# Identification of Melanoma From Hyperspectral Pathology Image Using 3D Convolutional Networks

Qian Wang, Li Sun, Yan Wang, *Member, IEEE*, Mei Zhou, Menghan Hu, Jiangang Chen,
Ying Wen, and Qingli Li, *Senior Member, IEEE*

*Abstract*— **Skin biopsy histopathological analysis is one of the primary methods used for pathologists to assess the presence and deterioration of melanoma in clinical. A comprehensive and reliable pathological analysis is the result of correctly segmented melanoma and its interaction with benign tissues, and therefore providing accurate therapy. In this study, we applied the deep convolution network on the hyperspectral pathology images to perform the segmentation of melanoma. To make the best use of spectral properties of three dimensional hyperspectral data, we proposed a 3D fully convolutional network named Hyper-net to segment melanoma from hyperspectral pathology images. In order to enhance the sensitivity of the model, we made a specific modification to the loss function with caution of false negative in diagnosis. The performance of Hyper-net surpassed the 2D model with the accuracy over 92%. The false negative rate decreased by nearly 66% using Hyper-net with the modified loss function. These findings demonstrated the ability of the Hyper-net for assisting pathologists in diagnosis of melanoma based on hyperspectral pathology images.**

*Index Terms*— **Microscopy, segmentation, skin, quantification and estimation, optical imaging.**

## I. INTRODUCTION

**A**S REPORTED [1]–[3], although taking up 5% of all skin cancers in the United States, melanoma is always in company with high mortality which accounts for almost 75% of all skin related deaths. Fortunately, the 5-year survival rate of melanoma can reach 99% if diagnosed in an early phase. Therefore, accurate and early diagnosis is the premise for guaranteeing highly survival rate. Clinically, skin surface inspection (e.g. dermoscopic analysis) [4] and skin biopsy histopathological analysis [5] are two primary and complementary diagnostic methods given the timing and expense. The advent of computer-aided diagnosis, which combines the advantages of high-precision imaging techniques and computational power, have already provided innovation in the field of clinical diagnosis [6]. Our previous work [7] has used the object based multiscale detection method to explore on this topic with a small dataset. With the rapidly developing technique, we are able to investigate more challenging task involving larger dataset with different magnitudes. In this paper, we focus on the on-going image processing technique implemented in skin biopsy pathological analysis, in terms of its potential in providing objective standards and improving efficiency for clinical screening. To this end, detecting the unhealthy even malign tissue in epidermis area is of vital importance, lining up with the goal of image segmentation from the perspective of image processing. This is because a comprehensive and reliable pathological analysis is the result of correct segmentation of the malign tissue and its interaction with the benign tissue; moreover, these analysis is the basis of accurate therapy.

Recently, the advance in deep learning is inspiring and encouraging process in pathology image analysis, which covers a wide range of applications including medical image classification, detection and segmentation [8] for a diversity of diseases [9]–[12]. In terms of segmentation, the task requires classification to be performed on each pixel of the image. Long *et al.* [13] proposed a novel approach which replaced fully connected layers by fully convolutional layers so that the network can output a complete segmentation image in a single forward process. The idea of fully convolutional network (FCN) has been applied in melanoma dermoscopic segmentation to generate irregular border [14], [15]; specifically, Yuan *et al.* [16] designed a loss function based on Jaccard distance to constrain the network for training imbalanced dataset. For pathology segmentation, Phillips *et al.* [17] trained a multi-stride FCN to measure the Breslow thickness for cutaneous melanoma. However, as the convolutional filters extract more abstract features layer by layer, the resolution of input image decreases gradually. To preserve the initial resolution, the most prevalent architecture termed U-net [18] was proposed, which constructs an encoding path and a symmetric decoding path to reach precise resolution. Given this,

studies have shown feasibility of adopting U-net ased networks to dermoscopic melanoma segmentation [19]–[22] and epidermis segmentation [23] from histopathological images. When it comes to melanoma histopathological segmentation, consideration should be taken to accommodate different scenarios because of differentially enriched structures among tissue types. Moreover, in order to assist pathologists in a more thorough way, such as estimating tumor stage, quantifying malignancy assessment, and even determining treatment options for patients [24], more efforts should be made to combined more techniques to provide important prognostic information in pathology image analysis.

Hyperspectral imaging technology (HSI), which integrates conventional imaging and spectroscopy [25], may shed light on handling these tasks in pathology image analysis. HSI combines two dimensional spatial images with hundreds of contiguous narrow spectral bands [26] to generate a cube of three dimensional data. Therefore, informative spectral resolution allows HSI to perform fine analysis, since it can discriminate subtle spectral characteristics that conventional imaging cannot capture [27]. Advances in HSI made it possible to develop from the area of remote sensing to a wide variety of life science fields. The medical field [28], [29] especially the cancer related analysis [30]–[33], has seen the promising results when integrated with HSI. An important hypothesis is that i) tissues with similar biochemical composition are likely to have similar spectrum; and ii) differences in spectrum can be measured to represent different tissues.

To this end, the hyperspectral pathology image can be used to identify the location of melanoma and separate them from unhealthy skin tissue, potentially providing malignancy assessment. However, the large volume of 3D data imposes challenges for this segmentation task. It's instinctive to employ 3D convolutional layer on the 3D imaging data in order to fully utilize spatial information along the third dimension. Such idea has already been verified on many 3D medical imaging modalities such as CT and MRI which require fine segmentation, where a lot of 3D architectures were inspired by U-net, such as 3D-Unet [34], recurrent residual U-Net [35] and V-net [36]. Moreover, a recent study [37] reported that 3D U-net outperformed 2D U-net on most of 3D medical image segmentation tasks in the context of sufficient datasets. Since then, lots of adaptions have been made to accommodate various segmentation scenarios. For instance, Li *et al.* [38] proposed a hybrid densely connected U-net composing of a 2D DenseUNet and 3D counterpart to perform liver and tumor segmentation on CT scans; Ibtehaz and Rahman [39] developed an architecture termed MultiResUNet which improved performance relatively on multimodal medical images on the benefit of multi-resolution analysis. On the other hand, deep learning based methods have been applied on hyperspectral imaging especially on remote sensing datasets where the biggest concern is the limited data for training 3D CNN networks. Therefore, the majority of studies focus on the patch-based method which means classifying the pixel based on its surrounding pixels [12]. This allows construction of light CNN architectures with fewer parameters to be learned. In this regard, the segmentation task may be seemed as a classification

task where the typical network is a feature extraction module followed by a fully connected layer [40], [41]. Many techniques derived from the 2D classification have been transferred into 3D and incorporated with hyperspectral imaging, such as the residual network [42], [43], dense network [44], multiscale module [45], attention module [46], etc. Although obtaining reasonable accuracy, this patch-based method suffers from repeated computation and limited local features. Recently, it was recognized that the U-net based architectures were useful for hyperspectral imaging segmentation. For instance, Ghassemi *et al.* [47] applied domain adaptation strategies on an encoder–decoder convolutional network for satellite image segmentation; Pan *et al.* [48] proposed an Copy Initialization Network (CoinNet) to utilize pre-trained architectures for multispectral imagery semantic segmentation. Again, to handle small sample problems, these studies required accompanying spectral reduction technique to account for very limited spectral bands.

Given the aforementioned state-of-art segmentation architectures in 3D medical imaging, we extended the trial of incorporating those models into hyperspectral imaging technique and making essential modifications to adapt to hyperspectral pathological images in this study. First of all, the differentially enriched structures of skin tissue in hyperspectral pathology images pose challenge on network architecture to preserve finer features as the network goes deeper. We construct a dual path at the last encoder-decoder module to extract multi-scale features and one of the path uses dilated convolution to preserve fine-grained features at the low resolution. Meanwhile, how to integrate features from different modules for the final segmentation is also of vital importance and, in our work, we employ a feature fusion path followed by an add layer to allow the model to adjust features instead of concatenating them directly. The last is to handle class imbalance issue in pathology images where the malign tissue often weighs more than healthy tissues since a patient will undergo more suffering if he/she is diagnosed as a healthy person. This specific scenario calls for modification of loss function so that we modify a self-adaptive loss function composed of weighted loss functions [34] and overlap metrics [36] with regard to false negative. To this end, we proposed a 3D fully convolutional neural network called Hyper-net to perform segmentation of melanoma from hyperspectral pathology image. With special care of melanoma and its spectral properties, the Hyper-net allows to segment melanoma from a huge volume of hyperspectral pathology image efficiently and precisely. Moreover, in order to provide more practical information to pathologists for clinical use, we also demonstrated how the Hyper-net could be used to show the possibility of melanoma deterioration.

## II. Methodology

### A. Overall Scheme of Melanoma Segmentation From Hyperspectral Pathology Image

The overall scheme of melanoma segmentation is summarized in Fig. 1. Firstly, we want to make a brief illustration of our hyperspectral pathology image. Each hyperspectral
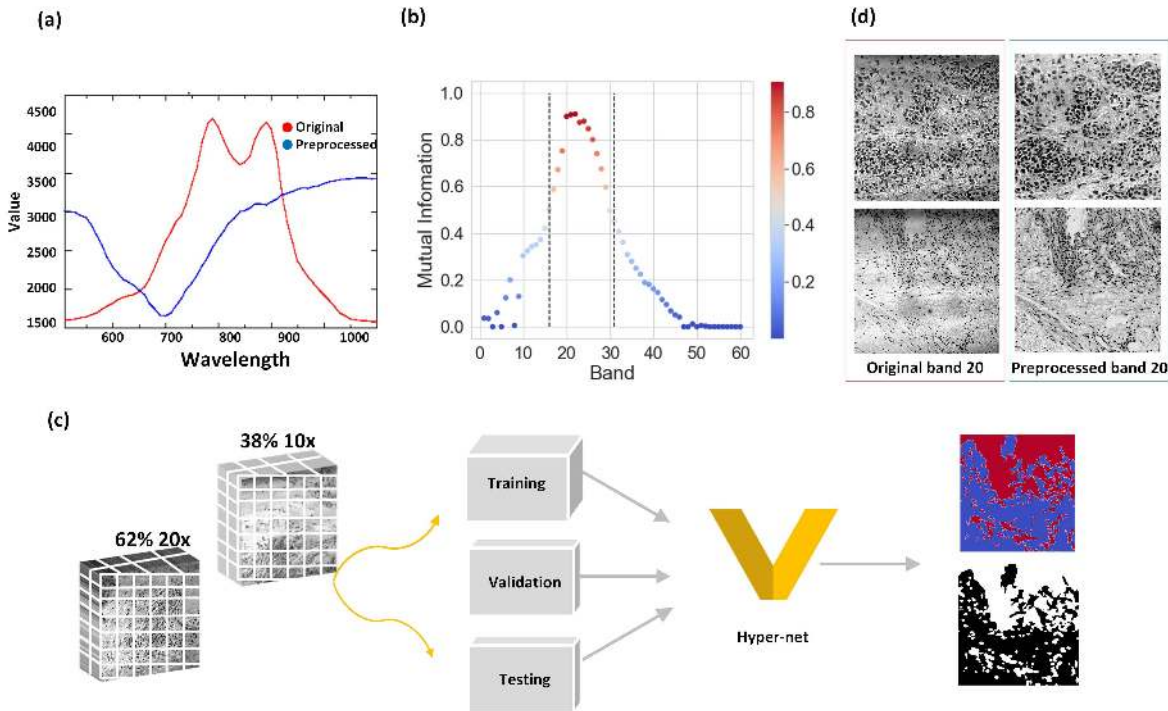
Fig. 1.    Overall scheme of melanoma segmentation from hyperspectral pathology image. (a) The spectra of melanoma before and after pre-processing. (b) The mutual information of each band in hyperspectral image. (c) Strategy for training Hyper-net: hyperspectral pathology images were separated into $256 \times 256 \times 16$ image cubes without overlapping; 66% of the dataset was used for training Hyper-net, 16% for validation and 16% for testing; probability heatmap of melanoma (red) and healthy skin tissue (blue) were derived from the binary segmentation map. (d) The 20th band of hyperspectral image before and after pre-processing.

pathology image is composed of line dimension, sample dimension, and wavelength dimension, and comes from our homemade MHSI system [49] which is designed to incorporate with microscopic system for pathology smear analysis, and ultimately for histopathology diagnosis. During image acquisition, one way of light enters the color CCD to take ordinary microscope image; the other way enters the AOTF filter, and then imaged on the grayscale CCD to collect hyperspectral image. The captured image data contained 60 bands, in the spectral range of 550-1000 nm with 7.5 nm spectral resolution. Each band contains $1024 \times 1024$ pixels$\times 12$ bit/pixel stored in BSQ format, with the spatial resolution of $3\mu$m under 20X magnification and $6\mu$m under 10X magnification. Each pixel of the 3D data cube presents an N-dimensional spectrum which can be viewed as the representation of spectrum feature for the specific substance that the pixel belongs to. The red curve in Fig. 1a shows a representation of transmittance spectrum feature of melanoma in a hyperspectral pathology image. When it comes to skin histopathology analysis, it requires comparison among different tissues such as melanoma, vessels or blood cells. Their spectra combined with spatial feature, are believed to make various tissues highly distinguishable, since those features represent the homogeneity within the same tissues and heterogeneity among different tissues.

In our datasets, patients were separated into 3 parts to make sure the relevant clinical features matched: 85 for training, 20 for validation and 20 for testing. Then, we used the MHSI system to scan each patient's biopsy specimen to capture 3 to 4 images of size $1024 \times 1024 \times 60$. To accommodate

special needs of pathologists for clinical application scenarios, we collected the raw dataset at the magnification of both 10X and 20X. Given some constrains in data acquisition and implementation, a series of pre-processing methods was applied and discussed in the following sections. Since the whole 3D data were too large to fit into the model, overlapping subsets of $256 \times 256 \times 16$ cubes were obtained from the pre-processed hyperspectral pathology images. Finally, 4760 patches of data cubes were used for training, 1116 for validation and 1120 for testing, among each dataset the distribution of magnification (62% 20X and 38% 10X) remained the same. Based on the outputs of the proposed 3D fully conventional network (Hyper-net), we illustrated a binary (normal-verse-melanoma) segmentation map and each class of the segmentation map is in company with a probability map, as shown in the right of Fig. 1c.

### B. Hyperspectral Pathology Image Pre-Processing

As mentioned, there are some constrains in data acquisition and implementation such as the emission spectra of the illumination sources, the transmission of the optics in the microscope and the detection sensitivity of the charge coupled device (CCD) camera, resulting in redundant and noisy data to some extent. We followed the band selection strategy in [50] using mutual information as the indicator. The Principle Component Analysis (PCA) was first used to select the most informative component as the reference band. The assessment of mutual information between each band and the
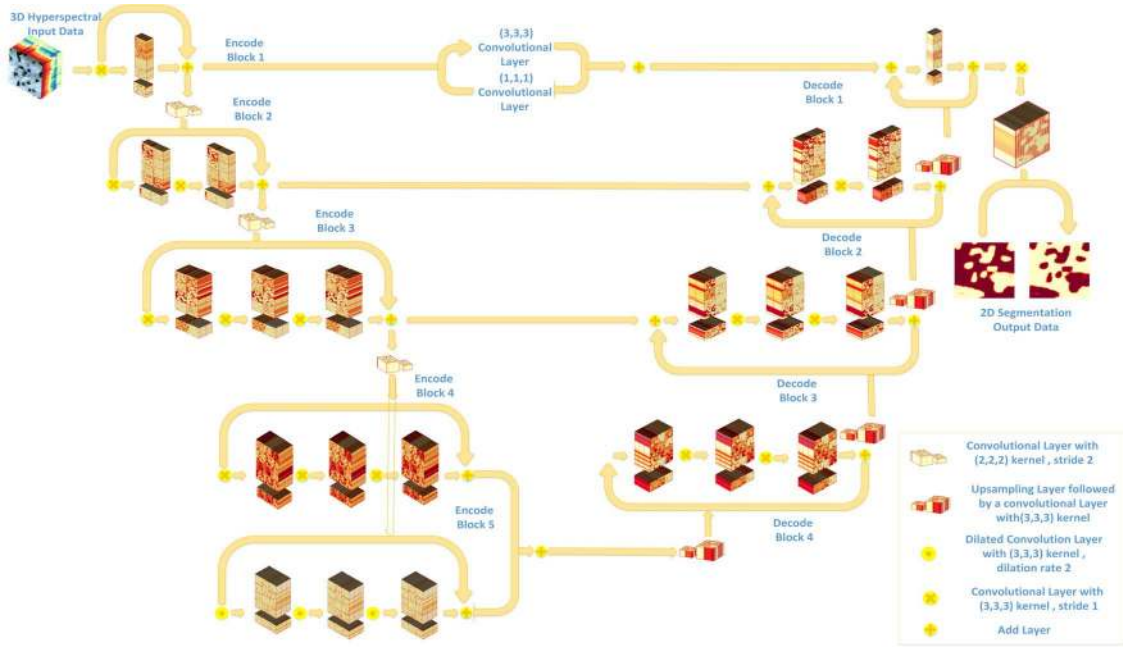
Fig. 2. The overall architecture of Hyper-net.

reference band was defined in equation (1).

$$MI_n(H_n, R) = \sum_H P(H_n, R) * \log \frac{P(H_n, R)}{P(H_n) * P(R)} \quad (1)$$

where $n \in [1, 60]$ represents the nth band, P() represents the probability distributions of grey scale of the nth band or the reference component R. Moreover, a calibration process is needed beforehand, aimed to eliminate noisy data and obtain the significant characteristic spectra of histopathology tissue. In the spectral imaging microscope under the transmitted illumination system, especially for biomedical applications, the typical method is to use the blank hyperspectral image as the reference image which is acquired at the predetermined wavelengths by imaging a coverslipped slide containing no section [51], [52]. In MHSI system, the blank image is used to characterize the electronic instrument noise. The calibration is calculated by equation (2), where R, B and D represents the raw image, the blank image and the dark image, respectively.

$$H_{i,j}^n = R_{i,j}^n - D_{i,j}^n \Big/ B_{i,j}^n - D_{i,j}^n \quad (2)$$

Fig. 1a illustrates the spectra of melanoma before (red curve) and after (blue curve) calibration. For comparative purpose, the pre-processed spectrum is multiplied by the average spectrum of the blank image. Figure 1b shows values of mutual information obtained from the pre-processed image along the band dimension compared to the reference band generated from PCA. The highest consecutive sixteen bands from 670nm to 783nm (band 16 to band 31) illustrated in Fig. 3a were selected and used for the segmentation task whereas the subset of unselected band were also shown in Fig. 3b for comparison. These unselected bands were excluded from the original data due to severe damage which may destroy the model's stability.

## C. Architecture of the Proposed 3D Convolutional Network

The network is built on the novel encoder-decoder structure but with elaborated adaption in respect to the enriched pathological melanoma tissue. The overall architecture is illustrated in Fig. 2. The role of the encoder network (left part) is to provide nonlinear local-to-global representation of the hyperspectral pathology cubes; contrastingly, decoder network (right part) is designed to supplement encoder network by reconstructing nonlinear global-to-local representation. One of the adaptions lies in the last encoder-decoder module which constructs a dual path to extract fine-grained multi-scale features at the low resolution. The other is the fusion path between these two parts, depicted by the long horizontal lines in the middle of Fig. 1, which allows the network to integrate multi-scale features from different paths to yield precise segmentations.

The encoder network consists of five successive encode blocks working at four different resolutions; specifically, the last two blocks compose a dual path module to extract multi-scale features at the same resolution. Along the encoder network, the number of feature channels is doubled. While the size of resolution is halved, since each encode block is followed by a "down sampling" layer except the last two blocks. In this way, more global representations are learned by the model which can be seen from the activation output extracted from a sample illustrated in each block in Fig. 2. However, as the resolution goes down, the global features are gained at the cost of some fine-grained features which cannot be completely compensated through the decode blocks; meanwhile, these fine-grained features are of vital importance especially for skin pathological images since they help the network to distinguish enriched skin tissues. To preserve these fine-grained features for the final segmentation map,

we removed the last down sampling layer to enable the last two blocks to work at the same resolution. In addition, the model could deal with multi-scale features, which is implemented by replacing one of convolutions in the dual-path module with dilated convolution. The dilated convolution has been verified in hyperspectral image to learn more global features without sacrificing image resolution [53]. Despite its advantage, dilated convolution will cause gridding artifacts when a feature map has higher-frequency content than the sampling rate of the dilated convolution [54]. Therefore, it's suggested to combine the dilated convolution with the standard convolution [55]. Theoretically, the dual-path module can be constructed at any layer to help preserve the fine-grained features. For the sake of accuracy, we tried to minimize the parameters of the model as much as possible to prevent it from overfitting, so the dual-path was constructed on the last layer. One of the practical strategies that we implemented in each block is the residual learning which has been inspired and confirmed by V-net [36] that significantly enhanced the training efficiency. The input of each block goes through a series of 3D convolutional layers until reaching the output; then, the output directly combines the input and the last convolutional layer (shown as the "+" sign in Fig. 2), so as to realize the residual learning. Another important component is the "down sampling" layer which is implied by a 3D convolutional layer with the kernel size of $2 \times 2 \times 2$ with stride 2. The idea of using the convolutional layer instead of the max-pooling layer in traditional CNN comes from the all convolutional net [56] in terms of significantly reducing amount of the computation but still keeping high accuracy. Also, the use of "down sampling" layer enables the model to learn global features in a wider receptive field.

The decoder network is composed of four decode blocks with each undergoing three layers in sequence: up sampling layer, add layer and a series of symmetric convolution layers. Also, given the characteristics of hyperspectral pathology images, specific concerns have been taken when decode blocks were assembled. We used the up sampling layer suggested by [57] instead of the de-convolution layer in V-net to avoid the checkerboard artifacts which are beyond serious in hyperspectral pathology images. The add layer is used to replace the concatenation layer to reduce the parameters so as to make the network compact enough to avoid overfitting. To this end, decoder network helped the model to reconstruct local features from global features and finally increase the resolution until it's the same with input data. After the up sampling layer with stride 2, the resolution is doubled. Then, the add layer integrates these features with the fine-grained, high-resolution features from the encoder network. However, it's found that directly joining features from encoder and decoder part can cause problem sometimes [39] since these features undergo different number of convolution layers. This observation exists in our hyperspectral pathology images and is extremely serious in the highest resolution encoder-decoder block since the gap of processing between two blocks is the largest. Given this, we implemented a fusion path to automatically adjust the disparity from the encoder block to the decoder block at the highest resolution. The fusion path consists of only two convolution layers with kernel size of $1 \times 1 \times 1$ and $3 \times 3 \times 3$, working

as the smallest residual module. We intended to minimize the fusion path for the concern that features from the encoder part should be preserved as completely as possible. Finally, the same strategy of residual learning is applied to a series of convolution layers to extract the most representative features.

The output of the decoder network is then convoluted by $1 \times 1 \times 1$ kernel size with stride 1, generating a 3D cube with the same size of input data. In order to output the 2D segmentation map, the 3D cube is computed by a 2D convolution layer followed by a softmax layer.

### D. Learning Strategy for the Proposed Network

Once the model structure has set up, for any batch of hyperspectral input data, the model outputs the batch of 2D segmentation prediction accordingly. The error between prediction and ground truth is where the model is trained to learn from. This is what we called "loss function" and therefore, the goal of training is to minimize the error or the loss function as much as possible. An appropriate loss function is of vital importance to the whole model since it offers a way of measuring how well the model can learn to represent the data. One of the most common loss function in image segmentation is cross entropy [58] which is based on the distribution of predicted and truth samples; another popular category is the overlap measure based loss functions which assess overlap between two samples. Recently, researchers found that for medical image segmentation the performance of overlap-based functions is unstable when segmenting small and large objects. But the compound loss functions of the cross entropy loss and the overlap measure functions show more stability [59], [60]. Given the variants of shapes and scales of melanoma in pathology, we applied the idea of compound loss function for training the Hyper-net with good robustness. In terms of pathological image analysis, it is not enough only to focus on the overall accuracy which means treating any kind of error equally. Considerations should be taken to the cases where the malignant tissue is identified as normal tissue, i.e., the false negative, since it may result in serious consequence if patients miss the best therapy timing. Based on the idea, we use the weighted cross entropy to make the network sensitive to false negative cases. The weights are regularized by a coefficient $\alpha$ which should be larger than 1 to reduce the false negative [13]. Instead of setting the coefficient $\alpha$ as a practical hyper parameter, we introduce a strategy to update $\alpha$ automatically which is inversely proportional to the ratio of false negative to all the false cases (defined in equation (3)). For the overlap based part, Lov´asz-Softmax loss [61] is used in terms of ease of training and converge. The compound loss function is shown in equation (4).

$$\alpha = \frac{(1-p, q) + (p, 1-q)}{(p, 1-q)} \tag{3}$$

$$F_{Loss}(p, q) = F_{WCE}^{\alpha}(p, q) + F_{Lovasz}(p, q) \tag{4}$$

where p, q represents the predicted label and ground truth, respectively.

Other empirical training strategies includes Batch Normalization [62] which is used before each convolutional layer to adjust and scale the activations to speed up training and

nonlinear ReLu activation [63] which is applied after convolutional layer throughout the network. Besides, the Adam optimizer [64] with learning rate of 0.001 is used with a batch size of 8 samples per epoch. To have a fair comparison, we adapted the early stopping strategy for each model. Training would be terminated if the validation loss remains the same or even decreases over three epochs, since this is actually one of the signals of overfitting.

## III. EXPERIMENTS AND RESULTS

In this study, skin tissues were collected from melanoma patients including children, adolescents, and adults with the age between 7 and 65 that underwent clinical examination at the Department of Dermatology, Wenzhou Hospital, Wenzhou, China. A total of 125 patients who were advised to skin biopsy examinations were clinically diagnosed with melanoma. The 125 patients were firstly separated into three independent datasets within which the age, gender, disease degree, and other relevant clinical features matched: 85 for training, 20 for validation and 20 for testing. Then, hyperspectral images and RGB images were taken within each dataset accordingly, which ensures that our model is never trained, validated and tested on the same patient and patches. After these tissues were dyed by hematoxylin and eosin (H&E) staining and prepared into biopsy specimen by clinician, our homemade MHSI system was operated for getting hyperspectral pathology image and corresponding RGB images of the same field of view by controlling a switch integrated on microscope, at the magnification of both 10X and 20X. These captured 3D image data were composed of 60 bands with $1024 \times 1024$ pixels $\times$ 12 bit/pixel per band stored in BSQ format. Each image's corresponding ground truth was generated directly on the hyperspectral images using self-developed software under the supervision of two professional pathologists. First, the 20th band of one hyperspectral image and RGB image in the same field of view were shown on the software interface. One of the pathologists defined the melanoma area on the RGB image. We depicted the contour on the single band image and save it as binary label image. Then, the label image and RGB image were shown and reviewed by the other pathologist. In case of disagreement, the image would be reviewed by one or more other expert pathologists. According to the pre-processing procedure, a subset of spectra was selected based on the value of mutual information. Therefore, consecutive sixteen bands from 670nm to 783nm (band 16 to band 31) were selected lining up with relatively higher value of mutual information. Because the entire hyperspectral pathology image was too large to be directly fed into the model, each was divided into $256 \times 256 \times 16$ image cubes with overlapping of 64 pixels by row and column. This overlapping extraction strategy benefits in two ways, working as a kind of augmentation method in training to gain more generality; in testing, solving patch edge artifacts by smoothing method when concatenating adjacent patches. The annotations were firstly performed by each image in the size of $1024 \times 1024$ and then were divided into patches in the size of $256 \times 256$ in accord with hyperspectral pathology image cubes. Finally, 4760 patches were used for training, 1116 for validation and 1120 for testing.

### TABLE I
ACCURACY PERFORMANCE OF 2D CNN AND HYPER-NET WITH DIFFERENT MODIFIED LOSS FUNCTIONS

| Model | 10x | 20X | Overall |
|---|---|---|---|
| 2D CNN (RGB) | 90.34±4.78 | 85.65±8.69 | 88.00±7.40 |
| 2D CNN (Hyper) | 88.38+8.95 | 90.15±3.21 | 89.27±6.78 |
| V-net (Hyper) | 91.02±4.98 | 91.00±3.78 | 91.01±4.42 |
| Hyper-net (Lov´asz-Softmax) | 89.51±5.44 | 88.04±4.19 | 88.77±4.91 |
| Hyper-net (cross entropy) | 90.71±5.88 | 92.74±2.08 | 91.73±4.53 |
| Hyper-net (compound) | 92.50±4.37 | 92.84±2.29 | 92.67±3.49 |

### TABLE II
PERFORMANCE OF 2D CNN AND HYPER-NET WITH DIFFERENT MODIFIED LOSS FUNCTIONS

| Model | FPR | TPR | Dice |
|---|---|---|---|
| 2D CNN (RGB) | 0.18±0.21 | 0.86±0.11 | 0.84±0.16 |
| 2D CNN (Hyper) | 0.09+0.12 | 0.81±0.14 | 0.86±0.11 |
| V-net (Hyper) | 0.16±0.16 | 0.90±0.12 | 0.88±0.12 |
| Hyper-net (Lov´asz-Softmax) | 0.10±0.17 | 0.78±0.17 | 0.84±0.14 |
| Hyper-net (cross entropy) | 0.08±0.17 | 0.85±0.12 | 0.90±0.10 |
| Hyper-net (compound) | 0.15±0.16 | 0.92±0.10 | 0.91±0.09 |

### A. Performance Among Different Imaging Modalities

In order to test the feasibility of using hyperspectral pathological image for melanoma segmentation, we constructed a 2D CNN model which is similar to the 2D structure of V-net, where the filters only move in two dimensions, and used RGB and hyperspectral image as input. In this way, the 2D CNN views the hyperspectral image as a normal 2D image with multiple channels. Therefore, when the image cube was fed into the model, these 2D filters averaged the value along the spectral axis, resulting in the reduction of most valuable information inherent in spectral dimension. For hyperspectral image, we employed a 3D CNN model which is similar to the structure of V-net. The comparison was conducted among 2D CNN (RGB), 2D CNN (Hyper), V-net (Hyper) and the proposed Hyper-net which were all trained using cross entropy loss function. As shown in Table I, the V-net with 3D hyperspectral image outperformed 2D CNN model for both RGB and 2D hyperspectral image in all magnifications and exhibited better generalization ability with an overall accuracy of $91.02 \pm 4.42$ compared to $88.00 \pm 7.40$ of RGB and $89.27 \pm 6.78$ of 2D hyperspectral image. The proposed Hyper-net model made a further enhancement to $91.73 \pm 4.53$. Meanwhile, Table II demonstrated that the dice score increased from $0.84 \pm 0.16$ of RGB to $0.86 \pm 0.11$ of 2D hyperspectral image and then increased from $0.88 \pm 0.12$ of V-net to $0.91 \pm 0.09$ of the proposed Hyper-net. One thing can be learned is the spectral information along with the spatial information enable hyperspectral imaging to be helpful in melanoma pathology segmentation. Moreover, it should be noted that since our current segmentation is a binary class normal-verse- melanoma task, the difference in performance
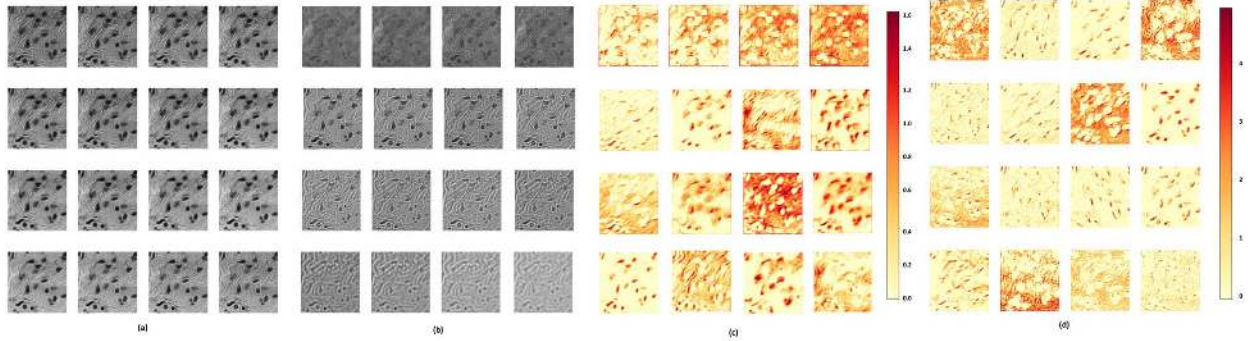
Fig. 3. Comparison of the activation output from the first convolutional layer of (c) Hyper-net and (d) 2D CNN. (a) band 16 to band 31 of the preprocessed hyperspectral images (b) subset of the preprocessed hyperspectral images that was not selected by the band selection method (c) 16 activation outputs were selected from $16 \times 256 \times 256 \times 16$ outputs of the first layer of Hyper-net (d) 16 activation outputs from $16 \times 256 \times 256$ outputs of the first layer of 2D CNN.

of the two models, which in this case was the accuracy, was not significant. But if we dig into these two models, it can be found that filters from 3D CNN captured more detailed and subtle features than that of 2D CNN, which can be seen from Fig. 3. Figure 3 compares the activation output of the first conventional layer from the two models since visualizing the output by the convolution layer given that a certain input helps us understand what are the important features that the network has learned. It's clearly that the 3D CNN model detected more specific features such as edge feature (the second and third row of Fig.3c), texture feature (the second column of Fig.3c) or spectral feature (the first row of Fig.3c) than that of 2D CNN. This was because pixels at the same location presented varying spectral values while the spatial features remained the same. Therefore, we believe that when it comes to handle more challenging and delicate segmentation tasks, for example, distinguishing different types of melanoma in skin pathology images, the proposed 3D CNN model still preserved the ability to perform well by the help of specific features captured by those various filters.

### B. Performance of the Modified Loss Function

We evaluated the performance of the proposed Hyper-net model in terms of loss function described in the method part. In this study, we proposed a compound loss function which was made up of the weighted cross entropy loss and the overlap based loss function. The weighted cross entropy loss was used to guide the network to reduce false negative rate where the weights can be regularized by a coefficient $\alpha$. Instead of setting the coefficient $\alpha$ as a practical hyper parameter, we introduced a strategy to update $\alpha$ automatically by the inverse ratio of false negative to all the false cases. Lov´asz-Softmax loss was added as the overlap based part for the sake of stability and converge. In order to test the influence of the proposed loss function, we set up three Hyper-net models which were trained with the Lov´asz-Softmax loss, the cross entropy loss and the compound loss function, respectively.

As shown in Table I, the Hyper-net (Lov´asz-Softmax) had the worst accuracy of $88.77 \pm 4.91$ which was even worse than $89.27 \pm 6.78$ of 2D CNN (Hyper); it may be explained that even though the overlap based losses worked well in

many medical image segmentation applications, they were not feasible for hyperspectral pathology images with more fine-grained tissue structures. The compound loss function outperformed the cross entropy loss in both magnifications and achieved the overall accuracy of $92.67 \pm 3.49$ and the dice score of $0.91 \pm 0.09$. Despite the enhancement in accuracy and dice score, we also compared changes in the percentage of false negative and false positive in the three models to see if the loss functions had guided the model to learn as strategized. As shown in Fig. 4c, the false negative rate of the model using the compound loss function has dropped nearly 66% and 54% respectively compared to the Hyper-net models trained with the Lov´asz-Softmax loss and the cross entropy loss. Nevertheless, we had to admit that there existed a trade-off between the false negative and false positive rate that the false positive rate increased about 43% and 58%. The influences on magnitude were plot in Fig.4b and Fig.4d. For an overall comparison, we plot the ROC curve. Again, the results confirmed that designing a loss function concerning a specific issue was essential for training a model to handle that specific issue and in our work of hyperspectral pathology segmentation. We proposed the compound loss function to make the Hyper-net more sensitive to false negative while maintaining the trade-off with false positive since the model is regularized by an adaptive coefficient $\alpha$.

In addition, we did three ablation studies to justify the modification of the proposed Hyper-net. In the first ablation study, we constructed the dual-path on all of four layers. The second ablation study was used to show the effect of replacing all convolution with dilated convolution. We did the third ablation study by using the max pooling layer instead of convolution. The performances of all three ablation study were recorded in the Table III and compared with the proposed Hyper-net. The ablation studies showed that the modifications applied in Hyper-net were effective adaptions.

### C. Visualization of Melanoma Segmentation

Usually, numerical measures can be used to reflect and compare global performance among different models; the technique of visualization, especially in medical imaging field, is another commonly used complementary measure to intuitively demonstrate outcomes. The results can be used in
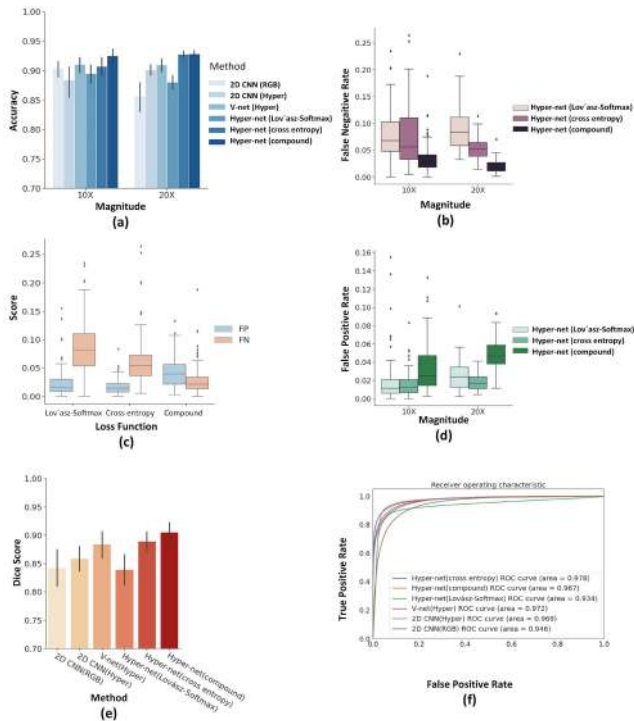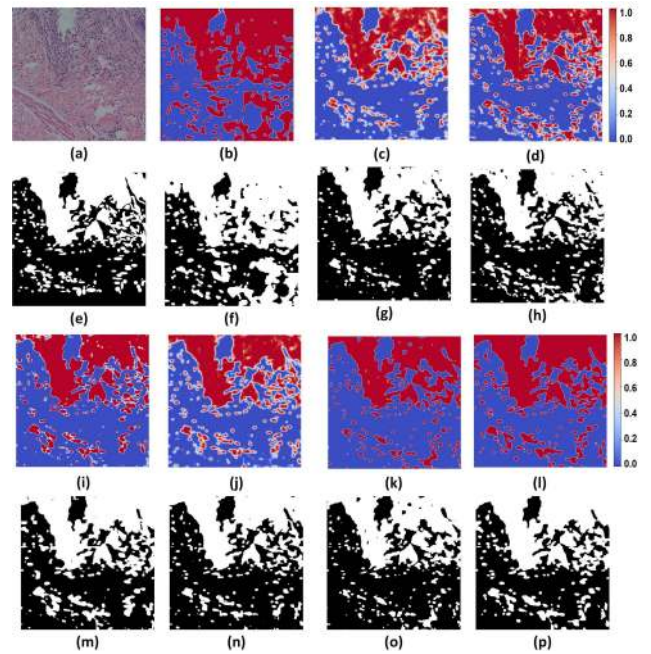
Fig. 4. Performance of CNN models with different imaging modalities (a) The overall accuracy of CNN models with different imaging modalities (b) The false negative rate of the three loss functions in terms of magnitude (c) The overall false negative rate and false positive rate of Hyper-net with modified loss function (d) The false positive rate of the three loss functions in terms of magnitude (e) The Dice score of CNN models with different imaging modalities (f) The ROC curve and AUC of CNN models with different imaging modalities.

TABLE III
PERFORMANCE OF HYPER-NET WITH DIFFERENT ABLATION STUDIES

| Model | Accuracy | FPR | TPR | Dice |
|---|---|---|---|---|
| Hyper-net(Max Pooling) | 88.08±6.87 | 0.115±0.19 | 0.778±0.18 | 0.830±0.16 |
| Hyper-net(All dual-path) | 90.60±4.24 | 0.113±0.19 | 0.839±0.13 | 0.871±0.12 |
| Hyper-net(All dilated convolution) | 91.27±4.06 | 0.106±0.19 | 0.840±0.16 | 0.874±0.12 |
| Hyper-net(proposed) | 92.67±3.49 | 0.15±0.16 | 0.92±0.10 | 0.91±0.09 |

turn to help understand models' ability. Therefore, we generated outputs from aforementioned modalities and models and exhibited in both segmentation map and probability map. Figures. 5 and 6 demonstrate the outputs derived from 10X and 20X pathology images predicted by aforementioned seven modalities and models along with corresponding RGB and ground truth images. The last logits layer of CNN model is a 2D convolution layer. It produces raw prediction values for each class, which is melanoma or healthy tissue in this scenario, as real numbers ranging from [−infinity, +infinity]. These raw prediction values for each class were constrained into the range of [0, 1] and then visualized by heatmap to plot the probability map. Then the final segmentation map was generated on each pixel by comparing values from the two classes and choosing the one with higher prediction value as the predicted class for this pixel. First, we fed the hyperspectral images without being preprocessed into Hyper-net model with



Fig. 5. Segmentation map and probability map s from 20X hyperspectral pathology images predicted by different CNN models with corresponding RGB image (a) and ground truth (e). Probability map (b) and segmentation map (f) from Hyper-net model with original hyperspectral image without being preprocessed. Probability map (c) and segmentation map (g) from 2D CNN with hyperspectral image. Probability map (d) and segmentation map (h) from 2D CNN with RGB image. Probability map (i) and segmentation map (m) from V-net with hyperspectral image. Probability map (j) and segmentation map (n) from Hyper-net with Lov´asz-Softmax loss function. Probability map (k) and segmentation map (o) from Hyper-net with cross entropy loss function. Probability map (l) and segmentation map (p) from Hyper-net with compound loss function.

the results shown in Fig. 5b, f and Fig. 6b, f. Compared to the ground truth, it can be seen that some tissues especially those on the boundary of the field of view were misclassified since there were artifacts and noise in the unprocessed raw data as shown in Fig. 1d. Given the current condition of our MHSI system, it's necessary to include the pre-processing step in our network. Next comes the visualized outputs of two 2D CNN models which were trained with RGB (c and g) and hyperspectral images (d and h), respectively. Although the accuracy of these two models was nearly the same as summarized in Table I, the RGB based model generated more clear boundary than that of 2D CNN (Hyper) since RGB images were able to reflect the difference among various tissues by the color than single band gray-scale hyperspectral images. We also obtained outputs (i an m) of V-net model trained with 3D hyperspectral images. Compared to 2D CNN models, V-net model performed better in determining the border whereas it failed in segmenting small objects compared with our Hyper-net model. To some extent, it was proved that the modification in architecture of Hyper-net worked in extracting fine-grained features. Last, the visualized outputs were compared among Hyper-net trained with different loss functions which were Lov´asz-Softmax (j and n), cross entropy (k and o) and the compound (l and p). As the Lov´asz-Softmax is the overlap measure based loss functions, it outputted more complete and compact segmentation but less clear border than the cross entropy loss; however, the cross entropy loss which
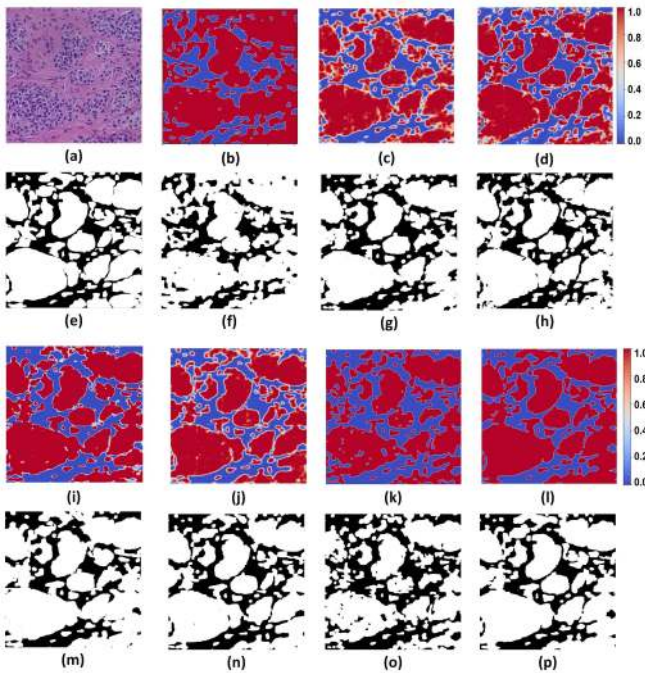
Fig. 6. Segmentation map and probability map s from 20X hyperspectral pathology images predicted by different CNN models with corresponding RGB image (a) and ground truth (e). Probability map (b) and segmentation map (f) from Hyper-net model with original hyperspectral image without being preprocessed. Probability map (c) and segmentation map (g) from 2D CNN with hyperspectral image. Probability map (d) and segmentation map (h) from 2D CNN with RGB image. Probability map (i) and segmentation map (m) from V-net with hyperspectral image. Probability map (j) and segmentation map (n) from Hyper-net with Lovász-Softmax loss function. Probability map (k) and segmentation map (o) from Hyper-net with cross entropy loss function. Probability map (l) and segmentation map (p) from Hyper-net with compound loss function.

is pixel-based, generated clearer border but more fragmentary segmentation map. Fortunately, the designed compound loss function exhibited benefits of focusing on small objects and fine-grained features as well as integrated and clear boundary. Nevertheless, the way of visualization is diverse and easy to implement as long as the segmentation is generated; therefore, it is up to pathologists to choose the best way to assist their diagnosis. From a researcher's perspective, the key value of visualization lies in the practical implementation of segmentation results and in providing a method that will allow us to adjudicate among alternative models.

## IV. CONCLUSION

Overall, our work demonstrated that the 3D convolutional neural network can be used to segment melanoma in hyperspectral pathology images, therefore assisting pathologists in determining melanoma deterioration. Melanoma can be identified from healthy tissue under both 10X and 20X magnification with accuracy more than 92%. Also, to improve the sensitivity of diagnosis, we proposed the Hyper-net neural network by making specific modification of loss function, leading to enhanced performance of greatly reduction in false positive and false negative predictions. By comparing with the 2D CNN model, we demonstrated that hyperspectral pathology images contained rich informative properties of tissues so

as to improve the segmentation results and assure diagnosis robustness. Although the accuracy was reasonable, our work only covered limited diversity of tissues in skin pathology. In addition, there are many other complex and rare tissues that may need pathologists' professional knowledge such as cyst, necrosis and inflammation. In terms of our MHSI system, we intend to upgrade the current MHSI system to enlarge the spectra range so as to include spectra between 400 and 750, and to consider more significant features to enhance the quality of these channels. Last but not least, we have to admit that there is a long way before the hyperspectral pathology system could be used in clinical, that's why researchers could not stop exploring various ways of probability. Obviously, deep networks opened a new insight into medical field with carefulness and consideration on clinical situations.

## REFERENCES

[1] C. Karimkhani et al., "The global burden of melanoma: Results from the global burden of disease study 2015," Brit. J. Dermatol., vol. 177, no. 1, pp. 134–140, Jul. 2017.

[2] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, p. 115, 2017.

[3] A. M. Glazer, R. R. Winkelmann, A. S. Farberg, and D. S. Rigel, "Analysis of trends in US melanoma incidence and mortality," JAMA Dermatol., vol. 153, no. 2, pp. 225–226, 2017.

[4] S. W. Menzies et al., "Dermoscopic evaluation of amelanotic and hypomelanotic melanoma," Arch. Dermatol., vol. 144, no. 9, pp. 1120–1127, 2008.

[5] D. C. Allen, "Malignant melanoma BT," in Histopathology Reporting: Guidelines for Surgical Cancer, D. C. Allen, Ed. London, U.K.: Springer, 2013, pp. 207–216.

[6] X. Bai et al., "Learning ECOC code matrix for multiclass classification with application to glaucoma diagnosis," J. Med. Syst., vol. 40, no. 4, p. 78, Apr. 2016.

[7] Q. Wang, Q. Li, M. Zhou, L. Sun, S. Qiu, and Y. Wang, "Melanoma and melanocyte identification from hyperspectral pathology images using object-based multiscale analysis," Appl. Spectrosc., vol. 72, no. 10, pp. 1538–1547, 2018.

[8] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," IEEE Access, vol. 6, pp. 9375–9389, 2018.

[9] J. Ker, S. P. Singh, Y. Bai, J. Rao, T. Lim, and L. Wang, "Image thresholding improves 3-Dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans," Sensors, vol. 19, no. 9, p. 2167, May 2019.

[10] J. Ker, Y. Bai, H. Y. Lee, J. Rao, and L. Wang, "Automated brain histology classification using machine learning," J. Clin. Neurosci., vol. 66, pp. 239–245, Aug. 2019.

[11] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," IEEE Trans. Med. Imag., vol. 35, no. 2, pp. 550–566, Feb. 2016.

[12] T. de Bel, M. Hermsen, B. Smeets, L. Hilbrands, J. van der Laak, and G. Litjens, "Automatic segmentation of histopathological slides of renal tissue using deep learning," Proc. SPIE, vol. 10581, Mar. 2018, Art. no. 1058112.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 3431–3440.

[14] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, "Stepwise integration of deep class-specific learning for dermoscopic image segmentation," Pattern Recognit., vol. 85, pp. 78–89, Jan. 2019.

[15] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," Comput. Methods Programs Biomed., vol. 162, pp. 221–231, Aug. 2018.

[16] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance," IEEE Trans. Med. Imag., vol. 36, no. 9, pp. 1876–1886, Sep. 2017.

[17] A. Phillips, I. Teo, and J. Lang, "Segmentation of prognostic tissue structures in cutaneous melanoma using whole slide images," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jun. 2019, pp. 2738–2747.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[19] G. M. Venkatesh, Y. G. Naresh, S. Little, and N. E. O'Connor, "A deep residual architecture for skin lesion segmentation," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. London, U.K.: Univ. College London, 2018, pp. 277–284.

[20] S. Vesal, N. Ravikumar, and A. Maier, "SkinNet: A deep learning framework for skin lesion segmentation," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Nov. 2018, pp. 1–3.

[21] E. Z. Chen, X. Dong, X. Li, H. Jiang, R. Rong, and J. Wu, "Lesion attributes segmentation for melanoma detection with multi-task U-Net," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 485–488.

[22] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: http://arxiv.org/abs/1802.06955

[23] K. R. J. Oskal, M. Risdal, E. A. M. Janssen, E. S. Undersrud, and T. O. Gulsrud, "A U-Net based approach to epidermal tissue segmentation in whole slide histopathological images," *Social Netw. Appl. Sci.*, vol. 1, no. 7, p. 672, Jul. 2019.

[24] N. Coudray *et al.*, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," *Nature Med.*, vol. 24, no. 10, pp. 1559–1567, 2018.

[25] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[26] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.

[27] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[28] M. A. Calin, S. V. Parasca, D. Savastru, and D. Manea, "Hyperspectral imaging in the medical field: Present and future," *Appl. Spectrosc. Rev.*, vol. 49, no. 6, pp. 435–447, Aug. 2014.

[29] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, p. 10901, Jan. 2014.

[30] M. Halicek *et al.*, "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging," *J. Biomed. Opt.*, vol. 22, no. 6, p. 60503, 2017.

[31] H. Fabelo *et al.*, "Deep learning-based framework for *in vivo* identification of glioblastoma tumor using hyperspectral images of human brain," *Sensors*, vol. 19, no. 4, p. 920, Feb. 2019.

[32] R. Pike, G. Lu, D. Wang, Z. G. Chen, and B. Fei, "A minimum spanning forest-based method for noninvasive cancer detection with hyperspectral imaging," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 653–663, Mar. 2016.

[33] G. Lu *et al.*, "Detection and delineation of squamous neoplasia with hyperspectral imaging in a mouse model of tongue carcinogenesis," *J. Biophotonics*, vol. 11, no. 3, Mar. 2018, Art. no. e201700078.

[34] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 424–432.

[35] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imag.*, vol. 6, no. 1, p. 14006, 2019.

[36] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[37] F. Isensee *et al.*, "NnU-Net: Self-adapting framework for U-net-Based medical image segmentation," 2018, *arXiv:1809.10486*. [Online]. Available: http://arxiv.org/abs/1809.10486

[38] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[39] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.

[40] Z. Niu, W. Liu, J. Zhao, and G. Jiang, "DeepLab-based spatial feature extraction for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 251–255, Feb. 2019.

[41] M. Halicek *et al.*, "Hyperspectral imaging for head and neck cancer detection: Specular glare and variance of the tumor margin in surgical specimens," *J. Med. Imag.*, vol. 6, no. 3, p. 35004, 2019.

[42] F. Cao and W. Guo, "Deep hybrid dilated residual networks for hyperspectral image classification," *Neurocomputing*, vol. 384, pp. 170–181, Apr. 2020.

[43] Y. Jiang, Y. Li, and H. Zhang, "Hyperspectral image classification based on 3-D separable ResNet and transfer learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1949–1953, Dec. 2019.

[44] C. Zhang, G. Li, and S. Du, "Multi-scale dense networks for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9201–9222, Nov. 2019.

[45] H. Gao, D. Yao, Y. Yang, C. Li, H. Liu, and Z. Hua, "Multiscale 3-D CNN based on spatial–spectral joint feature extraction for hyperspectral remote sensing images classification," *Proc. SPIE*, vol. 29, no. 1, p. 13007, 2020.

[46] H. Sun, X. Zheng, X. Lu, and S. Wu, "Spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, May 2020.

[47] S. Ghassemi, A. Fiandrotti, G. Francini, and E. Magli, "Learning and adapting robust features for satellite image segmentation on heterogeneous data sets," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6517–6529, Sep. 2019.

[48] B. Pan, Z. Shi, X. Xu, T. Shi, N. Zhang, and X. Zhu, "CoinNet: Copy initialization network for multispectral imagery semantic segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 816–820, May 2019.

[49] Q. Li, M. Zhou, H. Liu, Y. Wang, and F. Guo, "Red blood cell count automation using microscopic hyperspectral imaging technology," *Appl. Spectrosc.*, vol. 69, no. 12, pp. 1372–1380, Dec. 2015.

[50] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, Oct. 2006.

[51] R. L. Ornberg, B. M. Woerner, and D. A. Edwards, "Analysis of stained objects in histological sections by spectral imaging and differential absorption," *J. Histochem. Cytochem.*, vol. 47, no. 10, pp. 1307–1313, Oct. 1999.

[52] C. Rothmann, A. M. Cohen, and Z. Malik, "Chromatin condensation in erythropoiesis resolved by multipixel spectral imaging: Differentiation versus apoptosis," *J. Histochem. Cytochem.*, vol. 45, no. 8, pp. 1097–1108, Aug. 1997.

[53] X. Zhang, Y. Zheng, W. Liu, and Z. Wang, "A hyperspectral image classification algorithm based on atrous convolution," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–12, Dec. 2019.

[54] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 472–480.

[55] R. Hamaguchi, A. Fujita, K. Nemoto, T. Imaizumi, and S. Hikosaka, "Effective use of dilated convolutions for segmenting small object instances in remote sensing imagery," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 1442–1450.

[56] J. Tobias Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," 2014, *arXiv:1412.6806*. [Online]. Available: http://arxiv.org/abs/1412.6806

[57] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[58] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.

[59] S. A. Taghanaki *et al.*, "Combo loss: Handling input and output imbalance in multi-organ segmentation," *Computerized Med. Imag. Graph.*, vol. 75, pp. 24–33, Jul. 2019.

[60] S. Zhou, D. Nie, E. Adeli, J. Yin, J. Lian, and D. Shen, "High-resolution encoder–decoder networks for low-contrast medical image segmentation," *IEEE Trans. Image Process.*, vol. 29, pp. 461–475, Jun. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8741187, doi: 10.1109/TIP.2019.2919937.

[61] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4413–4421.

[62] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[63] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[64] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980