Superpixel-Based Conditional Random Fields (SuperCRF): Incorporating Global and Local Context for Enhanced Deep Learning in Melanoma Histopathology

Konstantinos Zormpas-Petridis [1,2†], Henrik Failmezger [3†], Shan E Ahmed Raza [3], Ioannis Roxanis [3,4], Yann Jamin [1,2] and Yinyin Yuan [3*]

[1] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, United Kingdom, [2] The Royal Marsden NHS Trust, Surrey, United Kingdom, [3] Division of Molecular Pathology, The Institute of Cancer Research, London, United Kingdom, [4] Royal Free London NHS Foundation Trust, London, United Kingdom

Computational pathology-based cell classification algorithms are revolutionizing the study of the tumor microenvironment and can provide novel predictive/prognosis biomarkers crucial for the delivery of precision oncology. Current algorithms used on hematoxylin and eosin slides are based on individual cell nuclei morphology with limited local context features. Here, we propose a novel multi-resolution hierarchical framework (SuperCRF) inspired by the way pathologists perceive regional tissue architecture to improve cell classification and demonstrate its clinical applications. We develop SuperCRF by training a state-of-art deep learning spatially constrained- convolution neural network (SC-CNN) to detect and classify cells from 105 high-resolution (20×) H&E-stained slides of The Cancer Genome Atlas melanoma dataset and subsequently, a conditional random field (CRF) by combining cellular neighborhood with tumor regional classification from lower resolution images (5, 1.25×) given by a superpixel-based machine learning framework. SuperCRF led to an 11.85% overall improvement in the accuracy of the state-of-art deep learning SC-CNN cell classifier. Consistent with a stroma-mediated immune suppressive microenvironment, SuperCRF demonstrated that (i) a high ratio of lymphocytes to all lymphocytes within the stromal compartment ($p = 0.026$) and (ii) a high ratio of stromal cells to all cells ($p < 0.0001$ compared to $p = 0.039$ for SC-CNN only) are associated with poor survival in patients with melanoma. SuperCRF improves cell classification by introducing global and local context-based information and can be implemented in combination with any single-cell classifier. SuperCRF provides valuable tools to study the tumor microenvironment and identify predictors of survival and response to therapy.

Keywords: deep learning, machine learning, conditional random fields, digital pathology, cell classification, melanoma, tumor microenvironment

INTRODUCTION

Cancer is a highly complex, non-autonomous disease. The interactions between microenvironmental selective pressures and cancer cells dictate how cancer progresses and evolves. Accurate and spatially explicit characterization of the tumor microenvironmental landscape including how cancer cells interact with the extra-cellular matrix and other cellular players such as stromal cells and immune cells within the tumoral niche, is needed to understand the context in which cancer evolves, and may also provide robust predictor of cancer behavior for risk-stratification (1). More specifically the recent success of cancer immunotherapy including the spectacular response observed in patients with previously incurable melanoma, a highly aggressive form of skin cancer, calls for a better understanding of the cancer-immune interface.

In the new era of digital pathology, advanced image analysis can objectively, consistently, and quantitatively characterize the different components of the tumor and how they spatially interact, and as a result assist pathologists in tasks such as tumor grading (2). Algorithms for cell detection and classification are key components of this process. Machine learning, and more recently deep learning algorithms, both exploiting the phenotypic differences in nuclear morphology between each cell type, revolutionized the field yielding significantly better cell detection, segmentation, and classification results (3–9).

However, even state-of-the-art deep learning algorithms can underperform especially in cases where different cell types appear morphologically similar. Current computed pathology tools focus on individual cell nuclei morphology with limited abstract local context features, whereas pathologists incorporate regional tissue architecture (in practice, by zooming in/out), together with cell morphological features to accurately classify cells.

Here, we hypothesize that robust tumor regional classification from lower resolution images can provide the contextual information that is key to further improve single cell classification algorithms. Our aim is to introduce dependencies on global tissue context and cell neighborhood and enhance learning results for cell classification from deep convolution neural networks (CNNs). Probabilistic graphical models have successfully been applied to improve cell classification in time-lapse imaging by taking into account the temporal context of a cell (10–15). Probabilistic graphical models have also been used successfully in histopathology images for pathology detection and segmentation (16–19), disease and tissue staging (20, 21), and nuclei segmentation (22). In our study, instead of time dependency, we apply graphical models to introduce the spatial context of a cell as additional information to improve single-cell classification. A multi-resolution hierarchical framework was proposed to mirror the way pathologists perceive tumor architecture, and applied to whole-slide images (WSI) hematoxylin and eosin (H&E)-stained slides of melanoma skin cancer (**Figure 1A**). We demonstrated that our new system is computationally efficient and significantly improves single cell classification. The increased accuracy in cell classification further enabled us to shed new light on the understanding of cancer-immune-stroma interface of melanoma.

MATERIALS AND METHODS

Datasets

In total, 105 full-face, H&E stained section images from formalin-fixed, paraffin-embedded (FFPE) diagnostic blocks of melanoma skin cancer from The Cancer Genome Atlas (TCGA) were used. We scaled all digitized (Aperio ImageScope) histology images to 20, 5, and 1.25× magnification with pixel resolution 0.504, 2.016, and 8.064 μm, respectively, using Bio-Formats (https://www.openmicroscopy.org/bio-formats/). WSIs at 20× magnification (representative size: 30,000 × 30,000 pixels), were split into sub-images (tiles) of 2,000 × 2,000 pixels each, for computational efficiency.

For the purpose of training and testing the different parts of our system we divided the dataset into sub-datasets, namely single-cell classification dataset, 5× sub-dataset, 1.25× sub-dataset and discovery sub-dataset (**Table 1**, also see **Supplementary Tables 1–4**).

Single-Cell Classification Using a Spatially Constrained Convolutional Neural Network

We used a Spatially Constrained Convolutional Neural Network (SC-CNN) (6) for single cell classification (**Figure 1E**). SC-CNN uses spatial regression in order to predict the probability of a pixel being the center of the nucleus. The nucleus is classified by a neighboring ensemble predictor (NEP) in conjunction with a standard softmax CNN. We randomly initialized the network's layers as we have found that to perform better than transfer learning from real-world datasets in our experiments with pathological samples.
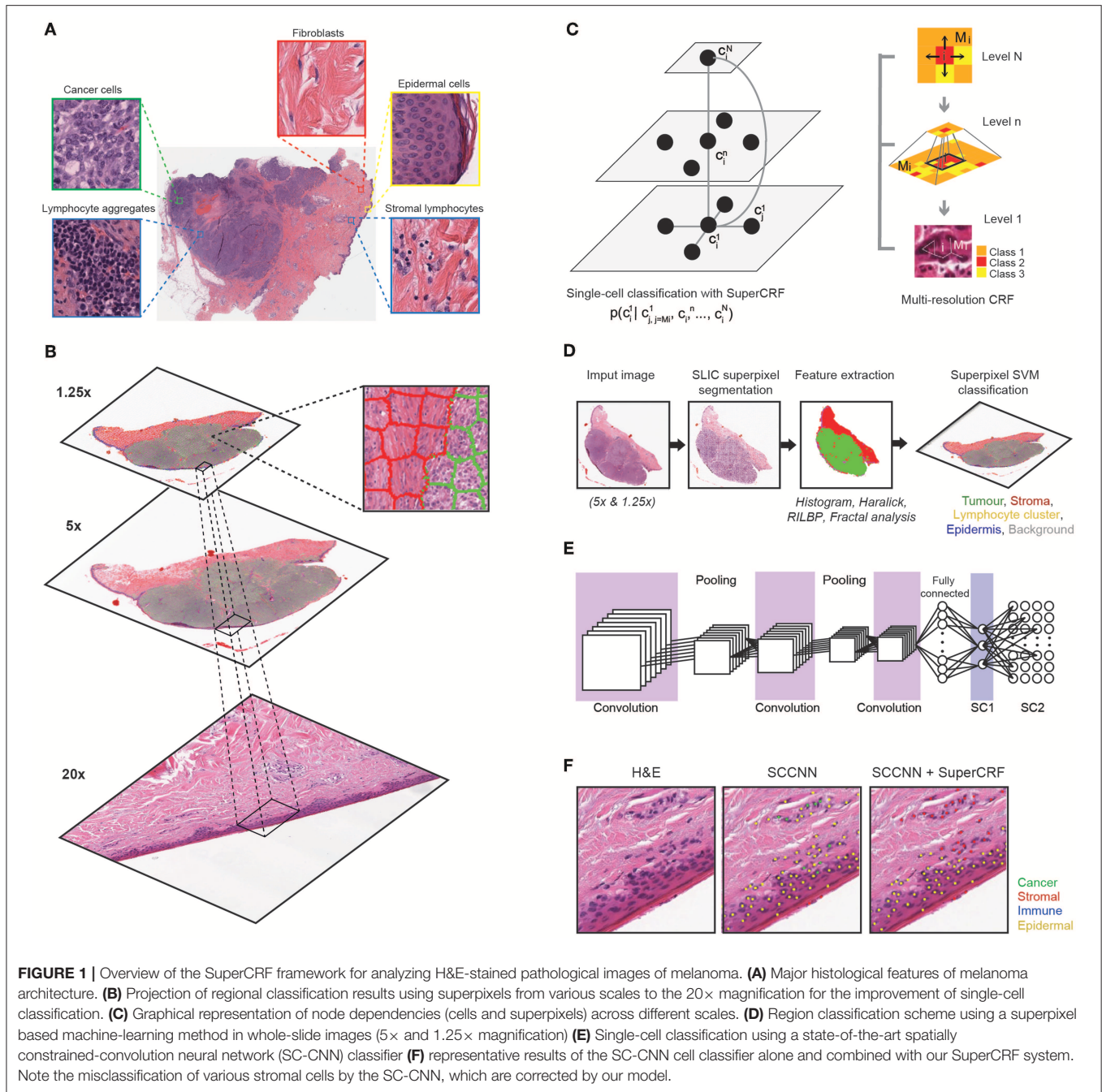
Superpixel-Based Tumor Region Classification

A machine learning superpixel-based framework was implemented in Matlab (23) to classify tumor tissue regions and was subsequently applied to low resolution (5 and 1.25×) images. Reinhard stain normalization (24) was applied separately on each of the 5 and 1.25× sub-datasets to account for stain variabilities that could affect the classification (25).

Downscaled images were segmented using the simple linear iterative clustering (SLIC) superpixels algorithm (26), which is designed to provide roughly uniform superpixels. Choosing the optimal number of superpixels is important to ensure that the superpixels capture homogeneous areas and adhere to image boundaries. With our pathologist's input, we visually identified a size of superpixels that met these criteria and chose the number of superpixels automatically based on each image's size (Equation 1).

$$N_i = ceiling\left(\frac{S_i}{U}\right) \tag{1}$$

where N_i is the number of superpixels in the ith image, S_i is the size of image i in pixels, and U (here $U = 1,250$) is a constant held across all images that defined a desired size of the superpixels. This means, on average, a superpixel occupies an area of approximately 35 × 35 pixels, equivalent to 280 × 280 mm^2. We identified the superpixels belonging to each area by

FIGURE 1 | Overview of the SuperCRF framework for analyzing H&E-stained pathological images of melanoma. **(A)** Major histological features of melanoma architecture. **(B)** Projection of regional classification results using superpixels from various scales to the 20× magnification for the improvement of single-cell classification. **(C)** Graphical representation of node dependencies (cells and superpixels) across different scales. **(D)** Region classification scheme using a superpixel based machine-learning method in whole-slide images (5× and 1.25× magnification) **(E)** Single-cell classification using a state-of-the-art spatially constrained-convolution neural network (SC-CNN) classifier **(F)** representative results of the SC-CNN cell classifier alone and combined with our SuperCRF system. Note the misclassification of various stromal cells by the SC-CNN, which are corrected by our model.

determining whether their central points fell within the regions annotated by the pathologist.

Overall, for the 1.25× training sub-dataset, we found 15,477 superpixels belonging in tumor areas, 6,989 in stroma areas, 141 in epidermis and 691 in lumen/white space, while for the 5× training sub-dataset we found 1,193 superpixels belonging in tumor, 1,324 in stroma areas, 360 in epidermis, 506 in lymphocyte clusters and 830 in lumen/white space.

Next, we extracted four types of features, 85 in total, from each superpixel, including seven histogram features (mean values of hue, saturation, and brightness, sum of intensities, contrast, standard deviation, and entropy), and well-established texture features [12 Haralick features (27), 59 rotation-invariant local binary patterns (RILBP), 7 segmentation-based fractal texture analysis (SFTA) features (28)]. Features were standardized into z-scores. The mean values and standard deviation of the features from the training set were used for the normalization of the test set. A support vector machine (SVM) with a radial basis function (RBF, $\gamma = 1/number\_of\_features$) was trained with these

TABLE 1 | Summary of the data used to train and test the different parts of the SuperCRF system, as well as study the cancer-immune-stroma interface (also, see **Supplementary Tables 1–4**).

Name	Number of WSIs	Purpose
Single-cell classification sub-dataset	**8** Training SC-CNN: 3 (348 tiles) Training SuperCRF: 2 (84 tiles) Testing: 3 (290 tiles)	Single-cell classification into four categories: cancer cells, lymphocytes, stromal cells, epidermal cells
5x sub-dataset	**16** Training: 10 Testing: 6	Region classification into five categories: tumor, normal stroma, lymphocyte cluster, normal epidermis, lumen/white space
1.25x sub-dataset	**58** Training: 21 Testing: 37	Region classification into four categories: tumor, normal stroma, normal epidermis, lumen/white space
Discovery dataset	**97**	Study of the tumor-stroma interface. To accelerate the analysis, 50 tiles (2,000 × 2,000 pixels) containing tumors were randomly sampled from every whole-slide image (WSI)

The values are bold for visual (illustration) purposes.

features to classify superpixels into different biologically meaningful categories.

For the 5× sub-dataset, superpixels were classified into five categories: tumor area, normal stroma, normal epidermis, lymphocytes cluster, and lumen/white space. We increased the penalty in the cost function for the epidermis and lumen/white space classes by a factor of 10 when training the SVM, to account for class imbalance. For the 1.25× sub-dataset superpixels classification consisted of four categories: tumor area, normal stroma, normal epidermis, and lumen/white space. We randomly selected a subset of 5,000 cancer and stroma superpixels and increased the penalty in the cost function for the epidermis and lumen/white space classes by a factor of 10, again to account for class imbalance (**Figure 1D**).

SuperCRF

Single-cell based classification approaches often assign a class label based on the morphology of -individual cells, regardless of their neighboring cells. However, these spatial relationships provide important information that is used by pathologists. Conditional random fields (CRF) are undirected graphical models that represent efficient ways to model dependences, by factorizing the probability density into a specific set of conditional dependence (29). Therefore, the tumor microenvironment can be modeled by a CRF by introducing nodes for cells and superpixels, as well as edges whenever there is a spatial relationship between nodes.

We excluded lymphocytes from the CRF assumption that neighboring cells have a higher probability to share the same class labels, since they infiltrate, in an inconsistent manner ranging from sparse to highly dense, in tumor as well as stromal tissue. Therefore, lymphocytes kept their label as assigned by the SC-CNN.

Let n be the total number of cells (besides lymphocytes) in the image and $c_i \in \{stromal, cancer, epidermis\}$, $i = 1, 2, \ldots, n$ the input labels of the cells as assigned by the SC-CNN. Let s_i, be the corresponding superpixel for a cell c_i with $s_i \in \{stromal, cancer, epidermis, white space\}$ for 1.25× superpixels and $s_i \in \{stromal, cancer, epidermis, lymphocyte, whitespace\}$

for 5× superpixels. $\mathbf{x} \in \{\mathbf{c}, \mathbf{s}\}$ comprises the nodes of the CRF. The CRF assigns output labels $y_i \in \{stromal, cancer, epidermis, lymphocyte, white space\}$ based on the input data. The joint probability distribution over input data and output labels, $p(y_1, y_2, \ldots, y_n \lfloor x_1, x_2, \ldots, x_n)$ can be modeled by factorizing the probability density into a specific set of conditional dependence relationships (**Figure 1C**).

$$p(y \mid x) = \prod_n^{i=1} p(y_i \mid x_i) = \frac{1}{Z} \exp(\sum E(x_i, y_i, xN_i, yN_i)) \quad (2)$$

where Z is a normalizing constant, w is a weight vector and

$$E(x_i, y_i, xN_i, yN_i,) = \sum \Phi(x_i, y_i) + \sum \psi_c(xN_i, yN_i) \quad (3)$$

defines the energy function of the CRF.

The node potentials $\Phi(x_i, y_i)$ represent the evidence that a cell i, with the input label x_i takes the class label y_i. The node potential can be defined as $\Phi(x_i, y_i) = f(x_i, y_i) + b$, with $f(x_i, y_i) = \begin{cases} 1 & if\ y_i = x_i \\ 0 & else \end{cases}$ and b representing the bias.

The edge potentials $\psi_c(xN_i, yN_i)$ model the probability that neighboring cells take a similar cell label. N_i is the neighborhood of cell i, defined as all the cells that can be found in a defined distance. The edge potentials are defined as: $\psi_c(x_i, y_i, xN_i, yN_i) = f(x_i, y_i) * f(xN_i, yN_i) + b$.

The CRF was trained with stochastic gradient descent and the decoding was applied using loopy belief propagation. The toolbox of M. Schmidt was used to train and decode the CRF (30).
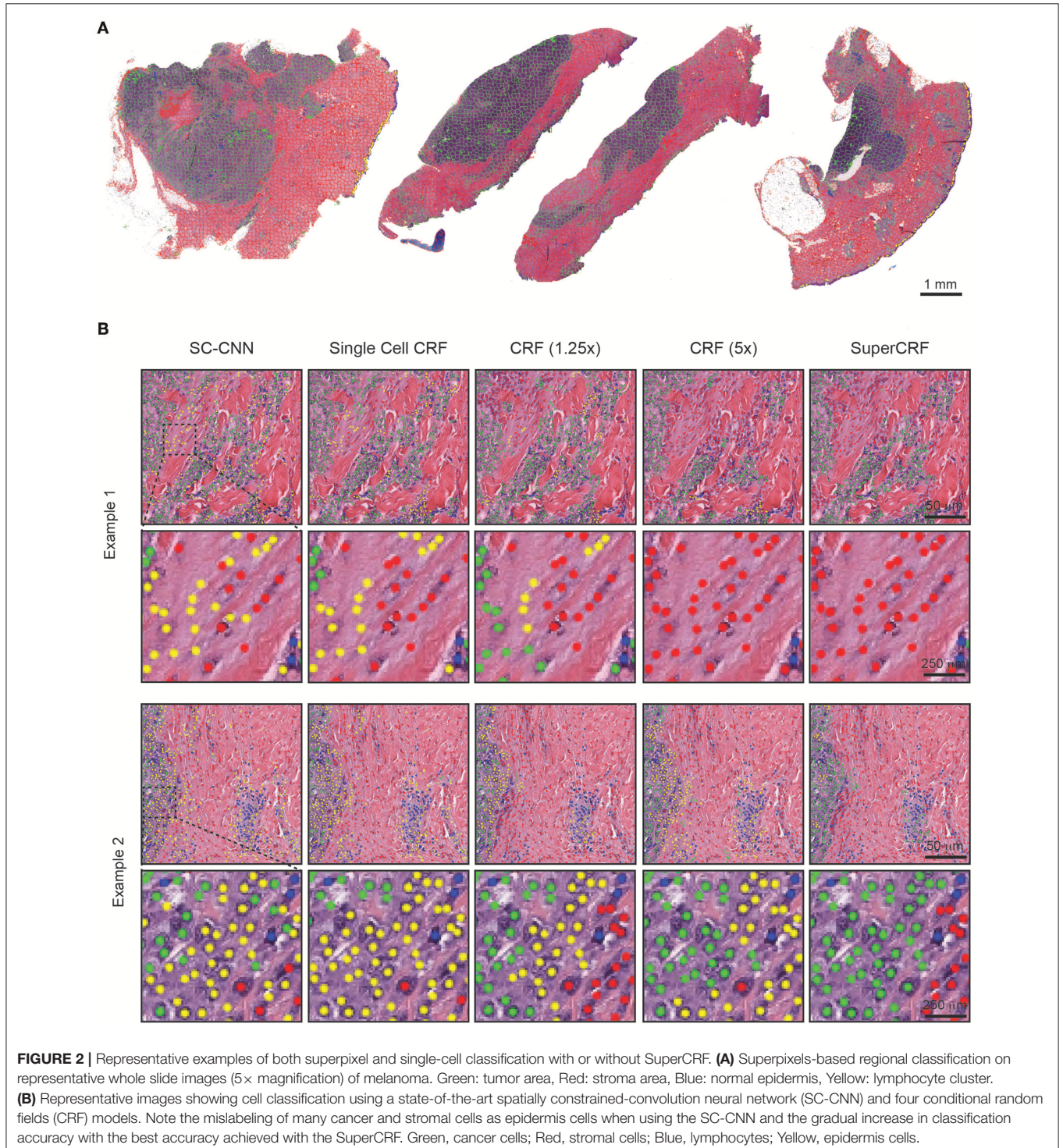
The source code for the study is available at Github (https://github.com/Henrik86/SuperCRF).

Survival Analysis

We evaluated the prognosis value of the abundance of stromal cells and location of lymphocytes in our discovery sub-dataset. The ratio of stromal cells to all cells, the ratio of lymphocytes in cancer areas to all lymphocytes, and the ratio of lymphocytes in stroma areas to all lymphocytes were calculated for each patient.

Patients were divided into high and low ratio groups, split at the median value of all scores. Patients with a ratio of lymphocytes being high inside the tumor area and low in the stroma were categorized as the "immune infiltration" group whereas patients with a ratio of lymphocytes being low in the tumor area and high in the stroma were categorized as "immune excluded," based on the recent classification of the main immune phenotypes of

anticancer immunity that predict response to immunotherapy (31). The number of patients belonging to neither of these two groups (high/high $n = 6$ and low/low $n = 5$) was too small to perform the survival analysis. Non-parametric Kaplan-Meier estimation was used to analyze overall survival in 94 patients. Differences between survival estimates were assessed with the log-rank test. Finally, Cox regression models were adjusted,

FIGURE 2 | Representative examples of both superpixel and single-cell classification with or without SuperCRF. **(A)** Superpixels-based regional classification on representative whole slide images (5× magnification) of melanoma. Green: tumor area, Red: stroma area, Blue: normal epidermis, Yellow: lymphocyte cluster. **(B)** Representative images showing cell classification using a state-of-the-art spatially constrained-convolution neural network (SC-CNN) and four conditional random fields (CRF) models. Note the mislabeling of many cancer and stromal cells as epidermis cells when using the SC-CNN and the gradual increase in classification accuracy with the best accuracy achieved with the SuperCRF. Green, cancer cells; Red, stromal cells; Blue, lymphocytes; Yellow, epidermis cells.

testing for the independent prognostic relevance of our risk scores. To test if Breslow-thickness (the distance between the upper layer of the epidermis and the deepest point of tumor penetration) was contributing to a high ratio of stromal cells, we created a multivariate model containing both stromal cells ratio and Breslow-thickness, as well as two univariate models containing the covariates separately. Pearson's correlation was used to test for linear relation between the two variables.

RESULTS

SuperCRF Improves Accuracy of Cell Classification

First, we trained the state-of-the-art deep learning method, spatially-constrained CNN (SC-CNN) algorithm, to detect and classify cells in high resolution ($20\times$) WSI into four categories: cancer cells, stroma cells, lymphocytes, and epidermis cells. The SC-CNN network yielded an accuracy of 84.63% over 4,059 cells in the independent test set (**Table 1**, **Supplementary Table 5**). Visual inspection revealed that the majority of false positives were misclassification of stromal and cancer cells as epidermis, which confirmed our initial motive for the incorporation of regional and spatial information to improve classification.

Subsequently, we trained a conditional random field (CRF) by combining the cellular neighborhood with tumor region classification (cancer area, normal stroma, normal epidermis, lymphocyte cluster, and lumen/white space) from low resolution images (5 and $1.25\times$, **Figure 1B**), given by the superpixel-based machine-learning framework. The SLIC superpixels algorithm has previously been shown to be computationally efficient, requiring only 3s on average to segment a single downscaled image of $2,500 \times 2,500$ pixels using a 2.9 GHz Intel core i7 processor. Performance of classification using individual and various combinations of feature sets was tested and the use of all 85 features, yielded the highest accuracy (23). It was then applied on the two datasets of 1.25 and $5\times$ magnification (**Figure 2A**) and achieved high accuracy in regional classification ($1.25\times$ sub-dataset: Overall accuracy 97.7% in the training set using 10-fold cross validation and 95.7% in 2,997 superpixels annotated in the 37 images of the independent test set; $5\times$ sub-dataset: Overall accuracy 97.1% in the training set using 10-fold cross validation and 95.2% in 1,798 superpixels annotated in the six images of the independent test set).

To train SuperCRF, we first introduced dependencies on cell neighborhood. Cells were considered neighbors in the CRF, if they were located in a spatial proximity of $15\,\mu m$ (or 30 pixels), which resulted in an average of 1.3 neighbors per cell. Subsequently, we integrated this local neighborhood with global context by connecting the CRF single-cell nodes to the regional classification results from superpixels. To determine the best configuration, we trained four different CRFs and compared their performance in terms of single-cell classification on a test set, including three samples, 290 tiles and 4,059 single-cell annotations (1,527 cancer cells, 676 lymphocytes, 837 normal epidermis cells, 1,019 stromal cells).

In detail, for the first CRF we did not use any context classification, just cell neighborhood dependencies, i.e., the

TABLE 2 | Evaluation of different conditional random fields (CRF) versions and a state-of-the-art spatially constrained-convolution neural network (SC-CNN) deep learning cell-classifier.

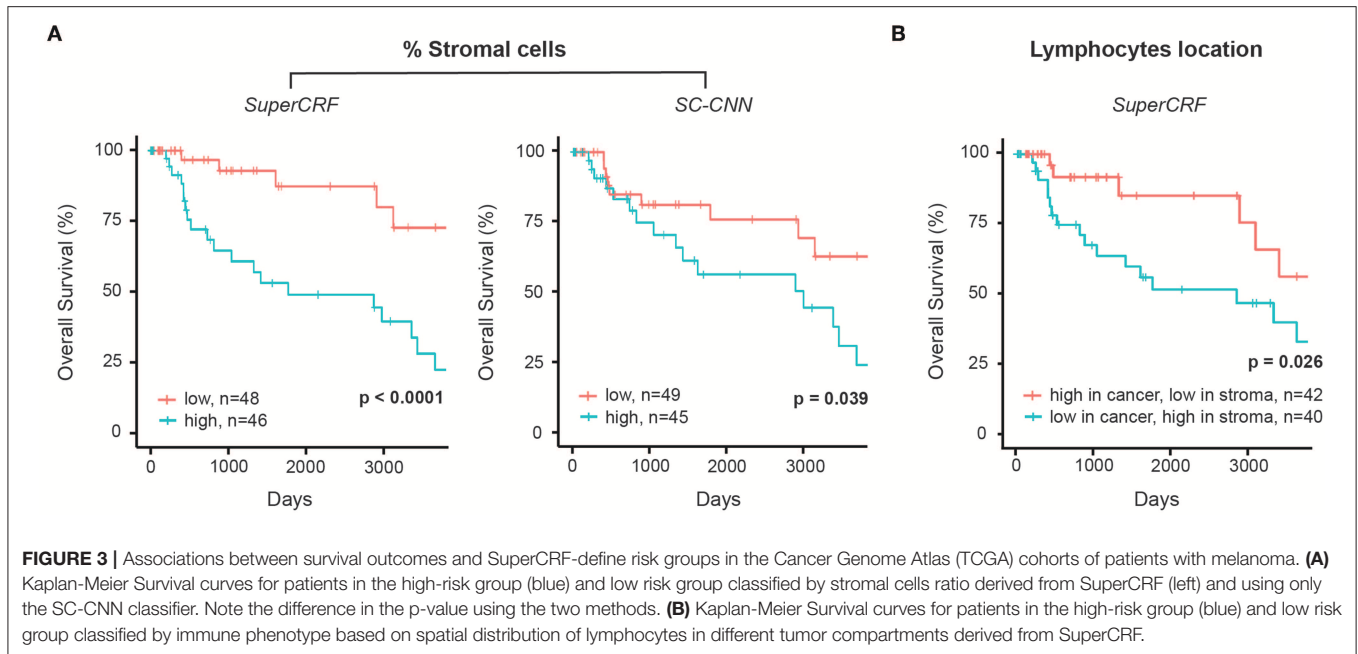| Method	Accuracy (%)	Precision	Recall
SC-CNN	84.63	0.8756	0.8808
singleCellCRF	87.61	0.8973	0.8946
CRF1.25×	90.79	0.9248	0.9110
CRF5×	91.70	0.9298	0.9126
SuperCRF	**96.48**	**0.9644**	**0.9629**

The values are bold to indicate the highest achieved accuracy, precision and recall.

only edges of the CRF were between neighboring cells (singleCellCRF). For the second and third CRF we introduced superpixel nodes. Now, single-cell nodes are not only connected to neighboring cells but every single-cell node is also connected to a superpixel node. We trained a CRF for $5\times$ superpixel classification (CRF5×) and $1.25\times$ superpixel classification (CRF1.25×). Furthermore, we trained a CRF in which every single-cell node was connected to two superpixel nodes in 5 and $1.25\times$ resolution (SuperCRF). Already the singleCellCRF (Accuracy: 87.6%, Precision: 89.7%, Recall: 89.5%, **Table 2**) improves the classification accuracy compared to the SC-CNN (84.6%, Precision: 87.6%, Recall: 88.1%, **Table 2**). However, the use of contextual information by the introduction of superpixel nodes, markedly improves the classification metrics (Accuracy 90.8%, Precision: 92.5%, Recall: 91.1%, **Table 2**) for CRF1.25× and (Accuracy 91.7%, Precision: 93%, Recall: 91.3%, **Table 2**) for the CRF5×. The SuperCRF, using nodes from superpixels in both 5 and $1.25\times$ resolution images, as well as the neighboring cells, resulted in the highest classification outcome (Accuracy 96.5%, Precision: 96.4%, Recall: 96.3%, **Table 2**, **Figures 1F**, **2B**, **Supplementary Tables 5–9**).

SuperCRF's Increased Accuracy of Cell Classification Improves Confidence in Stromal Cell Ratio as a Predictive Feature of Survival in Melanoma

The crosstalk between cancer cells and stromal cells play an active role in tumor invasion and metastasis, and controlling immune infiltration and is increasingly recognized as a hallmark of cancer (32). Tumor-stromal cell ratio has been shown to hold prognostic and predictive information in patient with solid tumors (31, 33, 34). Here, we demonstrate that a high stromal cell ratio is also a predictor of poor prognosis in melanoma using both values derived from the multivariate models of SC-CNN and SuperCRF in our discovery sub-dataset. Yet SuperCRF yields a significantly higher confidence in the predictive value of the stromal cell ratio (SuperCRF: $p < 0.0001$, Coxph-Regression (discretized by median): HR = 4.1, $p = 0.006$; SC-CNN: $p = 0.039$, Coxph-Regression (discretized by median): HR = 2.4, $p = 0.05$, **Figure 3A**).

Similar regression coefficients for both stromal cells ratio and Breslow-thickness covariates were observed between the multivariate and the two univariate survival models (1.404 and

FIGURE 3 | Associations between survival outcomes and SuperCRF-define risk groups in the Cancer Genome Atlas (TCGA) cohorts of patients with melanoma. **(A)** Kaplan-Meier Survival curves for patients in the high-risk group (blue) and low risk group classified by stromal cells ratio derived from SuperCRF (left) and using only the SC-CNN classifier. Note the difference in the p-value using the two methods. **(B)** Kaplan-Meier Survival curves for patients in the high-risk group (blue) and low risk group classified by immune phenotype based on spatial distribution of lymphocytes in different tumor compartments derived from SuperCRF.

0.171, respectively, for the multivariate model and 1.633 and 0.179 for the univariate models) of the SuperCRF. Pearson's correlation showed no correlation between stromal cells ratio and Breslow-thickness ($r = -0.05$), overall indicating that stromal cells ratio is independent to Breslow-thickness.

Combining Cell and Region Classification: Location of the Immune Infiltrate Is Predictive of Survival in Melanoma

There is increasing evidence of the value of immune infiltration to provide prognostic information and predictors of response in patient with melanoma [recently reviewed in (35)]. The spatial compartmentalization of immune cells afforded by our SuperCRF (by the cell and region classification results) was used to define the recently-described main immune phenotypes of anticancer immunity that predict response to immunotherapy (31). Patients with a classified "immune excluded" phenotype, defined by a low lymphocyte ratio inside the tumor area and high inside the stroma area, was associated with a significantly worse prognosis compared to "inflamed" tumors characterized by a high ratio of lymphocytes inside the tumor and a low ratio inside the stroma ($p = 0.026$, Cox PH –regression: HR = 2.57, $p = 0.032$, **Figure 3B**). Taken together, our data is consistent with the model of a stroma-mediated immune suppressive microenvironment that exclude T cells from the vicinity of cancer cells.

DISCUSSION

In this study, we implemented a framework which fuses traditional machine learning with deep learning to model the way pathologists incorporate large-scale tissue architecture

and context across spatial scales, to improve single-cell classification in large whole-section slide images. Using this approach, we demonstrated a marked 11.85% overall improvement in the accuracy of the state-of-art deep learning SC-CNN cell classifier. Also, the similar values of both precision and recall and their simultaneous increase in every step show the unbiased nature of our approach.

Computational pathology algorithms, typically exploit the inter-cell phenotypic differences for cell classification, yet even state-of-art deep learning algorithms tend to underperform in this task, mainly due to the disproportional numbers of cells sharing similar nuclear morphological features, or due to intra-class diversity, seen for example in tumor stroma (fatty tissue, necrosis, vessels, muscle, fibroblasts, and associated collagen). Whilst computers can quantify morphological differences in a considerably more complex way, pathologists still generally outperform computers in cell classification. An essential reason is that they incorporate key contextual information such as heterogeneous tissue architecture, together with cell morphological features.

The idea that a cancer cell is dependent on its neighboring cells and global context is comparable to the fundamental concept in landscape ecology that a living population depends on the existing habitats and is not equally spread on the terrain. A particular habitat could favor the development of specific organisms. In practice, landscape ecologists denote the habitats from satellite images and then "ground-truth" them by detailed small-scale sampling of the habitats of interest (36). This inspired the design of our framework by introducing CRF dependencies between (i) the cells and their neighbors and (ii) the cells and to the global context (i.e., habitats from low resolution captured by the classified superpixels).

Our proposed framework connects deep learning and classical image processing using probabilistic graphical models. All the information was combined using a CRF graphical model, which have been widely applied in image analysis for pathological images, yet mainly for semantic segmentation (16, 17, 37, 38). Here, (1) we introduce a new way to capture high-level spatial context using superpixels, (2) propose a new CRF model that introduces dependences over space and across different spatial scales, thereby modeling multiple cells and their associated superpixels simultaneously for more accurate classification, (3) introduce the concept of context-specific CRF modeling, given that the strength of dependence can be variable according to tumor compartments. There is an increasing interest in combining deep learning with different strategies, or "umbrella approaches," such as the use of traditional machine learning to spatially explicit context used in this study, with the aim to, not only refine and improve the overall existing deep learning network (17, 39–41), but also facilitate biological interpretation compared to the "black-box"-like approach of deep-learning-only methods. However, optimizing and inventing new and refined deep learning networks is of equal importance, as during experimentation we observed that the better we made our single-cell classifier baseline, the more effective our SuperCRF approach became.

We also showed that combining cell classification with the global context given by the region classification (both inherent parts of the SuperCRF architecture) can open new avenues to study the cancer microenvironment from histopathological slides. For example, the spectacular response observed in clinical trials of immunotherapy in patients with incurable melanoma calls for a better understanding of the tumor microenvironment and in particular the cancer-immune-stroma interface. Here, our approach and its ability to look at lymphocytes within their cellular and global context can predict melanoma patient survival and potentially provide biomarker stratification for immunotherapeutic approaches, by identifying the three main types of tumor immunophenotypes including (i) inflamed tumors which are characterized by infiltrated T Cells within the tumor, and associated with a generally good prognosis (ii) immune-excluded tumors, in which T cells are present but prevented to infiltrate the tumor due to stromal interaction, and associated with worse prognosis (and obviously (iii) immune desert tumors). This could also potentially be extended to provide quantitative biomarkers to characterize the immune infiltrating response to immunotherapy. We also demonstrated that in accordance with the immune-excluded phenotype, tumors rich in stromal cells had a marked poorer prognosis in patients with melanoma. With p-value lower by two orders of magnitude, our method provide stronger predictive power than by using deep-learning only method for cell classification.

In the future, we plan to extend our framework and include an upward optimization step for the superpixels which may include additional classes for cells, regions and structures in order to provide a complete characterization of the tumor microenvironment. This may include deriving further classes from higher resolution images as we did for lymphocyte clusters in this study which were difficult to visualize in

$1.25\times$ resolution images. Incorporating additional deep learning methods should also be explored to perfect the classification of superpixels, for example by incorporating features extracted from a DCNN or a deep autoencoder, or to provide a potential alternative to superpixels, which may not be appropriate for the characterization of complicated structures, such as glands (42).

The primary aim of this study was to demonstrate proof-of-principle that the introduction of global and local context as cell dependencies using a probabilistic graphical model as a post-processing step, like an "umbrella," can significantly improve the performance of deep learning or classical machine learning cell classifiers based only on cell-morphology and abstract local context information. We chose the SC-CNN architecture as our primary cell classification step due to its state-of-the-art performance in cell detection and classification compared to other well-established deep learning and classical machine learning approaches (6). Alternatively, other promising deep learning networks could potentially be used including Inception v3 (43), Inception v4 (44), or a VGG architecture (45).

Overall, our vision is to establish a network which will provide a complete characterization of every component of the tumor microenvironment where all the parts will interact with each other like an ecological landscape. Such system has immense potential and can be virtually transferred to any cancer type, to provide a better understanding of the cancer-immune cell interface, cell-stroma interactions, and predictive biomarkers of response to novel therapies, including immunotherapy, which has radically changed melanoma patient survival.

CONCLUSION

The novel general framework SuperCRF improves cell classification by introducing global and local context-based information much like pathologists do. SuperCRF can be implemented in combination with any single-cell classifier and represent valuable tools to study the cancer-stroma-immune interface, which we used to identify predictors of survival in melanoma patients from conventional H&E stained histopathology.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga.

AUTHOR CONTRIBUTIONS

KZ-P, HF, SR, IR, YJ, and YY: substantial contributions to the conception or design of the work; or the method development, analysis, or interpretation of data for the work; drafting the work or revising it critically for important intellectual content; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2019.01045/full#supplementary-material

REFERENCES

1. Gurcan MN, Boucheron L, Can A, Madabhushi A, Rajpoot N, Yener B. Histopathological image analysis: a review. *IEEE Rev. Biomed. Eng.* (2009) 2:147. doi: 10.1109/RBME.2009.2034865

2. Kothari S, Phan JH, Stokes TH, Wang MD. Pathology imaging informatics for quantitative analysis of whole-slide images. *J Am Med Inform Assoc.* (2013) 20:1099–108. doi: 10.1136/amiajnl-2012-001540

3. Jones TR, Kang IH, Wheeler DB, Lindquist RA, Papallo A, Sabatini DM, et al. CellProfiler analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinf.* 9:482. doi: 10.1186/1471-2105-9-482

4. Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin SF, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci. Trans. Med.* (2012) 4:157ra143. doi: 10.1126/scitranslmed.3004330

5. Chen CL, Mahjoubfar A, Tai L-C, Blaby IK, Huang A, Niazi KR, et al. Deep learning in label-free cell classification. *Sci Rep.* (2016) 6:21471. doi: 10.1038/srep21471

6. Sirinukunwattana K, Raza SEA, Tsang Y-W, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imag.* (2016) 35:1196–206. doi: 10.1109/TMI.2016.2525803

7. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep.* 7:16878. doi: 10.1038/s41598-017-17204-5

8. Khoshdeli M, Cong R, Parvin B. Detection of nuclei in H&E stained sections using convolutional neural networks. In: *Biomedical and Health Informatics (BHI), 2017 IEEE EMBS International Conference on.* Orlando, FL: IEEE (2017). p. 105–8.

9. Piccinini F, Balassa T, Szkalisity A, Molnar C, Paavolainen L, Kujala K, et al. Advanced cell classifier: user-friendly machine-learning-based software for discovering phenotypes in high-content imaging data. *Cell Syst.* (2017) 4:651–5.e655. doi: 10.1016/j.cels.2017.05.012

10. Held M, Schmitz MH, Fischer B, Walter T, Neumann B, Olma MH, et al. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. *Nat Method.* (2010) 7:747–54. doi: 10.1038/nmeth.1486

11. Zhong Q, Busetto AG, Fededa JP, Buhmann JM, Gerlich DW. Unsupervised modeling of cell morphology dynamics for time-lapse microscopy. *Nat Methods.* (2012) 9:711–3. doi: 10.1038/nmeth.2046

12. Failmezger H, Frohlich H, Tresch A. Unsupervised automated high throughput phenotyping of RNAi time-lapse movies. *BMC Bioinf.* (2013) 14:292. doi: 10.1186/1471-2105-14-292

13. Failmezger H, Praveen P, Tresch A, Frohlich H. Learning gene network structure from time laps cell imaging in RNAi Knock downs. *Bioinformatics.* (2013) 29:1534–40. doi: 10.1093/bioinformatics/btt179

14. Niederberger T, Failmezger H, Uskat D, Poron D, Glauche I, Scherf N, et al. Factor graph analysis of live cell-imaging data reveals mechanisms of cell fate decisions. *Bioinformatics.* (2015) 31:1816–23. doi: 10.1093/bioinformatics/btv040

15. Failmezger H, Dursun E, Dumcke S, Endele M, Poron D, Schroeder T, et al. Clustering of samples with a tree-shaped dependence structure, with an application to microscopic time lapse imaging. *Bioinformatics.* (2018). doi: 10.1093/bioinformatics/bty939

16. Karimaghaloo Z, Arnold DL, Arbel T. Adaptive multi-level conditional random fields for detection and segmentation of small enhanced pathology in medical images. *Med Image Anal.* (2016) 27:17–30. doi: 10.1016/j.media.2015.06.004

17. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Machine Intell.* (2018) 40:834–48. doi: 10.1109/TPAMI.2017.2699184

18. Li Y, Ping W. Cancer metastasis detection with neural conditional random field. In: *1st Conference on Medical Imaging With Deep Learning (MIDL).* Amsterdam (2018).

19. Zanjani FG, Zinger S. Cancer detection in histopathology whole-slide images using conditional random fields on deep embedded spaces. In: *Medical Imaging 2018: Digital Pathology.* Houston, TX: International Society for Optics and Photonics (2018). p. 105810I.

20. Rajapakse JC, Liu S. Staging tissues with conditional random fields. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* Boston, MA: IEEE (2011). p. 5128–31. doi: 10.1109/IEMBS.2011.6091270

21. Li C, Chen H, Xue D, Hu Z, Zhang L, He L, et al. Weakly supervised cervical histopathological image classification using multilayer hidden conditional random fields. In: *International Conference on Information Technologies in Biomedicine.* Kamień Śląski: Springer (2019). p. 209–21. doi: 10.1007/978-3-030-23762-2_19

22. Paramanandam M, O'Byrne M, Ghosh B, Mammen JJ, Manipadam MT, Thamburaj R, et al. Automated segmentation of nuclei in breast cancer histopathology images. *PLoS ONE.* 11:e0162053. doi: 10.1371/journal.pone.0162053

23. Zormpas-Petridis K, Failmezger H, Roxanis I, Blackledge M, Jamin Y, Yuan Y. Capturing global spatial context for accurate cell classification in skin cancer histology. *Comput Pathol Ophth Med Image Anal.* (2018) 11039:52–60. doi: 10.1007/978-3-030-00949-6_7

24. Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color transfer between images. *IEEE Comp Grap Appl.* (2001) 21:34–41. doi: 10.1109/38.946629

25. Khan AM, Rajpoot N, Treanor D, Magee D. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans Biomed Eng.* (2014) 61:1729–38. doi: 10.1109/TBME.2014.2303294

26. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell.* (2012) 34:2274–82. doi: 10.1109/TPAMI.2012.120

27. Haralick RM, Shanmugam K. Textural features for image classification. *IEEE Trans Syst Man Cyber.* (1973) 3:610–21. doi: 10.1109/TSMC.1973.4309314

28. Imran M, Hashim R, Khalid NEA. Segmentation-based fractal texture analysis and color layout descriptor for content based image retrieval. In: *Intelligent Systems Design and Applications (ISDA), 2014 14th International Conference on IEEE.* Okinawa (2014). p. 30–3.

29. Lafferty J, McCallum A, Pereira FC. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001).* Williamstown, MA: Morgan Kaufmann Publishers (2001). p. 282–9.

30. Schmidt M. *UGM: A Matlab Toolbox for Probabilistic Undirected Graphical Models.* (2007). Available online at: http://www.cs.ubc.ca/~schmidtm/Software/UGM.html

31. de Kruijf EM, van Nes JG, van de Velde CJ, Putter H, Smit VT, Liefers GJ, et al. 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 Treat.* (2011) 125:687–96. doi: 10.1007/s10549-010-0855-6

32. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013

33. Wu J, Liang C, Chen M, Su W. Association between tumor-stroma ratio and prognosis in solid tumor patients: a systematic review and meta-analysis. *Oncotarget.* (2016) 7:68954–65. doi: 10.18632/oncotarget.12135

34. Scheer R, Baidoshvili A, Zoidze S, Elferink MAG, Berkel AEM, Klaase JM, et al. Tumor-stroma ratio as prognostic factor for survival in rectal adenocarcinoma: a retrospective cohort study. *World J Gastroinest Oncol.* (2017) 9:466–74. doi: 10.4251/wjgo.v9.i12.466

35. Barnes TA, Amir E. HYPE or HOPE: the prognostic value of infiltrating immune cells in cancer. *Br J Cancer.* (2017) 117:451–60. doi: 10.1038/bjc.2017.220

36. Kim JY, Gatenby RA. Quantitative clinical imaging methods for monitoring intratumoral evolution. *Methods Mol Biol.* (2017) 1513:61–81. doi: 10.1007/978-1-4939-6539-7_6

37. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic image segmentation with deep convolutional nets and fully connected CRFS. In: *3rd International Conference on Learning Representations, ICLR 2015.* San Diego, CA (2014).

38. Kokkinos I. Ubernet: training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Honolulu, HI (2017). p. 6129–38.

39. Arnab A, Zheng S, Jayasumana S, Romera-Paredes B, Larsson M, Kirillov A, et al. Conditional random fields meet deep neural networks for semantic segmentation. *IEEE Signal Process Mag.* 35:37–52. doi: 10.1109/MSP.2017.2762355

40. Qin Y, Kamnitsas K, Ancha S, Nanavati J, Cottrell G, Criminisi A, et al. Autofocus layer for semantic segmentation. *arXiv:1805.08403* (2018). doi: 10.1007/978-3-030-00931-1_69

41. Roy AG, Navab N, Wachinger C. Concurrent spatial and channel 'Squeeze & Excitation' in fully convolutional networks. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, editors. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018. Lecture Notes in Computer Science, Vol 11070.* Cham: Springer (2018). doi: 10.1007/978-3-030-00928-1_48

42. Raza SEA, Cheung L, Shaban M, Graham S, Epstein D, Pelengaris S, et al. Micro-Net: a unified model for segmentation of various objects in microscopy images. *Med. Image Anal.* (2019) 52:160–73. doi: 10.1016/j.media.2018.12.003

43. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2016). p. 2818–26.

44. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-First AAAI Conference on Artificial Intelligence.* San Francisco, CA (2017).

45. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *3rd International Conference on Learning Representations, ICLR 2015.* San Diego, CA (2014).