Medical Image Analysis 35 (2017) 530–543

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

# When machine vision meets histology: A comparative evaluation of model architecture for classification of histology sections[☆]

Cheng Zhong[a], Ju Han[a], Alexander Borowsky[c], Bahram Parvin[b], Yunfu Wang[a,d,*], Hang Chang[a,*]

[a] Lawrence Berkeley National Laboratory, Berkeley CA USA
[b] Department of Electrical and Biomedical Engineering, University of Nevada, Reno, NV USA
[c] Center for Comparative Medicine, University of California, Davis,CA, USA
[d] Department of Neurology, Taihe Hospital, Hubei University of Medicine, Shiyan, Hubei, China

ARTICLE INFO

*Article history:*
Received 25 February 2016
Revised 12 August 2016
Accepted 26 August 2016
Available online 9 September 2016

*Keywords:*
Computational histopathology
Classification
Unsupervised feature learning
Sparse feature encoder

ABSTRACT

Classification of histology sections in large cohorts, in terms of distinct regions of microanatomy (e.g., stromal) and histopathology (e.g., tumor, necrosis), enables the quantification of tumor composition, and the construction of predictive models of genomics and clinical outcome. To tackle the large technical variations and biological heterogeneities, which are intrinsic in large cohorts, emerging systems utilize either prior knowledge from pathologists or unsupervised feature learning for invariant representation of the underlying properties in the data. However, to a large degree, the architecture for tissue histology classification remains unexplored and requires urgent systematical investigation. This paper is the first attempt to provide insights into three fundamental questions in tissue histology classification: I. Is unsupervised feature learning preferable to human engineered features? II. Does cellular saliency help? III. Does the sparse feature encoder contribute to recognition? We show that (a) in I, both Cellular Morphometric Feature and features from unsupervised feature learning lead to superior performance when compared to SIFT and [Color, Texture]; (b) in II, cellular saliency incorporation impairs the performance for systems built upon pixel-/patch-level features; and (c) in III, the effect of the sparse feature encoder is correlated with the robustness of features, and the performance can be consistently improved by the multi-stage extension of systems built upon both Cellular Morphmetric Feature and features from unsupervised feature learning. These insights are validated with two cohorts of Glioblastoma Multiforme (GBM) and Kidney Clear Cell Carcinoma (KIRC).

## 1. Introduction

Although molecular characterization of tumors through gene expression analysis has become a standardized technique, bulk tumor gene expression data provide only an average genome-wide measurement for a biopsy and fail to reveal inherent cellular composition and heterogeneity of a tumor. On the other hand, histology sections provide wealth of information about the tissue architecture that contains multiple cell types at different states of cell cycles. These sections are often stained with hematoxylin and eosin (H&E) stains, which label DNA (e.g., nuclei) and protein contents, respectively, in various shades of color. Furthermore, morphometric abberations in tumor architecture often lead to disease progression, and it is therefore desirable to quantify tumor architecture as well as the corresponding morphometric abberations in large cohorts for the construction of predictive models of end points, e.g., clinical outcome, which have the potential for improved diagnosis and therapy.

Despite the efforts by some researchers on reducing inter- and intra-pathologist variations (Dalton et al., 2000) during manual analysis, this approach is not a scalable solution, and therefore impedes the effective representation and recognition from large cohorts for scientific discoveries. With its value resting on capturing detailed morphometric signatures and organization, automatic quantitative analysis of a large collection of histological data is highly desirable, and is unfortunately impaired by a number of barriers mostly originating from the technical variations (e.g., fixation, staining) and biological heterogeneities (e.g., cell type, cell state) always presented in the data. Specifically, a histological tissue section refers to an image of a thin slice of tissue applied to a microscopic slide and scanned from a light microscope, and the

[☆] Related resources have been released for public consumption at http://bmihub.org.

[*] Corresponding authors.
*E-mail address:* hchang@lbl.gov (H. Chang).

http://dx.doi.org/10.1016/j.media.2016.08.010

technical variations and biological heterogeneities lead to significant color variations both within and across tissue sections. For example, within the same tissue section, nuclear signal (color) varies from light blue to dark blue due to the variations of their chromatin content; and nuclear intensity in one tissue section may be very close to the background intensity (e.g., cytoplasmic, macromolecular components) in another tissue section.

It is also worth to mention that alternative staining (e.g., fluorescence) and microscopy methods (multi-spectral imaging) have been proposed and studied in order to overcome the fundamental limitations/challenges in tissue histology (Stack et al., 2014; Levenson et al., 2015; Rimm, 2014; Huang et al., 2013; Ghaznavi et al., 2013); however, H&E stained tissue sections are still the gold standard for the assessment of tissue neoplasm. Furthermore, the efficient and effective representation and interpretation of H&E tissue histology sections in large cohorts (e.g., The Cancer Genome Atlas dataset) have the potential to provide predictive models of genomics and clinical outcome, and are therefore urgently required.

Although many techniques have been designed and developed for tissue histology classification (see Section 2), the architecture for tissue histology classification remains largely unexplored and requires urgent systematical investigation. To fulfil this goal, our paper provides insights to three fundamental questions in tissue histology classification: I. Is unsupervised feature learning preferable to human engineered features? II. Does cellular prior knowledge help? III. Does the sparse feature encoder contribute to recognition? The novelty of our work resides in three folds: (i) architecture design: we have systematically experimented the system architecture with various combinations of feature types, feature extraction strategies and intermediate layers based on sparsity/locality-constrained feature encoders, which ensures the extensive evaluation and detailed insights on impact of the key components during the architecture construction; (ii) experimental design: our experimental evaluation has been performed through cross-validation on two independent datasets with distinct tumor types, where both datasets have been curated by our pathologist to provide examples of distinct regions of microanatomy (e.g., stromal) and histopathology (e.g., tumor, necrosis) with sufficient amount of technical variations and biological heterogeneities, so that the architecture can be faithfully tested and validated against important topics in histopathology (see Section 4 for details). More importantly, such an experimental design (combination of cross-validation and validation on independent datasests), to the maximum extent, ensures the consistency and unbiasedness of our findings; and (iii) outcome: the major outcome of our work are well-justified insights in the architecture design/construction. Specifically, we suggest that the sparse feature encoders based on Cellular Morphometric Feature and features from unsupervised feature learning provide the best configurations for tissue histology classification. Furthermore, these insights also led to the construction of a highly scalable and effective system (CMF-**PredictiveSFE**-KSPM, see Section 4 for details) for tissue histology classification. Finally, we believe that our work will not only benefit the research in computational histopathology, but will also benefit the community of medical image analysis at large by shedding lights on the systematical study of other important topics.

Organization of this paper is as follows: Section 2 reviews related works. Section 3 describes various components for the system architecture during evaluation. Section 4 elaborates the details of our experimental setup, followed by a detailed discussion on the experimental results. Lastly, Section 5 concludes the paper.

## 2. Related work

Current work on histology section analysis is typically forumulated and performed at multiple scales for various end points, and several outstanding reviews can be found in Demir and Yener (2009); Gurcan et al. (2009). From our perspective, the trends are: (i) nuclear segmentation and organization for tumor grading and/or the prediction of tumor recurrence (Basavanhally et al., 2009; Doyle et al., 2011). (ii) patch level analysis (e.g., small regions) (Bhagavatula et al., 2010; Kong et al., 2010), using color and texture features, for tumor representation. and (iii) detection and representation of the auto-immune response as a prognostic tool for cancer (Fatakdawala et al., 2010).

While our focus is on the classification of histology sections in large cohorts, in terms of distinct regions of microanatomy (e.g., stromal) and histopathology (e.g., tumor, necrosis), the major challenge resides in the large amounts of technical variations and biological heterogeneities in the data (Kothari et al., 2012), which typically leads to techniques that are tumor type specific or even laboratory specific. The major efforts addressing this issue fall into two distinct categories: (i) fine-tuning human engineered features (Bhagavatula et al., 2010; Kong et al., 2010; Kothari et al., 2012; Chang et al., 2013a); and (ii) applying automatic feature learning (Huang et al., 2011; Chang et al., 2013c) for robust representation. Specifically, the authors in Bhagavatula et al. (2010) designed multi-scale image features to mimic the visual cues that experts utilized for the automatic identification and delineation of germ-layer components in H&E stained tissue histology sections of teratomas derived from human and nonhuman primate embryonic stem cells; the authors in Kong et al. (2010) integrated multiple texture features (e.g., wavelet features) into a texture-based content retrieval framework for the identification of tissue regions that inform diagnosis; the work in Kothari et al. (2012) utilized various features (e.g., color, texture and shape) for the study of visual morphometric patterns across tissue histology sections; and the work in Chang et al. (2013a) constructed the cellular morphometric context based on various cellular morphometric features for effective representation and classification of distinct regions of microanatomy and histopathology. Although many successful systems have been designed and developed, based on human engineered features, for various tasks in computational histopathology, the generality/applicability of such systems to different tasks or to different cohorts can sometimes be limited, as a result, systems based on unsupervised feature learning have been built with demonstrated advantages especially for the study of large cohorts, among which, both the authors in Huang et al. (2011) and Chang et al. (2013c) utilized sparse coding techniques for unsupervised charactorization of tissue morphometric patterns.

Furthermore, tissue histology classification can be considered as a specific application of image categorization in the context of computer vision research, where spatial pyramid matching(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. Meanwhile, sparsity/locality-constrained feature encoders, through dictionary learning, have also been widely studied, and the improvement in classification performance has been confirmed in various applications (Yang et al., 2009; Wang et al., 2010; Chang et al., 2013a).

The evolution of our research on the classification of histology sections contains several stages: (i) kernel-based classification built-upon human engineered feature (e.g., SIFT features) (Han et al., 2011); (ii) independent subspace analysis for unsupervised discovery of morphometric signatures without the constraint of being able to reconstruct the original signal (Le et al., 2012); (iii) single layer predictive sparse decomposition for unsupervised discovery of morphometric signatures with the constraint of being able to reconstruct the original signal (Nayak et al., 2013); (iv) combination of either prior knowledge (Chang et al., 2013a) or predictive sparse decomposition (Chang et al., 2013c) with spatial pyramid matching; and (v) more recently, stacking multiple pre-

**Table 1**
Annotation of abbreviations in the paper, where FE stands for feature extraction; SFE stands for sparse feature encoding; and SPM stands for spatial pyramid matching. Here, we also provide the dimension information about (i) original features (outcome of FE); (ii) sparse codes (outcome of SFE); (iii) final representation (outcome of final spatial pooling, i.e., SPM); and (iv) final prediction (outcome of architectures as a one-dimensional class label.).

| Category | Abbreviation | Description | Dimension |
|---|---|---|---|
| FE | CMF | Cellular Morphometric Feature | 15 |
| | DSIFT | Dense SIFT | 128 |
| | SSIFT | Salient SIFT | 128 |
| | DCT | Dense [Color,Texture] | 203 |
| | SCT | Salient [Color,Texture] | 203 |
| | DPSD | Dense PSD | 1024 |
| | SPSD | Salient PSD | 1024 |
| SFE | SC | Sparse Coding | 1024 |
| | GSC | Graph Regularized Sparse Coding | 1024 |
| | LLC | Locality-Constraint Linear Coding | 1024 |
| | LCDL | Locality-Constraint Dictionary Learning | 1024 |
| SPM | KSPM | Kernal SPM | (256, 512, 1024) |
| | LSPM | Linear SPM | (256, 512, 1024) |
| Architecture | FE-KSPM | Architectures without the sparse feature encoder | 1 |
| | FE-SFE-LSPM | Architectures with the sparse feature encoder | 1 |

**Table 2**
Annotation of important terms used in this paper.

| Term | Description |
|---|---|
| Human Engineered Features | Refers to features that are pre-determined by human experts, with manually fixed filters/kernels/templates during extraction. |
| Cellular Prior Knowledge | Refers to the morphometric information, in terms of shape, intensity, etc., that are extracted from each individual cell/nucleus |
| Cellular Saliency | Refers to perceptually salient regions corresponding to cells/nulei in tissue histology sections. |
| Multi-Stage System | Specifically refers to the architectures with multiple stacked feature extraction/abstratcion layers. |
| Single-Stage System | Specifically refers to the architectures with a single feature extraction layer. |

dictive sparse coding modules into deep hierarchy (Chang et al., 2013d). And this paper builds on our longstanding expertise and experiences to provide (i) extensive evaluation on the model architecture for the classification of histology sections; and (ii) insights on several fundamental questions for the classification of histology sections, which, hopefully, will shed lights on the analysis of histology sections in large cohorts towards the ultimate goal of improved therapy and treatment.

## 3. Model architecture

To ensure the extensive evaluation and detailed insights on impact of the key components during the architecture construction, we have systematically experimented the model architecture with various combinations of feature types, feature extraction strategies and intermediate layers based on sparsity/locality-constrained feature encoders. And this section describes how we built the tissue classification architecture for evaluation. Tables 1 and 2 summarize the aberrations and important terms, respectively, and detailed descriptions are listed in the sections as follows,

### 3.1. Feature extraction modules (FE)

The major barrier in tissue histology classification, in large cohorts, stems from the large technical variations and biological heterogeneities, which requires the feature representation to capture the intrinsic properties in the data. In this work, we have evaluated three different features from two different categories (i.e., human-engineered feature and unsupervised feature learning). Details are as follows,

**Cellular Morphometric Feature - CMF:** The cellular morphometric features are human-engineered biological meaningful cellular-level features, which are extracted based on segmented nuclear regions over the input image. It has been recently shown that tissue classification systems based on CMF are insensitive to segmentation strategies (Chang et al., 2013a). In this work, we employ the segmentation strategy proposed in Chang et al. (2013b), and simply use the same set of features as described in Table 3. It is worth to mention that although generic cellular features, e.g., Zernike monments (Apostolopoulos et al., 2011; Asadi et al., 2006), have been successfully applied in various biomedical applications, we choose to use CMF due to (i) its demonstrated power in tissue histology classification (Chang et al., 2013a); and (ii) the limited impact by including those generic cellular features on both evaluation and understanding of the benefits introduced by the sparse feature encoders.

**Dense SIFT - DSIFT:** The dense SIFT features are human-engineered features, which are extracted from regularly-spaced patches over the input image, with the fixed patch-size (16 × 16 pixels) and step-size (8 pixels).

**Salient SIFT - SSIFT:** The salient SIFT features are human-engineered features, which are extracted from patches centered at segmented nuclear centers (Chang et al., 2013b) over the input image, with a fixed patch-size (16 × 16 pixels).

**Dense [Color,Texture] - DCT:** The dense [Color,Texture] features are human engineered features, and formed as a concatenation of texture and mean color with the fixed patch-size (20 × 20 pixels) and step-size (20 pixels), where color features are extracted in the RGB color space, and texture features (in terms of mean and variation of filter responses) are extracted via steer-

**Table 3**
Cellular morphometric features, where the curvature values were computed with $\sigma = 2.0$, and the nuclear background region is defined to be the region outside the nuclear region, but inside the bounding box of nuclear boundary.

| Feature | Description |
|---|---|
| Nuclear Size | #pixels of a segmented nucleus |
| Nuclear Voronoi Size | #pixels of the voronoi region, where the segmented nucleus resides |
| Aspect Ratio | Aspect ratio of the segmented nucleus |
| Major Axis | Length of Major axis of the segmented nucleus |
| Minor Axis | Length of Minor axis of the segmented nucleus |
| Rotation | Angle between major axis and x axis of the segmented nucleus |
| Bending Energy | Mean squared curvature values along nuclear contour |
| STD Curvature | Standard deviation of absolute curvature values along nuclear contour |
| Abs Max Curvature | Maximum absolute curvature values along nuclear contour |
| Mean Nuclear Intensity | Mean intensity in nuclear region measured in gray scale |
| STD Nuclear Intensity | Standard deviation of intensity in nuclear region measured in gray scale |
| Mean Background Intensity | Mean intensity of nuclear background measured in gray scale |
| STD Background Intensity | Standard deviation of intensity of nuclear background measured in gray scale |
| Mean Nuclear Gradient | Mean gradient within nuclear region measured in gray scale |
| STD Nuclear Gradient | Standard deviation of gradient within nuclear region measured in gray scale |

**Table 4**
Properties of various features in evaluation. Note, all human-engineered features are pre-determined and dataset independent; while features from unsupervised feature learning are task/dataset-dependent, and are able to capture task/dataset-specific information, such as potentially meaningful morphometric patterns in tissue histology.

| FE | Design | Target | Biological Information |
|---|---|---|---|
| **CMF** | Human-Engineered | Cell (dataset independent) | Cellular morphometric information |
| **SIFT** | Human-Engineered | Generic (dataset independent) | NA |
| **CT** | Human-Engineered | Color and texture patterns (dataset independent) | NA |
| **PSD** | Learned | Generic (dataset dependent) | Dataset dependent |

able filters (Young and Lesperance, 2001) with 8 directions ($\theta \in \{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{1\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}\}$) and 5 scales ($\sigma \in \{1, 2, 3, 4, 5\}$) on the grayscale image.

**Salient [Color,Texture] - SCT:** The salient [Color,Texture] features are human-engineered features, which are extracted on patches centered at segmented nuclear centers (Chang et al., 2013b) over the input image, with a fixed patch-size (20 × 20 pixels).

**Dense PSD - DPSD:** The unsupervised features are learned by predictive sparse decomposition (PSD) on randomly sampled image patches following the protocol in Chang et al. (2013c), and the dense PSD features are extracted from regularly-spaced patches over the input image, with the fixed patch-size (20 × 20 pixels), step-size (20 pixels) and number of basis functions (1024). Briefly, given $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbb{R}^{m \times N}$ as a set of vectorized image patches, we formulated the PSD optimization problem as:

$$\min_{\mathbf{B},\mathbf{Z},\mathbf{W}} \|\mathbf{X} - \mathbf{BZ}\|_F^2 + \lambda\|\mathbf{Z}\|_1 + \|\mathbf{Z} - \mathbf{WX}\|_F^2$$
$$\text{s.t. } \|\mathbf{b}_i\|_2^2 = 1, \forall i = 1, \ldots, h \quad (1)$$

where $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_h] \in \mathbb{R}^{m \times h}$ is a set of the basis functions; $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N] \in \mathbb{R}^{h \times N}$ is the sparse feature matrix; $\mathbf{W} \in \mathbb{R}^{h \times m}$ is the auto-encoder; $\lambda$ is thee regularization constant. The goal of jointly minimizing Eq. (1) with respect to the triple $<$ **B, Z, W** $>$ is to enforce the inference of the regressor **WX** to be resemble to the optimal sparse codes **Z** that can reconstruct **X** over **B** (Kavukcuoglu et al., 2008). In our implementation, the number of basis functions (**B**) is fixed to be 1024, $\lambda$ was fixed to be 0.3, empirically, for the best performance.

**Salient PSD - SPSD:** The salient PSD features are extracted on patches centered at segmented nuclear centers (Chang et al., 2013b) over the input image, with the fixed patch-size (20 × 20 pixels) and fixed number of basis functions (1024).

The properties of aforementioned features are summarized in Table 4. Note that salient features are not included, given the fact that they only differ from their corresponding dense versions with extra saliency information. It is clear that, different from SIFT and CT, which are generic features designed for general purposes, both CMF and PSD can encode biological meaningful information, where the former works in a pre-determined manner while the latter has the potential to capture biological meaningful patterns in an unsupervised fashion. Therefore, within the context of tissue histology classification, CMF and PSD have the potential to work better due to these intrinsic properties, as shown in our evaluation.

### 3.2. Sparse feature encoding modules (SFE)

It has been shown recently (Yang et al., 2009; Wang et al., 2010) that the impose of the feature encoder through dictionary learning, with sparsity or locality constraint, significantly improves the efficacy of existing image classification systems. *The rationale is that the sparse feature encoder functions as an additional feature extraction/abstraction operation, and thus adds an extra layer (stage) to the feature extraction component of the system. Therefore, it extends the original system with multiple feature extraction/abstraction stages, which is able to capture intrinsic patterns at the higher-level, as suggested in Jarrett et al. (2009).* To study the impact of the sparse feature encoder on tissue histology classification, we adopt three different sparsity/locality-constrained feature encoders for evaluation. Briefly, let $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_M] \in \mathbb{R}^{a \times M}$ be a set of features, $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_M] \in \mathbb{R}^{b \times M}$ be the set of sparse codes, and $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_b] \in \mathbb{R}^{a \times b}$ be a set of basis functions for feature encoding, the feature encoders are summarized as follows,

**Sparse Coding - (SC):**

$$\min_{\mathbf{B},\mathbf{C}} \sum_{i=1}^{M} ||\mathbf{y}_i - \mathbf{B}\mathbf{c}_i||^2 + \lambda||\mathbf{c}_i||_1; \quad \text{s.t. } ||\mathbf{b}_i|| \leq 1, \forall i \quad (2)$$

where $||\mathbf{b}_i||$ is a unit $\ell_2$-norm constraint for avoiding trivial solutions, and $||\mathbf{c}_i||_1$ is the $\ell_1$-norm enforcing the sparsity of $\mathbf{c}_i$. In our implementation, the number of basis functions (**B**) is fixed to be 1024, $\lambda$ is fixed to be 0.15, empirically, for the best performance.

**Graph Regularized Sparse Coding - (GSC)** (Zheng et al., 2011)

$$\min_{\mathbf{B},\mathbf{C}} \sum_{i=1}^{M} ||\mathbf{y}_i - \mathbf{B}\mathbf{c}_i||^2 + \lambda||\mathbf{c}_i||_1 + \alpha\mathbf{Tr}(\mathbf{CLC}^\mathbf{T}); \quad \text{s.t. } ||\mathbf{b}_i|| \le 1, \forall i \tag{3}$$

where $||\mathbf{b}_i||$ is a unit $\ell_2$-norm constraint for avoiding trivial solutions, and $||\mathbf{c}_i||_1$ is the $\ell_1$-norm enforcing the sparsity of $\mathbf{c}_i$, $\mathbf{Tr}(\cdot)$ is the trace of matrix $\cdot$, $\mathbf{L}$ is the Laplacian matrix, and the third term encodes the Laplacian regularizer (Belkin and Niyogi, 2003). Please refer to Zheng et al. (2011) for details of the formulation. In our implementation, the number of basis functions ($\mathbf{B}$) is fixed to be 1024, the regularization parameters, $\lambda$ and $\alpha$ are fixed to be 1 and 5, respectively, for the best performance.

**Locality-Constraint Linear Coding - (LLC)** (Wang et al., 2010)**:**

$$\min_{\mathbf{B},\mathbf{C}} \sum_{i=1}^{M} ||\mathbf{y}_i - \mathbf{B}\mathbf{c}_i||^2 + \lambda||\mathbf{d}_i \odot \mathbf{c}_i||_1; \quad \text{s.t. } \mathbf{1}^\top\mathbf{c}_i = 1, \forall i \tag{4}$$

where $\odot$ denotes the element-wise multiplication, and $\mathbf{d}_i \in \mathbb{R}^b$ encodes the similarity of each basis vector to the input descriptor $\mathbf{y}_i$, Specifically,

$$\mathbf{d}_i = \exp\left(\frac{\text{dist}(\mathbf{y}_i, \mathbf{B})}{\sigma}\right) \tag{5}$$

where $\text{dist}(\mathbf{y}_i, \mathbf{B}) = [\text{dist}(\mathbf{y}_i, \mathbf{b}_1), \ldots, \text{dist}(\mathbf{y}_i, \mathbf{b}_b)]$, $\text{dist}(\mathbf{y}_i, \mathbf{b}_j)$ is the Euclidean distance between $\mathbf{y}_i$ and $\mathbf{b}_j$, $\sigma$ is used to control the weight decay speed for the locality adaptor. In our implementation, the number of basis functions ($\mathbf{B}$) is fixed to be 1024, the regularization parameters $\lambda$ and $\sigma$ are fixed to be 500 and 100, respectively, to achieve the best performance.

**Locality-Constraint Dictionary Learning - (LCDL)** (Zhou and Barner, 2013)**:** The LCDL optimization problem is formulated as:

$$\min_{\mathbf{B},\mathbf{C}} \|\mathbf{Y} - \mathbf{BC}\|_F^2 + \lambda \sum_{i=1}^{N}\sum_{j=1}^{K}\left[c_{ji}^2 \left\|\mathbf{y}_i - \mathbf{b}_j\right\|_2^2\right] + \mu\|\mathbf{C}\|_F^2$$
$$\text{s.t.} \begin{cases} \mathbf{1}^\top\mathbf{c}_i = 1 & \forall i \qquad\qquad\qquad (*) \\ c_{ji} = 0 & \text{if } \mathbf{b}_j \notin \Omega_\tau(\mathbf{y}_i) \quad \forall i, j \qquad (**) \end{cases} \tag{6}$$

where $\Omega_\tau(\mathbf{y}_i)$ is defined as the $\tau$-neighborhood containing $\tau$ nearest neighbors of $\mathbf{y}_i$, and $\lambda$, $\mu$ are positive regularization constants. $\mu\|\mathbf{C}\|_F^2$ is included for numerical stability of the least–squares solution. The sum-to-one constraint ($*$) follows from the symmetry requirement, while the locality constraint ($**$) ensures that $\mathbf{y}_i$ is reconstructed by atoms belonging to its $\tau$-neighborhood, allowing $\mathbf{c}_i$ to characterize the intrinsic local geometry. In our implementation, the number of basis functions ($\mathbf{B}$) is fixed to be 1024, the regularization parameters $\lambda$ and $\mu$ are fixed to be 0.3 and 0.001, respectively, and the neighborhood size $\tau$ is fixed to be 5, empirically, to achieve the best performance.

The major differences of aforementioned sparse feature encoders reside in two folds:

1. Objective:
   (a) SC: Learning sets of over-complete bases for efficient data representation, originally applied to modeling the human visual cortex;
   (b) GSC : learning the sparse representations that explicitly take into account the local manifold structure of the data;
   (c) LLC: generating descriptors for image classification by using efficient locality-enforcing term;
   (d) LCDL learning a set of landmark points to preserve the local geometry of the nonlinear manifold;
2. Locality Enforcing Strategy:
   (a) SC: None;
   (b) GSC: using graph Laplacian to enforce the smoothness of sparse representations along the geodesics of the data manifold;
   (c) LLC: using a locality adaptor which penalizes far-way samples with larger weights. During optimization, the basis functions are normalized after each iteration, which could cause the learned basis functions deviate from the original manifold and therefore lose locality-preservation property;
   (d) LCDL deriving an upper-bound for reconstructing an intrinsic nonlinear manifold without imposing any constraint of the energy of basis functions;

It is clear that SC is the most general approach for data representation purpose. Although various locality-constrained sparse coding techniques have demonstrated success in many applications (Zheng et al., 2011; Wang et al., 2010; Zhou and Barner, 2013), their distance metric in Euclidean Space has imposed implicit hypothesis on the manifold of the target feature space, which might potentially impair the performance, as reflected in our evaluation.

### 3.3. Spatial pyramid matching modules (SPM)

As an extension of the traditional Bag of Features (BoF) model, SPM has become a major component of state-of-art systems for image classification and object recognition (Everingham et al., 2012). Specifically, SPM consists of two steps: (i) vector quantization for the construction of dictionary from input; and (ii) histogram (i.e., histogram of dictionary elements derived in previous step) concatenation from image subregions for spatial pooling. Most recently, the effectiveness of SPM for the task of tissue histology classification has also been demonstrated in Chang et al. (2013a); 2013c). Therefore, we include two variations of SPM as a component of the architecture for tissue histology classification, which are described as follows,

**Kernel SPM (KSPM** Lazebnik et al., 2006**):** The nonlinear kernel SPM that uses spatial-pyramid histograms of features. In our implementation, we fix the level of pyramid to be 3.

**Linear SPM (LSPM** Yang et al., 2009 **):** The linear SPM that uses the linear kernel on spatial-pyramid pooling of sparse codes. In our implementation, we fix the level of pyramid to be 3, and choose the max pooling function on the absolute sparse codes, as suggested in Yang et al. (2009); Chang et al. (2013a).

The choice of spatial pyramid matching module is made to optimize the performance/efficiency of the entire classification architecture. Experimentally, we find that (i) **FE-KSPM** outperforms **FE-LSPM**; and (ii) **FE-SFE-LSPM** and **FE-SFE-KSPM** have similar performance, while the former is more computationally efficient than the latter. Therefore, we adopt **FE-SFE-LSPM** and **FE-KSPM** during the evaluation.

As suggested in Jarrett et al. (2009), the vector quantization component of SPM can be seen as an extreme case of sparse coding, and the local histogram construction/concatenation component of SPM can be considered as a special form of spatial pooling. As a result, SPM is conceptually similar to the combination of sparse coding with spatial pooling, and therefore is able to serve as an extra layer (stage) for feature extraction. Consequently, **FE-KSPM** can be considered as a single-stage system, and **FE-SFE-LSPM** can be considered as a multi-stage system with two feature extraction/abstraction layers.

### 3.4. Classification

For architecture: **FE-SFE-LSPM**, we employed the linear SVM for classification, the same as in Wang et al. (2010); Yang et al. (2009). For architecture: **FE-KSPM**, the homogeneous kernel map (Vedaldi and Zisserman, 2012) was first applied, followed by linear SVM for classification.

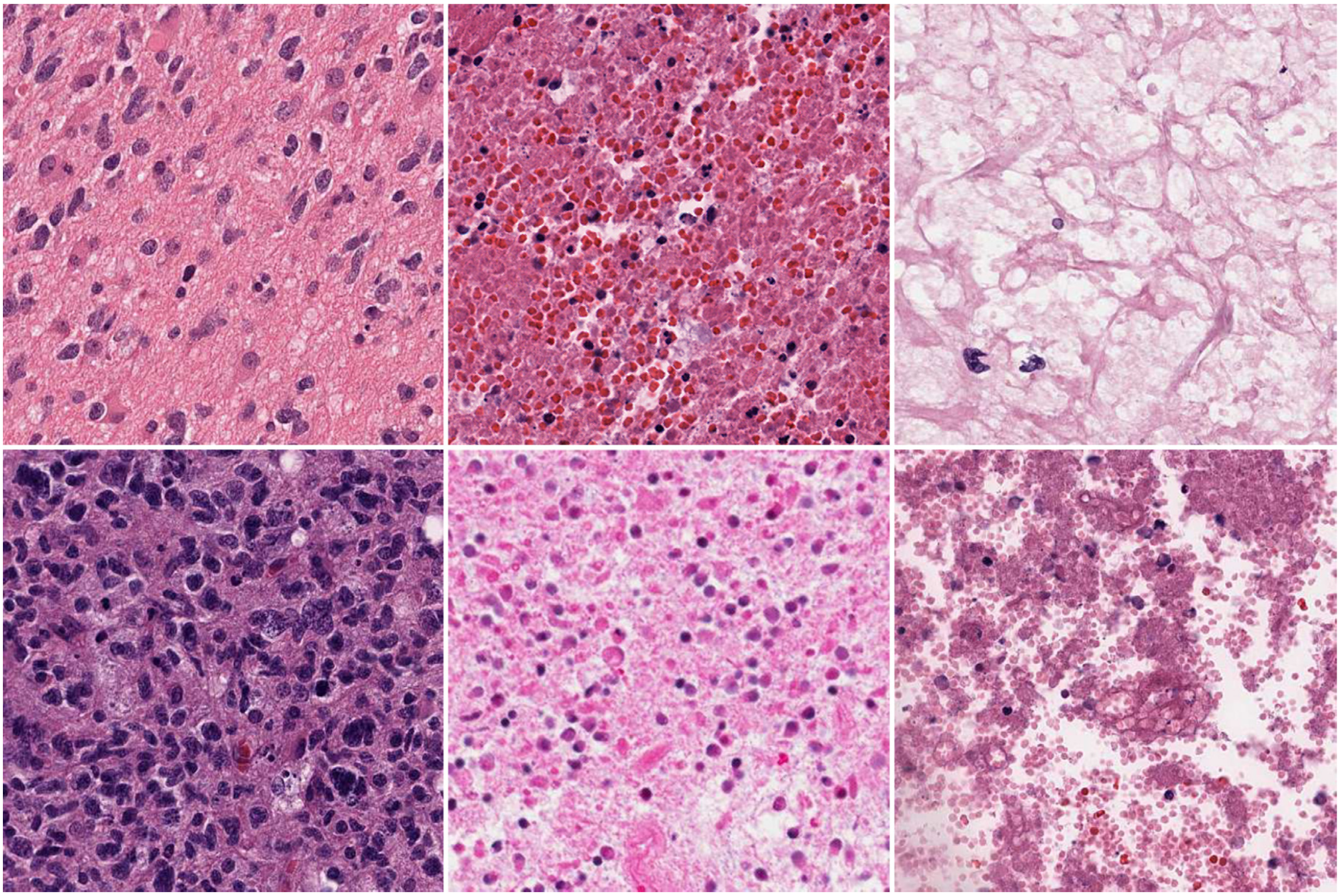

**Fig. 1.** GBM Examples. First column: Tumor; Second column: Transition to necrosis; Third column: Necrosis. Note that the phenotypic heterogeneity is highly diverse in each column.

## 4. Experimental evaluation of model architecture

### *4.1. Experimental setup*

Our extensive evaluation is performed based on the cross-validation strategy with 10 iterations, where both training and testing images are randomly selected per iteration, and the final results are reported as the mean and standard error of the correct classification rates with various dictionary sizes (256,512,1024) on the following two distinct datasets, curated from (i) Glioblastoma Multiforme (GBM) and (ii) Kidney Renal Clear Cell Carcinoma (KIRC) from The Cancer Genome Atlas (TCGA), which are publicly available from the NIH (National Institute of Health) repository. The curation is performed by our pathologist in order to provide examples of distinct regions of microanatomy (e.g., stromal) and histopathology (e.g., tumor, necrosis) with sufficient amount of biological heterogeneities and technical variations, so that the classification model architecture can be faithfully tested and validated against important studies. Furthermore, the combination of extensive cross-validation and independent validation on datasets with distinct tumor types, to the maximum extent, ensures the consistency and unbiasedness of our findings. The detailed description of our datasets as well as the corresponding task forumulation are described as follows,

**GBM Dataset:** In brain tumors, necrosis, proliferation of vasculature, and infiltration of lymphocytes are important prognostic factors. And, some of these analyses, such as the quantification of necrosis, have to be defined and performed as classification tasks in histology sections. Furthermore, necrosis is a dynamic process and different stages of necrosis exist (e.g., from cells initiating a necrosis process to complete loss of chromatin content). Therefore, the capability of identification/classification of these end points, e.g., necrosis-related regions, in brain tumor histology sections, is highly demanded. In this study, we aim to validate the model architecture for the three-category classification (i.e., Tumor, Necrosis, and Transition to Necrosis) on the GBM dataset, where the images are curated from the whole slide images (WSI) scanned with a 20*X* objective (0.502 micron/pixel). Representative examples of each class can be found in Fig. 1, which reveal a significant amount of intra-class phenotypic heterogeneity. Such a highly heterogenous dataset provides an ideal test case for the quantitative evaluation of the composition of model architecture and its impact, in terms of performance and robustness, on the classification of histology sections. Specifically, the number of images per category are 628, 428 and 324, respectively, and most images are 1000 × 1000 pixels. For this task, we train, with various model architectures, on 160 images per category and tested on the rest, with three different dictionary sizes: 256, 512 and 1024.

**KIRC Dataset:** Recent studies on quantitative histology analysis (Lan et al., 2015; Rogojanu et al., 2015; Huijbers et al., 2013; de Kruijf et al., 2011) reveal that the tumor-stroma ratio is a prognostic factor in many different tumor types, and it is therefore interesting and desirable to know how such an index plays its role in KIRC, which can be fulfilled with two steps as follows, (i) identification/classification of tumor/stromal regions in tissue histology sections for the construction of tumor-stroma ratio; and (ii) correl-

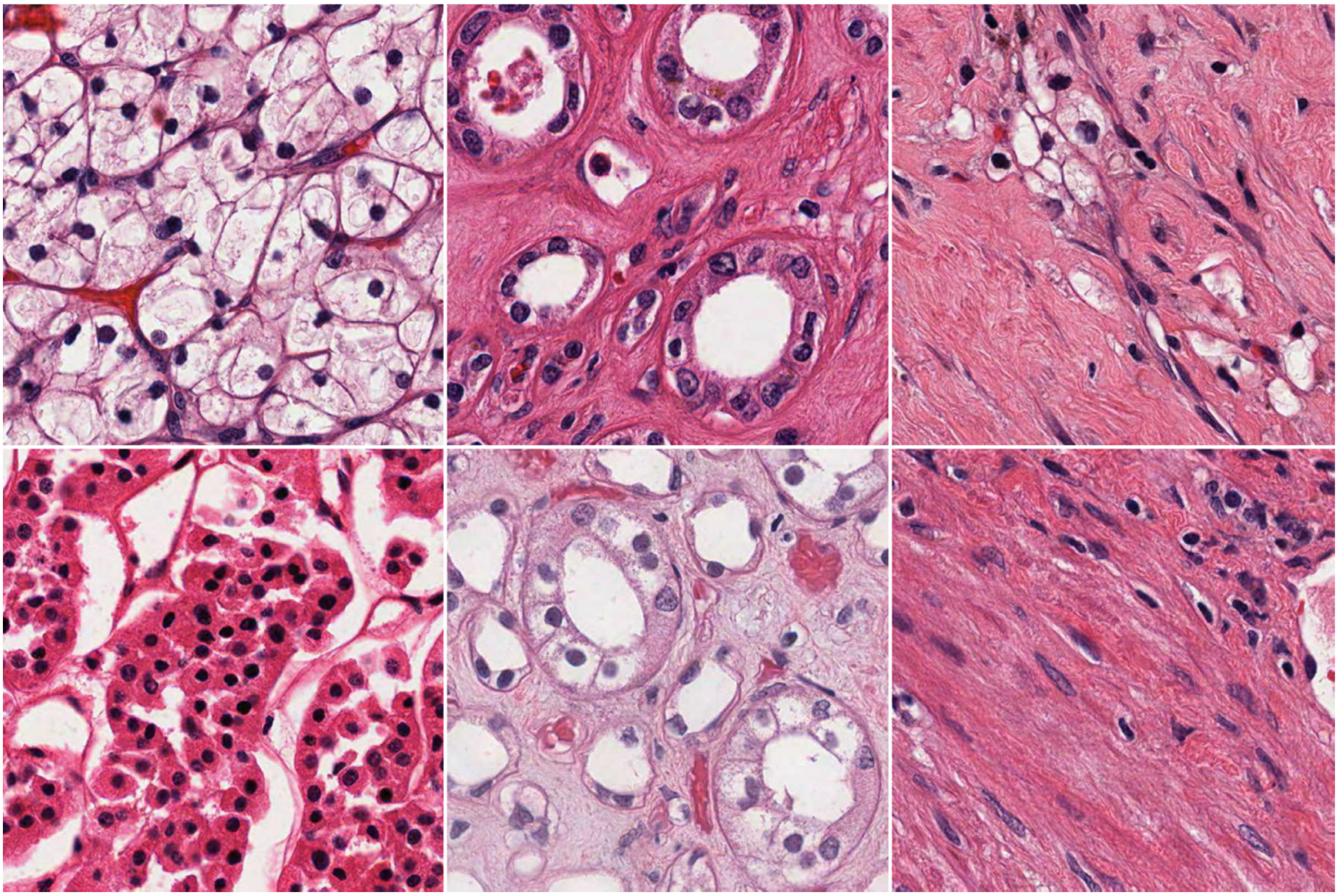

**Fig. 2.** KIRC examples. First column: Tumor; Second column: Normal; Third column: Stromal. Note that (a) in the first column, there are two different types of tumor corresponding to clear cell carcinoma, with the loss of cytoplasm (first row), and granular tumor (second row), respectively; and (b) in the second column, staining protocol is highly varied. The cohort contains a significant amount of tumor heterogeneity that is coupled with technical variation.

ative analysis of the derived tumor-stroma ratio with clinical outcome. Therefore, in this study, we aim to validate the model architecture for the three-category classification (i.e., Tumor, Normal, and Stromal) on the KIRC dataset, where the images are curated from the whole slide images (WSI) scanned with a 40X objective (0.252 micron/pixel). Representative examples of each class can be found in Fig. 2, which (i) contain two different types of tumor corresponding to clear cell carcinoma, with the loss of cytoplasm (first row), and granular tumor (second row), respectively; and (ii) reveal large technical variations (i.e., in terms of staining protocol), especially in the normal category. The combination of the large amount of biological heterogeneity and technical variations in this curated dataset provides an ideal test case for the quantitative evaluation of the composition of model architecture and its impact, in terms of performance and robustness, on the classification of histology sections. Specifically, the number of images per category are 568, 796 and 784, respectively, and most images are 1000 × 1000 pixels. For this task, we train, with various model architectures, on 280 images per category and tested on the rest, with three different dictionary sizes: 256, 512 and 1024.

### 4.2. Is unsupervised feature learning preferable to human engineered features?

Feature extraction is the very first step for the construction of classification/recogonition system, and is one of the most important factors that affect the performance. To answer this question, we evaluated four well-selected features based on two vastly different tumor types as described previously. The evaluation was carried out with the **FE-KSPM** architecture for its simplicity, and the performance was illustrated in Fig. 3 for the GBM and KIRC datasets. It is clear that the systems based on CMF (CMF-KSPM) and PSD (PSD-KSPM) have the top performances, which are due to i) the critical role of cellular morphometric context during the pathological diagnosis, as suggested in Chang et al. (2013a); and ii) the capability of unsupervised feature learning in capturing intrinsic morphometric patterns in histology sections.

### 4.3. Does cellular saliency help?

CMF differs from DSIFT, DCT and DPSD in that (1) CMF characterizes biological meaningful properties at cellular-level, while DSIFT, DCT and DPSD are purely pixel/patch-level features without any specific biological meaning; (2) CMF is extracted per nuclear region which is cellular-saliency-aware, while DSIFT, DCT and DPSD are extracted per regularly-spaced image patch without using cellular information as prior. An illustration of aforementioned feature extraction strategies can be found in Fig. 4. Recent study (Wu et al., 2013) indicates that saliency-awareness may be helpful for the task of image classification, thus it will be interesting to figure out whether SIFT, [Color,Texture] and PSD features can be improved by the incorporation of cellular-saliency as prior. Therefore, we design salient SIFT (SSFIT), salient [Color,Texture] and salient PSD (SPSD) features, which are only extracted at nuclear centroid locations. Comparison of classification performance between dense features and salient features, with the **FE-KSPM** architecture, is illustrated

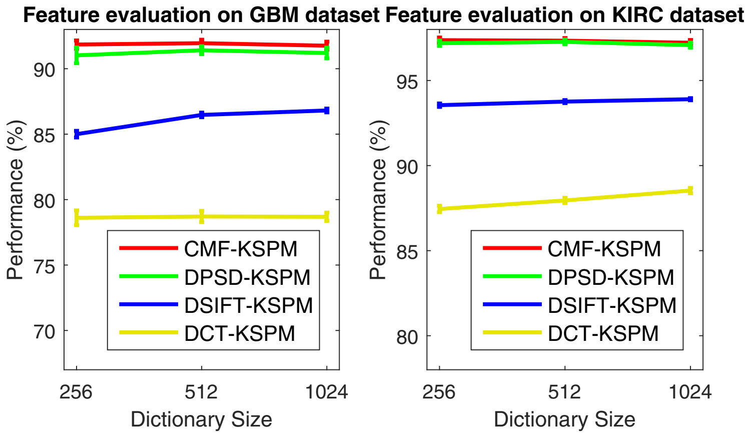

**Fig. 3.** Evaluation of different features with **FE-KSPM** architecture on both GBM (left) and KIRC (right) datasets. Here, the performance is reported as the mean and standard error of the correct classification rate, as detailed in Section 4.

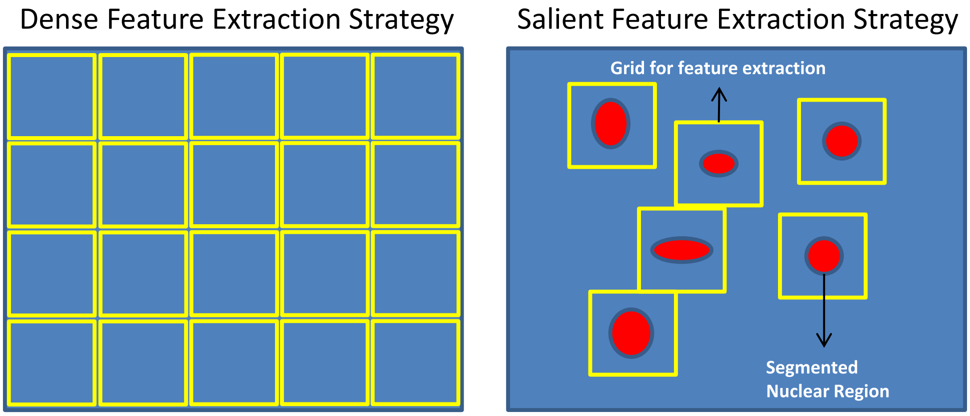
Dense Feature Extraction Strategy
Salient Feature Extraction Strategy
Grid for feature extraction
Segmented
Nuclear Region

**Fig. 4.** Illustration of dense feature extraction strategy (left) and salient feature extraction strategy (right), where dense features are extracted on regularly-spaced patches, while salient features are extracted on patches centered at segmented nuclear centers. Here, yellow rectangle and red blob represent feature extraction patch/grid and segmented nuclear region, respectively.

in Fig. 5 for GBM and KIRC datasets, which show that, for SIFT, [Color,Texture] and PSD features, cellular-saliency-awareness plays a negative role for the task of tissue histology classification. One possible explanation is that, different from CMF, which encodes specific biological meanings and summarizes tissue image with intrinsic biological-context-based representation, SIFT, [Color,Texture] and PSD lead to appearance-based image representation, and thus require dense sampling all over the place in order to faithfully assemble the view of the image.

### *4.4. Does the sparse feature encoder help?*

The evaluation of systems with the sparse feature encoder is carried out with the configuration **FE-SFE-LSPM**, where LSPM is used instead of KSPM for improved efficiency. Classification performance is illustrated in Fig. 6 and Fig. 7 for the GBM and KIRC datasets, respectively; and the results show that, compared to **FE-KSPM**,

1. For **FE**=CMF and **SFE** $\in$ {SC,GSC,LLC,LCDL}, **FE-SFE-LSPM** consistently improves the classification performance for both GBM and KIRC datasets;
2. For **FE** $\in$ {SIFT,[Color,Texture]} and **SFE** $\in$ {SC,GSC,LLC,LCDL}, **FE-SFE-LSPM** improves the performance for KIRC dataset; while impairs the performance for GBM dataset;
3. For **FE**=PSD, **FE-SFE-LSPM** improves the performance for both GBM and KIRC datasets, with **SFE** = SC; while, in general, impairs the performance for both datasets, with **SFE** $\in$ {GSC,LLC,LCDL}.

The observations above suggest that, the effect of the sparse feature encoder highly correlates with the robustness of the features being used, and significant improvement of performance can be achieved consistently across different datasets with the choice of CMF. It is also interesting to notice that, with the choice of PSD, the sparse feature encoder only helps improve the performance with sparse coding (SC) as the intermediate feature extraction layer. A possible explanation is that, compared to CMF which has real physical meanings, the PSD feature resides in a hyper space constructed from unsupervised feature learning, where Euclidean-distance, as a critical part of GSC, LLC and LCDL, may not apply.

Furthermore, it is also interesting and important to know the effect of incorporating deep learning for feature extraction. Therefore, for further validation, we have also evaluated two popular deep learning techniques, namely Stacked PSD (Chang et al., 2013d) and Convolutional Neural Networks (CNN) (Lecun et al., 1998; Huang and LeCun, 2006; Krizhevsky et al., 2012). Specifically,

1. StackedPSD-KSPM: for the evaluation of Stacked PSD, the same protocol as in Chang et al. (2013d) is utilized. Briefly, two layers of PSD, with 2048 (first layer) and 1024 (second layer) basis functions, respectively, are stacked to form a deep architecture for the feature extraction on 20 × 20 image-patches with a step-size fixed to be 20, empirically, for best performance. Af-

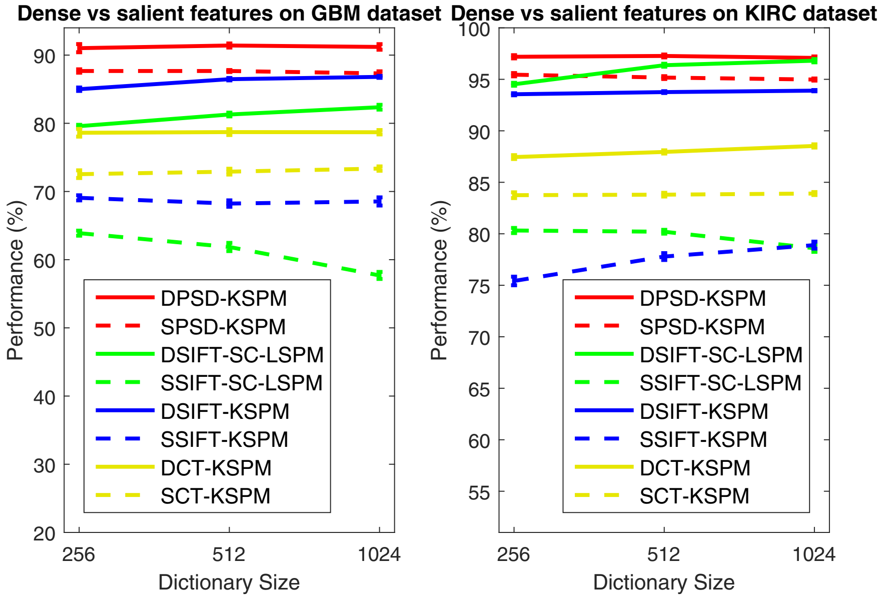

**Fig. 5.** Evaluation of dense feature extraction and salient feature extraction strategies with the **FE-KSPM** architecture on both GBM (left) and KIRC (right) datasets, where solid line and dashed line represent systems built upon dense feature and salient feature, respectively. Here, the performance is reported as the mean and standard error of the correct classification rate, as detailed in Section 4.

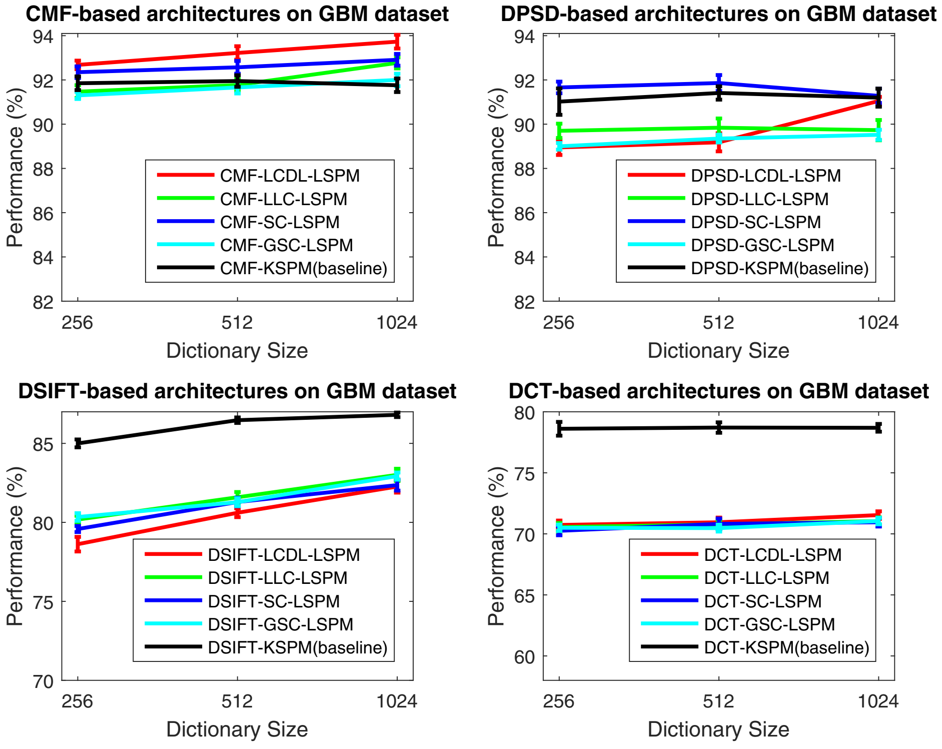

**Fig. 6.** Evaluation of the architectures with sparse feature encoders (**FE-SFE-LSPM**) on GBM dataset. Here, the performance is reported as the mean and standard error of the correct classification rate, as detailed in Section 4.

ter the patch-based extraction, the same protocol as shown in **FE-KSPM** is utilized for classification.

2. AlexNet-KSPM: for the evaluation of CNN, we adopt one of the most powerful deep neural network architecture: AlexNet (Krizhevsky et al., 2012) with the Caffe (Jia et al., 2014) implementation. Given (i) the extremely large scale (60 million parameters) of the AlexNet architecture; (ii) the significantly smaller data-scale of GBM and KIRC, compared to ImageNet (Deng et al., 2009) with one thousand categories and millions of images, where AlexNet is originally trained; and (iii) the significant decline of performance due to over-fitting that we experience with the end-to-end tuning of AlexNet on our dataset as a result of (i) and (ii), we simply adopt the pre-trained AlexNet for feature extraction on 224 × 224 image-patches with a step-size fixed to be 45, empirically, for best performance. After the patch-based extraction, the same protocol as shown in **FE-KSPM** is utilized for classification. It is worth to

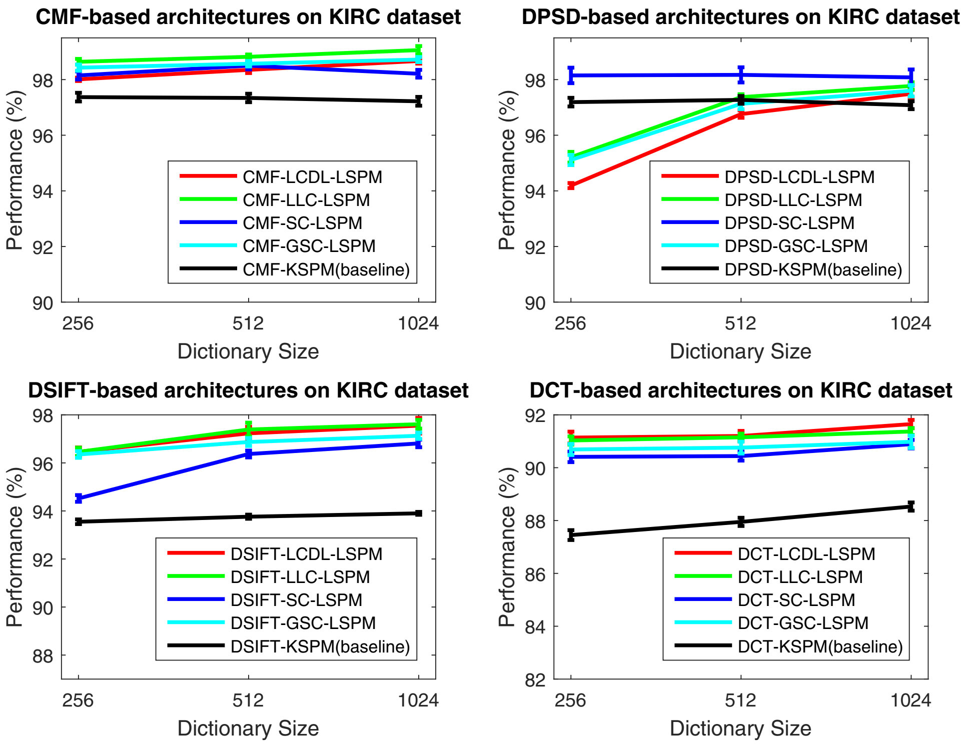

**Fig. 7.** Evaluation of the architectures with sparse feature encoders (**FE-SFE-LSPM**) on KIRC dataset. Here, the performance is reported as the mean and standard error of the correct classification rate, as detailed in Section 4.

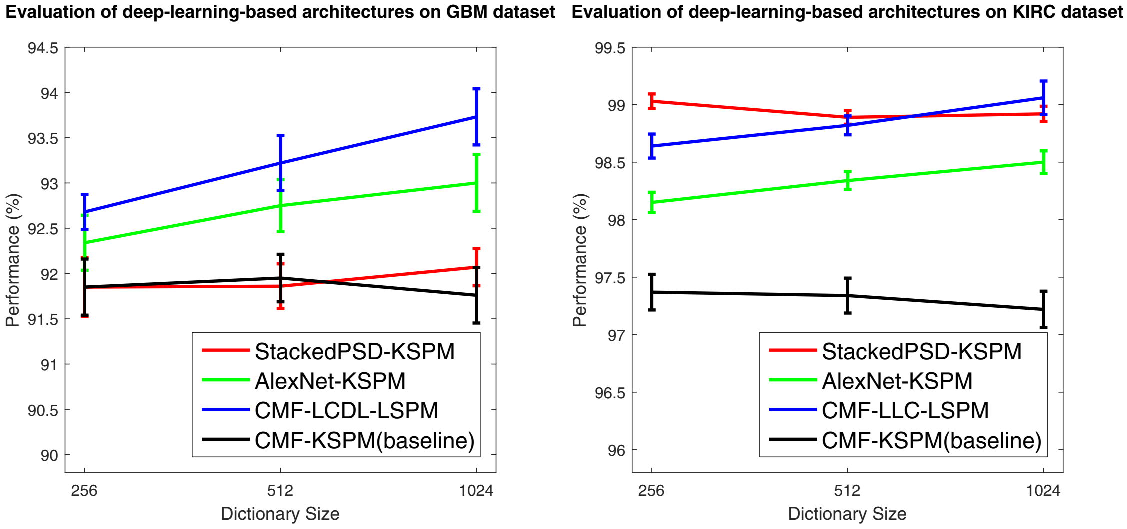

**Fig. 8.** Evaluation of the effect of incorporating deep learning for feature extraction on both GBM and KIRC datasets. Note that, given the various combinations of **FE-SFE-LSPM**, CMF-LCDL-LSPM and CMF-LLC-LSPM are chosen for GBM and KIRC datasets, respectively, for their best performance. Here, the performance is reported as the mean and standard error of the correct classification rate, as detailed in Section 4.

mention that such an approach falls into the categories of both deep learning and transfer learning.

Experimental results, illustrated in Fig. 8, suggest that,

1. Both sparse feature encoders and feature extraction strategies based on deep learning techniques consistently improve the performance of tissue histology classification;
2. The extremely large-scale convolutional deep neural networks (e.g., AlexNet), pre-trained on extremely large-scale dataset (e.g., ImageNet), can be directly applicable to the task of tissue histology classification due to the capability of deep neural networks in capturing transferable base knowledge across domains (Yosinski et al., 2014). Although the fine-tuning of AlexNet towards our datasets shows significant performance drop due to the problem of over-fitting, the direct deployment of pre-trained deep neural networks still provides a promising solution for tasks with limited data and labels, which is very common in the field of medical image analysis.

### 4.5. Revisit on spatial pooling

To further study the impact of pooling strategy, we also provide extensive experimental evaluation on one of the most popular pooling strategies (i.e., *max* pooling) in place of spatial pyramid

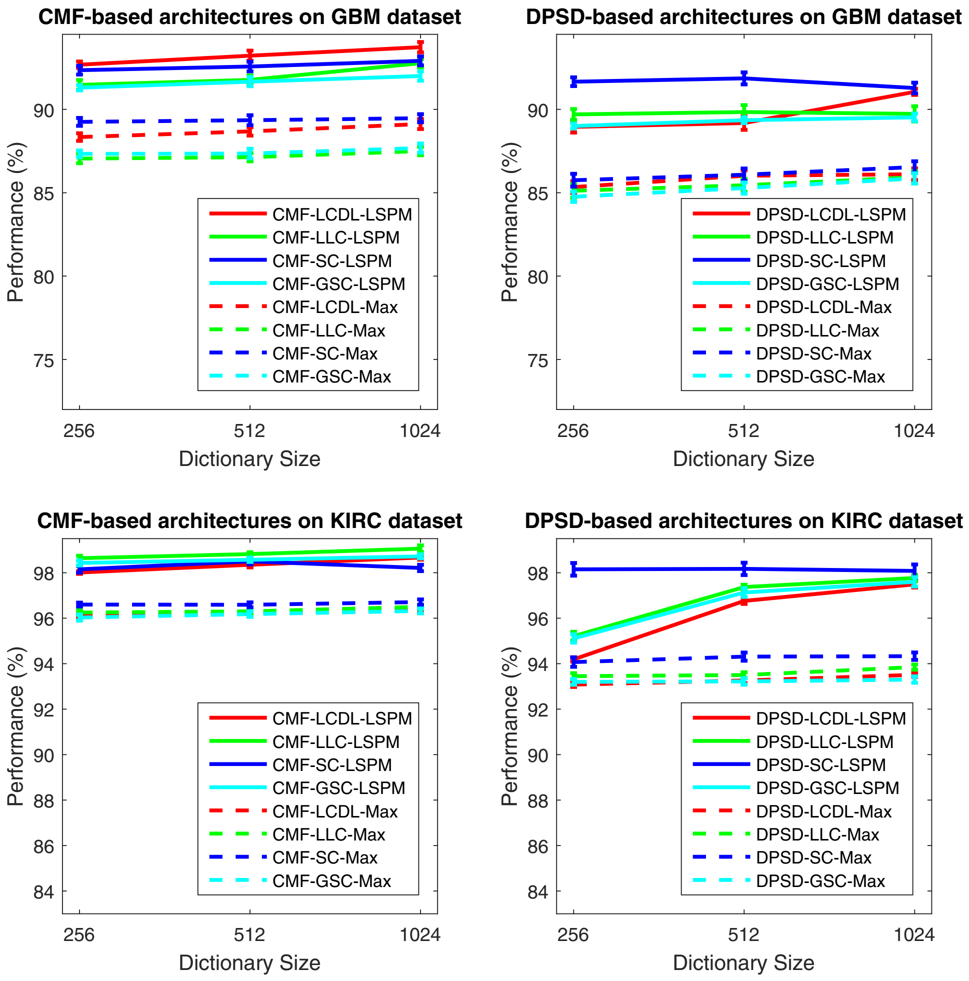

**Fig. 9.** Evaluation of the impact of different spatial pooling strategies with the **FE-SFE-LSPM** framework on both GBM and KIRC datasets. Note that, given many of the popular spatial pooling strategies, *max* pooling is chosen due to the extensive justification by both biophysical evidence in the visual cortex and researches in image categorization tasks. The derived architecture is described as **FE-SFE-Max**, and only the top-two-ranked features (i.e., DPSD and CMF) are involved during evaluation. Here, the performance is reported as the mean and standard error of the correct classification rate, as detailed in Section 4.

matching within **FE-SFE-LSPM** framework, which is defined as follows,

$$max : f_j = \max\{|c_{1j}|, |c_{2j}|, \ldots, |c_{Mj}|\} \tag{7}$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_M] \in \mathbb{R}^{b\times M}$ is the set of sparse codes extracted from an image, $c_{ij}$ is the matrix element at i-th row and j-th column of **C**, and $f = [f_1, \ldots, f_b]$ is the pooled image representation. The choice of max pooling procedure has been justified by both biophysical evidence in the visual cortex (Serre et al., 2005) and researches in image categorization (Yang et al., 2009), and the derived architecture is described as **FE-SFE-Max**. In our experimental evaluation, we focus on the top-two-ranked features (i.e., DPSD and CMF), where the corresponding comparisons of classification performance are illustrated in Fig. 9. It is clear that systems with SPM pooling consistently outperforms systems with *max* pooling with various combinations of feature types and sparse feature encoders. A possible explanation is that the vector quantization step in SPM can be considered as an extreme case of sparse coding (i.e., with a single non-zero element in each sparse code); and the local histogram concatenation step in SPM can be considered as a special form of spatial pooling. As a result, SPM is conceptually similar to an extra layer of sparse feature encoding and spatial pooling, as suggested in Jarrett et al. (2009), and therefore leads to an improved performance, compared to the architecture with *max* pooling.

### 4.6. Revisit on computational cost

In addition to classification performance, another critical factor, in clinical practice, is the computational efficiency. Therefore, in this section, we provided a detailed evaluation on computational cost of various systems. Given the fact that (i) training can always be carried out off-line; (ii) the classification of the systems in evaluation are all based on linear SVM, our evaluation on computational efficiency focuses on on-line feature extraction (including sparse feature encoding), which is the most time-consuming part during the testing phase. As shown in Table 5,

1. SIFT features are the most computational efficient features among all the ones in comparison. However, the systems built on SIFT features greatly suffer from the technical variations and biological heterogeneities in both datasets, and therefore are not good choices for the classification of tissue histology sections;
2. Given the fact that the nuclear segmentation is a prerequisite for salient feature extraction (e.g., SPSD, SSIFT and SCT), systems built upon salient features may not be necessarily more efficient than systems built upon dense features. Furthermore, since the salient features typically impair the tissue histology classification performance, they are therefore not recommended;

**Table 5**
Average computational cost (measured in second) for feature extraction (including sparse feature encoding) on images with size 1000 × 1000 pixels. The evaluation is carried out with Intel(R) Xeon(R) CPU X5365 @ 3.00GHz, and GeForce GTX 580.

| Feature Extraction Component(s) | Average Computational Cost (in second) |
|---|---|
| Nuclear Segmentation | 40 |
| CMF-**SFE** | 42 = Nuclear-Segmentation-Cost(40) + SFE-Cost(2) |
| DPSD-**SFE** | 115 = DPSD-Cost(95) + SFE-Cost(20) |
| SPSD-**SFE** | 70 = SPSD-Cost(60) + SFE-Cost(10) |
| DSIFT-**SFE** | 16 = DSIFT-Cost(10) + SFE-Cost(6) |
| SSIFT-**SFE** | 47 = SSIFT-Cost(45) + SFE-Cost(2) |
| DCT-**SFE** | 90 = DCT-Cost(80) + SFE-Cost(10) |
| SCT-**SFE** | 108 = SCT-Cost(105) + SFE-Cost(3) |
| CMF | 40 = Nuclear-Segmentation-Cost(40) |
| DPSD | 95 |
| SPSD | 60 = Nuclear-Segmentation-Cost(40) + PSD-Cost(20) |
| DSIFT | 10 |
| SSIFT | 45 = Nuclear-Segmentation-Cost(40)+SIFT-Cost(5) |
| DCT | 80 |
| SCT | 105 = Nuclear-Segmentation-Cost(40) + SCT-Cost(65) |
| StackedPSD | 100 |
| AlexNet | 1200/180 (CPU-Only/GPU-Acceleration) |

**Table 6**
**PredictiveSFE** achieved 40X speed-up, compared to **SFE**, in sparse cellular morphometric feature extraction. The evaluation was carried out with Intel(R) Xeon(R) CPU X5365 @ 3.00GHz .

| Sparse Cellular Morphometric Feature Extraction | Average Computational Cost (in second) |
|---|---|
| **PredictiveSFE** | 0.05 |
| **SFE** | 2 |

3. The gain in performance of sparse feature encoders in our evaluation is at the cost of computational efficiency. And the scalability of derived systems can be improved by (i) the development of more computational-efficient algorithms, which was demonstrated in Table 6; and (ii) the deployment of advanced computational techniques, such as cluster computing or GPU acceleration for clinical deployment, which was demonstrated in Table 5 for AlexNet.
4. Most interestingly, the sparse feature encoder, based on CMF-**SFE**, is much more efficient even compared to many shallow architectures based on PSD or CT features; and, it is only 5% slower compared to its corresponding shallow version, based on CMF. The computational efficiency are due to (i) the high sparsity of nuclei compared to dense image patches (e.g., 350 nuclei/image v.s. 2000 patches/image); and (ii) the extremely low dimensionality of cellular morphometric features compared to other features (e.g., 15 nuclear morphometric features v.s. 128 SIFT features, 203 CT features and 1024 PSD features). Furthermore, in computational histopathology, both nuclear-level information (based on nuclear segmentation) and patch-level information (based on tissue histology classification) are very critical components, which means the nuclear segmentation results can be shared across different tasks for the further improvement of the efficiency of multi-scale integrated analyses.

To further improve the scalability of systems built-upon CMF-**SFE**, as a demonstration of algorithmic-scaling-up of sparse feature encoders, we constructed a predictive sparse feature encoder (**PredictiveSFE**) in place of **SFE** as follows, to approximate the morphometric sparse codes, specifically, provided by Eq. 2,

$$\min_{\mathbf{B},\mathbf{C},\mathbf{G},\mathbf{W}} \|\mathbf{Y}-\mathbf{B}\mathbf{C}\|_F^2 + \lambda\|\mathbf{C}\|_1 + \|\mathbf{C}-\mathbf{G}\sigma(\mathbf{W}\mathbf{Y})\|_F^2$$
$$\text{s.t. } \|\mathbf{b}_i\|_2^2 = 1, \forall i = 1,\ldots,h \tag{8}$$

where $\mathbf{Y}=[\mathbf{y}_1,\ldots,\mathbf{y}_N]\in\mathbb{R}^{m\times N}$ is a set of cellular morphometric descriptors; $\mathbf{B}=[\mathbf{b}_1,\ldots,\mathbf{b}_h]\in\mathbb{R}^{m\times h}$ is a set of the basis functions; $\mathbf{C}=[\mathbf{c}_1,\ldots,\mathbf{c}_N]\in\mathbb{R}^{h\times N}$ is the sparse feature matrix; $\mathbf{W}\in\mathbb{R}^{h\times m}$ is the auto-encoder; $\sigma(\cdot)$ is the element-wise sigmoid function; $\mathbf{G}=\text{diag}(g_1,\ldots,g_h)\in\mathbb{R}^{h\times h}$ is the scaling matrix with diag being an operator aligning vector, $[g_1,\ldots,g_h]$, along the diagonal; and $\lambda$ is the regularization constant. Joint minimization of Eq. (8) w.r.t the quadruple $<$ **B, C, G, W** $>$, enforces the inference of the nonlinear regressor $\mathbf{G}\sigma(\mathbf{WY})$ to be similar to the optimal sparse codes, **C**, which can reconstruct **Y** over **B** (Kavukcuoglu et al., 2008). As shown in Algorithm 1, optimization of Eq. (8) is iterative, and it terminates when either the objective function is below a preset threshold or the maximum number of iterations has been reached. In our implementation, the number of basis functions (**B**) was fixed to be 128, and the SPAMS optimization toolbox (Mairal et al., 2010) is adopted for efficient implementation of OMP to compute the sparse code, **C**, with sparsity prior set to 30. The end result is a highly efficient (see Table 6) and effective (see Fig. 10) system, CMF-**PredictiveSFE**-KSPM, for tissue histology classification.

**Algorithm 1** Construction of the Predictive Sparse Feature Encoder (PredictiveSFE).

```
Require: Training set $\mathbf{Y}=[\mathbf{y}_1,\ldots,\mathbf{y}_N]\in\mathbb{R}^{m\times N}$
Ensure: Predictive Sparse Feature Encoder $\mathbf{W}\in\mathbb{R}^{h\times m}$
print Randomly initialize B, W, and G
repeat
    Fixing B, W and G, minimize Eq. (8) w.r.t C, where Ccan be either solved as a $\ell_1$-minimization problem Lee et al. (2007) or equivalently solved by greedy algorithms, e.g., Orthogonal Matching Pursuit (OMP) Tropp and Gilbert (2007).
    Fixing B, W and C, solve for G, which is a simple least-square problem with analytic solution.
    Fixing C and G, update B and W, respectively,using the stochastic gradient descent algorithm.
until Convergence (maximum iterations reached or objective function $\leq$ threshold)
```

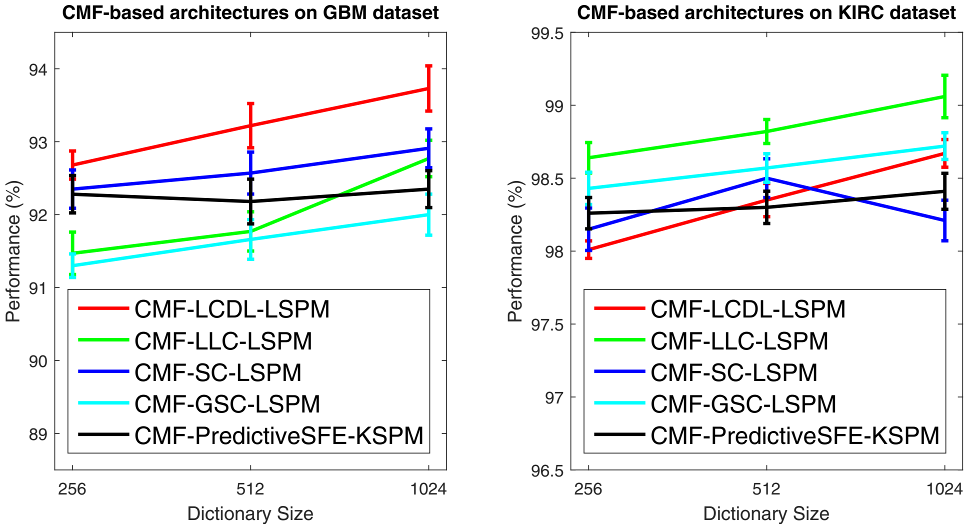

**Fig. 10.** Systems built-upon CMF-**PredictiveSFE** provide very competitive performance compared to systems built-upon CMF-**SFE**. Here, the performance is reported as the mean and standard error of the correct classification rate, as detailed in Section 4.

## 5. Conclusions

This paper provides insights to the following three fundamental questions for the task of tissue histology classification:

I. Is unsupervised feature learning preferable to human engineered features? The answer is that, CMF and PSD work the best, compared to SIFT and [Color,Texture] features, on two vastly different tumor types. The reasons are that (i) CMF encodes biological meaningful prior knowledge, which is widely adopted in the practice of pathological diagnosis; and (ii) PSD is able to capture intrinsic morphometric patterns in histology sections. As a result, both of them produce robust representation of the underlying properties preserved in the data.

II. Does cellular saliency help? The surprising answer is that cellular saliency does not help improve the performance for systems built upon pixel-/patch-level features. Experiments on both GBM and KIRC datasets confirm the performance-drop with salient feature extraction strategies, and one possible explanation is that both pixel-level and patch-level features are appearance-based representations, which require dense sampling all over the place in order to faithfully assemble the view of the image.

III. Does the sparse feature encoder contribute of recognition? The sparse feature encoder significantly and consistently improves the classification performance for systems built upon CMF; and meanwhile, it conditionally improves the performance for systems built upon PSD (PSD-SC-LSPM), with the choice of sparse coding (SC) as the intermediate feature extraction layer. It is believed that the consistency of performance highly correlates with the robustness of the feature being used, and the improvement of performance is due to the capability of the sparse feature encoder in capturing complex patterns at the higher-level. Furthermore, this paper provides a clear evidence that deep neural networks (i.e., AlexNet), pre-trained on large scale natural image datasets (i.e., ImageNet), is directly applicable to the task of tissue histology classification, which is due to the capability of deep neural networks in capturing transferable base knowledge across domains (Yosinski et al., 2014). Although the fine-tuning of AlexNet towards our datasets shows significant performance drop due to the problem of over-fitting, the direct deployment of pre-trained deep neural networks still provides a promising solution for tasks with limited data and labels, which is very common in the field of medical image analysis.

Besides the insights in the aforementioned fundamental questions, this paper also shows that the superior performance of the sparse feature encoder is at the cost of computational efficiency. However, the scalability of the sparse feature encoder can be improved by (i) the development of more computational-efficient algorithms; and (ii) the deployment of advanced computational techniques, such as cluster computing or GPU acceleration. As a demonstration, this paper provides an accelerated version of CMF-**SFE**, namely CMF-**PredictiveSFE**, which falls into the category of algorithmic-scaling-up and achieves 40X speed-up during sparse feature encoding. The end result is a highly scalable and effective system, CMF-**PredictiveSFE**-KSPM, for tissue histology classification.

Furthermore, all our insights are independently validated on two large cohorts, Glioblastoma Multiforme (GBM) and Kidney Clear Cell Carcinoma (KIRC), which, to the maximum extent, ensures the consistency and unbiasedness of our findings. To the best of our knowledge, this is the first attempt that systematically provides insights to the fundamental questions aforementioned in tissue histology classification; and there are reasons to hope that the configuration: **FE-SFE-LSPM** (**FE** $\in$ {CMF,PSD}) as well as its accelerated version: **FE-PredictiveSFE-KSPM** (**FE** $\in$ {CMF,PSD}), can be widely applicable to different tumor types.

## Acknowledgement

This work was supported by NIH R01 CA184476 (H.C) carried out at Lawrence Berkeley National Laboratory.

## References

Apostolopoulos, G., Tsinopoulos, S., Dermatas, E., 2011. Recognition and identification of red blood cell size using zernike moments and multicolor scattering images. In: 2011 10th International Workshop on Biomedical Engineering, pp. 1–4.
Asadi, M., Vahedi, A., Amindavar, H., 2006. Leukemia cell recognition with zernike moments of holographic images. In: NORSIG 2006, pp. 214–217.
Basavanhally, A., Xu, J., Madabhushu, A., Ganesan, S., 2009. Computer-aided prognosis of ER+ breast cancer histopathology and correlating survival outcome with oncotype DX assay. In: ISBI, pp. 851–854.
Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 15 (6), 1373–1396.

Bhagavatula, R., Fickus, M., Kelly, W., Guo, C., Ozolek, J., Castro, C., Kovacevic, J., 2010. Automatic identification and delineation of germ layer components in *h&e* stained images of teratomas derived from human and nonhuman primate embryonic stem cells. In: ISBI, pp. 1041–1044.

Chang, H., Borowsky, A., Spellman, P., Parvin, B., 2013a. Classification of tumor histology via morphometric context. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 2203–2210.

Chang, H., Han, J., Borowsky, A., Loss, L.A., Gray, J.W., Spellman, P.T., Parvin, B., 2013b. Invariant delineation of nuclear architecture in glioblastoma multiforme for clinical and molecular association. IEEE Trans. Med. Imaging 32 (4), 670–682.

Chang, H., Nayak, N., Spellman, P., Parvin, B., 2013c. Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching. Medical image computing and computed-assisted intervention–MICCAI.

Chang, H., Zhou, Y., Spellman, P.T., Parvin, B., 2013. Stacked predictive sparse coding for classification of distinct regions in tumor histopathology. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 502–507.

Dalton, L., Pinder, S., Elston, C., Ellis, I., Page, D., Dupont, W., Blamey, R., 2000. Histolgical gradings of breast cancer: Linkage of patient outcome with level of pathologist agreements. Modern Pathol. 13 (7), 730–735.

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

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A Large-Scale Hierarchical Image Database. In: CVPR09, pp. 248–255.

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

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., Zisserman, A., 2012. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results.

Fatakdawala, H., Xu, J., Basavanhally, A., Bhanot, G., Ganesan, S., Feldman, F., Tomaszewski, J., Madabhushi, A., 2010. Expectation-maximization-driven geodesic active contours with overlap resolution (EMagacor): Application to lymphocyte segmentation on breast cancer histopathology. IEEE Trans. Biomed. Eng. 57 (7), 1676–1690.

Ghaznavi, F., Evans, A., Madabhushi, A., Feldman, M.D., 2013. Digital imaging in pathology: Whole-slide imaging and beyond. Ann. Rev. Pathol. Mech. Dis. 8 (1), 331–359.

Gurcan, M., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., Bulent, Y., 2009. Histopathological image analysis: A review. IEEE Trans. Biomed. Eng. 2, 147–171.

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

Huang, C., Veillard, A., Lomeine, N., Racoceanu, D., Roux, L., 2011. Time efficient sparse analysis of histopathological whole slide images. Comput. med. imaging graphics 35 (7–8), 579–591.

Huang, F.J., LeCun, Y., 2006. Large-scale learning with svm and convolutional for generic object categorization. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1. IEEE Computer Society, Washington, DC, USA, pp. 284–291. doi:10.1109/CVPR.2006.164.

Huang, W., Hennrick, K., Drew, S., 2013. A colorful future of quantitative pathology: validation of vectra technology using chromogenic multiplexed immunohistochemistry and prostate tissue microarrays. Human Pathol. 44, 29–38.

Huijbers, A., Tollenaar1, R., v Pelt1, G., Zeestraten1, E., Dutton, S., McConkey, C., Domingo, E., Smit, V., Midgley, R., Warren, B., Johnstone, E.C., Kerr, D., Mesker, W., 2013. The proportion of tumor-stroma as a strong prognosticator for stage ii and iii colon cancer patients: Validation in the victor trial. Ann. Oncol. 24 (1), 179–185.

Jarrett, K., Kavukcuoglu, K., Ranzato, M., LeCun, Y., 2009. What is the best multi--stage architecture for object recognition? In: Proc. International Conference on Computer Vision (ICCV'09). IEEE, pp. 2146–2153.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv: 1408.5093.

Kavukcuoglu, K., Ranzato, M., LeCun, Y., 2008. Fast Inference in Sparse Coding Algorithms with Applications to Object Recognition. Technical Report CBLL-TR-2008-12-01. Computational and Biological Learning Lab, Courant Institute, NYU.

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

Kothari, S., Phan, J., Osunkoya, A., Wang, M., 2012. Biological interpretation of morphological patterns in histopathological whole slide images. In: ACM Conference on Bioinformatics, Computational Biology and Biomedicine, pp. 218–225.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3–6, 2012, Lake Tahoe, Nevada, United States., pp. 1106–1114.

de Kruijf, E.M., van Nes, J.G., van de Velde, C.J.H., Putter, H., Smit, V.T.H.B.M., Liefers, G.J., Kuppen, P.J.K., Tollenaar, R.A.E.M., Mesker, W.E., 2011. Tumor-stroma ratio in the primary tumor is a prognostic factor in early breast cancer patients, especially in triple-negative carcinoma patients. Breast Cancer Res. Treatment 125 (3), 687–696.

Lan, C., Heindl, A., Huang, X., Xi, S., Banerjee, S., Liu, J., Yuan, Y., 2015. Quantitative histology analysis of the ovarian tumour microenvironment. Scientific Reports 5 (16317).

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 2169–2178.

Le, Q., Han, J., Gray, J., Spellman, P., Borowsky, A., Parvin, B., 2012. Learning invariant features from tumor signature. In: ISBI, pp. 302–305.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. In: Proceedings of the IEEE, pp. 2278–2324.

Lee, H., Battle, A., Raina, R., Ng, A.Y., 2007. Efficient sparse coding algorithms. In: In NIPS. NIPS, pp. 801–808.

Levenson, R.M., Borowsky, A.D., Angelo, M., 2015. Immunohistochemistry and mass spectrometry for highly multiplexed cellular molecular imaging. Lab. Invest. 95, 397–405.

Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2010. Online learning for matrix factorization and sparse coding. J. Mach. Learn. Res. 11, 19–60.

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

Rimm, D.L., 2014. Next-gen immunohistochemistry. Nature Meth. 11, 381–383.

Rogojanu, R., Thalhammer, T., Thiem, U., Heindl, A., Mesteri, I., Seewald, A., Jger, W., Smochina, C., Ellinger, I., Bises, G., 2015. Quantitative image analysis of epithelial and stromal area in histological sections of colorectal cancer: An emerging diagnostic tool. BioMed. Res. Int. 2015 (569071), 179–185.

Serre, T., Wolf, L., Poggio, T., 2005. Object recognition with features inspired by visual cortex. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, 2, pp. 994–1000.

Stack, E.C., Wang, C., Roman, K.A., Hoyt, C.C., 2014. Multiplexed immunohistochemistry, imaging, and quantitation: A review, with an assessment of tyramide signal amplification, multispectral imaging and multiplex analysis. Methods 70 (1), 46–58.

Tropp, J., Gilbert, A., 2007. Signal recovery from random measurements via orthogonal matching pursuit. Inf. Theory, IEEE Trans. 53 (12), 4655–4666. doi:10.1109/TIT.2007.909108.

Vedaldi, A., Zisserman, A., 2012. Efficient additive kernels via explicit feature maps. IEEE Trans. Pattern Anal. Mach. Intell. 34 (3), 480–492.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y., 2010. Locality-constrained linear coding for image classification. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 3360–3367.

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

Yang, J., Yu, K., Gong, Y., Huang, T., 2009. Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 1794–1801.

Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8–13 2014, Montreal, Quebec, Canada, pp. 3320–3328.

Young, R.A., Lesperance, R.M., 2001. The gaussian derivative model for spatial-temporal vision. I. Cortical Model. Spatial Vision 2001, 3–4.

Zheng, M., Bu, J., Chen, C., Wang, C., Zhang, L., Qiu, G., Cai, D., 2011. Graph regularized sparse coding for image representation. IEEE Trans. Image Process. 20 (5), 1327–1336.

Zhou, Y., Barner, K.E., 2013. Locality constrained dictionary learning for nonlinear dimensionality reduction. IEEE Signal Process. Lett. 20 (4), 335–338.