

Texture-based Deep Learning for Effective Histopathological Cancer Image Classification

Nelson Zange Tsaku*, Sai Chandra Kosaraju[†], Tasmia Aqila[‡], Mohammad Masum[§],
Dae Hyun Song[¶], Ananda M. Mondal[‡], Hyun Min Koh^{¶||}, and Mingon Kang^{†||}

* Department of Computer Science, Kennesaw State University, Marietta, USA

[†] Department of Computer Science, University of Nevada, Las Vegas, Las Vegas, USA

[‡] School of Computing and Information Sciences, Florida International University, Miami, FL, USA

[§] Analytics and Data Science Institute, Kennesaw State University, Kennesaw, USA

[¶] Department of Pathology, Gyeongsang National University Changwon Hospital, Changwon, Republic of Korea

^{||} Correspondence: shekoh@hanmail.net, mingon.kang@unlv.edu

Abstract—Automatic histopathological Whole Slide Image (WSI) analysis for cancer classification has been highlighted along with the advancements in microscopic imaging techniques, since manual examination and diagnosis with WSIs are time- and cost-consuming. Recently, deep convolutional neural networks have succeeded in histopathological image analysis. However, despite the success of the development, there are still opportunities for further enhancements. In this paper, we propose a novel cancer texture-based deep neural network (CAT-Net) that learns scalable morphological features from histopathological WSIs. The innovation of CAT-Net is twofold: (1) capturing invariant spatial patterns by dilated convolutional layers and (2) improving predictive performance while reducing model complexity. Moreover, CAT-Net can provide discriminative morphological (texture) patterns formed on cancerous regions of histopathological images comparing to normal regions. We elucidated how our proposed method, CAT-Net, captures morphological patterns of interest in hierarchical levels in the model. The proposed method outperformed the current state-of-the-art benchmark methods on accuracy, precision, recall, and F1 score.

Index Terms—Whole Slide Images, Texture-based CNN

I. INTRODUCTION

The importance of histopathological image analysis for cancer classification has been dramatically increasing due to the advancements in microscopic imaging techniques that allow digitizing glass slides into Whole Slide Images (WSIs) [1]. Typically, medical specialists examine histopathological images from biopsies with various magnification factors, such as patterns, textures, and morphological characteristics of tissues that presents disease like cancer on WSIs. However, the manual examination and diagnosis with WSIs are time- and cost-consuming [2], [3].

However, although automatic histopathological image analysis using machine learning has strong potential for improving efficiency, there are still computational challenges to overcome. A number of studies have been proposed for automatically classifying cancer on histopathological images.

Recently, Convolutional Neural Networks (CNNs) have been widely explored in histopathological image analysis and have achieved significant improvement of performance on

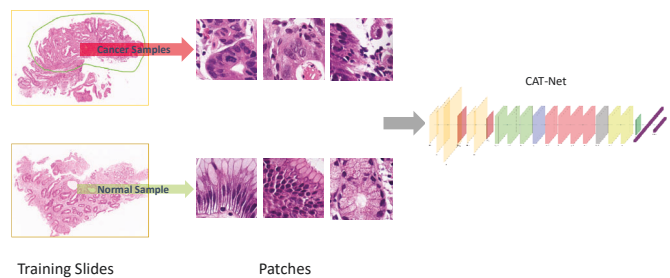


Fig. 1: Overall framework of cancer texture-based deep neural network

cancer classification [4], [5]. CNNs take an advantage of convolutional layers that can automatically train the optimal parametric kernels rather than using pre-defined features. Most CNNs have trained the models with patch images instead of directly introducing giga-pixel resolution WSIs [6]–[8]. Patch-based training significantly reduces model complexity. An Inception-based CNN model was proposed for detecting breast cancer metastasis in lymph nodes on gigapixel pathology images [6]. The deep neural network inputs a giga-pixel pathological image and classifies/localizes cancerous regions to assist pathologists. BiCNN was developed to classify breast cancer on histopathological images. BiCNN adopted a data augmentation method and a transfer learning technique to optimize model performance [9]. ImageNet was utilized to classify benign and malignant tumors in H&E-stained sections [10]. In the study, the CNN model was trained with patch images of two different sizes (32×32 and 64×64 pixels) extracted from WSIs and then classified cancer by combining the patches of WSIs using majority voting. A multiple instance learning-based CNN model was proposed to classify glioma subtypes with multi-gigapixel WSIs [11]. The model was trained by iteratively identifying discriminative patches in the WSIs. Furthermore, the model identified discriminative patches through Gaussian smoothing. These recent CNN approaches for tumor classification are all implemented with patch-based image extraction to avoid high computational cost

M. Kang and H. Koh are co-corresponding authors, and M. Kang and A. M. Mondal share senior authorship.

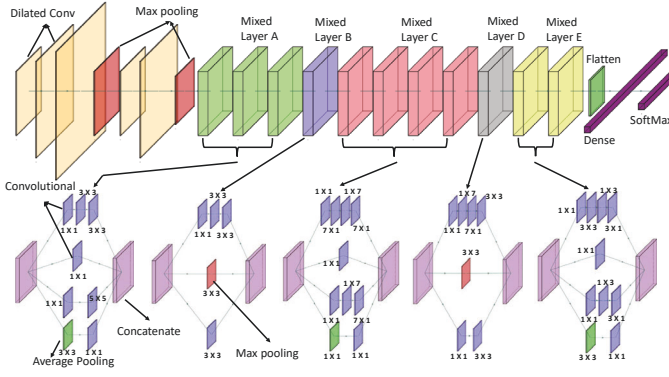


Fig. 2: The architecture of CAT-Net

and the limitations on preprocessing imposed by the nuclear-based approach.

In this study, we propose a novel Cancer Texture-based deep neural network model (CAT-Net). The innovations of CAT-Net are twofold: (1) capturing invariant spatial and morphological patterns by dilated convolutional layers and (2) improving predictive performance while reducing model complexity. CAT-Net can also provide discriminative texture patterns formed on cancerous regions of histopathological images comparing to normal regions.

II. METHODS

In this paper, we propose a cancer texture-based deep neural network (CAT-Net) that captures morphological patterns in cancer, while reducing model complexity comparing to existing state-of-the-art deep learning models such as Inception V3. We hypothesize that texture-based pattern analysis may improve histopathological cancer image classification, since morphological spatial patterns play an important role when a pathologist reads slides. CAT-Net trains the model with small-size patch images extracted by sliding a window on a given giga-pixel WSI and provides a score for cancer (see Fig. 1). CAT-Net is designed to effectively capture scalable morphological texture patterns from a WSI.

We leverage dilated layers to achieve two objectives in the model for the image texture analysis: (1) enlarging the field of view of filters to incorporate larger context and (2) employing varying values of dilation rate enabling object encoding at multiple scales. It offers a control of the field-of-view and finds the best trade-off between accurate localization (small field-of-view) and context assimilation (large field-of-view).

CAT-Net consists of 21 layers of dilated convolutional, max pooling, five types of mixed layers, fully connected, and softmax layers (see Fig. 2). In CAT-Net, eleven mixed layers are embedded in the architecture, inspired by the inception model [7]. The mixed layers, called Inception layers, are combinations of a set of conventional convolutional layers with various kernel sizes. The mixed layers allow internal layers that adaptively select the most appropriate filter size to a given image. CAT-Net enables spatial invariance with an dilation rate of 2×2 [12], [13].

To generalize our architecture for input texture and scale, CAT-Net includes five mixed layers with various texture scales: Mixed layer A, Mixed layer B, Mixed layer C, Mixed layer D and Mixed layer E in Fig. 2. The mixed layers share a similar structure but with different kernel sizes and numbers of convolution layers. Then, convolutional layers of various sizes with output filters are concatenated into a single output vector, forming the input of the next layer. Specifically, Mixed layer A contains three different filters of: 1×1 , 3×3 , and 5×5 filter sizes; Mixed layer B contains two filter sizes: 1×1 and 3×3 ; Mixed layer C contains three filter sizes: 1×1 , (1×7) , and 7×1 ; Mixed layer D contains 4 filters: 1×1 , 1×7 , 7×1 and 3×3 filters sizes; Mixed layer E filter contains 1×1 , 1×3 , 3×1 , and 3×3 filter sizes (see Fig. 2). The details of the architecture is described in Table I.

The multi-scaled CAT-Net can capture important morphological patterns of various sizes in cancer nuclei. CAT-Net does not include auxiliary classifiers unlike the Inception model, which makes model complexity less than the inception model but still improves accuracy.

III. EXPERIMENTAL RESULTS

We conducted intensive experiments with pathological image dataset obtained by gastroscopic biopsy specimen of 94 cases at the Gyeongsang National University Changwon Hospital (Changwon, Korea) between February 2016 and July 2017 to assess the performance of the proposed CAT-Net. The tissue specimens were stained with Hematoxylin and Eosin (H&E) using standard protocols in routine clinical care. Then, totally 94 whole slide images (WSIs) were generated. Specific cancer regions were annotated on the cancer WSIs by a pathologist at Gyeongsang National University Hospital. The WSIs were categorized into four: well (13 WSIs), moderately (11 WSIs), poorly-differentiated adenocarcinoma (20 WSIs), poorly cohesive carcinoma (20 WSIs) including signet-ring cell features, and normal gastric mucosa (30 WSIs), respectively. The histologic type and differentiation grade of the carcinoma were determined according to the classification system of the World Health Organization, fourth edition¹.

In this study, we considered only well (13 WSIs) and moderately differentiated (11 WSIs) WSIs as well as normal gastric mucosa WSIs (30 WSIs). Then, we randomly selected 24 WSIs from the 30 normal WSIs to balance the numbers of cancer and normal WSIs. The dataset was randomly split into 24, 16, and 8 WSIs for train, validation, and test respectively, while preserving the proportions of cancer and normal classes. The experiments were repeated ten times for model reproducibility.

Patches were extracted from the WSIs with the 40X magnification as an input to CAT-Net, due to the high resolution of a WSI. We generated patches by sliding a kernel window of size 256 and stride of 128 on a WSI. Image-morphological preprocessing, such as image thresholding and skew correction, were

¹This study was approved by the Institutional Review Board of Gyeongsang National University Hospital with a waiver for informed consent (2018-08-005-001).

TABLE I: Architecture of CAT-Net

Layer (Number)	Filter Size	Stride	Output (W×H×N)	
Dilated Conv (1)	3×3	2	256 × 256 × 3	
Dilated Conv (2)	3×3	1	149 × 149 × 32	
Dilated Conv (3)	3×3	1	147 × 147 × 32	
Max Pooling (4)	3×3	2	147 × 147 × 64	
Dilated Conv (5)	1×1	1	73 × 73 × 64	
Dilated Conv (6)	3×3	1	73 × 73 × 80	
Max Pooling (7)	3×3	2	71 × 71 × 192	
MixedA (8)	(1×1) (3×3) (5×5)	1	35 × 35 × 192	
Mixed layer A (9)	(1×1) (3×3) (5×5)	1	35 × 35 × 256	
Mixed layer A (10)	(1×1) (3×3) (5×5)	1	35 × 35 × 288	
Mixed layer B (11)	(1×1) (3×3)	2	35 × 35 × 288	
Mixed layer C (12, 13, 14, 15)	(1×1) (1×7)*2 (7×1)*2	1	17 × 17 × 768	
Mixed layer D (16)	(1×1) (1×7) (7×1) (3×3)		17 × 17 × 768	
Mixed layer E (17)	(1×1) (1×3) (3×1) (3×3)		8 × 8 × 1280	
Mixed layer E (18)	(1×1) (1×3) (3×1) (3×3)	1	8 × 8 × 2048	
Average Pooling (19)	(1×1)		8 × 8 × 2048	

TABLE II: Experimental results with patch images

Method	Accuracy*	Precision*	Recall*	F1 score*
VGG16	0.682 (0.0120)	0.674 (0.0091)	0.689 (0.0130)	0.681 (0.0110)
EM-CNN	0.704 (0.0707)	0.738 (0.1053)	0.689 (0.0831)	0.681 (0.1477)
N-Net50	0.693 (0.0080)	0.696 (0.0042)	0.692 (0.0080)	0.693 (0.0090)
N-Net256	0.662 (0.0034)	0.664 (0.0054)	0.663 (0.0052)	0.663 (0.0045)
GB-INCv3	0.937 (0.0042)	0.938 (0.0042)	0.936 (0.0033)	0.937 (0.0046)
CAT-Net	0.961 (0.0029)	0.965 (0.0052)	0.961 (0.0031)	0.962 (0.0055)

used to minimize noise on the patches. Cancer patches were extracted only from the annotated cancer regions in the cancer WSIs. Finally, 24,000, 16,000, and 8,000 patches of 256×256 pixels on average were generated for training, validation, and test respectively.

We compared the performance of CAT-Net with the current state-of-the-art benchmark methods. The benchmark classifiers included VGG16 [14], EM-CNN [15], N-Net [16], and Google Brain's Inception V3 (GB-INCv3) [17]. The experimental settings for the benchmark methods were as follows:

- **VGG16:** We trained with patches of 256 × 256 pixels using VGG16 architecture as proposed in [14]. We implemented VGG16 using Keras in Python, where SGD optimizer was used. The optimal hyper-parameters of learning rate, dropout, and weight decay were obtained by grid search with training and validation data on each experiment.
- **EM-CNN:** EM-CNN was implemented in Keras with the

proposed CNN structure, but we considered the same size of patch images (256×256). The model was trained by ADAM optimizer and batch Normalization with momentum of 0.99 and epsilon of 0.001. The optimal hyper-parameters of learning rate and dropout were obtained by grid search.

- **N-Net50:** To directly compare the performance, the non-overlapping window of 50×50 pixels was used to generate patches without nuclei segmentation. We implemented N-Net using Keras in Python, where Stochastic Gradient Descent (SGD) optimizer was used. The optimal hyper-parameters were obtained with the same manner with VGG16.
- **N-Net256:** N-Net was also trained with the same size's patches (256×256 pixels) inputs with CAT-Net. The optimal hyper-parameters were obtained with the same manner with VGG16.
- **GB-INCv3:** The kernel window of 299×299 pixels with

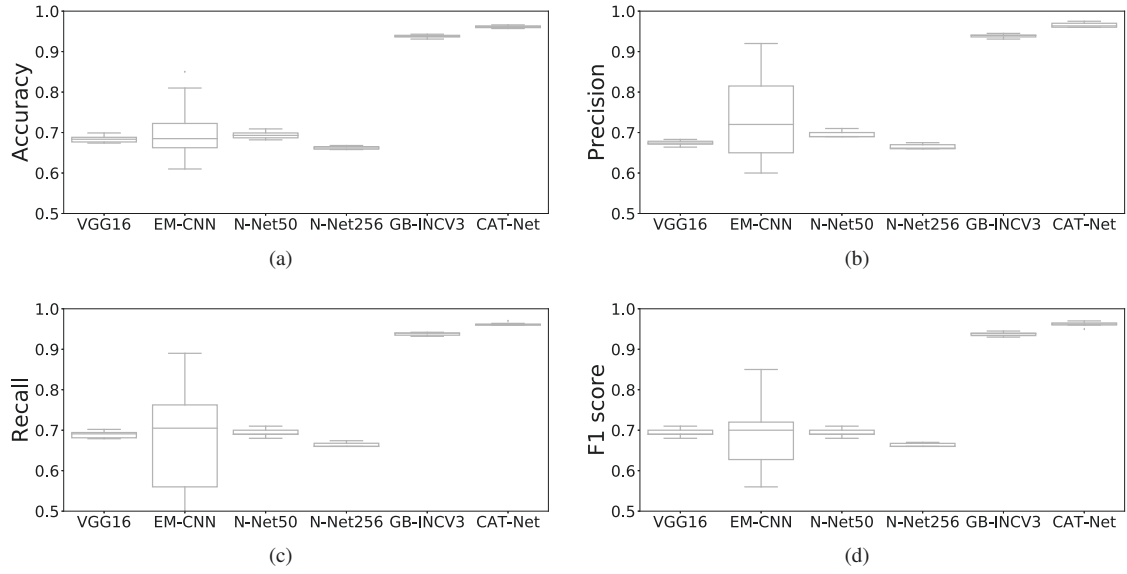


Fig. 3: Boxplots of the experimental results: (a) Accuracy, (b) Precision, (c) Recall, and (d) F1 score

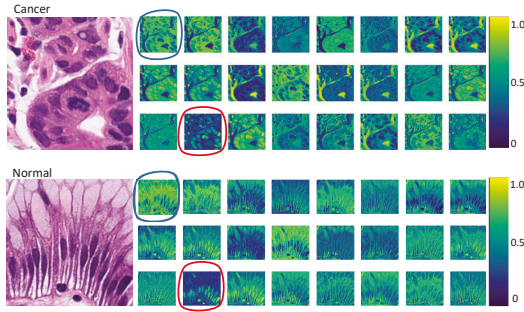


Fig. 4: Output feature maps of cancer and normal patches from Mixed Layer A

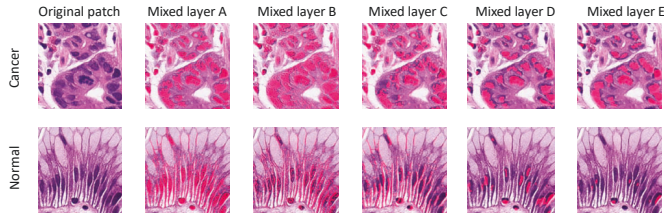


Fig. 5: Feature maps of cancer and normal patches at Mixed layers A–E. Morphological patterns of interest are captured in the hierarchical levels.

stride of 128 pixels was slid on the WSIs to generate patches, as proposed. GB-INCV3 was implemented by its open source in PyTorch provided by the authors, where Adaptive Moment Estimation (Adam) optimizer was used. The optimal parameters were obtained with same manner as VGG16.

- **CAT-Net:** CAT-Net was implemented in PyTorch, where SGD optimizer was used. We used batch-normalization to

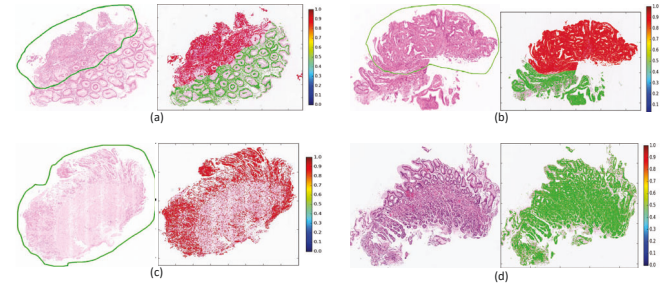


Fig. 6: Heatmaps on four examples of cancer and normal WSIs. The annotation (circles in green in the left side) indicates cancerous regions on the WSIs. For instance, (a) and (b) show partial cancerous regions on the WSIs, whereas the tissue in (c) is entirely cancerous and (d) is a totally normal tissue. The heatmaps in the right side show accurate prediction of cancerous regions on the WSIs.

achieve faster training and drop-out to avoid over-fitting. Three hyper-parameters of learning rate, weight decay, momentum, and kernel size were optimized automatically by grid search with the minimum validation loss.

We measured accuracy, precision, recall, and F1 score on the patch-based classification. The experimental results of the ten experiments are summarized in Table II. CAT-Net outperformed the other four benchmark methods, where CAT-Net produced the highest accuracy of 0.9613 ± 0.0029 (mean $\pm$ std), precision of 0.9653 ± 0.0052 , recall of 0.9617 ± 0.0031 and F1 score of 0.9625 ± 0.0055 , respectively. Figure 3 visualizes the experimental results of the benchmark methods, where our proposed method, CAT-Net, showed the best performance with high reproducibility.

We examined feature maps of CAT-Net on the mixed layers

to elucidate how the texture-based CAT-Net works for cancer detection and how the mechanism of the model is connected with knowledge in pathology. Figure 4 shows a subset of feature maps (24 out of 64) at the end of the Mixed layer A on patch examples of cancer and normal. Pixels in bright color (yellow) on the feature maps represent activated nodes by the optimally trained CAT-Net. Each feature maps were generated by multiple kernels and dilated/convolutional layers, which captures morphological patterns of interests in various aspects. For instance, in the cancer patch, the first feature map in the upper-left corner (circled in blue) highlights cytoplasm of a tissue, whereas the feature map circled in red captures most nuclei and mitosis of interests. In the same manner, the cytoplasm of the epithelial cells shows yellow color in the first feature map (circled in blue). Whereas, the feature map circled in red in the normal patch shows normal epithelial cells in yellow.

Figure 5 visualizes activated nodes in Mixed layers A–E, where the feature map scores (higher than the median) are colored in red in the figures. The feature maps also capture most nuclei of interest on the patches in the mixed layers, and morphological patterns of interests are identified in the hierarchical levels. Interestingly, in the cancer patch, inflammatory cells around cancer cells are identified in Mixed layer A, but the importance of the inflammatory cells tend to be diminished toward to Mixed layer E. It shows that both inflammatory cells and cancer cells are capture in the lower layers because they look similar each other but eventually more cancer cells are highlighted by the model at the end. As the same manner, a number of normal epithelial cells identified by the models decreased toward to Mixed layer E in the normal patch.

The optimally trained CAT-Net produced scores of cancer on the patches sliding a patch window over a WSI, and the scores are illustrated as a heat map, which indicates cancerous regions on the WSI. Heatmaps of four sample WSIs, including (a)–(b) partially cancerous tissues, (c) a entirely cancerous tissue, and (d) a entirely normal tissue, are visualized in Fig. 6. On each sub figure, annotation (circles in green in the left side) indicates cancerous regions on the WSI, whereas the right side shows its heatmap generated by CAT-Net. The heatmaps show accurate prediction of cancer corresponding to the ground truths.

IV. CONCLUSION

In this paper, we demonstrate a novel cancer texture-based deep learning model, named CAT-Net. CAT-Net improves the predictive performance by capturing multi-scale cancer morphological patterns from a WSI. CAT-Net leverages a part of the inception module without computing an auxiliary loss, which reduce model’s complexity while preserving the performance.

Through the intensive experiments, CAT-Net showed outstanding performance among the current state-of-the-art methods in cancer histopathological image classification. Moreover, feature maps which were extracted from CAT-Net illustrated

distinctive morphological patterns of cancer and normal on WSIs. The Morphological patterns of interest in a hierarchical manner in the deep neural network can be visualized. Heatmaps generated by CAT-Net showed substantial consensus with the annotations of the ground truth on a WSI, which can assist pathologists to efficiently diagnose WSIs.

V. ACKNOWLEDGMENTS

This research was supported by the Ministry of Science, ICT, Korea, under the High-Potential Individuals Global Training Program (2019–0–01601), supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP).

REFERENCES

- [1] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever, “Breast cancer histopathology image analysis: A review,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1400–1411, May 2014.
- [2] A. Rakhlin *et al.*, “Deep convolutional neural networks for breast cancer histology image analysis,” in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 737–744.
- [3] A. Belsare, “Histopathological Image Analysis Using Image Processing Techniques: An Overview,” *Signal & Image Processing : An International Journal*, 2012.
- [4] D. Bardou, K. Zhang, and S. M. Ahmad, “Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks,” *IEEE Access*, 2018.
- [5] A. Janowczyk and A. Madabhushi, “Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases,” *Journal of Pathology Informatics*, 2016.
- [6] M. Shah *et al.*, “Deep learning assessment of tumor proliferation in breast cancer histological images,” in *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, 2017.
- [7] C. Szegedy *et al.*, “Going deeper with convolutions,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June, 2015, pp. 1–9.
- [8] A. Rakhlin *et al.*, “Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018.
- [9] B. Wei *et al.*, “Deep learning model based breast cancer histopathological image classification,” in *2017 2nd IEEE International Conference on Cloud Computing and Big Data Analysis, ICCCBDA 2017*, 2017.
- [10] J. Xie, R. Liu, J. Luttrell, and C. Zhang, “Deep learning based analysis of histopathological images of breast cancer,” *Frontiers in Genetics*, 2019.
- [11] K. Kamijo *et al.*, “Automated gastric cancer diagnosis on H&E-stained sections; ltraining a classifier on a large scale with multiple instance machine learning,” in *Medical Imaging 2013: Digital Pathology*, 2013.
- [12] A. L. Yuille and others., “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- [13] T. Guan and H. Zhu, “Atrous Faster R-CNN for Small Scale Object Detection,” in *Proceedings - 2017 2nd International Conference on Multimedia and Image Processing, ICMIP 2017*, 2017.
- [14] D. Wang and others., “Deep Learning for Identifying Metastatic Breast Cancer,” *arXiv e-prints*, p. arXiv:1606.05718, Jun 2016.
- [15] L. Hou *et al.*, “Efficient multiple instance convolutional neural networks for gigapixel resolution image classification,” *CoRR*, vol. abs/1504.07947, 2015.
- [16] F. Sheikhzadeh *et al.*, “Automatic labeling of molecular biomarkers of immunohistochemistry images using fully convolutional networks,” *PLoS ONE*, 2018.
- [17] Y. Liu *et al.*, “Detecting cancer metastases on gigapixel pathology images,” *CoRR*, vol. abs/1703.02442, 2017. [Online]. Available: <http://arxiv.org/abs/1703.02442>