CHAPTER 8

Deep learning-based nuclei segmentation and classification in histopathology images with application to imaging genomics

Dimitris Metaxas*, Hui Qu*, Gregory Riedlinger†, Pengxiang Wu*, Qiaoying Huang*, Jingru Yi*, and Subhajyoti De†
*Rutgers University, Department of Computer Science, Piscataway, NJ, United States
†Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, United States

Contents

1. Joint nuclei segmentation and classification in histopathology images

Histopathological assessment remains a cornerstone of the clinical diagnosis and classification of cancer. The underlying tissue architectures in histopathological images provide a wealth of information about the nature of disease, cytogenetic abnormalities, and characteristics of the microenvironment [1]. For example, malignant tumor cells can be distinguished from benign cells by the features of their nuclei [1], and the extent of lymphocyte infiltration in the microenvironment often has prognostic significance [2]. Furthermore, phenotypic variations among tumor cells, which are indicative of intratumor heterogeneity, have consequences for treatment strategies for cancer patients [3]. Therefore, the development of algorithms for refined segmentation and classification of histopathological structures, such as lymphocytes and cancer nuclei, can help improve the clinical management of cancer.

Nuclei segmentation and classification are both challenging tasks. The size of nuclei is much smaller compared to that of glands or organs, and the nuclei are often close to each

Computer Vision for Microscopy Image Analysis
https://doi.org/10.1016/B978-0-12-814972-0.00008-4

other, making it hard to segment individual nuclei accurately. The fine-grained classification of nuclei is also difficult due to the large interclass and intraclass variances in nuclear shapes and textures. Traditional methods like thresholding, watershed, clustering, and region growing [4] are not able to cope with these challenges well. Early learning-based methods learn to segment or classify nuclei using low-level handcrafted features such as color, texture, and gradients of geometric features [5–7], or learn kernel and hashing functions for classification based on handcrafted features [8–10], which have limited representation capability.

Recently, deep convolutional neural networks (CNNs) have achieved great success for image classification and segmentation [11–14]. Furthermore, many deep learning-based methods have been proposed for histopathology image analysis [15, 16], such as metastasis detection [17, 18], invasive cancer localization [19], nuclei segmentation [15, 20–22], and nuclei classification [23, 24]. Xing et al. [20] utilized CNNs to obtain an initial shape of nuclei and then to separate individual nuclei using a deformable model. In Refs. [15, 21], nuclei segmentation is performed by classifying the pixels into classes using a patch around each pixel as the input to an image classification network. The computational cost is large because each patch predicts only 1 pixel. Fully convolutional neural networks (FCNs) [13, 14], which directly output the same size of segmentation map as the input image, are more efficient and effective for image segmentation tasks and have been used in nuclei segmentation [22]. Compared to nuclei segmentation, there are fewer works about nuclei fine-grained classification using deep learning. Sirinukunwattana et al. [23] built two CNNs to detect nuclei and then classified them into subcategories. Zhou et al. [24] proposed a sibling CNN with objectiveness prior to detect and classify nuclei simultaneously.

Although the current methods have achieved good accuracy, they focus on segmentation or classification. They are not able to produce the pixelwise masks of different types of nuclei at the same time, and thus they cannot generate both nucleus features and spatial distributions, which are important for histopathology image analysis. Actually, the network structures for two tasks are similar; both need to extract feature representations from the input image.

In this chapter, we propose a framework to solve these two tasks jointly. As opposed to previous methods, our model outputs the segmentation map for every type of nuclei and the background, which can segment individual nuclei, as well as classifying them into tumor, lymphocyte, and stroma subcategories. In addition, we use one more channel to predict the contours of nuclei, aiming at separating any touching nuclei. To improve the segmentation accuracy further, we take advantage of the perceptual loss [25] that can measure small differences in two images. In addition, transfer learning is utilized to promote the training due to the small size of the annotated dataset.

1.1 Training data generation

1.1.1 Dataset

To evaluate the performance of joint nuclei segmentation and classification algorithms, we annotated a dataset that consists of 40 H&E stained tissue images from eight lung adenocarcinoma or lung squamous cell carcinoma cases, with each case having five images about 900×900 in size. There are around 24,000 annotated nuclei in the dataset, and each nucleus is marked as one of the following three types: a tumor nucleus, a lymphocyte nucleus, and a stroma (e.g., fibroblasts, macrophages, neutrophils, and endothelial cells) nucleus. For each image, we use one labeled image to encode the segmentation mask and classify the class information of each nucleus. In a ground-truth label image, pixels of value 0 are in the background. Pixels that have the same positive integer belong to an individual nucleus. Therefore, each nucleus has a unique ID, as shown in Fig. 8.1B. The integer value ID also indicates the class of the nucleus, as follows (where *mod* is the modular operation):

- Tumor nucleus: $mod(id, 3) = 0$
- Lymphocyte nucleus: $mod(id, 3) = 1$
- Stroma nucleus: $mod(id, 3) = 2$

It is easy to extract the class information from this encoding. Fig. 8.1 shows an example of an original image and its labels.

In many machine learning tasks, a dataset consists of three parts: a training set that is used to train the model, a validation set for choosing the best model during training, and a test set that aims to evaluate the performance of the trained model. Following this rule, this dataset is split into three parts: 24 images for training, 8 for validation, and the remaining 8 for testing. The training set contains three images for each case, and the other two sets have one image for each case.

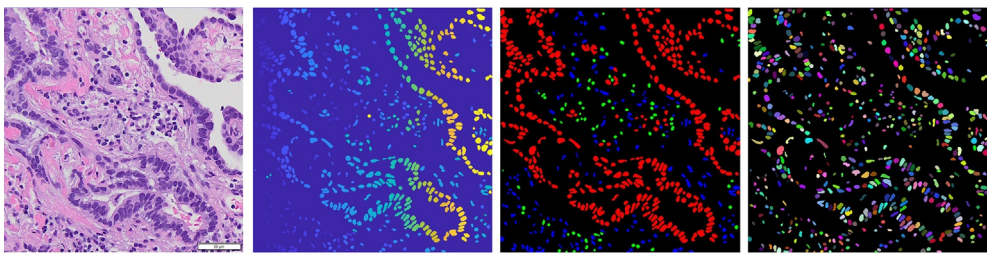

Fig. 8.1 Example of an image and its labels. (A) Original image; (B) ground-truth label; (C) classification label, red, green, and blue (dark gray, gray, and light gray) represent tumor, lymphocyte, and stroma nuclei, respectively); (D) segmentation label (distinct colors indicate different nuclei).

1.1.2 Preprocessing

As the tissue images come from different patients, there is a large variation in image colors, which has a negative effect on the segmentation results. Therefore, we use the color transfer method [26] to eliminate the color variation of these images. This method transforms the color of an image to a similar color as the referenced image in a three-step process:

1. Transform both images from the RGB color space to the $l\alpha\beta$ color space, aiming at removing the correlation of different color axes.
2. Adjust the mean and standard deviation of source image according to the referenced image.
3. Transform the corrected image from the $l\alpha\beta$ color space to the RGB color space.

This method is fast and effective to normalize all images in the dataset. Two images before and after color normalization are shown in Fig. 8.2.

Data augmentation is crucial for training deep neural networks when only a few training images are available, which is exactly the case in our task. For each large image in the training set, we extract 16 image patches with size 250×250 uniformly with overlap, resulting in 384 small image patches. For each patch, a 224×224 image is randomly cropped as the network input. Other augmentations include random scale, random horizontal and vertical flip, random affine transformation, random elastic transformation, random rotation, and normalization with mean and standard deviation by channel.

1.2 Deep learning-based segmentation and classification

1.2.1 Network structure

Our proposed framework is shown in Fig. 8.3. It consists of two parts: the prediction network, which generates the segmentation mask of each type of nuclei; and the perceptual loss network, which computes the perceptual loss between the predicted and ground-truth labels.

The prediction network is the routine encoder-decoder structure. We utilize the powerful representation ability of residual networks [12] to extract the features. The encoder is from ResNet34 [12], without the average pooling and fully connected layers,

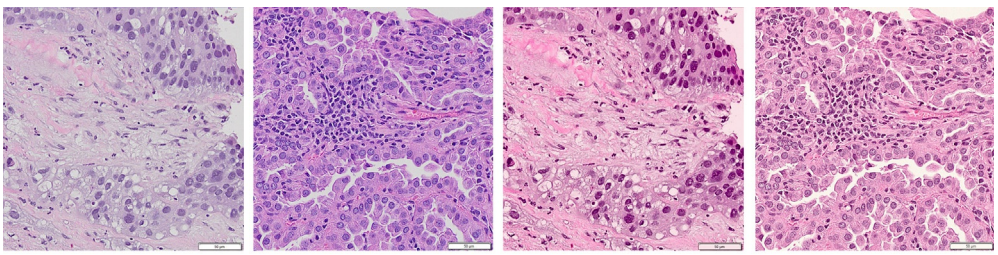

Fig. 8.2 Images before and after color normalization. (A) image 1, (B) image 2, (C) image 1 after color normalization, (D) image 2 after color normalization.

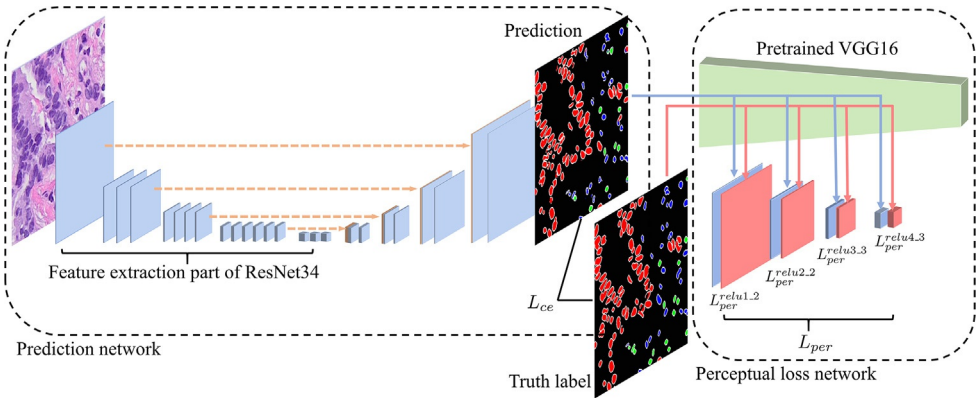

Fig. 8.3 System overview. Our framework consists of the prediction network and the perceptual loss network. The prediction network takes the feature extraction part of ResNet34 as the encoder and outputs the segmentation map of different types of nuclei. The loss network uses the fixed, pretrained VGG16 model as a feature extractor and computes the perceptual loss.

and is initialized with the pretrained parameters from image classification tasks. It extracts features of the input image layer by layer, from low level to high level. The decoder part recovers the resolution of feature maps and generates the segmentation results, with the help of the skip connections between the encoder and the decoder. This network outputs five probability maps: the background, the inner part of tumor nuclei, the inner part of lymphocyte nuclei, the inner part of stroma nuclei, and the contours of all nuclei. The contour map mainly aims to capture the contours of crowded and touching nuclei. As a result, the predicted inner parts of each nucleus are not connected, and we can obtain the instance segmentation of individual nuclei. The final nuclei mask is generated by a morphological dilation operation. In this way, we can obtain each individual nucleus without much extra effort.

The perceptual loss network is utilized to improve the segmentation accuracy of details in the image. It originates from Johnson et al. [25], in which the authors compute the loss between high-level features of the transformed image and the original image. The pretrained VGG16 model [27] is a feature extractor and is fixed during training and testing. Four levels of features are extracted using this network for the output of the prediction network and the ground-truth label (i.e., feature maps after the last ReLU layer of the first, second, third, and fourth blocks of the VGG16 model), denoted as *relu*1_2, *relu*2_2, *relu*3_3, and *relu*4_3. The mean squared loss is then computed between the feature sets of two inputs.

1.2.2 Loss function

The loss function of the method consists of two parts. The first part is the cross-entropy loss for five classes (i.e., background, inside tumor, inside lymphocyte, inside stroma, and contour). It is defined as

$$\mathcal{L}_{ce}(y, t, w) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{5} w_i t_i^{(m)} \log y_i^{(m)} \tag{8.1}$$

where N is the number of all the pixels, $y_i^{(m)}$ is the probability of pixel i belonging to class m, $t_i^{(m)} \in \{0, 1\}$ is the corresponding ground-truth label of class m, w_i is the optional weight for pixel i, and the default value is 1 for all pixels.

When training deep-learning models, a problem is the highly skewed frequencies of various classes in the dataset (i.e., the class imbalance problem). For example, in our dataset, the frequency of nuclear contour pixels is much less than that of noncontour pixels. One possible solution to this problem is to assign different values of w_i for different classes of pixels. We set the weight using a similar method in Ref. [14]:

$$w(x_i) = 1 + w_0 \cdot \exp\left(-\frac{(d_1(x_i) + d_2(x_i))^2}{2\sigma^2}\right) \tag{8.2}$$

where d_1 and d_2 are the distances to the nearest and the second-nearest nuclei, respectively. In the experiments, we set $\sigma = 5$ pixels and $w_0 = 10$. In this setting, pixels between close or touching nuclei are assigned much larger weights because those pixels are more important to split nuclei.

The second part is the perceptual loss. Let us denote the trained VGG16 model as a function f. The features after the ReLU layer of the kth block can be written as $f_k(x)$, where x is the input of VGG16. The size of kth-level features is denoted as $C_k \times H_k \times W_k$. The perceptual loss is

$$\begin{aligned} \mathcal{L}_{per}(\hat{y}, t) &= \frac{1}{4}\sum_{k=1}^{4} \mathcal{L}_{per}^{k}(\hat{y}, t) \\ \mathcal{L}_{per}^{k}(\hat{y}, t) &= \frac{1}{C_k H_k W_k} \|f_k(\hat{y}) - f_k(t)\|_2^2 \end{aligned} \tag{8.3}$$

where $\hat{y} = \arg\max y$ is the prediction map obtained from the output probability map y.

The loss function of the whole network is

$$\mathcal{L} = \mathcal{L}_{ce} + \beta \mathcal{L}_{per} \tag{8.4}$$

where β is a parameter that adjusts the weight of the perceptual loss and is set to 0.1 in the experiments.

1.2.3 Postprocessing

Because our model outputs the inside areas of each nucleus, we need postprocessing to get the final segmentation and classification results. The initial segmentation map is obtained by setting contour pixels as the background. Then we adopt several simple morphological operations (i.e., removal of small areas, connected component labeling, and

dilation with a disk filter) to generate the final results. The dilation operation aims to recover the whole mask of nuclei from the inside mask.

1.2.4 Evaluation methods

The performance of our method is evaluated using common metrics in segmentation and classification tasks. For nuclei segmentation, we use the F1-score to measure the detection accuracy. It is defined as the harmonic mean of precision and recall, as follows:

$$\mathrm{F1} = \frac{2 \times precision \times recall}{precision + recall} = \frac{2TP}{2TP + FP + FN} \tag{8.5}$$

where TP, FP, and FN are the number of true positives, false positives, and false negatives, respectively. A segmented nucleus is considered as a true positive if it overlaps with at least 50% of a ground-truth nucleus. Otherwise, it is a false positive. All ground-truth nuclei that have no corresponding segmented nuclei are treated as false negatives.

The object-level Dice coefficient [28], Jaccard index [29], and Hausdorff distance are used to measure the segmentation accuracy. The Dice coefficient and Jaccard index measure how well the ground-truth object G and predicted object S overlap with each other:

$$\mathrm{Dice}(G, S) = \frac{2|G \cap S|}{|G| + |S|}, \quad \mathrm{Jaccard}(G, S) = \frac{|G \cap S|}{|G \cup S|} \tag{8.6}$$

where $|\cdot|$ denotes set cardinality. The higher these values are, the better the segmented results overlap with the ground-truth objects. The Hausdorff distance is utilized to measure the shape similarity between G and S, and defined as

$$\mathrm{Haus}(G, S) = \max\left\{ \sup_{x \in G} \inf_{y \in S} d(x, y), \ \sup_{y \in S} \inf_{x \in G} d(x, y) \right\} \tag{8.7}$$

A lower Hausdorff distance indicates better shape similarity for two objects. Directly applying these metrics to the whole segmentation image results in pixel-level accuracies, which is not enough to represent the performance of instance-level segmentation. Therefore, we employ the object-level metrics defined in Ref. [30]:

$$M_{obj}(G, S) = \frac{1}{2}\left\{ \sum_{i=1}^{n_G} \gamma_i M(G_i, S_i) + \sum_{j=1}^{n_S} \sigma_j M\left(\widetilde{G}_j, \widetilde{S}_j\right) \right\} \tag{8.8}$$

$$\gamma_i = \frac{|G_i|}{\sum_{n=1}^{n_G} |G_n|}, \quad \sigma_j = \frac{|S_j|}{\sum_{n=1}^{n_G} |S_n|} \tag{8.9}$$

where M can be a Dice coefficient, Jaccard index, or Hausdorff distance; n_G and n_S are the number of objects in the ground-truth image and segmented image, respectively; G_i is a

ground-truth object; S_i is its corresponding true positive segmentation; $\widetilde{S}_j$ is a segmented object; and $\widetilde{G}_j$ is its true positive ground-truth. Here, true positive has the same meaning as that in F1-score. The object-level metrics measure not only how well the segmented objects overlap with ground-truth objects, but also how well the ground-truth objects overlap with segmented ones. The area of each object is also taken into consideration by applying the weights γ_i and σ_j.

For nuclei fine-grained classification, we only consider the accuracy among true-positive segmented nuclei because they have the true class labels. For those false, positively segmented nuclei, no ground-truth labels are available. The accuracy is not sufficient to represent the performance due to the various numbers of true positives. Therefore, we list the number of correctly classified nuclei for reference.

1.3 Experimental results

Here, we test the proposed method on the lung cancer dataset mentioned in Section 1.1 and compare it to two popular approaches for segmentation. One is the fully convolutional network proposed by Long et al. [13], which is the first FCN used for segmentation tasks. The other is U-Net [14], which has been widely used in medical image segmentation. Both networks output five probability maps as ours. For all models, we trained 300 epochs with the Adam optimizer. The learning rate, batch size, and weight decay are 0.0001, 8, and 0.0001, respectively.

The nuclei segmentation and classification results using FCN-8s, U-Net and our method are shown in Tables 8.1 and 8.2. It can be observed that all three models have achieved relatively good segmentation and fine-grained classification results, showing that our idea of combining the two tasks is feasible. Compared to FCN-8s and U-Net, our method makes improvements on the segmentation of all types of nuclei, especially lymphocytes. The improvements to subcategory nuclei are larger than those to all nuclei because the classification results also affect the subclass metrics (i.e., wrongly classified nuclei have no corresponding ground-truth ones, thus reducing the F1, Dice coefficient, and Jaccard values and increasing the Hausdorff distance). For nuclei classification, our method achieves the best accuracies except for lymphocytes. However, the number of correctly classified lymphocytes is 32% and 8% more than that of FCN-8s and U-Net, respectively. Therefore, our method also outperforms FCN-8s and U-Net on the fine-grained classification.

To illustrate the effects of transfer learning and the perceptual loss, we report the results for our model without perceptual loss or the pretrained weights of the encoder part. It is evident that both techniques can promote the performance of segmentation and classification. The results without transfer learning are worse than those without the perceptual loss, showing that transfer learning is more important than the perceptual loss for this small dataset. Without transfer learning, the training images are not sufficient to train the whole network well enough.

Table 8.1 Nuclei segmentation results of various types of nuclei on the test set using FCN-8s [13], U-Net [14], our method without perceptual loss (Ours w.o. $\mathcal{L}_{\text{per}}$), our method without transfer learning (Ours w.o. TL), and our method

Method		FCN-8s [13]	U-Net [14]	Ours w.o. $\mathcal{L}_{per}$	Ours w.o. TL	Ours
All	F1	0.8630	0.8735	0.8742	0.8652	**0.8859**
	Dice_{obj}	0.8418	0.8651	0.8695	0.8633	**0.8759**
	Jaccard_{obj}	0.7696	0.8092	0.8141	0.8041	**0.8205**
	Haus_{obj}	5.17	4.68	4.40	4.67	**4.14**
Tumor	F1	0.7775	0.8022	0.8059	0.7970	**0.8263**
	Dice_{obj}	0.8072	0.8313	0.8385	0.8311	**0.8459**
	Jaccard_{obj}	0.7417	0.7803	0.7877	0.7764	**0.7949**
	Haus_{obj}	8.52	7.58	6.93	7.43	**6.66**
Lymphocyte	F1	0.5274	0.6198	0.6204	0.6346	**0.6709**
	Dice_{obj}	0.5653	0.6220	0.6370	0.6257	**0.6677**
	Jaccard_{obj}	0.5219	0.5893	0.6039	0.5924	**0.6323**
	Haus_{obj}	41.52	36.99	28.13	31.75	**27.26**
Stroma	F1	0.5619	0.5928	0.6186	0.5850	**0.6223**
	Dice_{obj}	0.5281	0.5663	**0.5986**	0.5658	0.5889
	Jaccard_{obj}	0.4670	0.5150	**0.5458**	0.5125	0.5361
	Haus_{obj}	17.88	15.93	**14.36**	15.88	14.99

Note: Bold numbers indicates the best values.

Table 8.2 Nuclei fine-grained classification accuracies (%) of different types of nuclei on the test set using FCN-8s [13], U-Net [14], our method without perceptual loss (Ours w.o. $\mathcal{L}_{\text{per}}$), our method without transfer learning (Ours w.o. TL), and our method

Method	FCN-8s [13]	U-Net [14]	Ours w.o. $\mathcal{L}_{per}$	Ours w.o. TL	Ours
All	80.96 (3448)	83.00 (3653)	83.86 (3709)	83.19 (3618)	**84.75** (3735)
Tumor	85.14 (2103)	88.72 (2139)	89.56 (2144)	**90.66** (2059)	90.29 (2139)
Lymphocyte	**81.43** (421)	75.85 (515)	80.00 (488)	72.73 (544)	75.44 (556)
Stroma	72.64 (924)	76.20 (999)	75.90 (1077)	76.32 (1015)	**79.94** (1040)

Note: Bold numbers indicates the best value. The number of correctly classified nuclei is listed in parentheses for reference.

For the subcategory results of both tasks, the performance on tumor nuclei is the best because it is easier to distinguish. The shape and size of some lymphocyte and stroma nuclei are very similar, resulting in relatively lower segmentation and classification accuracies. Actually, the segmentation and classification of subcategory nuclei affect each other. High segmentation accuracy is beneficial for classification, and high classification accuracy reduces the number of unpaired segmented and ground-truth nuclei, which can increase the segmentation metrics. Some representative image results of segmentation and classification are shown in Fig. 8.4.

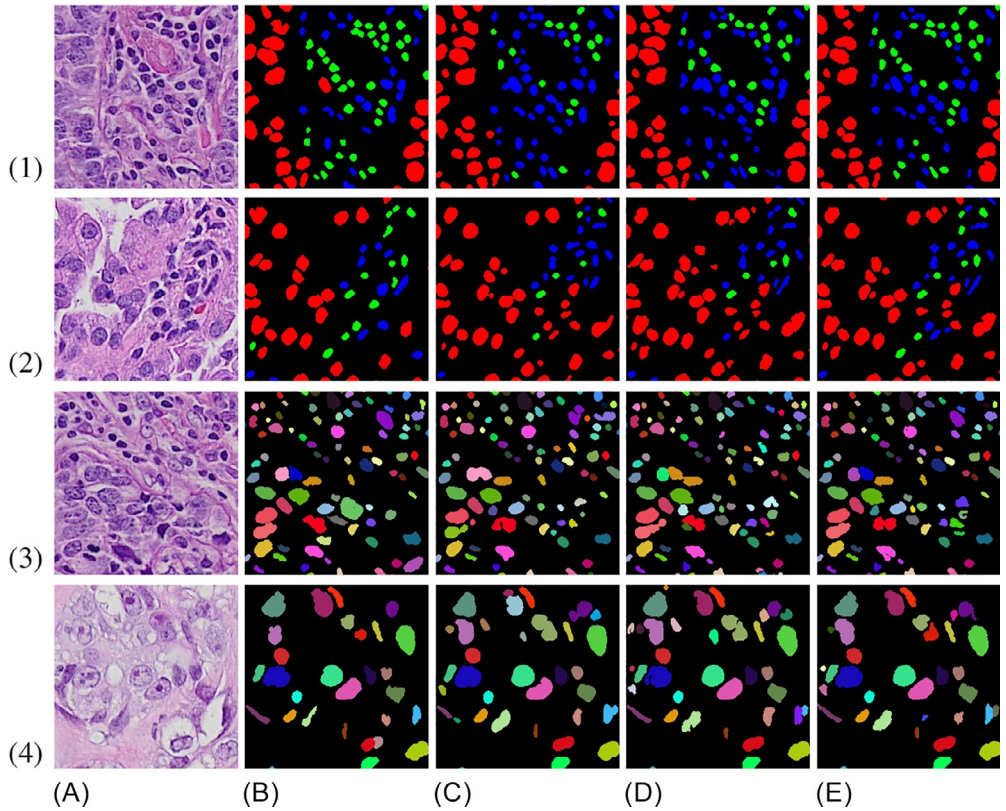

Fig. 8.4 Representative image results of FCN-8s, U-Net, and the proposed method. The top two rows (1) and (2) are the results of classification. Red, green, and blue (dark gray, gray, and light gray) represent tumor, lymphocyte, and stroma nuclei, respectively. The bottom two rows (3) and (4) are the results of instance-level segmentation. The different colors indicate individual nuclei. (A) Subimage, (B) true label, (C) FCN-8s, (D) U-net, and (E) Ours.

2. Applications to imaging genomics

Imaging genomics is an emerging field that explores the phenotype-genotype relationships (i.e., the relationships between imaging features of a disease and genomic features like genetic alternations, gene expression patterns, and other genome-related characteristics [31]). Imaging features can provide a comprehensive spatial view of the entire tumor, as well as the information on peritumoral regions [32]. Some features that seem irrelevant may have clinical significance. Genomic features, such as gene mutations, are at the molecular level and lead to the cause and development of cancer. However, it remains unclear how the genomic features affect many cancers because multiple gene mutations are often involved. Uncovering correlations between imaging and genetic features can

promote the understanding of some biologic mechanisms and pathways of gene expression and lead to finding more biomarkers that are predictive of clinical outcomes, which are beneficial for cancer diagnosis and treatment.

In imaging genomics, imaging features often come from modalities like computed tomography (CT) and magnetic resonance imaging (MRI). Many studies have tried to correlate such imaging features with genomic data [32]. However, features from histopathology images are important as well. The tissue structures from those images also contain underlying molecular profiles and are related to genetic alterations and gene expression patterns. Finding the genotype-phenotype correlations for histopathology images provide a better understanding of tumor biology and further improve the precision of clinical predictions [33]. The whole slide images (WSIs) obtained by digitalization of pathologic specimens often have high image quality and contain hundreds of millions of pixels; thus, the manual feature extraction is laborious and subjective, and may not be representative of the whole image. The automatic nuclei segmentation and classification method proposed in Section 1 can be applied here for imaging feature extraction. In this section, we introduce two possible applications for the extracted features in imaging genomics.

2.1 Intratumor heterogeneity

Intratumor heterogeneity (i.e., genetic, molecular, and phenotypic differences between tumor cells within a single tumor) is a major challenge for clinical management of cancer patients, contributing to therapeutic failure, disease relapses, and drug resistance. For example, in the work [34], the authors mention that a small part of subclones within chronic myeloid leukemia (CML) is resistant to the targeted drug tyrosine kinase inhibitors that should be effective, which may result from intratumor heterogeneity. The intratumor phenotypic heterogeneity has been observed since the earliest days of cancer biology [35], and recent findings suggest that there is extensive intratumor genetic heterogeneity in all major cancer types [36]. But it remains to be understood how the genetic heterogeneity relates to intratumor heterogeneity at the pathway and cell phenotype levels.

Based on the framework in Section 1, we are able to identify the accurate locations and shapes of tumor and nontumor nuclei in histopathological slides and then compute the spatial heterogeneity according to their locations, which can be associated with genetic heterogeneity.

2.1.1 Intratumor spatial heterogeneity

The term *spatial heterogeneity* refers to the spatial difference of tumor cells in a single tumor. Such a difference may lead to a situation in which a biopsy does not provide an adequate reflection of the phenotypic composition of the whole tumor [35]. With the locations of all nuclei in a WSI, we can use some measures in ecology, such as q-statistic [37] and diversity indices, to reflect the extent of local and regional heterogeneity.

The q-statistic measures the spatial stratified heterogeneity, which may imply the existence of distinct mechanisms in strata/areas (i.e., the various tumor subclones). It ranges from 0 to 1, where 0 means the spatial stratification of heterogeneity is not significant and 1 means high significance. There are two concepts when computing a q-value, unit, and stratum. A *unit* is the smallest block that contains tumor and/or nontumor cells. A *stratum* is a relatively large area that consists of a number of units. The split of strata decides which type of heterogeneity is revealed by the q-statistic. For example, three types of strata are illustrated in Fig. 8.5. The q-values computed from the patch, row, and column strata indicate how heterogeneous the tumor is on blocks, in the vertical and horizontal directions, respectively. The steps to compute a q-statistic of that region are:

1. Divide the tumor region into different strata.
2. Remove strata that are blank or contain a small part of tissue if the ratio of tissue patches in that stratum is less than a threshold.
3. For each remaining stratum h, compute the stratum variance σ_h^2 and the stratum size N_h.
4. Compute the variance and size of the whole population, σ^2, N.
5. Compute the q-statistic of each tumor region by

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} \tag{8.10}$$

After obtaining q-statistic values for all the tumor regions, the q-statistic of the whole slide image can be calculated by a weighted average of all regions' q-values according to the regions' areas.

There are several diversity indices that can be used for spatial heterogeneity. The Simpson diversity index [38], introduced by Edward H. Simpson, measures the degree of concentration when individuals are classified into types [38]. It is defined as

$$\lambda = \sum_{i=1}^{R} p_i^2 \tag{8.11}$$

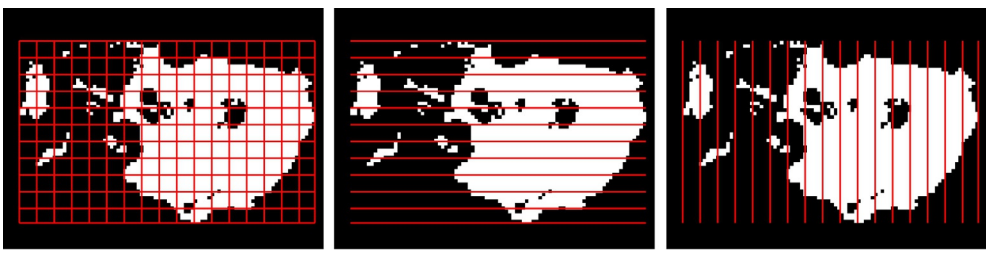

Fig. 8.5 Three types of strata. (A) Patch strata, (B) row strata, (C) column strata. White areas represent tissue regions of the whole slide image.

where R is the total number of species and p_i is the proportion of individuals belonging to the ith species. The measure equals the probability that two entities taken at random from the dataset of interest represent the same type [38]. The lower the value, the higher the diversity is. Another commonly used index is the Shannon index, proposed by Claude Shannon to measure the entropy in strings of text [39] and defined as

$$H' = -\sum_{i=1}^{R} p_i \ln p_i \tag{8.12}$$

The Shannon index measures the uncertainty in predicting the species identity of an individual that is taken at random from the dataset. A large Shannon index indicates high uncertainty in prediction, which means high diversity in the dataset; otherwise, it will be easy to predict the type of species. We can utilize these indices to compute the extent of spatial heterogeneity using the locations of tumor, lymphocyte, and stroma nuclei within a local region or for the whole image.

It is also possible to compute the q-statistic and diversity indices for various clones of tumor cells with the help of immuno-FISH and immunohistochemistry [40, 41], which can identify the known cancer gene mutation status in tumor cells on the slides, and those with similar cancer gene mutation status belong to the same subclone.

2.1.2 Intratumor genetic heterogeneity

During the formation of tumors, the driving genetic mutations are associated with the occurrence of many thousands of somatic genetic alterations [35], resulting in the genetic diversity of tumor cell populations. The clonal evolution in a branching manner may lead to clonal diversity as well [35], contributing to genetic heterogeneity within tumors.

We can use some metrics, such as clone numbers and sizes, to quantify the genetic heterogeneity. Clone numbers and sizes can be inferred using mutation and copy number data such as the following [36]:

1. Collect the somatic single-nucleotide variant (SNV) and copy number variant (CNV) data for the same tumors from TCGA.
2. Estimate the cellular prevalence of each SNV in consideration of CNVs and cluster the cellular prevalence into subpopulations using PyClone [42].
3. The inferred subpopulations are the clones, and the corresponding cellular prevalences are the clone sizes.

With measurements of both spatial and genetic heterogeneity, one can perform some analyses to explore the relationship between phenotype and genotype (e.g., the spatial heterogeneity scores and the number of clones). The relationships may help understand how genetic heterogeneity results in phenotype heterogeneity. Further, multiscale analysis integrating genetic, pathway, and phenotypic heterogeneity will provide

fundamental insights into "functional" variability within and across cancers, helping to refine precise approaches to improve the clinical management of cancer patients.

2.2 Tumor-Infiltrating Lymphocytes

Tumor-infiltrating lymphocytes (TILs) are a type of immune cells that have moved from the bloodstream to a tumor. Because lymphocytes can kill tumor cells, researchers have explored the relationship between TILs and clinical outcomes and have found that the presence of high TIL density is associated with better clinical outcomes [43]. In addition, the spatial statistics of TILs correlate with cancer diagnosis and prognosis [44, 45]. Much effort have been exerted to correlate TILs with clinical outcomes, but the relationships between TILs and genetic features have not been well studied. Kochi et al. [46] have shown that genomic markers are highly associated with TIL levels in breast cancer, and TIL-associated genomic signatures can predict chemotherapy responses in several breast cancer subtypes. Therefore, it is meaningful to explore further the correlations between TILs and genomic features in various types of cancers.

The nuclei segmentation and classification method can automatically extract the pixels belonging to lymphocytes in a whole slide image with relatively high accuracy. As a result, we can efficiently compute accurate TIL density/scores instead of evaluating a rough level of TILs manually in most studies. In addition, we are able to obtain the spatial maps of TILs, which can be used to generate local and global spatial structures of lymphocytes or perform hot spot analysis like Ref. [2]. Integrated with genomic features and clinical data, these imaging features can be utilized to find possible genomic biomarkers potentially informative of novel therapeutic strategies. For example, if some imaging features about TILs are favorable for survival, genes that are associated with these features may have similar effects. The integrative analysis also may help us understand the basic biological principles in gene expressions by finding the TIL and genomics features that both have similar effects on clinical outcomes.

3. Conclusion

In this chapter, we proposed a framework that jointly segments and classifies the various types of nuclei from histopathology images. The cross-entropy and perceptual losses are combined to enhance the segmentation of details in the image. We also use transfer learning to better train the model on a small dataset. Experiments show that our method is able to achieve good segmentation and fine-grained classification results simultaneously. We then briefly introduced how to apply the framework to imaging genomics. The segmentation maps of the various types of nuclei generated by our method can be used to analyze the nuclear features and their spatial distributions. These imaging features, integrated with genetic features and clinical outcomes, can be used in analyses of intratumor heterogeneity and tumor-infiltrating lymphocytes to achieve a better understanding of the genotype-phenotype relationship, as well as improving the clinical management of cancer.

References

[1] M.N. Gurcan, L. Boucheron, A. Can, A. Madabhushi, N. Rajpoot, B. Yener, Histopathological image analysis: a review, IEEE Rev. Biomed. Eng. 2 (2009) 147.

[2] S. Nawaz, A. Heindl, K. Koelble, Y. Yuan, Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer, Mod. Pathol. 28 (6) (2015) 766.

[3] S. Xx, Yu Q, Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment, Acta Pharmacol. Sin. 36 (10) (2015) 1219.

[4] H. Irshad, A. Veillard, L. Roux, D. Racoceanu, Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential, IEEE Rev. Biomed. Eng. 7 (2014) 97–114.

[5] M. Wang, X. Zhou, F. Li, J. Huckins, R.W. King, S.T. Wong, Novel cell segmentation and online svm for cell cycle phase identification in automated microscopy, Bioinformatics 24 (1) (2007) 94–101.

[6] H. Kong, M. Gurcan, K. Belkacem-Boussaid, Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting, IEEE Trans. Med. Imaging 30 (9) (2011) 1661–1677.

[7] Y. Al-Kofahi, W. Lassoued, W. Lee, B. Roysam, Improved automatic detection and segmentation of cell nuclei in histopathology images, IEEE Trans. Biomed. Eng. 57 (4) (2010) 841–852.

[8] X. Zhang, F. Xing, H. Su, L. Yang, S. Zhang, High-throughput histopathological image analysis via robust cell segmentation and hashing, Med. Image Anal. 26 (1) (2015) 306–315.

[9] X. Zhang, H. Su, L. Yang, S. Zhang, Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5361–5368.

[10] M. Jiang, S. Zhang, J. Huang, L. Yang, D.N. Metaxas, Scalable histopathological image analysis via supervised hashing with multiple features, Med. Image Anal. 34 (2016) 3–12.

[11] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[13] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[14] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.

[15] A. Janowczyk, A. Madabhushi, Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases, J. Pathol. Inform. 7 (2016).

[16] Z. Li, X. Zhang, H. Müller, S. Zhang, Large-scale retrieval for medical image analytics: A comprehensive review, Med. Image Anal. 43 (2018) 66–84.

[17] Y. Liu, K. Gadepalli, M. Norouzi, G.E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P.Q. Nelson, G.S. Corrado, et al., Detecting cancer metastases on gigapixel pathology images, *arXiv preprint arXiv*:1703.02442, 2017.

[18] B. Kong, X. Wang, Z. Li, Q. Song, S. Zhang, Cancer metastasis detection via spatially structured deep network, in: *International Conference on Information Processing in Medical Imaging*, Springer, 2017, pp. 236–248.

[19] B. Kong, S. Sun, X. Wang, Q. Song, S. Zhang, Invasive cancer detection utilizing compressed convolutional neural network and transfer learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 156–164.

[20] F. Xing, Y. Xie, L. Yang, An automatic learning-based framework for robust nucleus segmentation, IEEE Trans. Med. Imaging 35 (2) (2016) 550–566.

[21] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, A. Sethi, A dataset and a technique for generalized nuclear segmentation for computational pathology, IEEE Trans. Med. Imaging 36 (7) (2017) 1550–1560.

[22] P. Naylor, M. Laé, F. Reyal, T. Walter, Nuclei segmentation in histopathology images using deep neural networks, in: *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, IEEE, 2017, pp. 933–936.
[23] K. Sirinukunwattana, S.E.A. Raza, Y.W. Tsang, D.R. Snead, I.A. Cree, N.M. Rajpoot, Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images, IEEE Trans. Med. Imaging 35 (5) (2016) 1196–1206.
[24] Y. Zhou, Q. Dou, H. Chen, J. Qin, P.A. Heng, Sfcn-opi: Detection and fine-grained classification of nuclei using sibling fcn with objectness prior interaction, in: *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
[25] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: *European Conference on Computer Vision*, Springer, 2016, pp. 694–711.
[26] E. Reinhard, M. Adhikhmin, B. Gooch, P. Shirley, Color transfer between images, IEEE Comput. Graph. Appl. 21 (5) (2001) 34–41.
[27] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv*:1409.1556, 2014.
[28] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302.
[29] P. Jaccard, The distribution of the flora in the alpine zone, New Phytol. 11 (2) (1912) 37–50.
[30] K. Sirinukunwattana, D.R. Snead, N.M. Rajpoot, A stochastic polygons model for glandular structures in colon histology images, IEEE Trans. Med. Imaging 34 (11) (2015) 2366–2378.
[31] M.A. Mazurowski, Radiogenomics: what it is and why it is important, J. Am. Coll. Radiol. 12 (8) (2015) 862–866.
[32] H.X. Bai, A.M. Lee, L. Yang, P. Zhang, C. Davatzikos, J.M. Maris, S.J. Diskin, Imaging genomics in cancer research: limitations and promises, Br. J. Radiol. 89 (1061) (2016) 20151030.
[33] L.A. Cooper, J. Kong, D.A. Gutman, W.D. Dunn, M. Nalisnik, D.J. Brat, Novel genotype-phenotype associations in human cancers enabled by advanced molecular platforms and computational analysis of whole slide images, Lab. Investig. 95 (4) (2015) 366.
[34] J. Liu, H. Dang, X.W. Wang, The significance of intertumor and intratumor heterogeneity in liver cancer, Exp. Mol. Med. 50 (1) (2018) e416.
[35] A. Marusyk, V. Almendro, K. Polyak, Intra-tumour heterogeneity: a looking glass for cancer? Nat. Rev. Cancer 12 (5) (2012) 323.
[36] N. Andor, T.A. Graham, M. Jansen, L.C. Xia, C.A. Aktipis, C. Petritsch, H.P. Ji, C.C. Maley, Pan-cancer analysis of the extent and consequences of intratumor heterogeneity, Nat. Med. 22 (1) (2016) 105.
[37] J.F. Wang, T.L. Zhang, B.J. Fu, A measure of spatial stratified heterogeneity, Ecol. Indic. 67 (2016) 250–256.
[38] E.H. Simpson, Measurement of diversity, Nature 163 (1949) 688.
[39] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (3) (1948) 379–423.
[40] F.C. Martins, S. De, V. Almendro, M. Gönen, S.Y. Park, J.L. Blum, W. Herlihy, G. Ethington, S.J. Schnitt, N. Tung, et al., Evolutionary pathways in brca1-associated breast tumors, Cancer Discovery (2012). CD–11.
[41] S.Y. Park, M. Gönen, H.J. Kim, F. Michor, K. Polyak, Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype, J. Clin. Invest. 120 (2) (2010) 636–644.
[42] A. Roth, J. Khattra, D. Yap, A. Wan, E. Laks, J. Biele, G. Ha, S. Aparicio, A. Bouchard-Côté, S.P. Shah, Py-clone: statistical inference of clonal population structure in cancer, Nat. Methods 11 (4) (2014) 396.
[43] B. Mlecnik, M. Tosolini, A. Kirilovsky, A. Berger, G. Bindea, T. Meatchi, P. Bruneval, Z. Trajanoski, W.H. Fridman, F. Pagès, et al., Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction, J. Clin. Oncol. 29 (6) (2011) 610–618.
[44] J. Galon, A. Costes, F. Sanchez-Cabo, A. Kirilovsky, B. Mlecnik, C. Lagorce-Pagès, M. Tosolini, M. Camus, A. Berger, P. Wind, et al., Type, density, and location of immune cells within human colorectal tumors predict clinical outcome, Science 313 (5795) (2006) 1960–1964.

[45] J. Saltz, R. Gupta, L. Hou, T. Kurc, P. Singh, V. Nguyen, D. Samaras, K.R. Shroyer, T. Zhao, R. Batiste, et al., Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images, Cell Rep. 23 (1) (2018) 181.
[46] M. Kochi, T. Iwamoto, N. Niikura, G. Bianchini, S. Masuda, T. Mizoo, T. Nogami, T. Shien, T. Motoki, N. Taira, et al., Tumour-infiltrating lymphocytes (tils)-related genomic signature predicts chemotherapy response in breast cancer, Breast Cancer Res. Treat. 167 (1) (2018) 39–47.