2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)
Venice, Italy, April 8-11, 2019

A DUAL STREAM NETWORK FOR TUMOR DETECTION IN HYPERSPECTRAL IMAGES

P.J.C. Weijtmans[a, b], C. Shan[b], T. Tan[a], S.G. Brouwer de Koning[c], T.J.M. Ruers[c, d]

[a]Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands
[b]Philips Research, Eindhoven, The Netherlands
[c]Netherlands Cancer Institute, Antoni van Leeuwenhoek, Amsterdam, The Netherlands
[d]fMIRA Institute, University of Twente, Enschede, The Netherlands

ABSTRACT

Hyperspectral imaging has become an emerging imaging modality for medical applications. In this work, we propose to combine both the spectral and structural information in the hyperspectral data cube for tumor detection in tongue tissue. A dual stream network is designed, with a spectral and a structural branch. Hyperspectral data (480 to 920 nm) is collected from 7 patients with tongue squamous cell carcinoma. Histopathological analysis provided ground truth labels. The proposed dual stream model outperforms the pure spectral and structural approaches with areas under the ROC-curve of 0.90, 0.87 and 0.85, respectively.

***Index Terms*—** Hyperspectral imaging, machine learning, neural networks, tongue tumor

1. INTRODUCTION

Patients suffering from tumors in tongue tissue are generally treated by removing the tumor tissue surgically. Removal of the complete tumor is challenging. Currently there is no reliable way to provide real-time feedback to the surgeon during the tumor removal procedures.

Hyperspectral imaging (HSI) was originally developed for remote sensing by NASA/JPL [1] and has been successfully used in multiple fields such as food quality and safety, vegetation and resource control, archaeology and biomedicine [2]. With advancements in hardware and computational power, it has become an emerging imaging modality for medical applications. HSI has the potential advantages of low cost, relatively simple hardware and ease of use. This makes HSI a candidate for intra-operative support of a surgeon. Compared to regular RGB data it is challenging to process the HSI data due to the size of the data: hundreds of color bands in a multi-megapixel image results in large files with varying amounts of redundant information.

Fei et al. [3] have evaluated the use of HSI on specimen from patients with head and neck cancer. Multiple specimens have been taken from each of the 16 patients, and the specimen have been scanned with wavelengths from 450 to 900 nm. Specimen from 10 of the 16 patients are verified to be squamous cell carcinoma (SCCa). They achieved an area under the ROC-curve (AUC) of 0.94 for tumor classification with a linear discriminant analysis. However, their testing is done on specimens from the same patient as the classification was trained on. Halicek et al. [4] acquired multiple specimen from 50 head and neck cancer patients in 450-900 nm spectral range. 29 of the patients had SCCa tumors. Using patient held-out external validation, they applied a deep convolutional network and achieved an accuracy of 77%. An animal study by Ma et al. [5] achieved an accuracy of 91.36% using convolutional neural networks in a leave-one-out cross-validation. The specimen were taken from mice with induced tumors. In all three mentioned studies the focus lies entirely on spectral information.

In this paper a combination of both the spectral and structural information in the HSI data cube is explored. Particularly a dual stream model is proposed, with a spectral and a structural branch. Data is collected from 7 patients with tongue SCCa using a HSI system (480 to 920 nm). Histopathological analysis provided ground truth labels. The proposed method is compared with pure spectral and structural approaches.

2. METHODS

The HSI scan results in a data cube for each tissue sample. Using small patches spanning all bands the spectral information can be considered. By selecting bigger patches, structural information becomes available. By combining both inputs, the full spectral and some structural information is available for the network. Figure 1 shows the architecture of the dual stream network.

2.1. Spectral and structural branch

To exploit the full extent of the spectral information, a network is designed for using small patches including the full spectrum of the HSI data. To filter out noise, the average

This study was funded by grant H2020, ECSEL-04-2015-Smart Health, Advancing Smart Optical Imaging and Sensing for Health (ASTONISH).

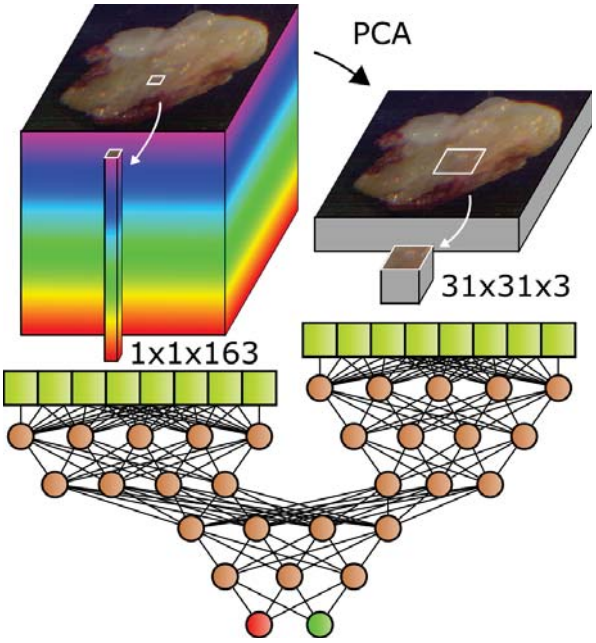

Fig. 1. From the HSI data a 1x1x163 spectral signature is fed into the spectral branch of the combined network. After applying PCA, resulting in three channels, 31x31 patches are used for the structural branch. Branches have been trained and tested individually.

spectrum of 5x5 pixels is calculated and the resulting vector is used as input for a neural network. The network has two hidden layers of 163 and 80 units and finally a softmax output layer. During training a dropout layer with rate 0.5 is used before the final softmax layer. This results in 40,014 trainable parameters. To be able to compare the dual stream network with just a spectral method, the spectral network is separately trained and tested.

By increasing the patch size, morphological features in the HSI data can be used to classify the tissue. Due to the high number of parameters in the network needed for those patches, the number of channels has been reduced using a principle component analysis (PCA). With 3 channels approximately 92% of variance is contained in the resulting 31x31x3 patches. Larger numbers of channels have been considered in the same network architecture, but did not increase performance significantly. The neural network used for the structural method flattens the input and has three fully connected layers of 1024, 128 and 64 units, a dropout layer during training and a final softmax layer. This gives 3,084,674 trainable parameters.

2.2. Dual stream network

When combining the spectral and structural information, the classifier has many features available, without the need of a 31x31x163 data cube. To use 1x1x163 and 31x31x3 patches the previous networks have been combined into a dual stream network. The spectral and structural branches are unchanged in both architecture and weights using transfer learning. Instead of the softmax output layers, the second fully connected layer of both networks is concatenated and fed into two new dense layers of 128 and 64 units. These are connected to a dropout layer and softmax for the output. The combined network has 26,946 trainable parameters.

3. EXPERIMENTS

3.1. Clinical data set

Tissue of 7 patients with tongue squamous cell carcinoma has been resected by the surgeon. Directly after resection, the specimen was brought to the pathology department, where it is processed according to standard pathological protocol. The pathologist localized the tumor by palpation and cut the specimen in two parts, right trough the middle of the tumor. Images are then taken from these new surfaces. All processes performed at the hospital follow the ethical guidelines for ex vivo human studies. An RGB picture of the tissue is taken to function as an intermediate step in the registration process.

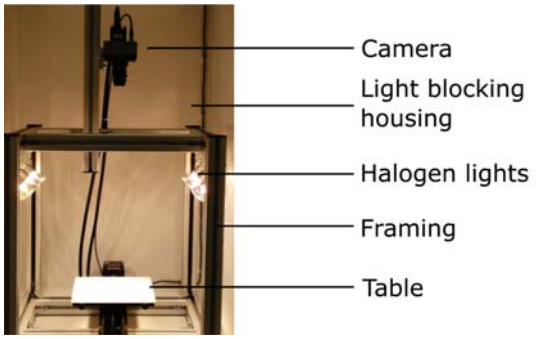

Fig. 2. The imaging system used to collect the HSI data. Tissue is placed on the Table in the bottom. The Table shifts the tissue under the sensor.

To get a hyperspectral image of the tissue, an imaging system with a line-scan sensor manufactured by IMEC (Leuven, Belgium) is used to capture the diffuse reflectance with wavelengths ranging from 480 to 920 nm in 163 channels. The imaging system is shown in Figure 2 The image is recorded line-by-line with 2048 samples per line while shifting the tissue. The system has been calibrated using a light reference image before use and after scanning the data was cropped to a region of interest.

In order to label the HSI data, a histopatological slide is taken from the surface that has been scanned. The slide is digitized and delineated to mark the tumor (red), healthy muscle (green) and epithelium (blue). This is the first step shown in Figure 3. From the delineation a mask is created.

During histopathological processing the specimen was deformed and to correct this, a non-rigid registration algorithm is used. Obvious matching points in the histopathological and RGB images were visually selected. Using these points, the mask is transformed to match the RGB picture. This is depicted in Figure 3 in middle row as transformation T1. The point-selection is done again on the RGB and HSI data to acquire transformation T2, which is used to transform the mask again to match the HSI data. Using the mask the data can be explored. Table 1 shows the number of pixels for patients and classes in the mask. Figure 4 shows the mean spectra of patients 2 and 4 for the tumor and muscle classes. At a wavelength of 630 nm is a clear drop in reflectance, which is present in all HSI images and caused by the hardware. Generally the muscle mean is lower then the tumor mean, but there is a clear overlap of the two classes. To work with the limited data set, leave-one-patient-out cross validation is used during training. To have a balanced training process a 50/50 balance for the tumor and healthy classes has been set for the patch selection. To evaluate the experiments, the area under the ROC-curve (AUC) is used. This gives a reasonable indication of performance, without choosing a threshold for the classification.

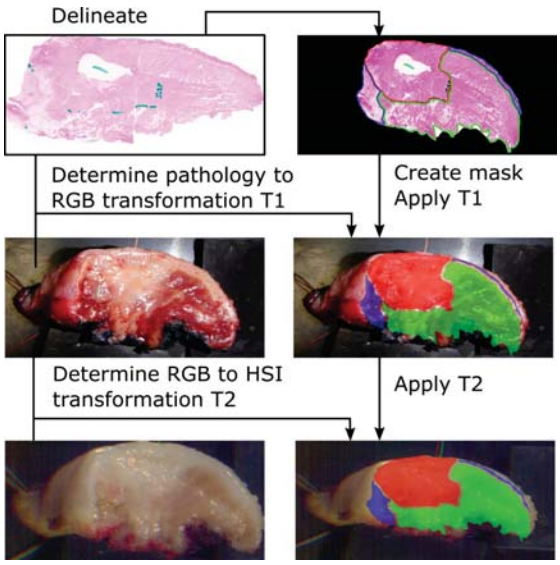

Fig. 3. Annotation of the hyperspectral data: tumor (red), healthy tongue muscle (green) and healthy epithelium (blue).

Table 1. Pixel counts of the data, in thousands

Patient	1	2	3	4	5	6	7	%
Total	498	148	861	247	181	176	146	100
Tumor	40	6	116	32	13	5	13	9.4
Muscle	34	48	71	45	56	70	42	17.1

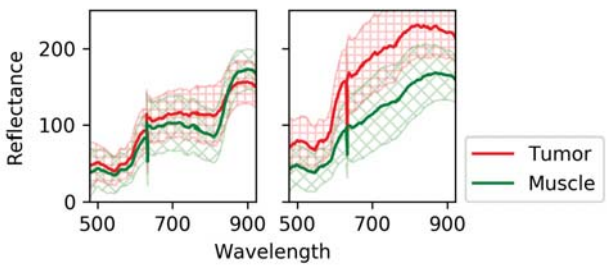

Fig. 4. Reflectance mean and standard error (vertical) versus wavelength in nm for patients 2 (left) and 4.

3.2. Experiment results

The results of our proposed method are shown in Table 2. The combination of spectral and structural streams gives a mean AUC of 0.904±0.053. A comparison with the individual features is summarized in Figure 5, showing the AUC for the three methods applied to all patients. Using the spectral input, a relatively small network achieves a mean AUC of 0.872 ± 0.050 after training for 10 epochs. Predictions of full HSI images are shown in the second row in Figure 6. The color intensities are direct representations of the predictions. This means that when colors are shown with their maximal intensity, the model makes a clear distinction. The model cannot make a strong distinction between classes in significant areas of the predictions of patients 4 and 7. The model performs the worst on patient 2, leaving much of the tumor undetected. With the structural network, using larger patches but fewer spectral bands, results are similar compared to the spectral network. The AUC is 0.850±0.077. In patient 2 large areas are misclassified as tumor but overall the result is somewhat smoother, suggesting the structural context is used by the model. The dual stream model, shown on the last row, generates the best predictions. It can clearly be seen in Figure 6 that the model makes confident decisions, despite misclassified regions. For patient 4, it is verified that most of the misclassified area is epithelia, which might have similar features as tumor.

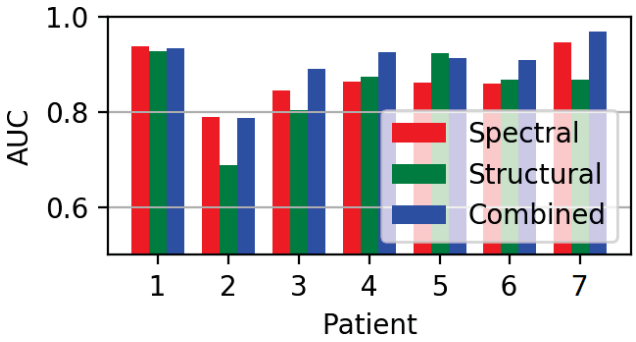

Fig. 5. Plot with the AUC in bars for all patients

Table 2. Results of the dual stream experiment

Patient	AUC	Accuracy	Sensitivity	Specificity
1	0.933	0.789	0.974	0.605
2	0.787	0.746	0.615	0.876
3	0.891	0.636	0.938	0.335
4	0.926	0.840	0.922	0.758
5	0.913	0.843	0.924	0.763
6	0.910	0.807	0.733	0.880
7	0.970	0.922	0.941	0.903
Mean	**0.904**	0.798	0.864	0.731
Stdev	0.053	0.083	0.125	0.188

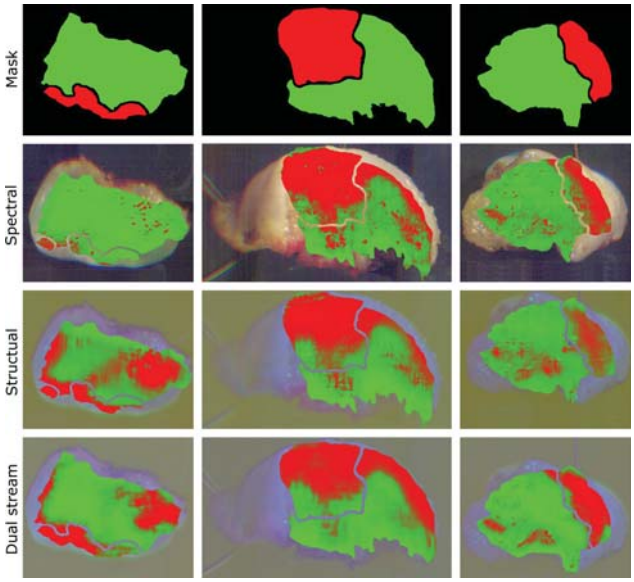

Fig. 6. The ground truth and predictions of the spectral, structural and dual stream approach of patients 2 (left), 4 (middle) and 7 (right). Tumor (red), healthy (green). Best viewed in color.

3.3. Discussion

In this work, for the first time, we make explicit use of both the structural and spectrum information by applying a dual stream neural network. Considering a small-size dataset, we strategically train our model by two steps. First, we train each stream of our model separately. Second, we add fully connected layers (FCN) to the streams and only the parameters of FCNs need to be trained. In the future, we will investigate the possibility of applying other network types such as a Long short-term memory (LSTM) model for the spectrum domain stream and convolutional networks for the structural stream. All models have been trained for 10 epochs, but performance barely increased over the epochs. This indicates that the model reaches the full capacity in an early stage. That might be caused by the limited data set, or by the fact that the network is very shallow. In the future, more data will be collected to improve the effectiveness of the model. Many of the false positives are found at the edge of the samples. This can be explained by the fact that these tumors often originate in the epithelia, and therefore the spectral features can have a resemblance. The imaging system uses a line-by-line scanning method, which makes real time applications difficult. In the future, channel selection can be performed instead of PCA to enable the manufacturer to construct a sensor with only specified wavelengths to reduce imaging time.

4. CONCLUSION

Given the limited data, it is possible to train networks that combine spectral and structural information and have a good performance on the classification of healthy and tumor tissue. From the data processing perspective, this opens the possibility of intra-operative feedback to surgeons. More elaborate networks can be studied to increase performance and efficiency.

5. REFERENCES

[1] Alexander F.H. Goetz, "Three decades of hyperspectral remote sensing of the earth: A personal view," *Remote Sensing of Environment*, vol. 113, pp. S5 – S16, 2009, Imaging Spectroscopy Special Issue.

[2] Guolan Lu and Baowei Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19, no. 1, pp. 010901, 2014.

[3] Baowei Fei, Guolan Lu, Xu Wang, Hongzheng Zhang, James V. Little, Mihir R. Patel, Christopher C. Griffith, Mark W. El-Diery, and Amy Y. Chen, "Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients," *Journal of Biomedical Optics*, vol. 22, no. 8, pp. 1–7, 2017.

[4] Martin Halicek, Guolan Lu, James V. Little, Xu Wang, Mihir Patel, Christopher C. Griffith, Mark W. El-Deiry, Amy Y. Chen, and Baowei Fei, "Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging," *Journal of Biomedical Optics*, vol. 22, no. 6, pp. 60503, 2017.

[5] Ling Ma, Guolan Lu, Dongsheng Wang, Xu Wang, Zhuo Georgia Chen, Susan Muller, Amy Chen, and Baowei Fei, "Deep learning based classification for head and neck cancer detection with hyperspectral imaging in an animal model," *Proc.SPIE*, vol. 10137, 2017.