The latest project that I worked on is the unsupervised segmentation of H&E stained histological images. We use progressive superpixel refinement to learn how to group similar pixels together.

Our result outperforms several traditional unsupervised segmentation approaches like K-means, mean shift, minimum cut.

CS 838 Final Report: Unsupervised Segmentation of H&E Stained Histological Image Using Deep Clustering and Superpixel Refinement

Jimmy Chang
cchang253@wisc.edu

Zhiyu Ji
zji29@wisc.edu

Bin Li
bli346@wisc.edu

Abstract

We explore the opportunity of using deep pixel clustering for histology image segmentation as a tool for pathology studies. By investigating the state-of-art deep learning models and unsupervised deep clustering strategies, we design a fully unsupervised model for differentiating histology tissue matrices in the context of histopathological images. The training of the model is driven by superpixel refinement. Over-segmented superpixels are first computed for the dataset and pseudo labels are generated by ensuring the spatial continuity in the network output defined by each superpixel. The model alternates between pseudo label assignment and prediction. We applied the model to a histology data acquired in standard clinical practices as well as the BSDS500 dataset. Promising quantitative and qualitative results are obtained. The source code and demo of this work is available at https://github.com/binli123/cs838projectdemo.

1. Introduction

Histopathological examination of Hematoxylin and Eosin (H&E) stained tissue biopsies is the gold standard for oncology care. Histopathological diagnosis usually requires pathologists to study the tissue morphology with respect to tissue physiological context and microenvironment on microscopic images.

Attributing to the recent advances in computer sciences and machine learning, computer-aided diagnosis (CAD) has merged as a powerful tool for pathological studies [29, 27, 6]. Pathologists could potentially be freed of some labor-intensive work and focus on more high-level tasks and carry out faster and more accurate decisions [13, 42, 9, 36, 26]. Algorithms such as image segmentation and classification have opened new opportunities for automatic histopathology diagnosis [39, 24, 31]. Many of these methods require the model to be trained with millions of labeled images and the efficacy is thus suffer from several limitations: (i) the datasets are limited in size, due to the amount of expert labor required; (ii) the datasets hardly represent the very high variability introduced by different imaging conditions, staining processes, tissue types and annotation biases; (iii) difficulty of integration into clinical workflow due the dynamic and interactive nature of diagnosis procedure. However, we identify the opportunity that, major histological matrices in biopsies presents distinguishing image statistics such as color, compactness, gradients, textures, and representative features could potentially be learned using a deep model driven by unsupervised pixel-level clustering [35, 7, 23, 32, 5]. Such model trained with unlabeled data could potentially be used to differentiate histology content, or as part of semi-supervised workflow to address the challenges such as histopathology where the number of labelled data is limited or labelling is technically difficult.

We design a convolutional neural network (CNN) with dilated convolution [44, 10] as the deep feature extractor which is followed by a classifier to produce pixel-level dense predictions. Over segmented superpixels are generated by Simple Linear Iterative Clustering (SLIC) and the network is trained to ensure local spatial continuity in each class based on the SLIC results [1, 2, 22]. Several constrains are used to delineate the boundaries of output and prevent the model from giving trivial solutions. The output of the model is evaluated on a histology dataset collected from clinical practices by visual observation and quantitative comparisons against shallow unsupervised clustering methods. We also show the result of using the model in BSDS500 dataset [3].

2. Related work

Most of the unsupervised image segmentation method pursue pixel clustering using feature such as color, brightness, or other hand-crafted low-level features. These methods includes K-means clustering [11], graph-based method [16, 14], Mean Shift [12] and superpixel based methods [2]. Recently, deep neural networks have emerged as very powerful tools for segmentation tasks where the networks give pixel-level dense prediction on the input image [30, 15, 17, 18]. Some of the state-of-the-art structures include fully convolutional netowrks (FCNs) and dilated convolutions [28, 44, 10]. Since pixel-level annotations for image segmentation are difficult to obtain, semi-supervised approaches have been used where the networks are trained using mostly unlabelled data and very limited labelled data [45, 8, 33, 37, 38]. Several unsupervised segmentation or clustering using CNN has been proposed, such as embedded clustering [41], k-means based deep clustering [7], and soft normalized cut [40]. The work presented here is mostly related to framework proposed in [22] where the training is driven by superpixel refinement. [20] presented a superpixel generating network that produces superpixels based on deep feature clustering, but the feature extractor is still trained in a supervised manner. Here we explore the opportunity to of using CNN to perform pixel-level clustering in the scope of the whole dataset, such that optimal decision boundaries can be learned using the information from the whole dataset. This is especially the case for specialized dataset such as medical images and industrial images.

3. Method

3.1. Problem formulation

We first briefly outline the problem set in the section. Consider a deep feature extractor $\mathcal{F}$ parameterized by θ which is the parameter set of the network. The feature extractor takes an input image x and produces some features which are fed to a classifier Φ to produce a class-wise probability map. Each channel is a pixel-level probability map of assigning pixel to a certain class. The indices of maximum probability along the channels are taken as the pixel level labels to form the output c_f, as described as:

$$c_f = \arg\max \Phi(\mathcal{F}(x)) \tag{1}$$

Given a training set $\{x_i, c_i\}_{i=1}^N$ where $x_i \in X$ is a sample in the input image set and $c_i \in C$ is the corresponding sample in the label set, an optimal solution of θ can be found by optimizing the following problem:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(\mathcal{F}_\theta(x_i), c_i) \tag{2}$$

Where $\mathcal{L}$ is a loss function used to measure the difference between the network output c_f and label c_i. For supervised problems, the label set C is usually created manually and the network can be trained by directly comparing the output to the ground truth label.

However, for unsupervised scenarios, ground truth labels are not predefined and often need to be created during the training by incorporating some constrains and priors that explore the intrinsic structural characteristics of the dataset. We notice several unique properties in histology data. First, the staining technique such as hemotaxylin and eosin (H&E) used in histology renders different colors to two major populations in tissue sections, where cells and extracellular matrix (ECM) are stained with blue and pink, respectively. Second, features such as color, gradient, compactness, and power spectrum are intrinsically distinct in different components in tissue sections. Based on these observations, we present a progressive pseudo labelling strategy driven by superpixel refinement that takes into account both features of pixel and spatial continuity.

3.2. Progressive refinement pseudo labelling

Our model alternates between assigning pixel-level pseudo labels to the dataset and training the deep network to predict these labels. The model comprises two branches. First, over-segmented superpixels are generated for each input image using Simple Linear Iterative Clustering (SLIC) [1, 2] which groups pixels into a overly large number of classes (typically 5000) based on pixel value and spatial constrains. Second, the input image is also fed to the network and an output c_f containing pixel-level labels are generated. In the first iteration, the results depend on the priors of network architecture and network parameters which are initialized randomly. The corresponding output labels in c_f are converted to the most frequent labels within each individual region defined by the superpixel. The intuition is that, the over-segmented results at least indicates that the pixels in each superpixel should be predicted the same label. By doing this, an output c_f^* can be obtained which contains an agglomerated version of c_f. c_f^* is then assigned to the image as its pixel-level pseudo labels. The process is iterated through all images in the data set and pseudo labels are obtained for each image. The network is then trained to predict the pseudo labels by back-propagating the cross entropy loss between c_f^* and the channel wise probability map that produces c_f. The network output is progressively merged to produced more refined pseudo labels during this iterative procedure, as illustrated in Figure 1.

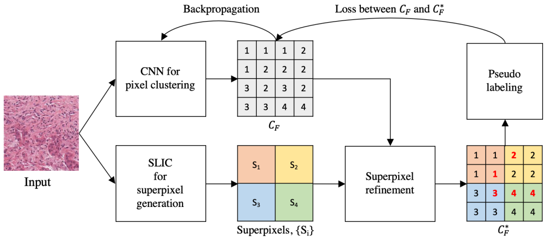

Figure 1. The flow of the network training strategy. Note that no hand-made labels are required in this process.

The network is trained for multiple epochs until certain stop criterion is met, such as loss per epoch or number of class per epoch is less than a threshold.

3.3. Network architecture

The network consists of three dilated convolution blocks with increasing receptive field size [44, 10]. Each dilated convolution block consists of a dilated convolution operation with kernel size of 3 × 3 and two normal convolution operations. Residual connection is created to connect the input and output which allow gradients to skip through the block [19], as shown in Figure 2.

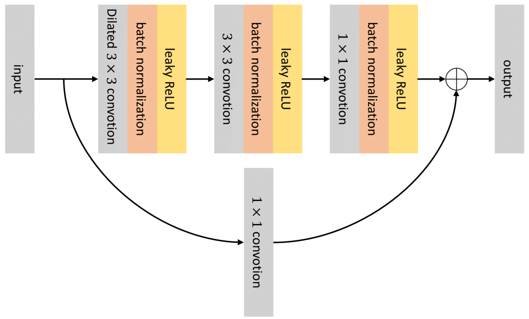

Figure 2. Dilated convolutional block

The feature maps of all dilated blocks are concatenated and fed to another convolutional layer to produce a pixel-level probability map with the same size of the input image. We empirically found that the extra 1 × 1 convolution layer following the concatenation is beneficial to reduce boundary effects. Dilated convolution has the advantage of preserving resolution of the features while keeping the size of the model relatively small. Plus, dilated convolutions also allow features to be extracted with different receptive field size as well as dense feature extraction [44, 10]. The network structure is shown in Figure 3.

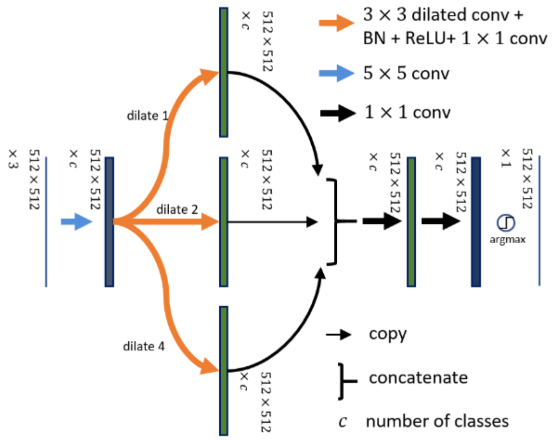

Figure 3. Structure of the network.

3.4. Avoiding trivial solutions

As pointed out in many unsupervised learning studies, the existence of trivial solutions is observed in any method that jointly learns a discriminative classifier and the labels, not specifically in the unsupervised training of neural networks [43]. Here we briefly discuss the main reasons that give rise to trivial solutions in the presented case and proposed three intuitive yet efficient workarounds.

During the training procedure, the number of class in the pseudo labels tends to decrease as the output is progressively merged, which in turn results in empty clusters and ultimately the network will collapse to too few classes. The extreme case will be the network learns a decision boundary that assign all of the inputs to a single cluster. Another reason is the unbalance in class sizes which leads to trivial parameterization of the model, such that the network always predicts the same labels regardless of input [43, 4, 21]. We present three strategies used to prevent the model from giving trivial solutions.

Autoencoder and decoder. Autoencoder is one of the most widely used structure in unsupervised deep learning. The encoder maps the the input image to a compact feature representation, which is taken by the decoder to reproduce the input. We implemented the auto-encoder structure by adding a Unet-like network as the decoder $\mathcal{G}$ which takes the pixel-level dense predictions produced by the encoder and reconstructs the input image by backpropagating the ℓ_1 loss between decoder output and the input image [34]. The parameters of the encoder is updated solely first for dense prediction, and then the parameters of both the encoder and the decoder are updated based on the reconstruction loss. This process is illustrated in Figure 4.

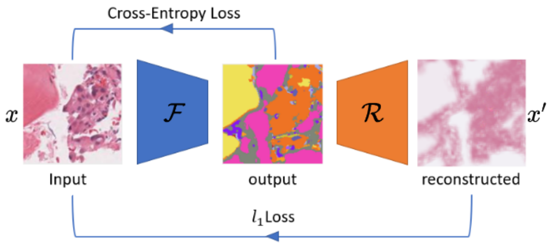

Figure 4. Autoencoder. The encoder produces segmentation output and decoder reconstructs the input image by taking the segmentation as input

The reconstruction from segmentation results to the original image is an ill-posed problem and gets more ill-posed as the information in the segmentation results decreases. Thus, the reconstruction poses a penalization on the content loss of the encoder, which in turn helps maintaining the number of class in the encoder output as well as helps delineate the boundaries between the segmentation result and the input image.

Splitting non-empty class. We implemented the strategy used in [7] to avoid empty clusters. Specifically, when one cluster becomes empty, we randomly select a cluster centroid from those non-empty clusters and then add some random perturbation on it to create the new centriod for the empty cluster. To make clusters more balanced, we select the largest cluster at that time as the reference of splitting.

2D instance normalization. A 2D instance normalization operation is performed on the channel-wise probability map before the the argmax operation which is useful to avoid trivial parameterization. For an arbitrary input image, the number of classes should be determined by the content of the input image and it is not guaranteed that class sizes are balanced in each image, nor in the training set. Thus, optimizing Eq. (2) leads to a trivial parametrization where the values in less populated channels will eventually become very small and the network will predict the same output regardless of the input. The 2D instance normalization operation centers the values of each class channel to have zero mean and unit variance, and thus, argmax can have an unbiased selection of maximum indices across channels.

3.5. Post-processing

The segmentation result can be further improved by using a fully connected Conditional Random Field (CRF) model to process the output of the trained model. CRF can delineate the boundary of the segmentation result to be more consistent with the original image and further remove isolated false segments. Details of CRF model and its deployment for post-processing can be found [25, 10].

4. Results

We evaluate the results of the presented model by both visual observation and quantitative evaluation metrics. To demonstrate the performance of our method, we compared the presented model to several other widely-used/recently-proposed unsupervised clustering algorithms. The algorithm outputs are evaluated against hand-labelled ground truth using Mean Intersection-Over-Union (mIOU).

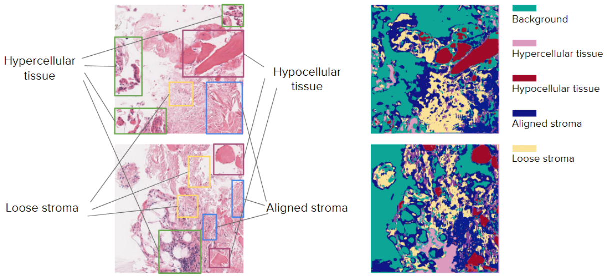

Figure 5. Representative results of the proposed method.

Figure 5 shows some representative results of our method. Five main categories of tissue components, namely hypocellular tissue, hypercellular tissue, aligned stroma, loose stroma, and background are separated in the output. Figure 6 shows the results of using CRF post-processing.

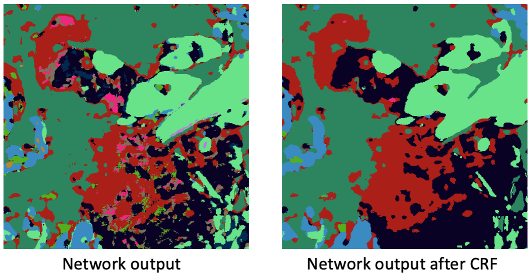

Figure 6. Post-processsing using CRF

Table 1 shows the comparison to four unsupervised segmentation methods: k-means, mean shift, min-cut, and a recently proposed superpixel refinement fuzzy c-means clustering method. Classes are matched by determine the highest overlap class with respect to the ground truth labels.

Overall, the presented model outperforms the traditional method, and the other superpixel refinement based method has comparable performance to our method. Plus, once trained, the presented method has much less running time compared to other methods.

To explore the generalizability of our approach, we also train the presented model on BSDS500 dataset for 750 epochs until the model converges where the number of classes remains 3. Some results are shown in Figure 7.

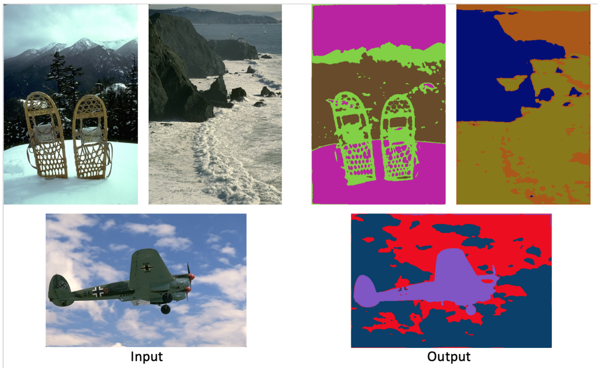

Figure 7. Some results from BSDS500 dataset

5. Conclusion, discussion and future plans

The keyways the our method differs from other unsupervised methods are: first, the feature extractor learned from the dataset can potentially outperform the hand-crafted features. Second, the deep model is trainable with respect to all images in the dataset. Though the training is driven by a heuristic loss with pseudo labels in each iteration, the boundaries decisions are learned in the scope of the whole dataset, such that similar pixels are grouped as the same class in each of the images. For specialized dataset such as histology data where there are only several major classes and the truth label of the output clusters can be identified once a look-up table is provided. Also we can further fine-tune the model using small amount of labelled data.

Method	Overall	Background	Loose stroma	Aligned stroma	Hypocellular	Hypercellular	Running time
Mean shift	0.365	0.829	0.382	0.436	0.064	0.116	<10 s
k-means (k=5)	0.392	0.836	0.384	0.352	0.353	0.055	<10 s
k-means (k=10)	0.385	0.759	0.228	0.265	0.251	0.425	<2 min
Minimum cut	0.207	0.386	0.061	0.058	0.091	0.442	>30 min
Superpixel	0.540	0.750	0.465	0.539	0.434	0.514	<10 s
Presented	0.608	0.817	0.508	0.472	0.564	0.679	<1 s

Table 1. Comparison of performance. Image size is 512 × 512

The learning of a deep network that jointly learns decision boundaries, labels, as well as the feature extractor is tricky. Although multiple workarounds are added to prevent the model from giving trivial solutions, the model still tends to collapse during the training process and the performance is highly sensitive to the trade-off between regularization weights and refinement weights. The model becomes more stable and performs better when the pixel class sizes are balanced in the dataset, which is usually not the case for natural image datasets. Model trained on these dataset such as BSDS500 dataset tends to collapse to only differentiate the objects from the background.

Future plan will be using trainable operation instead of argmax as the classifier by taking the channel-wise probabilities as feature vectors. We also plan to visualize the feature maps to see if high-level features are learned by the network. Following plans would be using our method as a self-supervised training strategy and further train the model with labelled data. We would like to explore the opportunity to address some real-world clinical application challenges where the number of labelled data is very limited or the labelling is technically hard.

6. Team Members and Contributions

Bin Li: network design, model implementation, method comparisons, dataset preparation.

Jimmy Chang: workflow design, training and evaluation implementation, model optimization, data loader implementation.

Zhiyu Ji: CRF post-processing implementation, network optimization, literature review.

References

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurélien Lucchi, Pascal Fua, and Sabine Süsstrunk. SLIC Superpixels, 2010.

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, Nov. 2012.

[3] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(5):898–916, May 2011.

[4] Francis R Bach and Zaïd Harchaoui. Diffrac: a discriminative and flexible framework for clustering. In *Advances in Neural Information Processing Systems*, pages 49–56, 2008.

[5] A. Basavanhally, S. Ganesan, M. Feldman, N. Shih, C. Mies, J. Tomaszewski, and A. Madabhushi. Multi-Field-of-View Framework for Distinguishing Tumor Grade in ER #x002b; Breast Cancer From Entire Histopathology Slides. *IEEE Transactions on Biomedical Engineering*, 60(8):2089–2099, Aug. 2013.

[6] N. Bayramoglu, J. Kannala, and J. Heikkilä. Deep learning for magnification independent breast cancer histopathology image classification. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 2440–2445, Dec. 2016.

[7] Mathilde Caron, Piotr Bojanowski, Armand Joulin, and Matthijs Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 132–149, 2018.

[8] Feng-Ju Chang, Yen-Yu Lin, and Kuang-Jui Hsu. Multiple structured-instance learning for semantic segmentation with uncertain training data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 360–367, 2014.

[9] Hao Chen, Xiaojuan Qi, Lequan Yu, Qi Dou, Jing Qin, and Pheng-Ann Heng. DCAN: Deep contour-aware networks for object instance segmentation from histology images. *Medical Image Analysis*, 36:135–146, Feb. 2017.

[10] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, Apr. 2018.

[11] Guy Barrett Coleman and Harry C Andrews. Image segmentation by clustering. *Proceedings of the IEEE*, 67(5):773–785, 1979.

[12] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (5):603–619, 2002.

[13] Nicolas Coudray, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyö, Andre L. Moreira, Narges Razavian, and Aristotelis Tsirigos. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature Medicine*, 24(10):1559–1567, Oct. 2018.

[14] Timothee Cour, Florence Benezit, and Jianbo Shi. Spectral segmentation with multiscale graph decomposition. In

2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 2, pages 1124–1131. IEEE, 2005.

[15] Clement Farabet, Camille Couprie, Laurent Najman, and Yann LeCun. Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1915–1929, 2012.

[16] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2):167–181, 2004.

[17] Bharath Hariharan, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. Simultaneous detection and segmentation. In *European Conference on Computer Vision*, pages 297–312. Springer, 2014.

[18] Bharath Hariharan, Pablo Arbeláez, Ross Girshick, and Jitendra Malik. Hypercolumns for object segmentation and fine-grained localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 447–456, 2015.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*, Dec. 2015. arXiv: 1512.03385.

[20] Varun Jampani, Deqing Sun, Ming-Yu Liu, Ming-Hsuan Yang, and Jan Kautz. Superpixel sampling networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 352–368, 2018.

[21] Armand Joulin and Francis Bach. A convex relaxation for weakly supervised classifiers. *arXiv preprint arXiv:1206.6413*, 2012.

[22] A. Kanezaki. Unsupervised Image Segmentation by Backpropagation. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1543–1547, Apr. 2018.

[23] Jakob Nikolas Kather, Cleo-Aron Weis, Francesco Bianconi, Susanne M. Melchers, Lothar R. Schad, Timo Gaiser, Alexander Marx, and Frank Gerrit Zöllner. Multi-class texture analysis in colorectal cancer histology. *Scientific Reports*, 6:27988, June 2016.

[24] Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim. Deep learning applications in medical image analysis. *Ieee Access*, 6:9375–9389, 2017.

[25] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in neural information processing systems*, pages 109–117, 2011.

[26] James S. Lewis, Sahirzeeshan Ali, Jingqin Luo, Wade L. Thorstad, and Anant Madabhushi. A quantitative histomorphometric classifier (QuHbIC) identifies aggressive versus indolent p16-positive oropharyngeal squamous cell carcinoma. *Am. J. Surg. Pathol.*, 38(1):128–137, Jan. 2014.

[27] Geert Litjens, Clara I. Sánchez, Nadya Timofeeva, Meyke Hermsen, Iris Nagtegaal, Iringo Kovacs, Christina Hulsbergen-van de Kaa, Peter Bult, Bram van Ginneken, and Jeroen van der Laak. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific Reports*, 6:26286, May 2016.

[28] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully Convolutional Networks for Semantic Segmentation. *arXiv:1411.4038 [cs]*, Nov. 2014. arXiv: 1411.4038.

[29] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis*, 33:170–175, Oct. 2016.

[30] Mohammadreza Mostajabi, Payman Yadollahpour, and Gregory Shakhnarovich. Feedforward semantic segmentation with zoom-out features. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3376–3385, 2015.

[31] Muhammad Khalid Khan Niazi, Anil V Parwani, and Metin N Gurcan. Digital pathology and artificial intelligence. *The Lancet Oncology*, 20(5):e253–e261, 2019.

[32] Fanny Orlhac, Benoit Thézé, Michaël Soussan, Raphaël Boisgard, and Irène Buvat. Multiscale Texture Analysis: From 18f-FDG PET Images to Histologic Images. *Journal of Nuclear Medicine*, 57(11):1823–1828, Nov. 2016.

[33] Deepak Pathak, Philipp Krahenbuhl, and Trevor Darrell. Constrained convolutional neural networks for weakly supervised segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1796–1804, 2015.

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Lecture Notes in Computer Science, pages 234–241. Springer International Publishing, 2015.

[35] C. T. Sari and C. Gunduz-Demir. Unsupervised Feature Extraction via Deep Learning for Histopathological Classification of Colon Tissue Images. *IEEE Transactions on Medical Imaging*, 38(5):1139–1149, May 2019.

[36] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, 19(1):221–248, 2017.

[37] Zhiyuan Shi, Yongxin Yang, Timothy M Hospedales, and Tao Xiang. Weakly-supervised image annotation and segmentation with objects and attributes. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2525–2538, 2016.

[38] Wataru Shimoda and Keiji Yanai. Distinct class-specific saliency maps for weakly supervised semantic segmentation. In *European Conference on Computer Vision*, pages 218–234. Springer, 2016.

[39] Eric J Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, 25(1):44–56, 2019.

[40] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *International MICCAI Brainlesion Workshop*, pages 178–190. Springer, 2017.

[41] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pages 478–487, 2016.

[42] Jun Xu, Xiaofei Luo, Guanhao Wang, Hannah Gilmore, and Anant Madabhushi. A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing*, 191:214–223, May 2016.

[43] Linli Xu, James Neufeld, Bryce Larson, and Dale Schuurmans. Maximum margin clustering. In *Advances in neural information processing systems*, pages 1537–1544, 2005.

[44] Fisher Yu and Vladlen Koltun. Multi-Scale Context Aggregation by Dilated Convolutions. *arXiv:1511.07122 [cs]*, Apr. 2016. arXiv: 1511.07122.

[45] Jun Zhu, Junhua Mao, and Alan L Yuille. Learning from weakly supervised data by the expectation loss svm (e-svm) algorithm. In *Advances in neural information processing systems*, pages 1125–1133, 2014.