineered Features?

Feature extraction is the very first step for the construction of classification/recognition system, and is one of the most important factors that affect the performance. In our evaluation, SIFT, color and texture are all human engineered features, which are widely used in various applications, including but not limited to image classification, object detection and segmentation. However, these prefixed human engineered features typically suffer from large data variations in the big dataset, and, as a result, have limited generalization ability. As shown in Fig. 7, systems based on unsupervised feature learning (e.g., NR-PSD^{1}SPM and LR-PSD^{1}SPM) generate better performance compared with the ones based on human engineered features (e.g., ScSPM, KSPM and CTSPM), which indicates that unsupervised feature learning can better capture the intrinsic properties in histological datasets, and therefore, leads to systems that are more tolerant to batch effects in the data.

5.2 Does Saliency Improve Classification?

Salient-PSD differs from PSD in that PSD features are densely extracted per regularly-spaced image patch without using saliency information as prior. Recent studies Wu et al. (2013) indicate that saliency-awareness may be helpful for the task of image classification, thus it will be interesting to figure out whether regular PSD features can be improved by the incorporation of saliency as prior. Therefore, we designed salient PSD (Salient-PSD) features, which were only extracted on patches aligned with nuclear centroid locations. Comparison of classification performance, between PSD features and salient-PSD features are shown in Fig. 8 for both GBM and KIRC datasets, which shows that, for PSD features, saliency-awareness plays a negative role for the task of tissue histology classification. One possible explanation is that, in the task of tissue histology classification, PSD leads to appearance-based image representation, thus requires dense sampling all over the place in order to faithfully assemble the view of the image.

5.3 Is There a Preference for Linear and Non-linear Regressors?

As pointed out in Kavukcuoglu et al. (2008), nonlinear regressor is required to produce the sparse representations using an over-complete set due to the non-orthogonality of the filters. To validate the choice of nonlinear regressor against the linear one for the task of tissue histology classification, we made comparisons between PSDs with linear regressor and nonlinear regressor. The experimental results suggest that, without any surprise, PSD with nonlinear regressor outperforms PSD with linear regressor in terms of both reconstruction and classification, as shown in Figs. 9 and 10, respectively.

5.4 Does Multilayer PSD Performs Better than a Single Layer PSD?

The work in Jarrett et al. (2009) suggests that multi-stage feature extraction system generally outperforms single-stage system. Also, it is worth to mention that, as pointed out in Yang et al. (2009), spatial pyramid matching kernel is conceptually similar to an extra layer of sparse coding and spatial pooling, which suggests that, structurally, PSD^1SPM and PSD^2SPM are similar to two-stage and three-stage systems, respectively. The comparisons among PSD (single-stage system), PSD^1SPM (two-stage system) and PSD^2SPM (three-stage system), as shown in Tables 1, 2 and Fig. 11, indicates that:

1. Multi-stage feature extraction system outperforms single-stage system for the classification of histology sections;
2. By stacking multiple unsupervised feature learning module (PSD) into hierarchy, we experienced a slightly improved performance, which might due to the fact that multi-stage feature learning system can capture higher-level patterns in the tissue images;
3. The improvement of performance might decay with the increase of the depth (stages) of feature extraction system. As a result, the two-stage system (PSD^1SPM) is superior to the single-stage system (PSD); while the three-stage system (PSD^2SPM) is only slightly better than the two-stage one.

From our intensive evaluation, as shown in Tables 3 and 4, some other important insights associated with the multi-stage architecture are listed as follows,

1. Local contrast normalization significantly impairs the performance. Local contrast normalization enhances the structures in the image, and, as a result, ignores the variation of intensity (color). However, for tissue histology images, variations in intensity (color) typically correspond to different biological processes (e.g., cell proliferation, cell death, etc). Therefore, with the choice of PSD as the unsupervised feature learning module, local contrast normalization is not desired during the construction of multi-stage feature extraction system, for the task of tissue histology classification.
2. Extra pooling step is not necessary. As pointed out in Yang et al. (2009), spatial pyramid matching kernel is conceptually similar to an extra layer of sparse coding and spatial pooling, the adoption of extra pooling step (e.g., max-pooling) during the concatenation of PSD modules does not improve the performance, and thus is not necessary.
3. Absolute value rectification is not necessary. As pointed out in Jarrett et al. (2009), the use of max-pooling alleviates the need for Abs rectification. This statement also seems to hold when spatial pyramid matching kernel is applied, at least for the application of tissue histology classification.

5.5 What are the Merits of (PSD)^{n}SPM for Classification of Tumor Histopathology?

Experimental results and discussions suggests that deep learning has the following merits for classification of tissue histology:

1. *Extensibility to different tumor types* Tables 1 and 2 demonstrate the superiority and consistency in the performance of the proposed approach on two vastly different tumor types, which confirms that unsupervised feature learning has better generalization capability compared to human engineered features (e.g., SIFT), and ultimately ensures the extensibility of proposed approach to different tumor types.
2. *Robustness in the presence of large amounts of technical variations and biological heterogeneities* Tables 1 and 2 indicate that the performance of our approach, based on small number of training samples, is comparable to or better than the performance of ScSPM, KSPM and CTSPM, which are based on large number of training samples. Given the fact that TCGA datasets contain large amounts of technical variations and biological heterogeneities, these results clearly verify the robustness of our approach, which improves the scalability with varying training sample sizes, and the reliability of further analysis on large cohort of whole mount tissue sections.

6 Conclusion and Future Work

In this paper, we proposed a multi-stage PSD framework for classification of distinct regions of tumor histopathology, which outperforms traditional methods that are typically based on pixel- or patch-level features. Our analysis indicates that the proposed approach is (i) extensible to different tumor types; (ii) robust in the presence of large amounts of technical variations and biological heterogeneities; and (iii) scalable with varying training sample sizes. Future research will focus on (i) further comparative study between our approach and the state-of-art convolutional neural networks (CNNs) LeCun et al. (1998); Huang and LeCun (2006); (ii) further validation of our approach on other tumor types; and (iii) further application of our approach on the discrimination of phenotypic responses in multicellular systems.

Acknowledgments

This work was supported by National Institute of Health (NIH) U24 CA1437991 and NIH R01 CA140663 carried out at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231.

References

Acar E, Plopper G, Yener B. Coupled analysis of in vitro and histology samples to quantify structure-function relationships. PLoS One. 2012; 7:e32227. [PubMed: 22479315]

Axelrod D, Miller N, Lickley H, Qian J, Christens-Barry W, Yuan Y, et al. Effect of quantitative nuclear features on recurrence of ductal carcinoma in situ (DCIS) of breast. Cancer Informatics. 2008; 4:99–109. [PubMed: 18779878]

Basavanhally A, Xu J, Madabhushu A, Ganesan S. Computer-aided prognosis of ER+ breast cancer histopathology and correlating survival outcome with oncotype DX assay. ISBI. 2009:851–854.

Bhagavatula, R.; Fickus, MC.; Kelly, W.; Guo, C.; Ozolek, JA.; Castro, CA.; Kovacevic, J. Automatic identification and delineation of germ layer components in H&E stained images of teratomas derived from human and nonhuman primate embryonic stem cells. Proceedings of the 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 14–17 April 2010; Rotterdam, The Netherlands. 2010. p. 1041-1044.

Bilgin C, Ray S, Baydil B, Daley W, Larsen M, Yener B. Multiscale feature analysis of salivary gland branching morphogenesis. PLoS One. 2012; 7:e32906. [PubMed: 22403724]

Boiman, O.; Shechtman, E.; Irani, M. In defense of nearest-neighbor based image classification; Proceedings of the conference on computer vision and pattern recognition; 2008. p. 1-8.

Bosch A, Zisserman A, Muñoz X. Scene classification using a hybrid generative/discriminative approach. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2008; 30:712–727. [PubMed: 18276975]

Chang, H.; Borowsky, A.; Spellman, P.; Parvin, B. Classification of tumor histology via morphometric context; Proceedings of the conference on computer vision and pattern recognition; 2013a.

Chang H, Nayak N, Spellman P, Parvin B. Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching. Medical Image Computing and Computed-Assisted Intervention-MICCAI. 2013b

Chang H, Han J, Borowsky A, Loss LA, Gray JW, Spellman PT, et al. Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association. IEEE Transactions on Medical Imaging. 2013c; 32:670–682. [PubMed: 23221815]

Ciresan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. MICCAI. 2013; 2:411–418. [PubMed: 24579167]

Cruz-Roa AA, Ovalle JEA, Madabhushi A, Osorio FAG. A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. MICCAI. 2013; 8150:403–410. [PubMed: 24579166]

Datar, M.; Padfield, DR.; Cline, H. Color and texture based segmentation of molecular pathology images using HSOMs. Proceedings of the 2008 IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 14–17 May 2008; Paris. 2008. p. 292-295.

Demir, C.; Yener, B. Technical report. Rensselaer Polytechnic Institute, Department of Computer Science; 2009. Automated cancer diagnosis based on histopathological images: A systematic survey.

Doyle, S.; Feldman, MD.; Tomaszewski, J.; Shih, N.; Madabhushi, A. Cascaded multi-class pairwise classifier (CascaMPa) for normal, cancerous, and cancer confounder classes in prostate histology. Proceedings of the 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 30 March–2 April 2011; Chicago, IL. 2011. p. 715-718.

Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing. 2006; 15:3736–3745. [PubMed: 17153947]

Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The PASCAL visual object classes challenge 2012 (VOC2012) results. 2012 http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

Fatakdawala H, Xu J, Basavanhally A, Bhanot G, Ganesan S, Feldman F, et al. Expectation-maximization-driven geodesic active contours with overlap resolution (EMaGA-COR): Application to lymphocyte segmentation on breast cancer histopathology. IEEE Transactions on Biomedical Engineering. 2010; 57:1676–1690. [PubMed: 20172780]

Gurcan M, Boucheron L, Can A, Madabhushi A, Rajpoot N, Bulent Y. Histopathological image analysis: A review. IEEE Transactions on Biomedical Engineering. 2009; 2:147–171.

Han, J.; Chang, H.; Loss, LA.; Zhang, K.; Baehner, FL.; Gray, JW.; Spellman, PT.; Parvin, B. Comparison of sparse coding and kernel methods for histopathological classification of glioblastoma multiforme. Proceedings of the 8th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 30 March–2 April 2011; Chicago, IL. 2011. p. 711-714.

Huang, FJ.; LeCun, Y. Large-scale learning with SVM and convolutional for generic object categorization. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR); 17–22 June 2006; New York. 2006. p. 284-291.

Huang C, Veillard A, Lomeine N, Racoceanu D, Roux L. Time efficient sparse analysis of histopathological whole slide images. Computerized Medical Imaging and Graphics. 2011; 35:579–591. [PubMed: 21145705]

Jarrett, K.; Kavukcuoglu, K.; Ranzato, M.; LeCun, Y. What is the best multi-stage architecture for object recognition?. Proceedings international conference on computer vision (ICCV'09); IEEE. 2009.

Kavukcuoglu, K.; Ranzato, M.; LeCun, Y. Technical report CBLL-TR-2008-12-01, Computational and Biological Learning Lab. NYU: Courant Institute; 2008. Fast inference in sparse coding algorithms with applications to object recognition.

Kong J, Cooper L, Sharma A, Kurk T, Brat D, Saltz J. Texture based image recognition in microscopy images of diffuse gliomas with multi-class gentle boosting mechanism. ICASSAP. 2010:457–460.

Kothari, S.; Phan, J.; Osunkoya, A.; Wang, M. Biological interpretation of morphological patterns in histopathological whole slide images; ACM conference on bioinformatics, computational biology and biomedicine; 2012.

Lazebnik, S.; Schmid, C.; Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories; Proceedings of the conference on computer vision and pattern recognition; 2006. p. 2169-2178.

Le, QV.; Han, J.; Gray, JW.; Spellman, PT.; Borowsky, A.; Parvin, B. Learning invariant features from tumor signatures. Proceedings of the 9th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 2–5 May 201; Barcelona. 2012. p. 302-305.

Lee, H.; Battle, A.; Raina, R.; Ng, AY. Efficient sparse coding algorithms. Advances in neural information processing systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems; 4–7 December 2006; Vancouver, BC. 2006. p. 801-808.http://books.nips.cc/papers/files/nips19/NIPS2006_0878.pdf

Lee, H.; Ekanadham, C.; Ng, AY. Sparse deep belief net model for visual area v2. Advances in neural information processing systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems; 3–6 December 2007; Vancouver, BC. 2007. http://books.nips.cc/papers/files/nips20/NIPS2007_0934.pdf

LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998; 86(11):2278–2324.

Mairal J, Bach F, Ponce J, Sapiro G. Online learning for matrix factorization and sparse coding. The Journal of Machine Learning Research. 2010; 11:19–60.

Moosmann F, Nowak E, Jurie F. Randomized clustering forests for image classification. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2008; 30(9):1632–1646. [PubMed: 18617720]

Nayak N, Chang H, Borowsky A, Spellman P, Parvin B. Classification of tumor histopathology via sparse feature learning. Proceedings of ISBI. 2013:410–413.

Poultney, C.; Chopra, S.; Lecun, Y. Advances in neural information processing systems. Cambridge: MIT Press; 2006. Efficient learning of sparse representations with an energy-based model. NIPS 2006

Ranzato, M.; Boureau, YL.; LeCun, Y. Sparse feature learning for deep belief networks. In: Platt, J.; Koller, D.; Singer, Y.; Roweis, S., editors. Advances in Neural Information Processing Systems. Vol. 20. Cambridge, MA: MIT Press; 2008. p. 1185-1192.

Tropp J, Gilbert A. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transactions on Information Theory. 2007; 53:4655–4666.

Vedaldi A, Zisserman A. Efficient additive kernels via explicit feature maps. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012; 34:480–492. [PubMed: 21808094]

Wu R, Yu Y, Wang W. Scale: Supervised and cascaded laplacian eigenmaps for visual object recognition based on nearest neighbors. CVPR. 2013:867–874.

Yang, J.; Yu, K.; Gong, Y.; Huang, T. Linear spatial pyramid matching using sparse coding for image classification; Proceedings of the conference on computer vision and pattern recognition; 2009. p. 1794-1801.

Young RA, Lesperance RM. The gaussian derivative model for spatial-temporal vision. I Cortical Model Spatial Vision. 2001; 2001:3–4.

Yu, K.; Zhang, T.; Gong, Y. Nonlinear learning using local coordinate coding. Advances in neural information processing systems. 22 & 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7–10 December 2009; Vancouver, British Columbia. 2009. p. 2223-2231.http://books.nips.cc/papers/files/nips22/NIPS2009_0719.pdf

Zhou, Y.; Chang, H.; Barner, KE.; Spellman, PT.; Parvin, B. Classification of histology sections via multispectral convolutional sparse coding. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 23–28 June 2014; Columbus. 2014. p. 3081-3088.

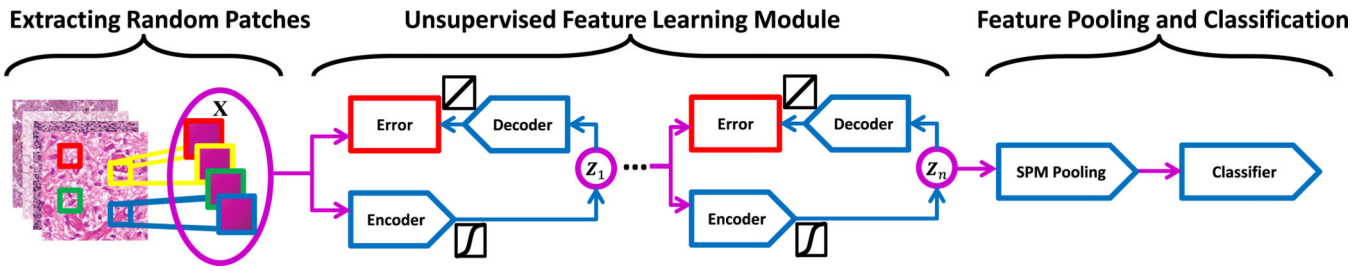

Fig. 1.

Computational workflow of our approach (PSD^{n}SPM), where **X** is a set of vectorized image patches, randomly selected from input tissue images; $\mathbf{Z}_k$ is a set of sparse codes from the kth layer; and the final representation of each tissue image is the summarization of sparse codes, from the last layer, through spatial pyramid kernels

Fig. 2.
Computed dictionary elements (**B**) from GBM (*top row*) and KIRC (*bottom row*) datasets, respectively

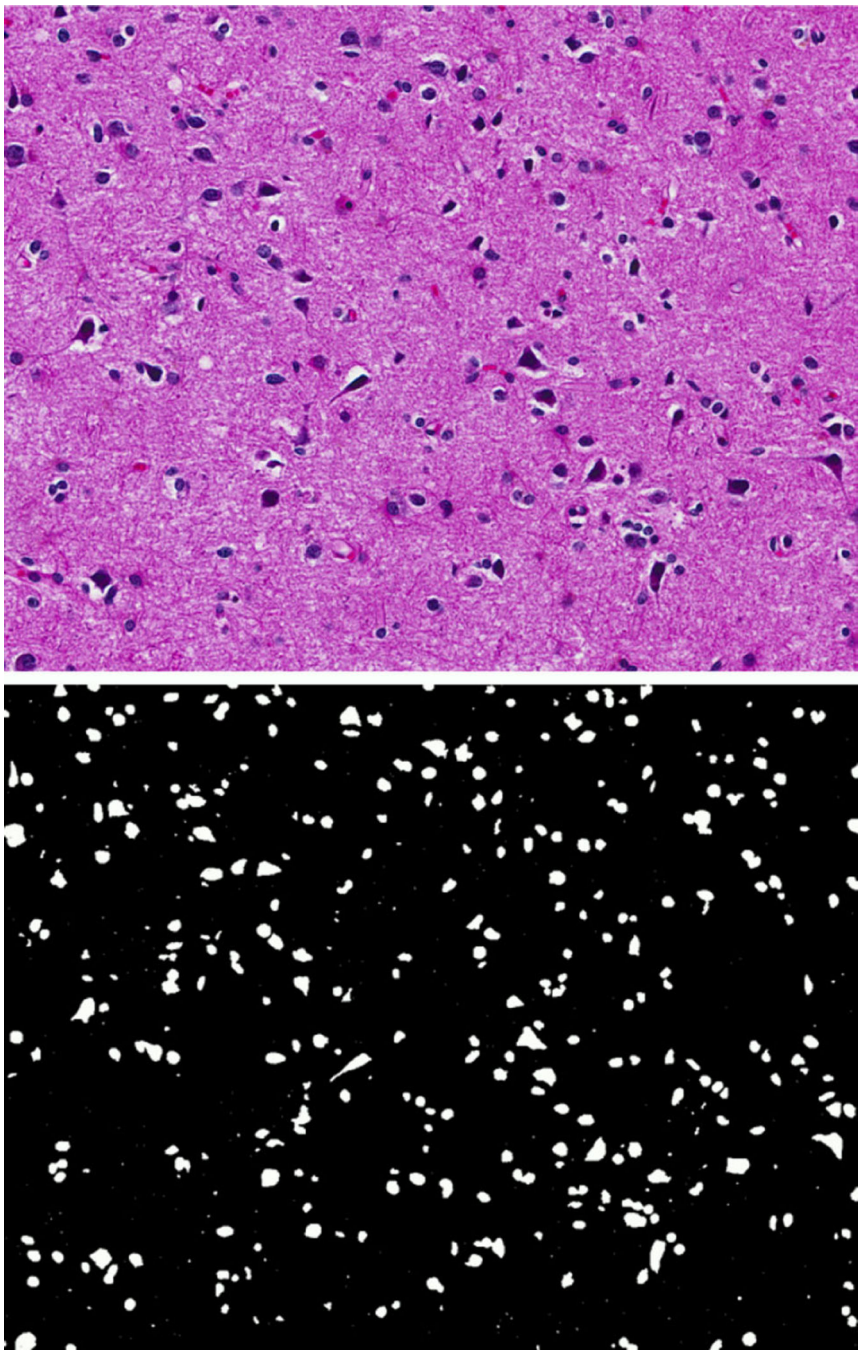

Fig. 3.
Top row original image; *bottom row* saliency mask

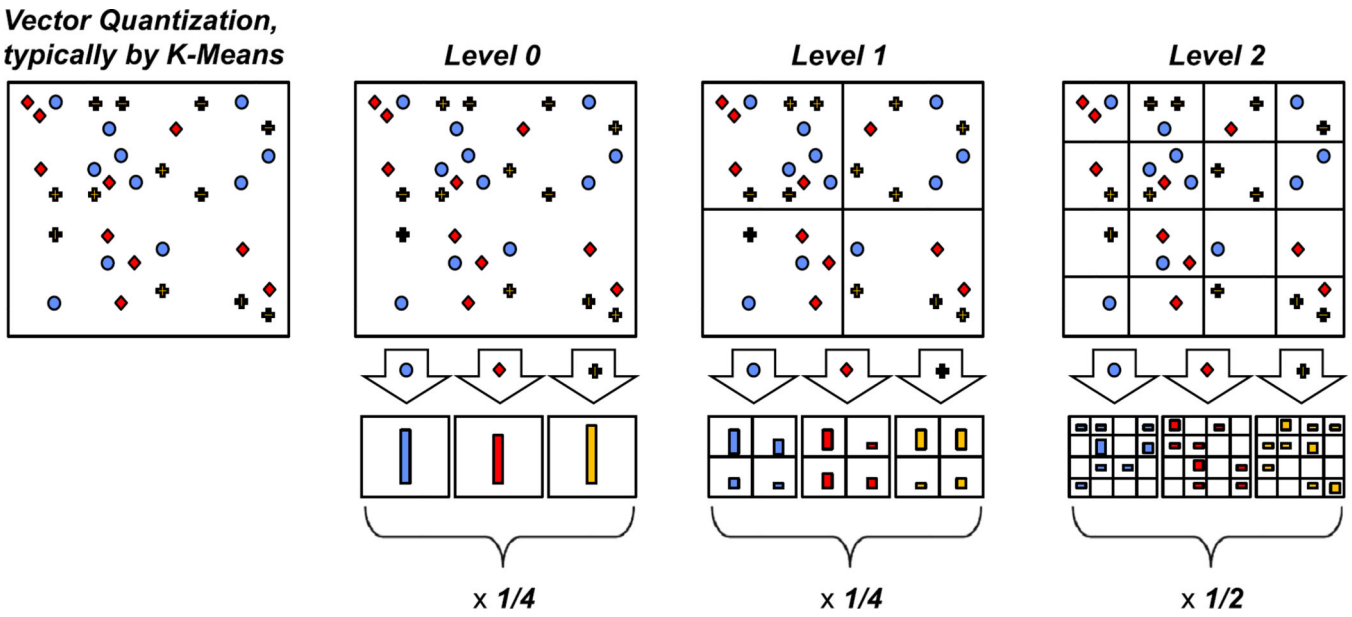

Fig. 4.

Toy example of constructing a three-level pyramid. After quantization (Eq. 3), there are three feature types, represented by *diamonds*, *circles* and *crosses*, respectively. First, we subdivide the image at three different level of resolution. Next, for each level of resolution and each feature type, we count the features that fall in each spatial bin. Finally, we weight and concatenate each spatial histogram according to Eq. 4

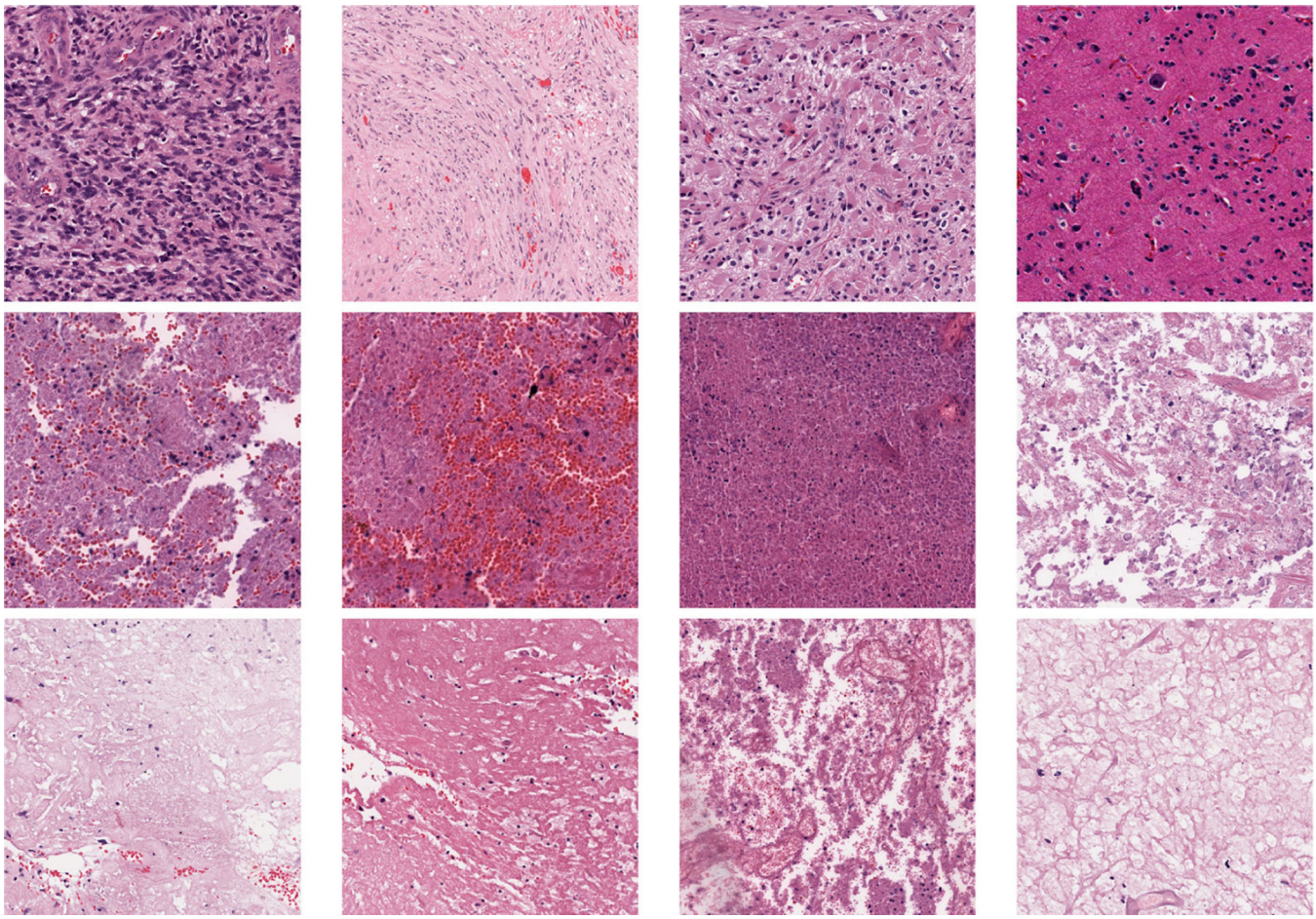

Fig. 5.

GBM examples. *First row* tumor; *second row* transition to necrosis; *third row* necrosis. In GBM, important phenotype for necrosis is loss of nuclear DNA and cell structure. However, necrosis is not instantaneous and is a dynamic process where a subset of cells gradually change shape, rupture, and release their contents. Transition to necrosis is represented as subset of cells being at different stages of loosing their DNA contents and various organelles

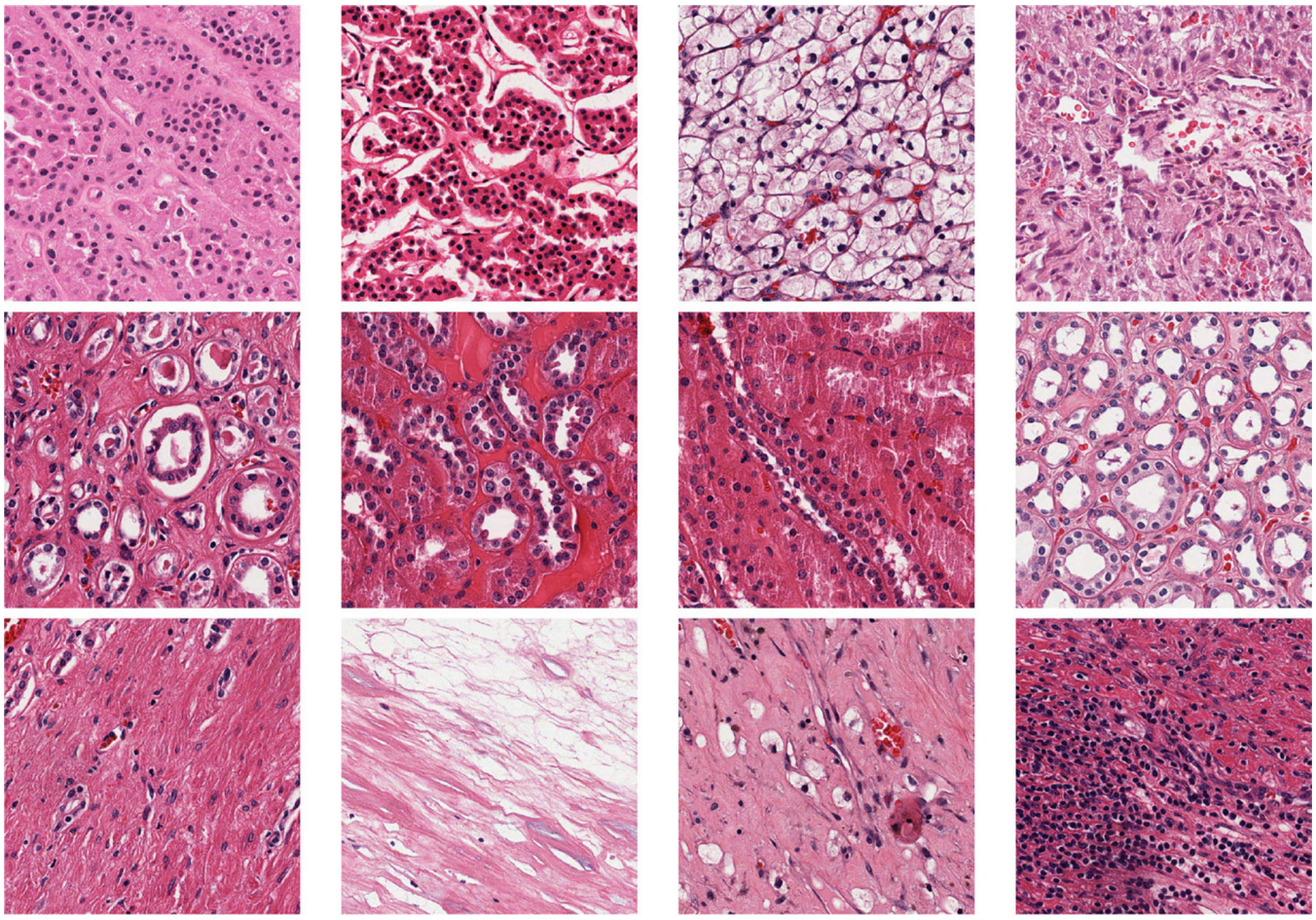

Fig. 6.

KIRC Examples. *First row* tumor; *second row* normal; *third row* stromal. In KIRC, normal is defined by self organization of cell in a “gland-like” structure and stroma refers to one of many supporting cells and scaffolding such as collagen, infiltrating immune cells, fibroblasts, and fat cells. Tumor can have two distinct phenotypes as follows: (i) cells are represented by the loss of proteins and macromolecules contained within the cytoplasm; thus, a clear region between the nucleus and cell membrane is formed; and (ii) cells can form aberrant organization that leads to the loss of normal gland-like structures

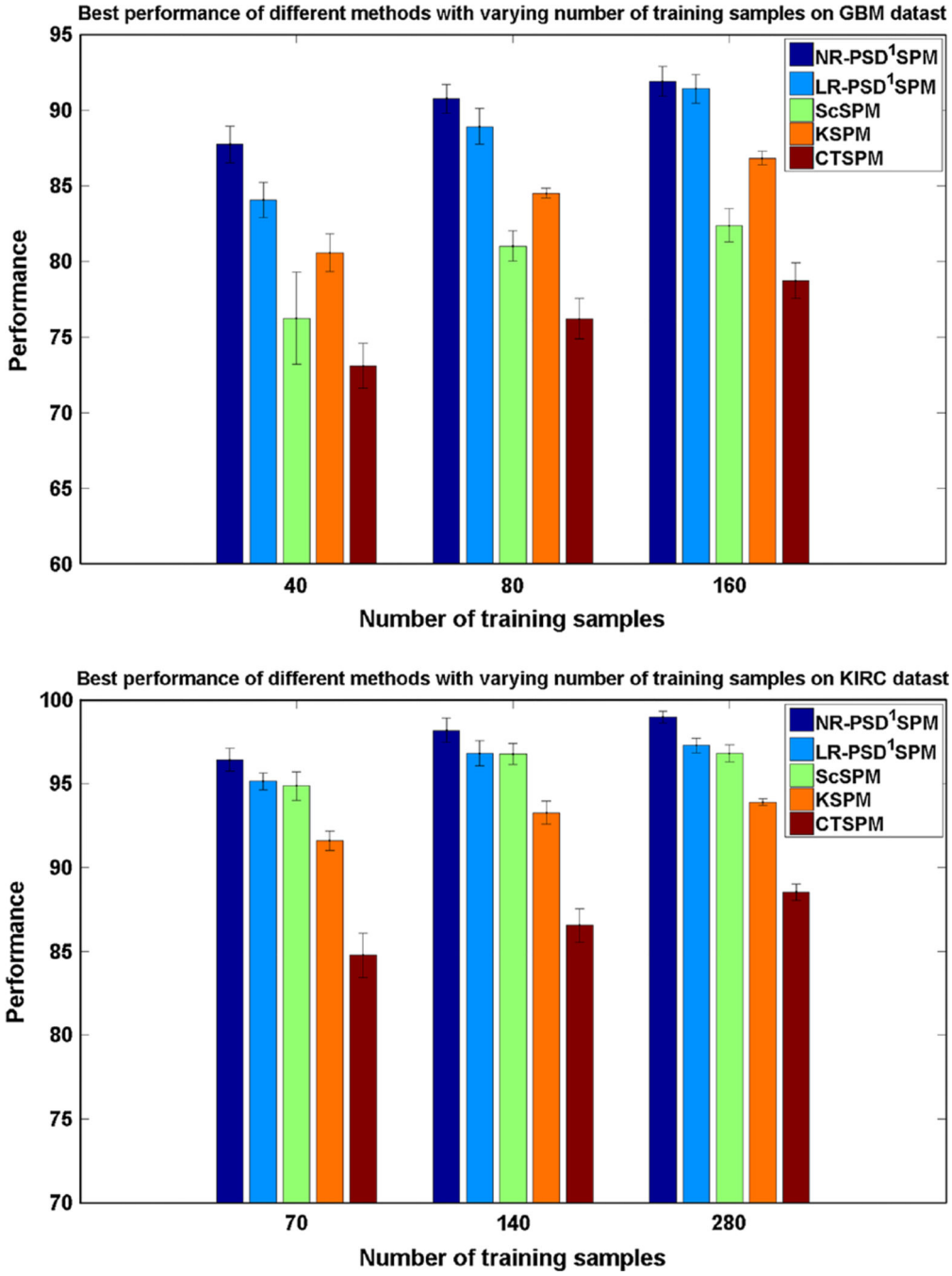

Fig. 7.

Comparison of performance between systems based on human engineered features and systems based on unsupervised feature learning

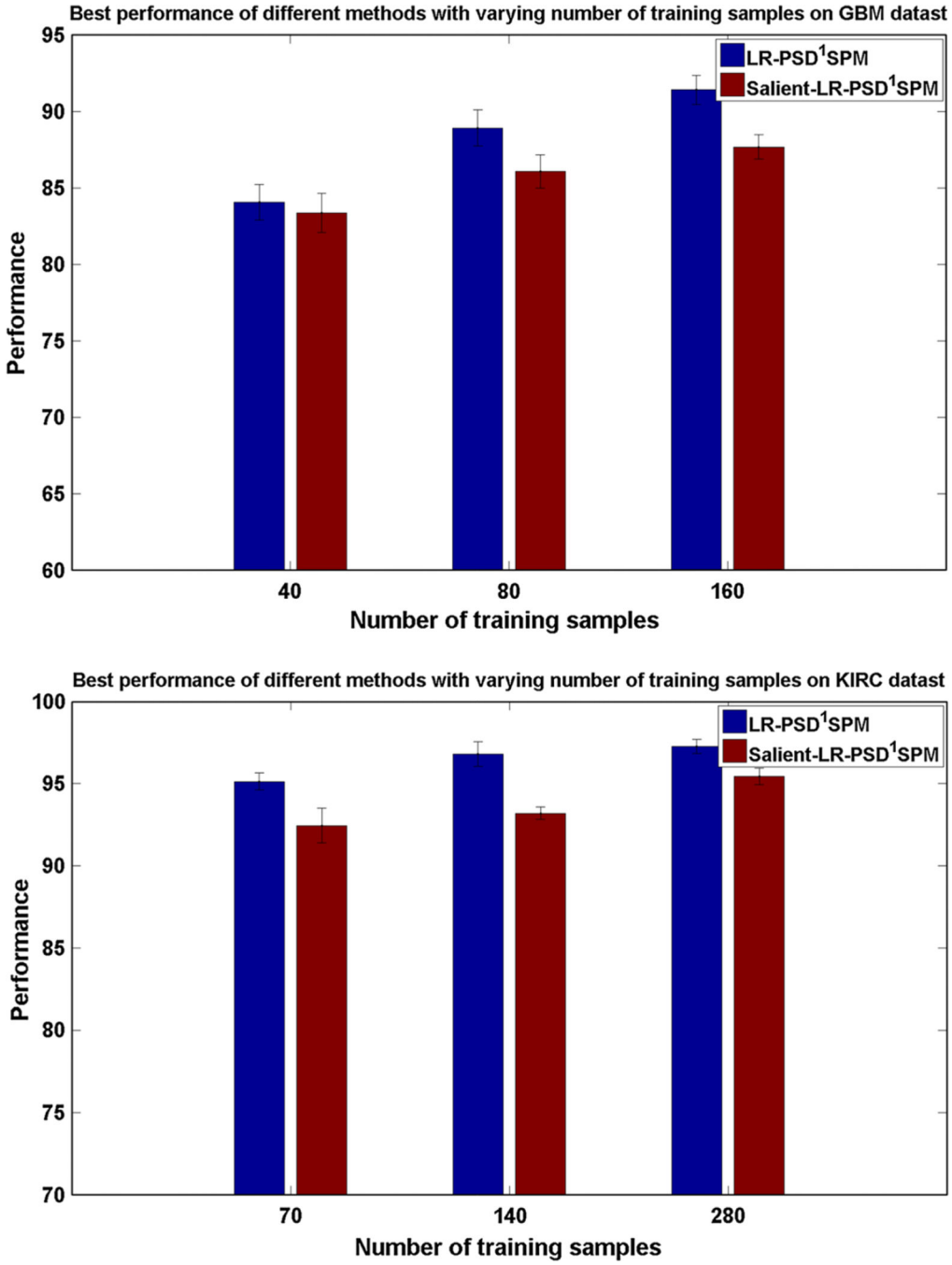

Fig. 8.

Comparison of performance between systems with/without the incorporation of saliency information during feature extraction

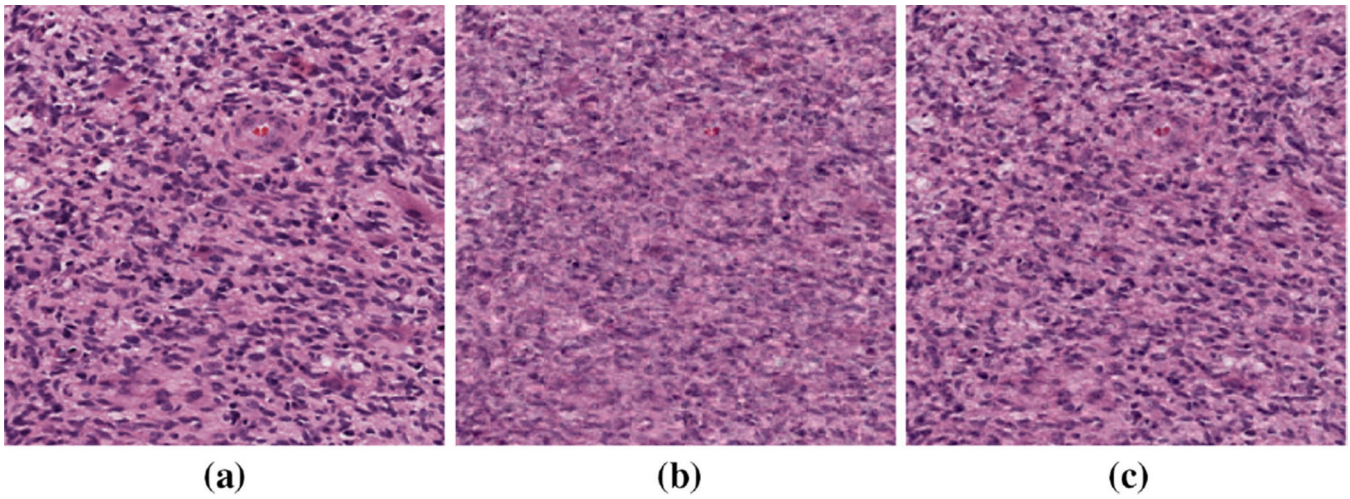

Fig. 9.

Comparison of PSD with linear and nonlinear regressors in terms of reconstruction. **a** Original image; **b** reconstruction by PSD with linear regressor (SNR = 4.9429); **c** Reconstruction by PSD with nonlinear regressor (SNR = 9.3436)

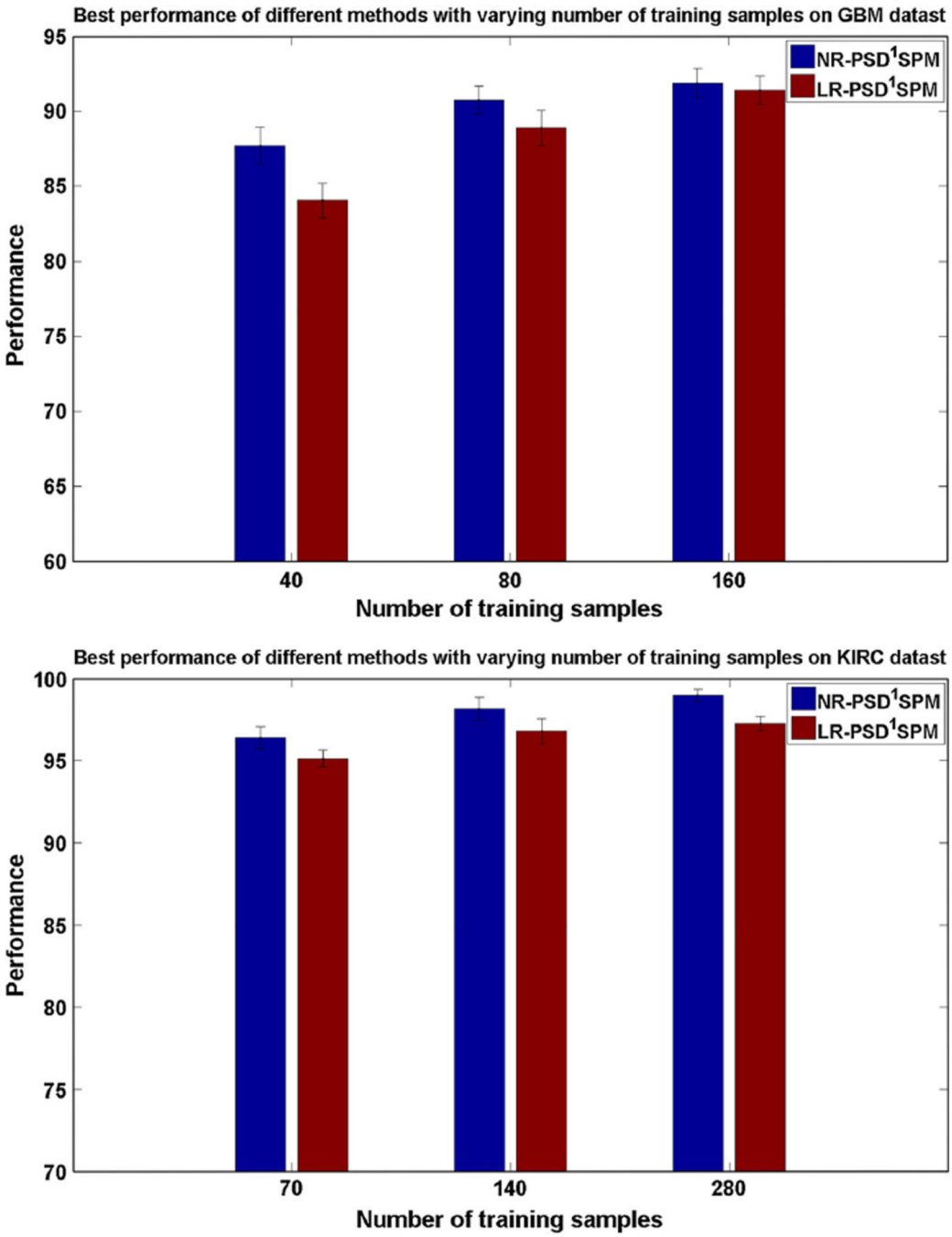

Fig. 10.

Comparison of PSD with linear and nonlinear regressors for sparse feature approximation in terms of classification

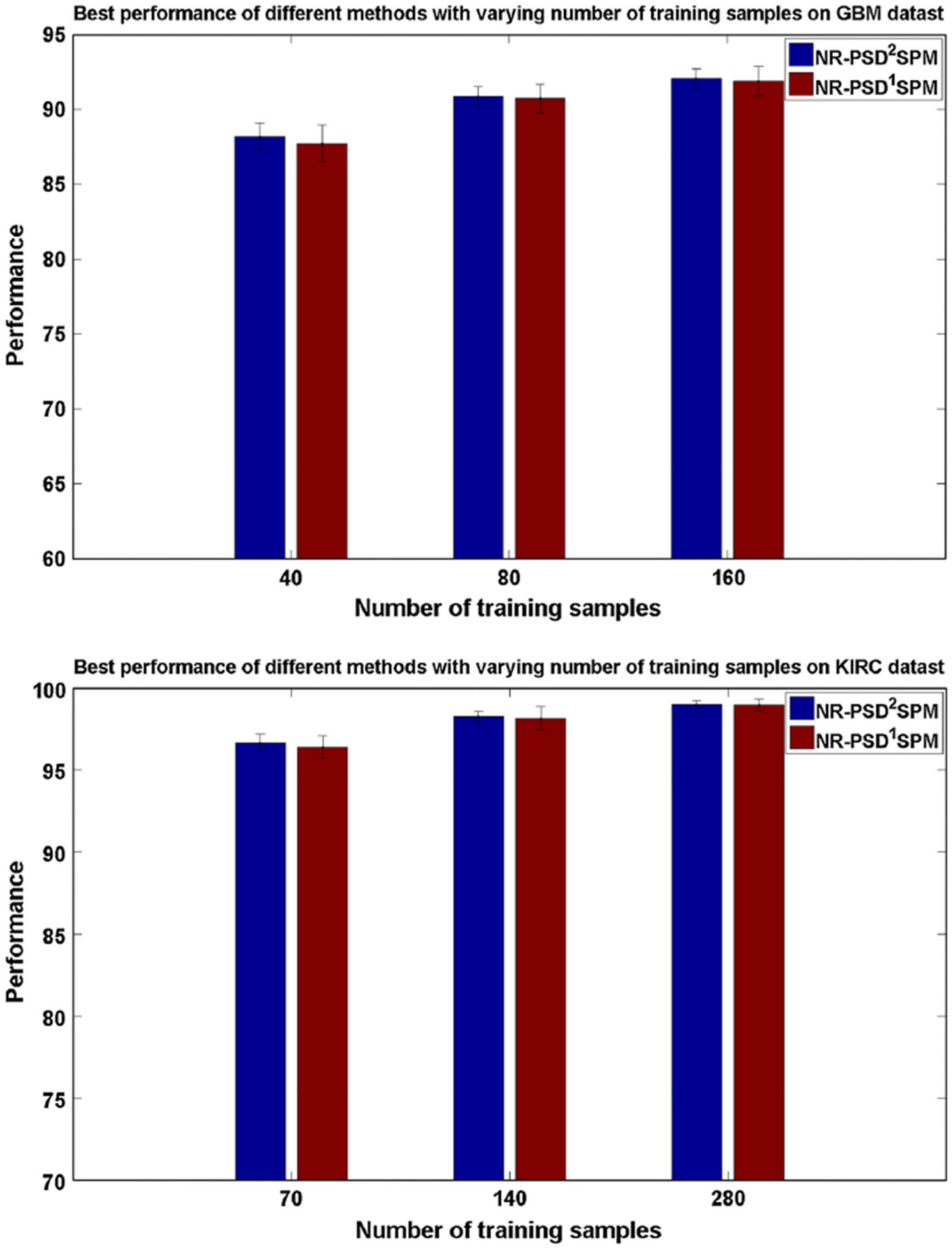

Fig. 11.
Comparison of performance between systems with various number of stages

Table 1

Performance of different methods on the GBM dataset

Method	Dictionary size=256	Dictionary size=512	Dictionary size=1,024
160 Training			
NR-PSD^{2}SPM	91.85 ± 1.03	91.86 ± 0.78	**92.07 ± 0.65**
NR-PSD^{1}SPM	91.85 ± 0.69	91.89 ± 0.99	91.74 ± 0.85
LR-PSD^{1}SPM Chang et al. (2013b)	91.02 ± 1.89	91.41 ± 0.95	91.20 ± 1.29
Salient-LR-PSD^{1}SPM	87.67 ± 0.78	87.67 ± 0.44	87.32 ± 0.96
LR-PSD Chang et al. (2013b)	86.07 ± 1.42	86.32 ± 1.14	86.15 ± 1.33
ScSPM Yang et al. (2009)	79.58 ± 0.61	81.29 ± 0.86	82.36 ± 1.10
KSPM Lazebnik et al. (2006)	85.00 ± 0.79	86.47 ± 0.55	86.81 ± 0.45
CTSPM	78.61 ± 1.33	78.71 ± 1.18	78.69 ± 0.81
80 Training			
NR-PSD^{2}SPM	90.51 ± 1.06	**90.88 ± 0.66**	90.51 ± 1.06
NR-PSD^{1}SPM	90.74 ± 0.95	90.42 ± 0.94	89.70 ± 1.20
LR-PSD^{1}SPM Chang et al. (2013b)	88.63 ± 0.91	88.91 ± 1.18	88.64 ± 1.08
Salient-LR-PSD^{1}SPM	86.07 ± 1.08	86.22 ± 0.76	85.75 ± 1.26
LR-PSD Chang et al. (2013b)	81.73 ± 0.98	82.08 ± 1.23	81.55 ± 1.17
ScSPM Yang et al. (2009)	77.65 ± 1.43	78.31 ± 1.13	81.00 ± 0.98
KSPM Lazebnik et al. (2006)	83.81 ± 1.22	84.32 ± 0.67	84.49 ± 0.34
CTSPM	75.93 ± 1.18	76.06 ± 1.52	76.19 ± 1.33
40 Training			
NR-PSD^{2}SPM	87.90 ± 0.91	**88.21 ± 0.90**	87.71 ± 0.81
NR-PSD^{1}SPM	87.72 ± 1.21	86.99 ± 1.76	86.33 ± 1.32
LR-PSD^{1}SPM Chang et al. (2013b)	84.06 ± 1.16	83.72 ± 1.46	83.40 ± 1.14
Salient-LR-PSD^{1}SPM	83.37 ± 1.28	83.19 ± 1.08	82.52 ± 1.28
LR-PSD Chang et al. (2013b)	78.28 ± 1.74	78.15 ± 1.43	77.97 ± 1.65
ScSPM Yang et al. (2009)	73.60 ± 1.68	75.58 ± 1.29	76.24 ± 3.05
KSPM Lazebnik et al. (2006)	80.54 ± 1.21	80.56 ± 1.24	80.46 ± 0.56
CTSPM	73.10 ± 1.51	72.90 ± 1.09	72.65 ± 1.41

The best performances with different training samples are printed in bold, which indicate the superior performance of proposed method over others

Table 2

Performance of different methods on the KIRC dataset

Method	Dictionary size=256	Dictionary size=512	Dictionary size=1,024
280 Training			
NR-PSD^{2}SPM	**99.03 ± 0.20**	98.89 ± 0.19	98.92 ± 0.21
NR-PSD^{1}SPM	98.98 ± 0.35	98.81 ± 0.45	98.69 ± 0.41
LR-PSD^{1}SPM Chang et al. (2013b)	97.19 ± 0.49	97.27 ± 0.44	97.08 ± 0.45
Salient-LR-PSD^{1}SPM	95.45 ± 0.50	95.17 ± 0.43	94.98 ± 0.32
LR-PSD Chang et al. (2013b)	90.72 ± 1.32	90.18 ± 0.88	90.43 ± 0.80
ScSPM Yang et al. (2009)	94.52 ± 0.44	96.37 ± 0.45	96.81 ± 0.50
KSPM Lazebnik et al. (2006)	93.55 ± 0.31	93.76 ± 0.27	93.90 ± 0.19
CTSPM	87.45 ± 0.59	87.95 ± 0.49	88.53 ± 0.49
140 Training			
NR-PSD^{2}SPM	**98.26 ± 0.34**	98.07 ± 0.46	97.85 ± 0.56
NR-PSD^{1}SPM	98.17 ± 0.72	98.05 ± 0.71	97.99 ± 0.82
LR-PSD^{1}SPM Chang et al. (2013b)	96.80 ± 0.75	96.52 ± 0.76	96.55 ± 0.84
Salient-LR-PSD^{1}SPM	93.20 ± 0.37	93.18 ± 0.65	92.78 ± 0.53
LR-PSD Chang et al. (2013b)	88.75 ± 0.37	88.93 ± 0.45	87.98 ± 0.86
ScSPM Yang et al. (2009)	93.46 ± 0.55	95.68 ± 0.36	96.76 ± 0.63
KSPM Lazebnik et al. (2006)	92.50 ± 1.12	93.06 ± 0.82	93.26 ± 0.68
CTSPM	86.55 ± 0.99	86.40 ± 0.54	86.49 ± 0.58
70 Training			
NR-PSD^{2}SPM	**96.67 ± 0.53**	96.20 ± 0.54	95.57 ± 0.66
NR-PSD^{1}SPM	96.42 ± 0.68	96.41 ± 0.59	96.03 ± 0.69
LR-PSD^{1}SPM Chang et al. (2013b)	95.12 ± 0.54	95.13 ± 0.51	95.09 ± 0.40
Salient-LR-PSD^{1}SPM	92.45 ± 1.06	92.32 ± 1.13	92.31 ± 0.97
LR-PSD Chang et al. (2013b)	87.56 ± 0.78	87.93 ± 0.67	87.13 ± 0.97
ScSPM Yang et al. (2009)	91.93 ± 1.00	93.67 ± 0.72	94.86 ± 0.86
KSPM Lazebnik et al. (2006)	90.78 ± 0.98	91.34 ± 1.13	91.59 ± 0.97
CTSPM	84.76 ± 1.32	84.29 ± 1.53	83.71 ± 1.42

The best performances with different training samples are printed in bold, which indicate the superior performance of proposed method over others

Table 3

Best performance of different architectures within the NR-PSD^{1}SPM framework on the GBM dataset with 160 training images per category, where ABS means absolute value rectification; LCN means local contrast normalization; MP means max-pooling

Structure	Best performance
NR-PSD^{1}SPM	91.89 ± 0.99
$\mathrm{NR-PSD}^1_{ABS}\mathrm{SPM}$	91.85 ± 0.83
$\mathrm{NR-PSD}^1_{LCN}\mathrm{SPM}$	89.13 ± 0.94
$\mathrm{NR-PSD}^1_{MP}\mathrm{SPM}$	91.87 ± 0.90
$\mathrm{NR-PSD}^1_{ABS+LCN+MP}\mathrm{SPM}$	89.35 ± 0.77

Table 4

Best performance of different architectures within the NR-PSD^{1}SPM framework on the KIRC dataset with 280 training images per category, where ABS means absolute value rectification; LCN means local contrast normalization; MP means max-pooling

Structure	Best performance
NR-PSD^{1}SPM	98.98 ± 0.35
$\mathrm{NR-PSD}^1_{ABS}\mathrm{SPM}$	98.95 ± 0.43
$\mathrm{NR-PSD}^1_{LCN}\mathrm{SPM}$	95.97 ± 0.88
$\mathrm{NR-PSD}^1_{MP}\mathrm{SPM}$	98.97 ± 0.50
$\mathrm{NR-PSD}^1_{ABS+LCN+MP}\mathrm{SPM}$	95.81 ± 0.67