# Frequency-based convolutional neural network for efficient segmentation of histopathology whole slide images

Wei Luo[1,2], Yushan Zheng[2(✉)], Dingyi Hu[1,2], Jun Li[1,2], Chenghai Xue[3,4], and Zhiguo Jiang[1,2]

[1] Image Processing Center, School of Astronautics, Beihang University,Beijing 100191, China
[2] Beijing Advanced Innovation Center for Biomedical Engineering,Beihang University, Beijing 100191, China
[3] Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
[4] Wankangyuan Tianjin Gene Technology, Inc., Tianjin 300220, China
{luowei0701,yszheng}@buaa.edu.cn

**Abstract.** CNN-based methods for WSI segmentation are time-consuming under the limits of communication bandwidth and memory usage, due to the high pixel resolution of WSIs. In this paper, we propose a novel framework for accelerating the segmentation of digital histopathology WSIs in the frequency domain. Based on the characteristics of the JPEG format in data storage and transmission on the existing digital histopathological diagnosis cloud platform, we extract DCT coefficients from the JEPG decoding and compress them into the DCT feature cubes by a frequency selection block. Based on the DCT feature cubes, we propose an extremely light-weighted model named Efficient DCT-Network (EDCT-Net). The size of the input data, as well as the bandwidth requirement for CPU-GPU transmitting, for EDCT-net reduces by 96 % compared to the common CNN-based methods. And, the number of model parameters and the floating-point operations (FLOPs) for EDCT-Net decreases by 98% and 94% compared to the baseline method. The experimental results have demonstrated that our method achieves a Dice score of 0.811 with only 8 frequency channels in the task of endometrial histopathology WSI segmentation, which is comparable with state-of-the-art methods.

**Keywords:** Histopathology WSI · Segmentation · DCT · Image compression.

## 1 Introduction

Digital histopathology whole slide images (WSIs) become popular in the modern cancer diagnosis. The diagnosis based on digital WSIs is time-consuming and laborious for pathologists due to the high pixel resolution of WSIs. In this case, an increasing number of computer-aided histopathology diagnosis technologies are developed to assist the pathologists in clinical diagnosis.

Recently, automatic methods based on deep learning have widely used for WSI analysis [23,3,18,6,22,21,2]. And cancerous region segmentation is always a popular and challenging topic. In the case of high pixel resolution, it is hard to directly segment the entire WSIs by common deep learning models for the limitations of recent CPU and GPU capacities. Therefore, the WSIs are generally divided into image patches and then fed into the deep learning model for inference. There are two typical strategies to achieve the segmentation of WSIs. The one is using end-to-end fully convolutional networks (FCNs) [10,14] such as U-Net [12], CRFs [20], DeepLab [4] and SegNet [1], which produce correspondingly-sized segmentation results for these patches and then mosaicked to obtain the WSI segmentation result. The other is based on convolutional neural networks (CNNs) [9,13,17], e.g. GoogleNet [15], ResNet [5] and DenseNet [8]. The classes of the patches are predicted by CNNs, then aggregated together into the segmentation result. Whereas, for both of the above strategies, the computation amounts are heavy when facing the large-scale whole slide images. Simultaneously, because the RGB-format patches are generally pre-processed on the CPUs and then transmitted to the GPUs/AI accelerator in the convolutional inference, the communication bandwidth between the CPU and the GPU/AI limits the analysis speed. To reduce the bandwidth and computational amount, the general and simplest way is to down-sample the input image. But it will result in a significantly decrease in the segmentation accuracy [14].

In the practical applications, the tiles of WSIs are commonly stored and transmitted on cloud platforms as compression data following the JPEG standard. The compressed data stream needs to be decoded into the RGB-format data in the sequence of unzigzag, Huffman decoding, IDCT, etc. following the JPEG standard before used for image analysis and processing. It is notable that the discrete cosine transformation (DCT) coefficients generated in the process of JPEG decoding are informative and potential to develop efficient deep learning models for histopathology patterns recognition. Especially for digital WSIs, all the image content nearly locates on one imaging plane and thereby the same structures, such as nuclei, occupies a consistent number of pixels in the WSI. In this case, one specific DCT frequency indicates a fixed image pattern. Based on the prior knowledge, we could greatly reduce the computational complexity of histopathology whole slide image analysis and meanwhile achieve a comparable accuracy by building a framework in the frequency domain.

In this paper, we propose a novel framework for fast segmentation of digital histopathology WSIs in the frequency domain. The proposed framework is illustrated in Fig. 1. Compared to the existing methods [3], we truncate the decoding process and directly extract the DCT coefficients from the JPEG-compressed data as the input of the deep learning model. Meanwhile, a frequency selection block is designed to compress the DCT coefficients into DCT feature cube, which is tens of times smaller than the corresponding RGB-format data. Furthermore, since the DCT feature cube is with smaller width and height, we build an extremely light-weighted convolutional neural network named Efficient DCT-Network (EDCT-Net) for fast classification of histopathology image

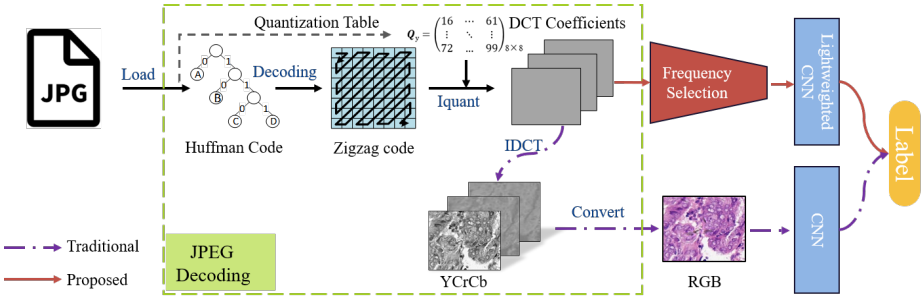


**Fig. 1.** The illustration of the proposed method, where the purple dotted arrow is the data flow of the deep learning methods based on RGB-format data, the red solid arrow is the data flow of the proposed Efficient DCT-Network (EDCT-Net), and the green dotted box is the process of JEPG decoding.

patches and also for the segmentation of digital WSIs. Our methods were evaluated on a large-scale endometrial WSI dataset. The experimental results have demonstrated that the proposed framework greatly reduces the computational complexity of whole slide image analysis and meanwhile achieve a comparable performance with the existing methods. The contribution of this paper can be summarized as:

1) We proposed a novel framework for efficient segmentation of digital histopathology WSIs in the frequency domain based on the JPEG compression standard. The JPEG-compressed image data is first decoded to obtain the DCT coefficients, then DCT feature cube is generated based on frequency selection of the DCT coefficients, and finally a light-weighted CNN, EDCT-Net, is trained to classify the DCT feature cubes. Compared to the existing methods based on RGB-format CNNs, the size of the input data, as well as the bandwidth requirement for CPU-GPU transmitting, is reduced by tens of times, the number of model parameters and the floating-point operations (FLOPs) decreases by 98% and 94% compared to the baseline method.

2) We studied the frequency selection for compression in the frequency domain. The comprehensive experiments have shown that a small part of low frequency data stored most of the discriminative information for histopathology patterns recognition. The proposed EDCT-Net achieved a Dice score of 0.811 with only 8 frequency channels in the task of endometrial histopathology WSI segmentation, which is comparable with the state-of-the-art RGB-based CNNs.

The remainder of this paper is organized as follows. Section 2 introduces the methodology of the proposed method. The experiment are presented in Section 3. Section 4 summarizes the contribution.

## 2 Method

The proposed framework involves 1) DCT coefficients extraction, 2) the frequency selection, 3) the Efficient DCT-Network, and 4) WSI segmentation based on EDCT-Net, which are detailed in this section.

### 2.1 DCT coefficients extraction

The JPEG standard is widely used for photographs and internet graphics compression. The standard encoding process defined by JPEG is as follows: 1) Convert RGB to YCbCr for each pixel, 2) Convert the Y, Cb and Cr channels into chunks of blocks in $8 \times 8$ pixels, then apply Discrete Cosine Transformation (DCT) to each block and use quantization to compress the resulting block, 3) Use Zig-zag encoding to flatten the $8 \times 8$ matrix into one-dimensional data, then apply run-length and Delta encoding, and 4) Use Huffman encoding to generate the binary stream. As shown in Fig. 1, the decoding process is the reverse of the encoding process.

In this paper, we truncate the decoding process of histopathology images to obtain the DCT coefficients. Letting $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$ denote an image patch in size of $H \times W$ in the WSI, the DCT coefficients of Y, Cb and Cr channels for $\mathbf{I}$ are represented as $\mathbf{D}^Y$, $\mathbf{D}^{Cb}$ and $\mathbf{D}^{Cr}$. Specifically,

$$\mathbf{D}^c = \begin{pmatrix} \mathbf{B}^c_{11} & \cdots & \mathbf{B}^c_{1N_c} \\ \vdots & \ddots & \vdots \\ \mathbf{B}^c_{M_c1} & \cdots & \mathbf{B}^c_{M_cN_c} \end{pmatrix}, c = Y, Cb, Cr, \tag{1}$$

where $\mathbf{B}^c_{ij} \in \mathbb{R}^{8\times8}$ represents the 64 frequency channels transformed from the $8 \times 8$ block in the $i$-th row and $j$-th column, and $M_c$ and $N_c$ denote the rows and columns in the $c$-th channel. Notice that the size of DCT coefficients matrix for the Y channel is larger than that for the Cr and Cb channels because of different sampling mode. Specifically in our study, the height and width of $\mathbf{D}^Y$ are the same as $\mathbf{I}$, while the height and width of $\mathbf{D}^{Cb}$ and $\mathbf{D}^{Cr}$ are half of $\mathbf{D}^Y$.

### 2.2 Frequency selection block

In the frequency selection block, the $8 \times 8$ blocks in $\mathbf{D}^Y$, $\mathbf{D}^{Cb}$ and $\mathbf{D}^{Cr}$ are first flattened into one-dimensional vectors while keep their spatial relations (as shown in Fig. 2). Then, the DCT coefficients matrics are rescaled into 192 channels (as shown in Fig. 2c), where each channel represents one frequency. Afterwards, the DCT frequency cubes can be defined as $\widetilde{\mathbf{D}}^Y \in \mathbb{R}^{H/8\times W/8\times 64}$, $\widetilde{\mathbf{D}}^{Cb} \in \mathbb{R}^{H/16\times W/16\times 64}$ and $\widetilde{\mathbf{D}}^{Cr} \in \mathbb{R}^{H/16\times W/16\times 64}$. Finally, these cubes are concatenated after upsampling the $\widetilde{\mathbf{D}}^{Cb}$ and $\widetilde{\mathbf{D}}^{Cr}$ to the same size as $\widetilde{\mathbf{D}}^Y$. The concatenated DCT frequency cube is formulated as:

$$\widetilde{\mathbf{D}} = \left[d^Y_0, d^Y_1, \ldots, d^Y_{63}, d^{Cb}_0, d^{Cb}_1, \ldots, d^{Cb}_{63}, d^{Cr}_0, d^{Cr}_1, \ldots, d^{Cr}_{63}\right] \in \mathbb{R}^{H/8\times W/8\times 192} \tag{2}$$

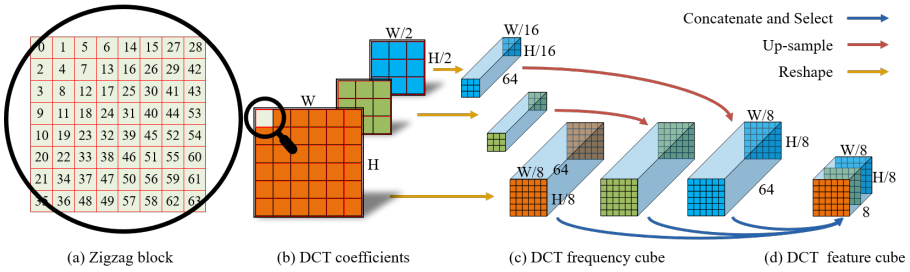


**Fig. 2.** The data processing pipeline of frequency selection block, where (a) is the Zigzag block of the each pixel in the original (b) DCT coefficient, the orange, green and blue cubes are from Y, Cb and Cr respectively, (c) is the DCT frequency cube generated after reshaping and up-sampling, (d) is the DCT frequency cube

where $d_k^c \in \mathbb{R}^{H/8 \times W/8}$ is the $k$-th frequency channel from $\widetilde{\mathbf{D}}^c$. Notice that $\widetilde{\mathbf{D}}$ maintains the same size with the image $\mathbf{I}$. The main different is that majority frequency channels of $\widetilde{\mathbf{D}}$, especially the high-frequency channels, are occupied by zero. We suppose that these zero-occupied channels are less informative and therefore less effective for histopathology image recognition. To reduce the size of the data, we propose keeping the informative channels and meanwhile discard the less-informative channels in the frequency selection stage. Generally, the final frequency feature cubes are formulated as $\mathbf{F} \in \mathbb{R}^{H/8 \times W/8 \times M}$, where $M = M_Y + M_{Cb} + M_{Cr}$ is the total number of selected frequency channels with $M_c, c = Y, Cb, Cr$ denotes the number of selected channel for the $c$-th color channel.

In our experiments, an original $224 \times 224 \times 3$ RGB-format image is finally converted into a DCT feature cube of size $28 \times 28 \times M$. The compression ratio is $M/192$. This conversion can be efficiently processed by a CPU during the data preparation for deep learning frameworks.

### 2.3 Efficient DCT-Network (EDCT-Net)

EfficientNets are a family of models which can be scaled up to any target resource constraints in a more principled way while maintaining model efficiency [16]. EfficientNet-b0 is the base network structure, which is a popular light-weighted model to develop efficient applications. In this paper, the EfficientNet-b0 is used as the baseline model for histopathology image patch classification. Moreover, considering the shape of the proposed DCT feature cube ($28 \times 28 \times M$), we defined a more light-weighted version of EfficientNet-b0 by discarding the first 4 MB-Conv blocks and named it as Efficient DCT-Network (EDCT-Net). EDCT-Net has much fewer layers and narrower width than EfficientNet-b0, but can achieve higher accuracy and faster speed than EfficientNet-b0 for the DCT feature cubes introduced in the previous section.

### 2.4 CNN-based segmentation for histopathology images

After training the EDCT-Net, the segmentation of cancerous regions from a WSI can be achieved by feeding the patches generated by the sliding window paradigm into the EDCT-Net. Specifically, the size of the window is $224 \times 224$ with the step of 224. Then, the probability map for the WSI is generated by aggregating the prediction scores of all the patches. Next, the segmentation result is obtained by the threshold operation after up-sampling the probability map to the same size of the WSI. Finally, to refine the segmentation result, an open operation followed by a hole filling operation was used as the post-processing.

## 3 Experiments

### 3.1 Dataset and settings

To study the effectiveness and feasibility of the proposed framework for efficient WSI segmentation, we collected a large-scale endometrial dataset containing 1288 WSIs. These WSIs were categorized to 5 types of endometrial pathology, including Normal, Well-differentiated Endometrioid adenocarcinoma (WDEA), Moderately differentiated Endometrioid adenocarcinoma (MDEA), Low differentiated Endometrioid adenocarcinoma (LDEA), and Serous. All the WSIs were scanned under a lens of 20x, and were well annotated by pathologists, as shown in Fig. 3. We randomly selected 772 samples for training, 193 for validation and 323 samples for testing.

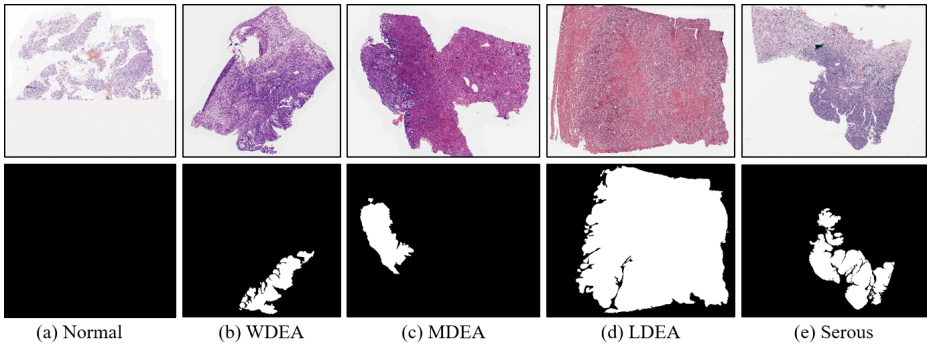


**Fig. 3.** Instances in the endometrial WSI dataset, where (a) is Normal, (b) is Well-differentiated Endometrioid adenocarcinoma, (c) is Moderately differentiated Endometrioid adenocarcinoma, (d) is Low differentiated Endometrioid adenocarcinoma, (e) is Serous, and the ground-truth for the cancerous regions are provided on the second row.

Image patches in size of $224 \times 224$ pixels were sampled from the WSIs and labeled according to the pathologists' annotations. Specifically, 728,659 patches

are obtained from the training WSIs, 184,905 patches from the validation WSIs, and 393,870 patches frome testing WSIs. In the experiments, the Normal samples were regarded as the negative and WDEA, MDEA, LDEA, and Serous were regarded as the positive. The sensitivity, specificity, and accuracy for the binary classification were used as the metrics. The segmentation of the WSI was achieved according to Section 3.3. The Dice Score (DSC) was used as the segmentation metric, which is formulated as

$$DSC = \frac{2TP}{FP + 2TP + FN}, \tag{3}$$

where TP, FP, and FN denotes the number of true positive pixels, false positive pixels and false negative pixels, respectively.

All the algorithms were implemented in python with Pytorch. The experiments conducted on a computer with 2 CPUs of Intel Xeon and 2 GPUs of Nvidia Geforce 2080Ti.

### 3.2 Patch classification based on DCT coefficients

We first conducted experiment to verify the hypothesis that the frequency data within the DCT coefficients represents consistent information of histopathology WSIs. As mentioned in the Section 2, each $224 \times 224 \times 3$ RGB-format patch is transformed by the feature selection block into DCT feature cube in size of $28 \times 28 \times M$ and used as the input to CNNs.

**Table 1.** The classification results on patches using DCT feature cubes as inputs with different $M$ settings, where $M = 192, 64, 40, 32, 24, 16, 8$ is the number of selected frequency channels.

| No. | Channel settings | | | | Metrics | | |
|---|---|---|---|---|---|---|---|
| | $M$ | $M_Y$ | $M_{Cb}$ | $M_{Cr}$ | Accuracy (%) | Specificity | Sensitivity |
| 1 | 192 | 64 | 64 | 64 | 90.67 | **0.90** | 0.92 |
| 2 | 64 | 64 | 0 | 0 | 89.39 | 0.86 | **0.93** |
| 3 | | 0 | 64 | 0 | 88.73 | 0.89 | 0.88 |
| 4 | | 0 | 0 | 64 | 84.94 | 0.87 | 0.83 |
| 5 | 40 | 40 | 0 | 0 | 89.33 | 0.88 | 0.91 |
| 6 | 32 | 32 | 0 | 0 | 89.25 | 0.86 | **0.93** |
| 7 | 24 | 24 | 0 | 0 | 89.68 | 0.89 | 0.90 |
| 8 | | 8 | 8 | 8 | **91.06** | 0.89 | **0.93** |
| 9 | 16 | 16 | 0 | 0 | 89.45 | **0.90** | 0.89 |
| 10 | 8 | 8 | 0 | 0 | 89.02 | 0.87 | 0.91 |
| 11 | | 4 | 2 | 2 | 90.67 | 0.89 | 0.92 |

**The effect of the frequency channels** In this experiment, we explored the importance of each frequency extracted from Y, Cr, Cb channel. We selected different number of frequency channels for patch classification by setting $M$ to 192, 64, 40, 32, 24, 16, 8, respectively. The detailed channels used for each color channel and the corresponding classification performance for the test set are shown in Table 1. The accuracy of $M_Y, M_{Cb}, M_{Cr} = [64, 64, 64]$ achieves at 90.67% when using all the frequency channels for the classification, which can be regarded as the baseline of the DCT-based models. The results for $M = 64$ indicate that the frequency channels from Y play a more important role in the classification than those from Cb and Cr. Interestingly, $M_Y, M_{Cb}, M_{Cr} = [8, 8, 8]$ achieves an accuracy of 91.06%, which is the best among all the experimental settings. Moreover, when we further reduced the channels to 8, it still maintains an accuracy of 90.67%, which is equivalent to the baseline. It demonstrates that in these 192 frequency channels, the features contained in the lower-frequency channels are more informative for classification. In contrast, higher-frequency channels contain less important information and may perform as the noise in the classification task.

**Table 2.** The classification results on patches when the number of selected frequency channels is $M = 8$, and $K_Y, K_{Cb}, K_{Cr}$ defines the index of the selected frequency channels from Y, Cr, Cb respectively.

| No. | Channel settings | | | Metrics | | |
|---|---|---|---|---|---|---|
| | $K_Y$ | $K_{Cb}$ | $K_{Cr}$ | Accuracy (%) | Specificity | Sensitivity |
| 1 | 0,1,2,3,4,5 | 0 | 0 | 90.56 | 0.88 | **0.93** |
| 2 | 0,1,2,3 | 0,1 | 0,1 | 90.67 | **0.89** | 0.92 |
| 3 | 0,1,2,3 | 0,2 | 1,3 | 90.34 | **0.89** | 0.91 |
| 4 | 0,2,4,6 | 0,1 | 0,1 | 90.48 | 0.88 | **0.93** |
| 5 | 0,1,3,5 | 0,1 | 0,1 | **90.75** | **0.89** | 0.92 |
| 6 | 0,1 | 0,1,2 | 0,1,2 | 90.42 | 0.90 | 0.91 |

**Channel selection for the Efficient DCT-Net** For the reason of building efficient segmentation framework, we determined retaining 8 frequency channels as the input setting of the EDCT-Net. However in the above experiments, we simply selected the top-lower-frequency channels. To pursue a higher accuracy, we conducted experiments to decide the eight specific channels. The results are presented in Table 2. It is obvious that the best accuracy is 90.75% when $K^Y = \{0, 1, 3, 5\}$, $K^{Cb} = \{0, 1\}$, $K^{Cr} = \{0, 1\}$, where $K^Y$, $K^{Cb}$ and $K^{Cr}$ denote the indexes of the selected channels for Y, Cb, and Cr, respectively. Finally, we used this solution that achieves the best accuracy in the following experiments.

As shown in Fig. 4, we reconstruct the RGB-format images from different remaining frequency channels of the DCT feature cubes. It illustrates that more

higher frequency channels are reserved, more edge details will be in the reconstructed image. The images in Fig. 4 (e-h) are reconstructed by only 8 frequency channels. It shows only the rough structure of the histopathology images can be reconstructed. Nevertheless, such rough information delivers comparable recognition performance with the baseline strategy. The results indicate the rough but crucial structures and spatial position relations of the nuclei could be more important than the high-frequency information in the considered task.

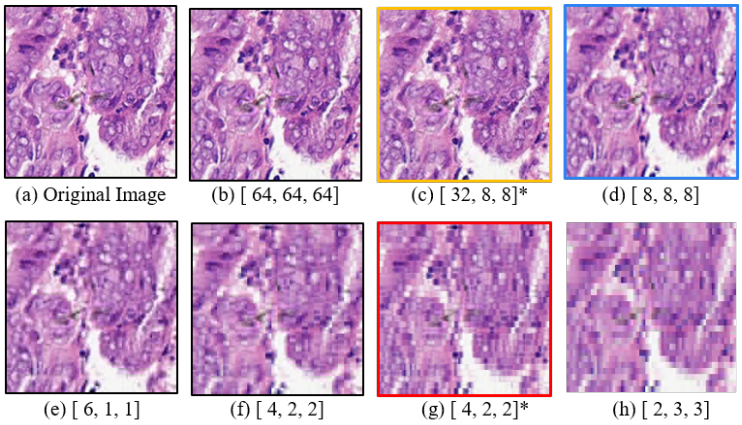


**Fig. 4.** The RGB-format images reconstructed by the remaining frequency channels. (a) is the original image, (b), (d), (e), (f) and (h) are constructed by the top $M_Y, M_{Cb}, M_{Cr}$ lower frequency channels from Y, Cb and Cr respectively. The yellow bounding box (c)is the selection scheme of DCTNet, while blue bounding box (d) is the selection scheme which achieves highest accuracy, and the red bounding box (g) is the selection scheme of the Efficient DCT-Network (EDCT-Net). The * denotes that the selection scheme of it is based on a certain subset of frequency index.

**Table 3.** The comparison of classification performance between the proposed EDCT-Net and baselines.

| Model | #Params($\times 10^6$) | #FLOPs($\times 10^8$) | Accuracy | Specificity | Sensitivity |
|---|---|---|---|---|---|
| Efficientnet-b0 [16] | 4.0 | 3.9 | **92.60** | **0.91** | 0.94 |
| Mobilenet-v3 [7] | 2.7 | 0.6 | 92.21 | 0.90 | **0.95** |
| Shufflenet-v2 [11] | 1.3 | 1.5 | 92.22 | **0.91** | 0.94 |
| DCTNet [19] | 23.50 | 12 | 90.34 | 0.87 | 0.93 |
| EDCT-Net (ours) | 0.08 | 0.23 | 91.43 | 0.90 | 0.93 |

**Comparison with the existing classification methods** Then, we compared the proposed EDCT-Net with the state-of-the-art methods, including the light-weighted networks (EfficientNet-b0 [16], Mobilenet-v3 [7], Shufflenet-v2 [11]) and another model based on DCT input (DCTNet [19]). Except DCTNet, all methods take RGB-format image as inputs. The input size of the RGB-based CNNs is $224 \times 224 \times 3$. The input of DCTNet is DCT feature cubes of $28 \times 28 \times 48$ size. The construction of DCT feature cubes refer to the [19]. As shown in the Table 3, the proposed EDCT-Net achieves an accuracy of 91.43%, which is 1.17% inferior to the EfficientNet-b0. However, the EDCT-Net contains only $0.08 \times 10^6$ parameters, which is 2.0% of Efficientnet-b0, 3.0% of Mobilenet-v3 and 6.2% of Shufflenet-v2. Meanwhile, the FLOPs of EDCT-Net is $0.23 \times 10^8$, which is about 5.8% of the EfficientNet-b0 and 2% of DCTNet [19]. All the above results have demonstrated that the proposed method is effective in histopathology image classification and is rather computational efficient than the existing light-weighted methods.

### 3.3 WSI segmentation based on Efficient DCT-Net

Finally, we evaluated the segmentation performance of the model for the 323 testing WSIs. The input format, normalized input size, dice score (DSC), accuracy for different models, average inference time for each WSI are presented in Table 4. It shows that our method achieves a DSC of 0.811 and a pixel accuracy of 0.810, which are comparable with the other RGB-based methods and superior to the DCT-based DCTNet. It is notable that the normalized input size of the proposed is 0.04 while the EfficientNet-b0 [16], Mobilenet-v3 [7], and Shufflenet-v2 [11] are 1.00 and DCTNet [19] is 0.25. Namely, the proposed EDCT-Net reduces the data scale required to be uploaded to the GPU by 96%, when compared to the RGB-based methods. This substantial reduction breaks through the bandwidth bottleneck between the CPU and the GPU, and thereby accelerating the speed of segmentation for histopathology WSIs. With the same batchsize setting, our method is more than 50% faster than the baselines and still has the potential to increase batchsize for a faster analysis. It is very significant to develop high data-efficient applications for WSI analysis.

**Table 4.** The segmentation performance of the proposed EDCT-Net, where normalized input size is the ratio of input sizes of the methods to that of EfficientNet-b0.

| Model | Input Format | DSC | Accuracy | Normalized Input Size | Time |
|---|---|---|---|---|---|
| Efficientnet-b0 [16] | $224 \times 224 \times 3$ RGB | **0.832** | **0.811** | 1.00 | 23.3s |
| Mobilenet-v3 [7] | $224 \times 224 \times 3$ RGB | 0.827 | 0.809 | 1.00 | 19.6s |
| Shufflenet-v2 [11] | $224 \times 224 \times 3$ RGB | 0.821 | **0.811** | 1.00 | 19.3s |
| DCTNet [19] | $28 \times 28 \times 48$ DCT | 0.804 | 0.806 | 0.25 | 17.7s |
| EDCT-Net | $28 \times 28 \times 8$ DCT | 0.811 | 0.810 | **0.04** | 11.3s |

Fig. 5 visualizes several segmentation instances of WSIs for the EfficientNet-b0 [16], Mobilenet-v3 [7], Shufflenet-v2 [11], DCTNet [19] and the proposed EDCT-Net. As shown in the last row, our model can be found better on segmenting isolated small tissue areas.

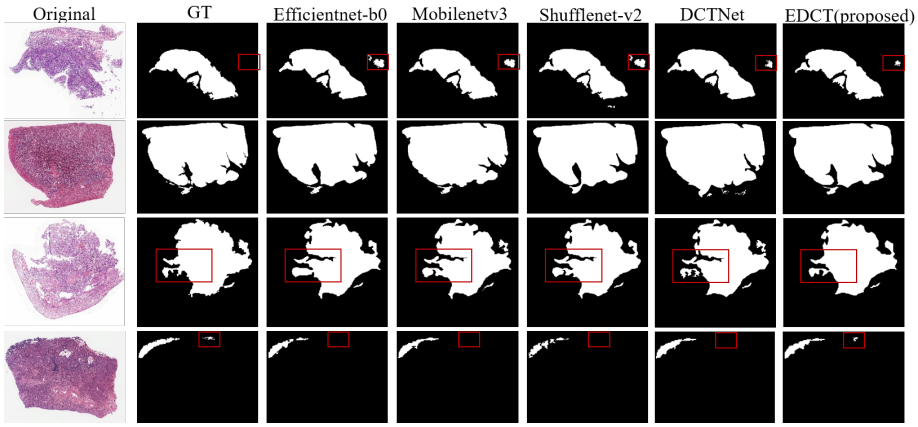


**Fig. 5.** Examples of segmentation results on the endometrial dataset. The first column is the original WSIs, the second column is the ground-truth, and the remaining columns are the segmentation results of the five methods respectively.

## 4 Conclusion

In this paper, we propose an efficient framework for recognition and segmentation of digital histopathology WSIs in the frequency domain. Our method utilizes the DCT coefficients from the JPEG standard to achieve the classification and segmentation for WSIs. The experimental results have demonstrated that the lower frequency channels contain the sufficient information for recognition while the higher frequency has less information and may introduce noise. The proposed EDCT-Net achieves a Dice score of 0.811 with only 8 frequency channels in the task of endometrial histopathology WSI segmentation. More comprehensive experimental results show that the bandwidth requirement for CPU-GPU transmitting, is reduced by 96% and the number of model parameters and the floating-point operations (FLOPs) decreases by 98% and 94% when compared to the baseline method. Our future work will focus on applying the DCT coefficients to the compression and classification of WSIs.

## Acknowledgment

This work was partly supported by the National Natural Science Foundation of China (Grant No. 61771031, 61901018, and 61906058), partly by China Postdoctoral Science Foundation (No. 2019M650446) and partly by Tianjin Science and Technology Major Project (Grant No. 18ZXZNSY00260).

## References

1. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence pp. 1–1 (2017)
2. Bandi, P., et al.: From detection of individual metastases to classification of lymph node status at the patient level: The camelyon17 challenge. IEEE Transactions on Medical Imaging **38**(2), 550–560 (2019)
3. Bejnordi, B.E., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama **318**(22), 2199–2210 (2017)
4. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: 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 (2018)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
6. Hollon, T.C., et al.: Near real-time intraoperative brain tumor diagnosis using stimulated raman histology and deep neural networks. Nature Medicine **26**(1), 52–58 (2020)
7. Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R.: Searching for mobilenetv3. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (2020)
8. Huang, G., Liu, Z., Laurens, V., Weinberger, K.Q.: Densely connected convolutional networks. IEEE Computer Society (2016)
9. Lerousseau, M., Vakalopoulou, M., Classe, M., Adam, J., Battistella, E., Carré, A., Estienne, T., Henry, T., Deutsch, E., Paragios, N.: Weakly supervised multiple instance learning histopathological tumor segmentation (2020)
10. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(4), 640–651 (2015)
11. Ma, N., Zhang, X., Zheng, H.T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: European Conference on Computer Vision (2018)
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. Springer, Cham (2015)
13. Sun, S., Jiang, B., Zheng, Y., Xie, F.: A Comparative Study of CNN and FCN for Histopathology Whole Slide Image Analysis. Image and Graphics (2019)
14. Sun, S., Yuan, H., Zheng, Y., Zhang, H., Hu, D., Jiang, Z.: Cancer sensitive cascaded networks (csc-net) for efficient histopathology whole slide image segmentation. In: IEEE 17th International Symposium on Biomedical Imaging (ISBI) (2020)

15. Szegedy, C., Wei, L., Jia, Y., Sermanet, P., Rabinovich, A.: Going deeper with convolutions. IEEE Computer Society (2014)
16. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. pp. 6105–6114 (2019)
17. Tokunaga, H., Teramoto, Y., Yoshizawa, A., Bise, R.: Adaptive weighting multi-field-of-view cnn for semantic segmentation in pathology. pp. 12589–12598 (06 2019). https://doi.org/10.1109/CVPR.2019.01288
18. Veta, M., et al.: Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge. Medical image analysis **54**, 111–121 (2019)
19. Xu, K., Qin, M., Sun, F., Wang, Y., Chen, Y.K., Ren, F.: Learning in the frequency domain (2020)
20. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.: Conditional random fields as recurrent neural networks. 2015 IEEE International Conference on Computer Vision (ICCV) (2015)
21. Zheng, Y., Jiang, B., Shi, J., Zhang, H., Xie, F.: Encoding histopathological wsis using gnn for scalable diagnostically relevant regions retrieval. In: MICCAI 2019. pp. 550–558. Springer International Publishing, Cham (2019)
22. Zheng, Y., Jiang, Z., Shi, J., Xie, F., Zhang, H., Huai, J., Cao, M., Yang, X.: Diagnostic regions attention network (dra-net) for histopathology wsi recommendation and retrieval. IEEE Transactions on Medical Imaging **40**(3), 1090–1103
23. Zheng, Y., Jiang, Z., Xie, F., Zhang, H., Ma, Y., Shi, H., Zhao, Y.: Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification. Pattern Recognition **71**, 14–25 (2017)