Classification of Histology Sections via Multispectral Convolutional Sparse Coding[*]

Yin Zhou[1†]   Hang Chang[3,4†⋆]   Kenneth Barner[1]   Paul Spellman[2]   Bahram Parvin[3,4⋆]

[1]ECE Department, University of Delaware, Newark, Delaware, U.S.A

{zhouyin, barner}@udel.edu

[2] Center for Spatial Systems Biomedicine, Oregon Health Sciences University, Portland, Oregon, U.S.A

spellmap@ohsu.edu

[3]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A
[4]University of California, Riverside, U.S.A

{hchang, b_parvin}@lbl.gov, † Equal-Contribution Authors, ⋆ Co-Corresponding Authors

Abstract

Image-based classification of histology sections plays an important role in predicting clinical outcomes. However this task is very challenging due to the presence of large technical variations (e.g., fixation, staining) and biological heterogeneities (e.g., cell type, cell state). In the field of biomedical imaging, for the purposes of visualization and/or quantification, different stains are typically used for different targets of interest (e.g., cellular/subcellular events), which generates multi-spectrum data (images) through various types of microscopes and, as a result, provides the possibility of learning biological-component-specific features by exploiting multispectral information. We propose a multispectral feature learning model that automatically learns a set of convolution filter banks from separate spectra to efficiently discover the intrinsic tissue morphometric signatures, based on convolutional sparse coding (CSC). The learned feature representations are then aggregated through the spatial pyramid matching framework (SPM) and finally classified using a linear SVM. The proposed system has been evaluated using two large-scale tumor cohorts, collected from The Cancer Genome Atlas (TCGA). Experimental results show that the proposed model 1) outperforms systems utilizing sparse coding for unsupervised feature learning (e.g., PSD-SPM [5]); 2) is competitive with systems built upon features with biological prior knowledge (e.g., SMLSPM [4]).
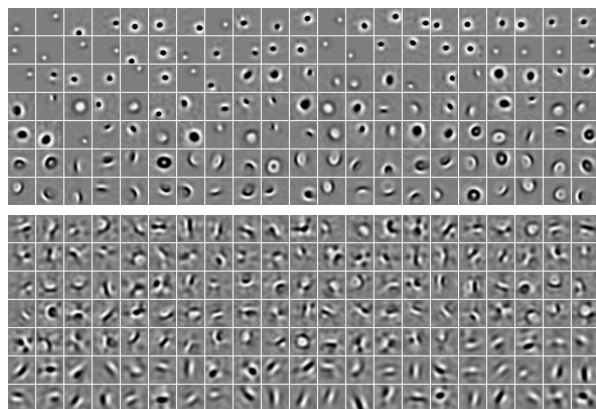
Figure 1. 27×27 multispectral filters learned from the Glioblastoma Multiforme (GBM) dataset, where each tissue image is decomposed into two channels corresponding to the nuclei and protein contents with the learned filters shown in top and bottom figures, respectively.

1. Introduction

Histology sections contain significant information about the tissue architecture. Hematoxylin and eosin (H&E) are two commonly used histological stains, which respectively label DNA (*e.g.,* nuclei) and protein contents, with various color shades. Abberations in the histology architecture are often seen as an indicator of the disease progression and subtypes. Therefore, computed indices, for each aberrant phenotypic signature, enable the prediction of clinical outcomes *e.g.,* survival, response to therapy. However, as an essential ground on which outcome-based analysis is established, large cohorts usually contain large technical variations and biological heterogeneities, which greatly under-

mines the performance of existing techniques [4, 5].

To solve such problems, several researchers [4, 15, 16] have proposed to design and fine tune the human-engineered features. These approaches are usually task-specific, which limits their cross-domain applicability. Not until recently has the potential of unsupervised feature learning been exploited in tissue classification [5]. These methods demonstrate very encouraging results compared to manually designed features. Yet, their underlying feature learning module is sparse coding, which suffers two major drawbacks, *viz.,* 1) yielding only Gabor-like low-level feature detectors (filters), and 2) having high redundance in the feature representation.

In this paper, we propose a multispectral unsupervised feature learning model (MCSCSPM) for tissue classification, based on convolutional sparse coding (CSC) [14] and spatial pyramid matching (SPM) [17]. The multispectral features are learned in an unsupervised manner through CSC, followed by the summarization through SPM at various scales and locations. Eventually, the image-level tissue representation is fed into linear SVM for efficient classification [9]. Compared with sparse coding, CSC possesses two merits: 1) invariance to translation; and 2) producing more complex filters, which contribute to more succinct feature representations. Meanwhile, the proposed approach also benefits from: 1) the biomedical intuitions that different color spectrums typically characterize distinct structures; and 2) the utilization of context, provided by SPM, which is important in diagnosis. In short, our work is the first attempt using convolutional sparse coding for tissue classification, and achieves superior performance compared to patch-based sparse feature learning algorithms, *e.g.,* PSD-SPM [5]. Moreover, MCSCSPM is capable of generating very competitive results compared to systems built upon biological prior knowledge, *i.e.,* SMLSPM [4]. Finally, our study further indicates that learning features over multiple spectra can potentially generate biological-component-specific filters. For example, the filters learned from the nuclear channel and protein/extracellular matrix channel respectively capture various nuclear regions and the structural connectivity within tissue sections.

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

2. Related Work

There are several excellent reviews, in the literature, for the analysis of H&E stained sections [7, 11]. Generally speaking, efforts in histology section analysis can be divided into three different directions: 1) some researchers [1, 2, 6, 8] advocate nuclear segmentation and organization

for tumor grading and/or the prediction of tumor recurrence; 2) some groups [12, 15] focus on patch level analysis (*e.g.,* small regions), using color and texture features, for tumor representation; 3) there is also a research branch [10] suggesting detection and representation of the auto-immune response as a prognostic tool for cancer.

Tissue classification is a challenging task due to the presence of significant technical variations and biological heterogeneities in the data [4, 16], which typically results in techniques that are tumor-type specific. To overcome this problem, recent studies have focused on either fine tuning human engineered features [15, 16], or applying automatic feature learning [5], for robust representation.

In recent years, convolutional sparse coding has received increasing research interest in computer vision and machine learning communities [3, 14, 21, 24, 25], due mainly to its capability of learning shift-invariant filters with complex patterns. Kavukcuoglu *et al.* [14] proposed to improve the feature extraction efficiency by jointly learning a feed-forward encoder with the convolutional filter bank, and applied the algorithm to Convolutional Networks (ConvNets), achieving impressive results on object recognition. Zeiler *et al.* [24] developed the Deconvolutional Networks for learning top-bottom feature hierarchies to reconstruct the original image, and further extended it by incorporating a set of latent switch variables and max-pooling, which allows unified training of multiple layers [25]. Bristow *et al.* [3] came up with an efficient method for convolutional sparse coding in Fourier domain, using the Alternating Direction Method of Multipliers approach. In addition to object recognition, convolutional sparse coding has also achieved state-of-the-art performances in pedestrian detection [21] and retinal blood vessels segmentation [19], etc.

3. Proposed Approach

In this paper, we adopt CSC [14] as the fundamental module for learning filter banks, based on which the proposed multispectral unsupervised feature learning system (MCSCSPM) is constructed. As noted by several researchers [3, 14], sparse coding typically assumes that training image patches are independent from each other, and thus neglects the spatial correlation among them. In practice, however, this assumption typically leads to filters that are simply translated versions of each other, and, as a result, generates highly redundant feature representation. In contrast, CSC generates more compact features due to its intrinsic shift-invariant property. Moreover, CSC is capable of generating more complex filters capturing higher-older image statistics, compared to sparse coding that basically learns edge primitives [14].

In the proposed multispectral feature learning framework, CSC is applied to each separate spectral channel, yielding target-specific filter banks. For instance, some bi-

ologically meaningful filters are learned from the nuclear channel and the protein/extracellular matrix channel respectively, as illustrated in Figure 1. Features extracted from multiple spectra are summarized by SPM [17] at various scales and locations, and ultimate tissue representations are fed into linear SVM [9] for classification.

3.1. Convolutional Sparse Coding

Let $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ be a training set containing N 2D images with dimension $m \times n$. Let $\mathbf{D} = \{\mathbf{d}_k\}_{k=1}^K$ be the 2D convolutional filter bank having K filters, where each $\mathbf{d}_k$ is an $h \times h$ convolutional kernel. Define $\mathbf{Z} = \{\mathbf{Z}^i\}_{i=1}^N$ be the set of sparse feature maps such that subset $\mathbf{Z}^i = \{\mathbf{z}_k^i\}_{k=1}^K$ consists of K feature maps for reconstructing image $\mathbf{x}_i$, where $\mathbf{z}_k^i$ has dimension $(m+h-1) \times (n+h-1)$. Convolutional sparse coding aims to decompose each training image $\mathbf{x}_i$ as the sum of a series of sparse feature maps $\mathbf{z}_k^i \in \mathbf{Z}^i$ convolved with kernels $\mathbf{d}_k$ from the filter bank $\mathbf{D}$, by solving the following objective function:

$$\min_{\mathbf{D},\mathbf{Z}} \quad \mathcal{L} = \sum_{i=1}^N \left\{ \left\| \mathbf{x}_i - \sum_{k=1}^K \mathbf{d}_k * \mathbf{z}_k^i \right\|_{\mathrm{F}}^2 + \alpha \sum_{k=1}^K \left\| \mathbf{z}_k^i \right\|_1 \right\}$$
$$\text{s.t.} \quad \|\mathbf{d}_k\|_2^2 = 1, \forall k = 1, \ldots, K \qquad (1)$$

where the first and the second term represent the reconstruction error and the ℓ_1-norm penalty respectively; α is a regularization parameter; $*$ is the 2D discrete convolution operator; and filters are restricted to have unit energy to avoid trivial solutions. Note that here $\|\mathbf{z}\|_1$ represents the entrywise matrix norm, *i.e.,* $\|\mathbf{z}\|_1 = \sum_{i,j} |z_{ij}|$, where z_{ij} is the entry at location (i,j) of a feature map $\mathbf{z} \in \mathbf{Z}$. The construction of $\mathbf{D}$ is realized by balancing the reconstruction error and the ℓ_1-norm penalty.

Note that the objective of Eq. (1) is not jointly convex with respect to (w.r.t.) $\mathbf{D}$ and $\mathbf{Z}$ but is convex w.r.t. one of the variables with the other remaining fixed [18]. Thus, we solve Eq. (1) by alternatively optimizing the two variables, *i.e.,* iteratively performing the two steps that first compute $\mathbf{Z}$ and then update $\mathbf{D}$. We use the Iterative Shrinkage Thresholding Algorithm (ISTA) to solve for the sparse feature maps $\mathbf{Z}$. The updating policy for the convolutional dictionary $\mathbf{D}$ uses the stochastic gradient descent for efficient estimation of the gradient by considering one training sample at a time [14]. The optimization procedure is sketched in Algorithm 1. Alternative methods for updating the dictionary can be found in [3, 24, 25].

3.2. Feature Extraction

In the field of biomedical imaging, different spectra usually capture distinct targets of interest. Specifically, in our case, color decomposition [20] produces two separate spectra (channels) which characterize the nuclear chromatin and

---

Algorithm 1 CSC Algorithm
___
Input: Training set $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$, K, α
Output: Convolutional filter bank $\mathbf{D} = \{\mathbf{d}_k\}_{k=1}^K$
1: **Initialize:** $\mathbf{D} \sim \mathcal{N}(0,1)$, $\mathbf{Z} \leftarrow \mathbf{0}$
2: **repeat**
3:     **for** $i = 1$ to N **do**
4:         Normalize each kernel in $\mathbf{D}$ to unit energy
5:         Fixing $\mathbf{D}$, compute sparse feature maps $\mathbf{Z}^i$ by solving
$$\mathbf{Z}^i \leftarrow \arg\min_{\mathbf{z}_k^i \in \mathbf{Z}^i} \|\mathbf{x}_i - \sum_{k=1}^K \mathbf{d}_k * \mathbf{z}_k^i\|_{\mathrm{F}}^2 + \alpha \sum_{k=1}^K \|\mathbf{z}_k^i\|_1$$
6:         Fixing $\mathbf{Z}$, update $\mathbf{D}$ as
        $\mathbf{D} \leftarrow \mathbf{D} - \mu \nabla_{\mathbf{D}} \mathcal{L}(\mathbf{D}, \mathbf{Z})$
7:     **end for**
8: **until** Convergence (maximum iterations reached or objective function $\leq$ threshold)

---

the protein/extracellular matrix, respectively (as shown in Figure 2). Therefore, in the filter learning phase, we propose to apply convolutional sparse coding to each spectrum, separately, for the purpose of learning biological-component-specific feature detectors. Without the loss of generality, we assume that the number of filters for each spectrum (channel) is K and there are S spectra (channels) after decomposition; the 2D feature map $\mathbf{y}_k^s$ is then defined as: $\mathbf{y}_k^s = \mathbf{d}_k^s * \hat{\mathbf{x}}^s$, for $1 \leq k \leq K$ and $1 \leq s \leq S$, where $\hat{\mathbf{x}}^s$ is the s-th spectrum component of input image $\mathbf{x}$ and $\mathbf{d}_k^s \in \mathbf{D}^s$ is the k-th convolutional kernel in filter bank $\mathbf{D}^s$ learned over spectrum with index s.

Upon learning the filter bank, we extract multispectral tissue histology features using the proposed framework illustrated in Figure 2, where an input image is first decomposed and divided into several spectral channels and then each decomposed component is convolved with the channel-specific filter bank followed by three cascaded layers, namely, element-wise absolute value rectification (Abs), local contrast normalization (LCN), and max-pooling (MP) [13]. Note that for specificity, the model in Figure 2 shows only two spectra, but it is straightforward to generalize to hyperspectral image-based applications. The Abs layer computes absolute value element wisely in each feature map, $\mathbf{y}_k^s$, to avoid the cancelation effect in sequential operations. The LCN layer aims to enhance the stronger feature responses and suppress weaker ones across feature maps, $\{\mathbf{y}_k^s\}_{k=1}^K$, in each spectrum, by performing local subtractive and divisive operations[1]. The MP layer partitions each feature map into non-overlapping windows and extracts the maximum response from each of the pooling window. The MP operation allows local invariance to trans-

---
[1]Limited by space, we refer readers to [13, 21] for detailed discussions on local contrast normalization.
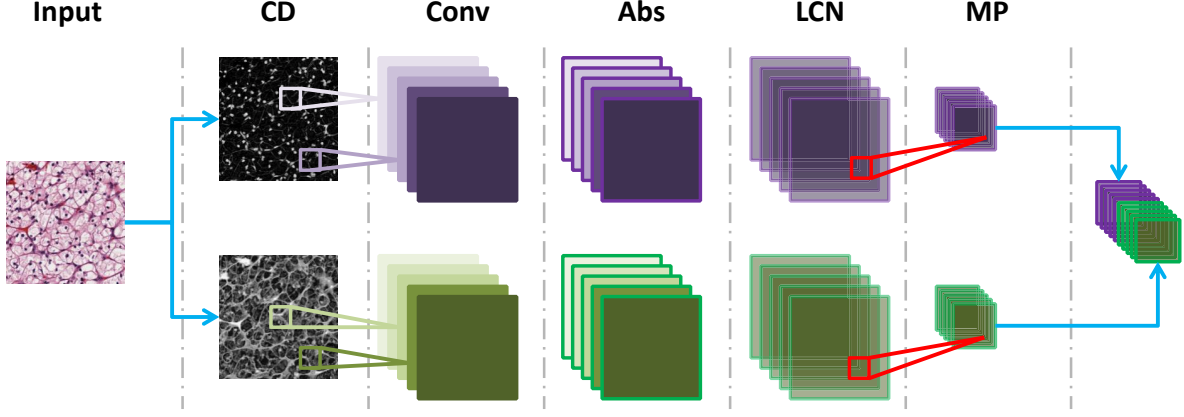
Figure 2. The proposed multispectral feature extraction framework. CD means color decomposition; Abs means absolute value rectification; LCN means local contrast normalization; MP means max-pooling. The figure is best viewed in color at 150% zoom-in.

lation [13]. Finally, the multispectral tissue features are formed by aggregating feature responses from all spectra.

We further denote the multispectral tissue features of image, $\mathbf{x}$, as a 3D array, $\mathbf{U} \in \mathbb{R}^{a \times b \times KS}$, where the first two dimensions indicate the horizontal and vertical locations of a feature vector in the image plane and the third dimension represents the length of feature vectors. The multispectral tissue features are then fed into SPM framework for classification as detailed in the following section.

3.3. SPM

Let $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_T] \in \mathbb{R}^{KS \times T}$ be the feature set of T feature vectors having dimension KS. In the standard SPM framework [17], the first step is to construct a codebook $\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_P] \in \mathbb{R}^{KS \times P}$, which includes P multispectral tissue morphometric types, by solving the following optimization problem:

$$\min_{\mathbf{B}, \mathbf{C}} \quad \sum_{i=1}^{T} \|\mathbf{v}_i - \mathbf{B}\mathbf{c}_i\|^2 \qquad (2)$$
$$\text{s.t.} \quad card(\mathbf{c}_i) = 1, \|\mathbf{c}_i\|_1 = 1, \mathbf{c}_i \succeq 0, \forall i$$

where $\mathbf{C} = [\mathbf{c}_1, ..., \mathbf{c}_T] \in \mathbb{R}^{P \times T}$ is a set of codes for reconstructing $\mathbf{V}$, cardinality constraint $card(\mathbf{c}_i)$ enforces $\mathbf{c}_i$ to have only one nonzero element, $\mathbf{c}_i \succeq 0$ is a non-negative constraint on all vector elements. Eq. (2) is optimized by alternating between the two variables, *i.e.,* minimizing one while keeping the other fixed. After training, the query signal set $\mathbf{V}$ is encoded via Vector Quantization (VQ) based on codebook $\mathbf{B}$, *i.e.,* assigning each $\mathbf{v}_i$ to its closest multispectral tissue type in $\mathbf{B}$.

The second step is to construct the spatial histogram for SPM [17]. This is done by dividing an image into increasingly finer subregions and computing local histograms of different multispectral tissue types falling into each of the subregions. The spatial histogram, H, is then formed

by concatenating the appropriately weighted histograms of multispectral tissue types at all resolutions, *i.e.,*

$$
\begin{aligned}
H_0 &= H_0^0 \\
H_l &= (H_l^1, ..., H_l^{4^l}), 1 \leq l \leq L \qquad (3) \\
H &= (\frac{1}{2^L} H_0, \frac{1}{2^L} H_1, ..., \frac{1}{2^{L-l+1}} H_l, ..., \frac{1}{2} H_L)
\end{aligned}
$$

where $(\cdot)$ is the vector concatenation operator, $l \in \{0, ..., L\}$ is the resolution level of the image pyramid, and H_l represents the concatenation of histograms for all image subregions at pyramid level l. In tissue classification, SPM intrinsically summarizes tissue morphometric contexts by computing and aggregating local histograms at various scales and locations. This is analogous to the fact that pathologists use "contexts" to determine a disease state [4]. For the final classification, a homogeneous kernel map [22] is employed to approximate χ^2 kernel, which enables efficient linear SVM [9] training and classification.

4. Experiments

In this section, we present detailed experimental design and evaluation of the proposed approach in tissue histopathology classification. The two distinct tumor datasets, for evaluation, are curated from The Cancer Genome Atlas (TCGA), namely (i) Glioblastoma Multiforme (GBM) and (ii) Kidney Renal Clear Cell Carcinoma (KIRC), which are publicly available from the NIH (National Institute of Health) repository.

4.1. Experimental Setup

We have evaluated the proposed method (MCSCSPM) in three different variations:

1. MCSCSPM-HE: Convolutional filter banks are learned from / applied to decomposed spectrum (channel) separately. Here, we have two spectra after de-

composition, which correspond to nuclear chromatin (stained with hematoxylin) and protein/extracellular matrix (stained with eosin), respectively.

2. MCSCSPM-RGB: Convolutional filter banks are learned from / applied to R, G, and B channels separately.

3. MCSSPM-Gray: Convolutional filter banks are learned from / applied to the grayscale image.

and compared its performance with other four classification methods on the GBM and KIRC datasets. Implementation details of all approaches involved are listed as follows:

1. MCSCSPM: the nonlinear kernel SPM that uses spatial-pyramid histograms of multispectral tissue types and homogeneous kernel map. In the multispectral case, an input tissue image was decomposed into two spectra (*i.e.,* $S = 2$) corresponding to the nuclear chromatin and the protein/extracellular matrix respectively, based on the optical density matrix established in [20]. In the RGB and grayscale case, each color channel was treated as one spectrum. For each spectrum, images were preprocessed with a 13×13 Gaussian filter. During training, we set K to 150 and 300 per spectrum for the GBM and KIRC datasets, respectively. The filter dimension was 27×27 for both datasets. The sparsity regularization parameter α was set to 0.1 for best performance. During multispectral feature extraction, we used the same 13×13 Gaussian filter for local contrast normalization and empirically set the max-pooling stepsize to be 27.

2. PSDSPM [5]: the nonlinear kernel SPM that uses spatial-pyramid histograms of sparse tissue morphometric types and homogeneous kernel map. The image patch size was set to 20×20, the number of basis function was set to 1024 and the sparsity regularization parameter was set to 0.3 for best performance.

3. ScSPM [23]: the linear SPM that uses linear kernel on spatial-pyramid pooling of SIFT sparse codes. The dense SIFT features was extracted on 16×16 patches sampled from each image on a grid with stepsize 8 pixels. The sparsity regularization parameter λ was set to 0.15, to achieve the best performance;

4. KSPM [17]: the nonlinear kernel SPM that uses spatial-pyramid histograms of SIFT features and homogeneous kernel map. The dense SIFT features was extracted on 16×16 patches sampled from each image on a grid with stepsize 8 pixels;

5. SMLSPM [4]: the linear SPM that uses linear kernel on spatial-pyramid pooling of cellular morphometric sparse codes.
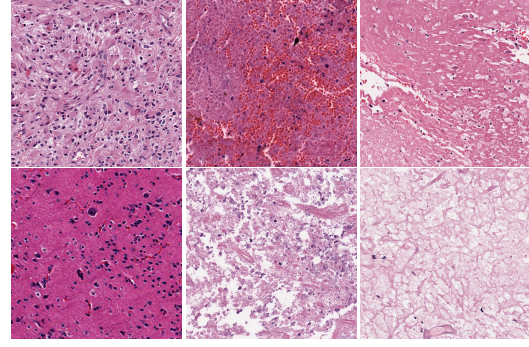

Figure 3. GBM Examples. First column: Tumor; Second column: Transition to necrosis; Third column: Necrosis. Note that the phenotypic heterogeneity is highly diverse in each column.

On the implementation of SPM for MCSCSPM, PSDSPM, KSPM and SMLSPM, we use the standard K-means clustering for constructing the dictionary and set the level of pyramid to be 3. Following the conventional evaluation procedure, we repeat all experiments 10 times with random splits of training and test set to obtain reliable results. The final results are reported as the mean and standard deviation of the classification rates on the following two distinct datasets, which include vastly different tumor types:

1. GBM Dataset. It contains 3 classes: Tumor, Necrosis, and Transition to Necrosis, which were curated from whole slide images (WSI) scanned with a 20X objective (0.502 micron/pixel). Examples can be found in Figure 3. The number of images per category are 628, 428 and 324, respectively. Most images are 1000×1000 pixels. In this experiment, we train on 40, 80 and 160 images per category and tested on the rest, with three different dictionary sizes: 256, 512 and 1024. Detailed comparisons are shown in Table 1.

2. KIRC Dataset. It contains 3 classes: Tumor, Normal, and Stromal, which were curated from whole slide images (WSI) scanned with a 40X objective (0.252 micron/pixel). Examples can be found in Figure 4. The number of images per category are 568, 796 and 784, respectively. Most images are 1000×1000 pixels. In this experiment, we train on 70, 140 and 280 images per category and tested on the rest, with three different dictionary sizes: 256, 512 and 1024. Detailed comparisons are shown in Table 2.

4.2. Discussion

1. Multispectral (HE) vs. RGB v.s. Gray. For GBM dataset, K was fixed to be 150 per spectrum (channel), which led to a total number of 300, 450 and 150 filters for MCSCSPM-HE, MCSCSPM-RGB and MCSCSPM-Gray, respectively. For the KIRC dataset,

	Method	DictionarySize=256	DictionarySize=512	DictionarySize=1024
160 training	MCSCSPM-HE	92.71 ± 0.91	**93.01 ± 1.10**	92.65 ± 0.75
	MCSCSPM-RGB	**92.58 ± 0.94**	92.50 ± 0.86	92.47 ± 0.73
	MCSCSPM-Gray	86.33 ± 1.12	**86.74 ± 0.91**	86.69 ± 0.81
	PSDSPM [5]	91.02 ± 1.89	**91.41 ± 0.95**	91.20 ± 1.29
	SMLSPM [4]	92.35 ± 0.83	92.57 ± 0.91	**92.91 ± 0.84**
	ScSPM [23]	79.58 ± 0.61	81.29 ± 0.86	**82.36 ± 1.10**
	KSPM [17]	85.00 ± 0.79	86.47 ± 0.55	**86.81 ± 0.45**
80 training	MCSCSPM-HE	**91.41 ± 1.07**	91.19 ± 0.91	91.13 ± 0.93
	MCSCSPM-RGB	90.88 ± 1.06	**91.28 ± 0.82**	90.85 ± 0.67
	MCSCSPM-Gray	**84.67 ± 1.63**	84.53 ± 1.58	84.56 ± 1.62
	PSDSPM [5]	88.63 ± 0.91	**88.91 ± 1.18**	88.64 ± 1.08
	SMLSPM [4]	90.82 ± 1.28	90.29 ± 0.68	**91.08 ± 0.69**
	ScSPM [23]	77.65 ± 1.43	78.31 ± 1.13	**81.00 ± 0.98**
	KSPM [17]	83.81 ± 1.22	84.32 ± 0.67	**84.49 ± 0.34**
40 training	MCSCSPM-HE	89.16 ± 1.04	**89.21 ± 0.75**	88.84 ± 0.83
	MCSCSPM-RGB	89.24 ± 1.03	89.46 ± 1.14	**89.53 ± 1.20**
	MCSCSPM-Gray	**81.37 ± 1.55**	81.31 ± 1.19	80.80 ± 1.71
	PSDSPM [5]	**84.06 ± 1.16**	83.72 ± 1.46	83.40 ± 1.14
	SMLSPM [4]	88.05 ± 1.38	87.88 ± 1.04	**88.54 ± 1.42**
	ScSPM [23]	73.60 ± 1.68	75.58 ± 1.29	**76.24 ± 3.05**
	KSPM [17]	80.54 ± 1.21	**80.56 ± 1.24**	80.46 ± 0.56

Table 1. Performance of different methods on the GBM dataset.

	Method	DictionarySize=256	DictionarySize=512	DictionarySize=1024
280 training	MCSCSPM-HE	97.39 ± 0.36	**97.51 ± 0.41**	97.48 ± 0.40
	MCSCSPM-RGB	97.11 ± 0.44	**97.49 ± 0.46**	97.44 ± 0.43
	MCSCSPM-Gray	88.76 ± 0.59	90.50 ± 0.70	**91.28 ± 0.72**
	PSDSPM [5]	97.19 ± 0.49	**97.27 ± 0.44**	97.08 ± 0.45
	SMLSPM [4]	98.15 ± 0.46	**98.50 ± 0.42**	98.21 ± 0.44
	ScSPM [23]	94.52 ± 0.44	96.37 ± 0.45	**96.81 ± 0.50**
	KSPM [17]	93.55 ± 0.31	93.76 ± 0.27	**93.90 ± 0.19**
140 training	MCSCSPM-HE	96.73 ± 0.84	**96.89 ± 0.48**	96.84 ± 0.67
	MCSCSPM-RGB	96.14 ± 1.17	96.46 ± 1.06	**96.64 ± 0.76**
	MCSCSPM-Gray	86.79 ± 0.98	88.26 ± 0.59	**88.50 ± 0.80**
	PSDSPM [5]	**96.80 ± 0.75**	96.52 ± 0.76	96.55 ± 0.84
	SMLSPM [4]	97.40 ± 0.50	**97.98 ± 0.35**	97.35 ± 0.48
	ScSPM [23]	93.46 ± 0.55	95.68 ± 0.36	**96.76 ± 0.63**
	KSPM [17]	92.50 ± 1.12	93.06 ± 0.82	**93.26 ± 0.68**
70 training	MCSCSPM-HE	95.32 ± 0.67	**95.62 ± 0.29**	95.40 ± 0.44
	MCSCSPM-RGB	94.45 ± 0.84	**94.64 ± 0.72**	94.45 ± 0.77
	MCSCSPM-Gray	84.04 ± 1.10	**85.13 ± 0.79**	84.66 ± 1.14
	PSDSPM [5]	95.12 ± 0.54	**95.13 ± 0.51**	95.09 ± 0.40
	SMLSPM [4]	96.20 ± 0.85	**96.37 ± 0.85**	96.19 ± 0.62
	ScSPM [23]	91.93 ± 1.00	93.67 ± 0.72	**94.86 ± 0.86**
	KSPM [17]	90.78 ± 0.98	91.34 ± 1.13	**91.59 ± 0.97**

Table 2. Performance of different methods on the KIRC dataset.

K was fixed to be 300 per spectrum (channel), which led to a total number of 600, 900 and 300 filters for MCSCSPM-HE, MCSCSPM-RGB and MCSCSPM-Gray, respectively. Table 1 and Table 2 show that, even with smaller number of filters, MCSCSPM-HE outperforms MCSCSPM-RGB in most cases. This is
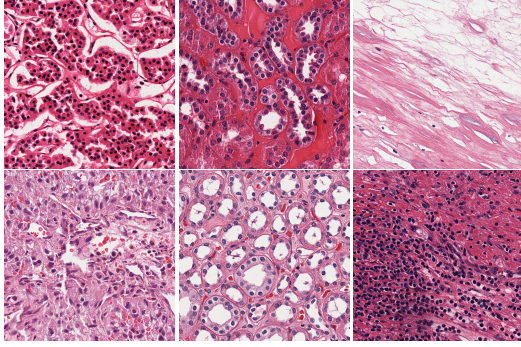
Figure 4. KIRC Examples. First column: Tumor; Second column: Normal; Third column: Stromal. Note that the phenotypic heterogeneity is highly diverse in each column.

due to the fact that, after color decomposition, the resulting two spectra are biological-component-specific, such that specialized filters can be obtained from each spectrum characterizing nuclear architecture and tissue structural connectivities, respectively, as demonstrated in Figure 1. Although the stain information (biological component information) leaks across channels for H&E stained tissue sections in its original RGB presentation, target-specific property can still be preserved to some extent (*e.g.,* most of the nuclear information resides in blue (B) channel); and this explains why MCSCSPM-RGB still has reasonable performance. However, when such a property is completely lost in grayscale, MCSCSPM-Gray sees a dramatic performance drop.

2. Convolutional v.s. patch-based sparse modeling. As listed in Table 1 and Table 2, the proposed approach, MCSCSPM-HE/MCSCSPM-RGB outperforms patch-based sparse feature learning models, *e.g.,* PSD-SPM [5], with fewer filters than PSDSPM. These facts indicate that, in tissue classification, convolutional sparse coding is more effective than traditional sparse coding in terms of using more succinct representations and producing better results, which has already been confirmed in other applications [14].

3. Unsupervised feature learning v.s. hand-engineered features. As shown in Table 1 and Table 2, the proposed approach significantly outperforms systems that are built on hand-engineered features for general image classification purpose (*e.g.,* KSPM, ScSPM). Even compared to the recently proposed system, SML-SPM [4], which is built upon features with biological prior knowledge, the proposed approach, MCSCSPM, robustly achieves very competitive performance over the two different tumor types, where MCSCSPM-HE performs better on the GBM dataset, while worse on the KIRC dataset. This confirms that the proposed ap-

proach, MCSCSPM, is a useful tool for analyzing large cohorts with substantial technical variations and biological heterogeneities.

5. Conclusion

In this paper, we propose a multispectral convolutional sparse coding framework for classification of histology sections with diverse phenotypic signatures. Our approach is benefited by exploiting multiple spectra, which potentially contain target-specific information for learning highly diversified feature detectors. We show that by decomposing images into nuclei and protein/extra-cellular content, biological-component-specific filters can be learned, which capture the nuclear architecture of distinct shapes and the structural connectivity within tissue sections, respectively. The multispectral features are then summarized within distinct tissue contexts at various scales and locations through SPM for classification. Experimental results show that the proposed approach outperforms patch-based sparse feature learning models (*e.g.,* PSDSPM) and human-engineered features (*e.g.,* SIFT); while generates very competitive performance compared to the dedicated system incorporating biological prior knowledge (*i.e.,* SMLSPM).

Future work will mainly focus on stacking the model into hierarchies with the aim to learn phenotypic concepts. In addition, it is also desirable to incorporate the learning of color decomposition matrix into the overall learning objective, which will potentially enable its extensibility to different applications. Lastly, we also plan to conduct investigation of our approach in other vision tasks, such as object recognition and segmentation.

Disclaimer

References

[1] D. Axelrod, N. Miller, H. Lickley, J. Qian, W. Christens-Barry, Y. Yuan, Y. Fu, and J. Chapman. Effect of quantitative nuclear features on recurrence of ductal carcinoma in situ (DCIS) of breast. *Cancer Informatics*, 4:99–109, 2008. 2

[2] A. Basavanhally, J. Xu, A. Madabhushu, and S. Ganesan. Computer-aided prognosis of ER+ breast cancer histopathology and correlating survival outcome with oncotype DX assay. In *ISBI*, pages 851–854, 2009. 2

[3] H. Bristow, A. Eriksson, and S. Lucey. Fast convolutional sparse coding. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 391–398, 2013. 2, 3

[4] H. Chang, A. Borowsky, P. Spellman, and B. Parvin. Classification of tumor histology via morphometric context. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 2013. 1, 2, 4, 5, 6, 7

[5] H. Chang, N. Nayak, P. Spellman, and B. Parvin. Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching. *Medical image computing and computed-assisted intervention–MICCAI*, 2013. 1, 2, 5, 6, 7

[6] M. Datar, D. Padfield, and H. Cline. Color and texture based segmentation of molecular pathology images using HSOMs. In *ISBI*, pages 292–295, 2008. 2

[7] C. Demir and B. Yener. Automated cancer diagnosis based on histopathological images: A systematic survey. *Technical Report, Rensselaer Polytechnic Institute, Department of Computer Science.*, 2009. 2

[8] S. Doyle, M. Feldman, J. Tomaszewski, N. Shih, and A. Madabhushu. Cascaded multi-class pairwise classifier (CASCAMPA) for normal, cancerous, and cancer confounder classes in prostate histology. In *ISBI*, pages 715–718, 2011. 2

[9] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008. 2, 3, 4

[10] H. Fatakdawala, J. Xu, A. Basavanhally, G. Bhanot, S. Ganesan, F. Feldman, J. Tomaszewski, and A. Madabhushi. Expectation-maximization-driven geodesic active contours with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology. *IEEE Transactions on Biomedical Engineering*, 57(7):1676–1690, 2010. 2

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

[12] J. Han, H. Chang, L. Loss, K. Zhang, F. Baehner, J. Gray, P. Spellman, and B. Parvin. Comparison of sparse coding and kernel methods for histopathological classification of glioblastoma multiforme. In *ISBI*, pages 711–714, 2011. 2

[13] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2146–2153, 2009. 3, 4

[14] K. Kavukcuoglu, P. Sermanet, Y.-L. Boureau, K. Gregor, M. Mathieu, and Y. L. Cun. Learning convolutional feature hierarchies for visual recognition. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems 23*, pages 1090–1098. 2010. 2, 3, 7

[15] J. Kong, L. Cooper, A. Sharma, T. Kurk, D. Brat, and J. Saltz. Texture based image recognition in microscopy images of diffuse gliomas with multi-class gentle boosting mechanism. In *ICASSAP*, pages 457–460, 2010. 2

[16] S. Kothari, J. Phan, A. Osunkoya, and M. Wang. Biological interpretation of morphological patterns in histopathological whole slide images. In *ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, 2012. 2

[17] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 2169–2178, 2006. 2, 3, 4, 5, 6

[18] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 689–696, New York, NY, USA, 2009. ACM. 3

[19] R. Rigamonti and V. Lepetit. Accurate and efficient linear structure segmentation by leveraging ad hoc features with learned filters. In N. Ayache, H. Delingette, P. Golland, and K. Mori, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, volume 7510 of *Lecture Notes in Computer Science*, pages 189–197. Springer Berlin Heidelberg, 2012. 2

[20] A. Ruifork and D. Johnston. Quantification of histochemical staining by color decomposition. *Anal Quant Cytol Histology*, 23(4):291–299, 2001. 3, 5

[21] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. Lecun. Pedestrian detection with unsupervised multi-stage feature learning. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3626–3633, 2013. 2, 3

[22] A. Vedaldi and A. Zisserman. Efficient additive kernels via explicit feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3):480–492, 2012. 4

[23] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 1794–1801, 2009. 5, 6

[24] M. Zeiler, D. Krishnan, G. Taylor, and R. Fergus. Deconvolutional networks. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2528–2535, 2010. 2, 3

[25] M. Zeiler, G. Taylor, and R. Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2018–2025, 2011. 2, 3