ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Hyperspectral Wavelength Analysis with U-Net for Larynx Cancer Detection

Felix Meyer-Veit[1], Rania Rayyes[1], Andreas O. H. Gerstner[2] and Jochen Steil[1] *

1- Technische Universität Braunschweig - Inst. für Robotik und Prozessinformatik
2- Klinikum Braunschweig - ENT - Holwedestr. 16, 38118 Braunschweig - Germany

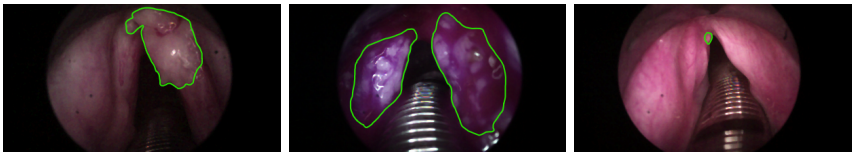
Abstract.
Early detection of laryngeal tumors is critical for their successful therapy. In this paper, we investigate how hyperspectral (HS) imaging can contribute to this aim based on an in-vivo data set of 13 HS image cubes recorded in clinical practice. We perform semantic segmentation with a tailored U-Net trained on labels provided by the clinicians. We specifically investigate the influence of exposure time during image acquisition, the suitable wavelengths to determine the most informative image channels, and present quantitative results on accuracy and the AUC measure.

1 Introduction

The larynx is the site of the majority of head and neck malignancies and early detection of laryngeal tumors is critical for therapy [1]. In order to detect cancer early and from images, the long-wavelength range is recommended because the penetration depth into the tissue is higher [2]. Hyperspectral Imaging (HSI) potentially supports capturing this relevant spectral information due to extending spectral resolution into the visible-near infrared. The HS image cubes then can be processed to indicate suspicious regions by deep learning for semantic segmentation. In this respect, Convolutional Neural Networks (CNN) have shown promising results for cancer detection [3], specifically the well-known U-Net architecture [4]. Although HSI thus is promising, there are relatively few works of Deep Learning in HSI. Eggert et al. [5] were able to predict in-vivo laryngeal cancer with a 2D CNN, and normal tissue within a spectral range of 380 nm to 680 nm in the visible spectrum and obtain an average AUC of 79%. However, longer wavelengths are favorable according to [6, 7]. Halicek et al. [8] developed a CNN classifier to diagnose head and neck cancer using pre-processed HSI with 78% AUC across a spectral range of 450 nm to 900 nm. But in contrast to [5] and our investigation, the tumor categorization was done ex-vivo. Furthermore, none of these past works used HS cube at a low spatial resolution as they are realistic in the clinical application and used in our study, nor a U-Net. In all approaches, the cost of the training and test time is still high which is a considerable challenge for medical applications like the one considered here, where ideally the medical practitioner should see the result in vivo immediately during the endoscopic examination of the larynx. In our previous work [9], we proposed

*The data set collection was funded by the German Cancer Aid within the project framework "Early detection of laryngeal cancer by means of Hyperspectral Imaging (109825110275)".

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.



(a) Low complexity. (b) Medium complexity. (c) High complexity.

Fig. 1: Examples for lesions of different difficulties, indicated by green lines.

an efficient HSI Deep Learning System for laryngeal cancer prediction, which shows great potential to be implemented as a clinical application.

In this paper, we investigate how to get information-rich images from HS imaging as well as how to determine the most relevant information in the captured images to speed up the prediction and training process. It turns out that several factors are crucial including the spectral resolution of the captured images, the time of exposure to the light source during image acquisition, and the wavelength range that carries the most relevant information. The latter is important as specifically for HS images training is very time-consuming. Furthermore, we use images directly from clinical practice that consequently display a large variance and are of varying quality (see also Fig. 1). To the best of our knowledge, this is the first work that includes a wavelength analysis as a function of the exposure time and considers diverse, clinical images in form of HS cubes at a low spatial resolution in the visible-near infrared spectrum to detect laryngeal cancer with a U-Net architecture.

2 Clinical HSI Data Set

The HSI data set was collected from 13 different patients during clinical practice and in a limited time frame, such that recording conditions could not be controlled. Different equipment was used as available and different physicians were involved. This results in a complex, but clinically very realistic data set. In particular, due to the varying power of the used halogen light sources and the varying diameter of the rigid endoscopes, the exposure time had to be adjusted accordingly, so that the bands are illuminated similarly across different HSIs. Hence, each HS image was captured at two different exposure times, with mean exposure times of 57.5 ms and 77.5 ms.

Images were captured using a hyperspectral snapshot camera[1], which has 25 spectral bands ranging from 665 to 975 nm in a 5×5 mosaic with a spatial resolution of 409×217 pixels. Ex-post analysis shows that HS images display very different conditions and for the further analysis below, we grouped images by means of visual inspection to [high, medium, low] complexity (cf. Fig. 1) representing $\approx 1/3$ of the dataset each. For all used images, the ground truth malignant property of the tumor lesion was verified through pathological investigation, and images were hand-labeled by the head physician (cf. Fig. 1).

[1] *MV0-D2048x1088-C01-HS02-160-G2*, Photonfocus AG, Switzerland

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

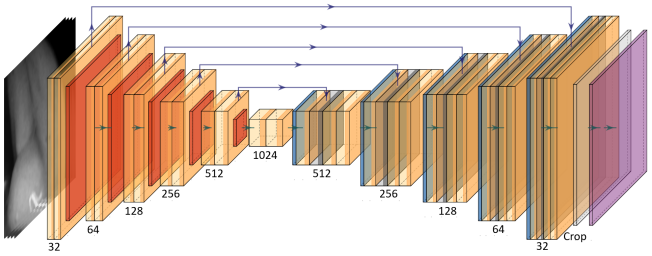


Fig. 2: Modified U-Net architecture for 4 channels, visualized using [10].

3 Methods

Pre-Processing and Data Augmentation: First, the HS cube is generated from the captured image by demosaicing, reflectance calculation, and spectral correction. Hence, the aligned and noise-reduced HS cube has 24 bands within the wavelength range of 668.86 nm to 948.54 nm. Second, glare points are detected by bi-dimensional histograms and removed by inpainting. Third, patch generation and data augmentation are performed to increase the size of the data set and avoid overfitting. Half of all patches are cut out of the data cube with an overlap of 50%. The other half is taken randomly around the perimeter of the lesion, which counteracts the imbalance between pixels of a lesion and background pixels. The patches are then randomly flipped, mirrored, and rotated to obtain 100 patches from each of the 13 HS cubes. Lastly, binary masks were created from the labeled images to provide the pixel-wise classification labels.
Network Architecture: Fig. 2 shows the used modified U-Net. Two convolutional, two batch normalization and one dropout layer are added in a convolutional block to enhance generalization and reduce overfitting. We also implement reflective padding and insert a final cropping layer.
Loss Function: In the data set, the distribution between the background and lesion pixels is unbalanced. Thus, a weighted sum of binary cross-entropy (BCE) and dice loss (DL) function is implemented, called Combo Loss [11] (cf. Eq. 1). It leverages the flexibility of DL with regard to class imbalance and the curve smoothing of BCE. The factor $\alpha = 0.7$ is chosen heuristically. For parameter updates, the Adam optimizer with an initial learning rate of 5×10^{-5} is utilized.

$$CL(\hat{y}, y) = \alpha DL(\hat{y}, y) + (1 - \alpha) L_{BCE}(\hat{y}, y) \tag{1}$$

where $L_{BCE}(\hat{y}, y) = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y}))$ and $DL(\hat{y}, y) = 1 - \frac{2y\hat{y}+1}{y+\hat{y}+1}$.

4 Experimental Results

As semantic segmentation addresses a pixel-wise classification problem, the output of the U-Net can pixel-wise be interpreted as the probability for assignment of the class malignant vs. non-malignant. A classification threshold on the probability value parameterizes the outcome and a respective ROC curve over

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

Ch-Group	Exposure Time of 57.5 ms		Exposure Time of 77.5 ms	
	Accuracy	**AUC**	**Accuracy**	**AUC**
$No.1-4$	0.82 ± 0.11	0.83 ± 0.20	0.81 ± 0.14	0.73 ± 0.27
$No.5-8$	0.83 ± 0.13	0.85 ± 0.17	0.85 ± 0.15	0.79 ± 0.21
$No.9-12$	0.84 ± 0.13	0.81 ± 0.19	0.80 ± 0.22	0.80 ± 0.22
$No.13-16$	0.84 ± 0.14	0.80 ± 0.20	0.75 ± 0.24	0.75 ± 0.24
$No.17-20$	$\mathbf{0.85 \pm 0.12}$	$\mathbf{0.86 \pm 0.14}$	0.76 ± 0.20	0.76 ± 0.20
$No.21-24$	0.81 ± 0.11	0.84 ± 0.17	0.76 ± 0.21	0.76 ± 0.21
$No.1-24$	0.82 ± 0.15	0.80 ± 0.24	0.76 ± 0.21	0.76 ± 0.21

Table 1: Average on leave-one-patient-out testing for threshold 0.5 and AUC.

all thresholds and the corresponding Area Under Curve (AUC) [12] can be derived as a summarizing measure to compare classifications with different training units. While AUC has issues and in future work we plan a more detailed analysis, the AUC makes the results comparable to the previous work [5] and [8]. For the metrics, only the pixels that are located within the field of view of the respective endoscope are considered to ensure a veritable impression of the skewed data set. To investigate the relevance of certain wavelength ranges, we divided the 24 available channels into six 4-channel groups and compared these subsets to the full HS cube. The evaluation was repeated for both exposure times.

First, Table 1 shows respective results for the prediction accuracy at the threshold 0.5, as well as the AUC, averaged over leave-one-patient-out cross-validation. These results show a tendency that the mean AUC value is smaller for larger exposure time, which we explain by partly overlighted channels at the edges of the cube. We assume that this is why the 24-channel group $No.1-24$ at 57.5 ms performs better than the ones at 77.5 ms. Further, the channel group $No.17-20$ ($877-911$ nm) performs best for the mean exposure time of 57.5 ms.

However, these numerical scores are ultimately inconclusive, because of the rather high standard deviation (SD). This was to some degree expected, given the high variance in clinical conditions when recording the initial data (cf. Fig. 1 and Sec. 2). On the other hand, visual inspection of the full segmented images, which summarize many of the pixel-wise classification results in one visual panel, showed excellent results in the majority of the cases (cf. e.g. the heatmaps in Fig 3). It turns out that most of the variance is caused by a few HS cubes only.

Thus, in a second step, we analyze results wrt to complexity where on less-complex data higher accuracy is expected and the higher the complexity, the less test accuracy is accepted [13]. Accordingly, Table 2 shows the same results as above only for medium and low complex HS cube (e.g. in Fig. 1). These confirm the tendency that a lower exposure time leads to an improved segmentation of the tumor lesion and the channel group $No.17-20$ achieved the highest AUC overall with 0.95 and a very low SD of 0.03 indicating excellent generalization.

Finally, Table 3 shows the average test results for low complex samples only. It confirms that the lower exposure time yields higher accuracy and better AUC, regardless of the wavelength range considered, which is illustrated with high accuracy and very low SD for all channel groups.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

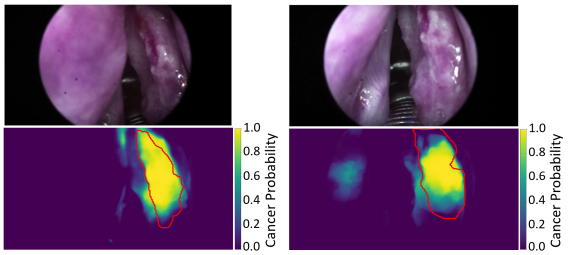
Ch-Group	Exposure Time of 57.5 ms		Exposure Time of 77.5 ms	
	Accuracy	**AUC**	**Accuracy**	**AUC**
No.1 – 4	0.86 ±0.11	0.93 ±0.07	0.79 ±0.15	0.84 ±0.18
No.5 – 8	0.86 ±0.15	0.90 ±0.16	0.85 ±0.15	0.90 ±0.11
No.9 – 12	0.85 ±0.13	0.89 ±0.15	0.86 ±0.16	0.88 ±0.17
No.13 – 16	0.84 ±0.14	0.86 ±0.19	0.82 ±0.16	0.84 ±0.21
No.17 – 20	0.86 ±0.11	**0.95 ± 0.03**	0.84 ±0.13	0.83 ±0.17
No.21 – 24	0.82 ±0.13	0.89 ±0.14	0.81 ±0.16	0.84 ±0.17
No.1 – 24	**0.88 ± 0.13**	0.91 ±0.17	0.84 ±0.13	0.80 ±0.19

Table 2: Average test error for medium and low complexity samples.

Ch-Group	Exposure Time of 57.5 ms		Exposure Time of 77.5 ms	
	Accuracy	**AUC**	**Accuracy**	**AUC**
No.1 – 4	0.93 ±0.03	0.97 ±0.01	0.87 ±0.07	0.95 ±0.05
No.5 – 8	0.93 ±0.01	0.97 ±0.02	0.92 ±0.02	0.92 ±0.11
No.9 – 12	0.91 ±0.03	0.97 ±0.02	0.92 ±0.04	0.96 ±0.04
No.13 – 16	0.93 ±0.02	0.97 ±0.02	0.90 ±0.06	0.95 ±0.08
No.17 – 20	0.92 ±0.03	0.95 ±0.04	0.88 ±0.04	0.86 ±0.15
No.21 – 24	0.87 ±0.08	0.96 ±0.03	0.87 ±0.05	0.95 ±0.04
No.1 – 24	**0.93 ± 0.01**	**0.98 ± 0.01**	0.89 ±0.08	0.91 ±0.10

Table 3: Average test error for low complexity samples.

This is complementarily visualized in the heatmaps in Fig. 3 which show the probability of cancer ranging from 0.0 (blue) to 1.0 (yellow) for each pixel. Accordingly, more true positives with a high probability can be observed in Fig. 3a and additional false positives on the left vocal band can be seen in Fig. 3b. Although the best accuracy, in this case, was achieved by the full HS cube with all 24-channels, still the 4-channel group from $17-20$ reaches a very good accuracy of 0.92 with SD of 0.03. Hence, the channel group No.$17-20$ achieved on average the highest accuracy and AUC as well as the lowest SD over all the results regarding the different complexity degrees. Beyond, the training time is reduced from ~ 53 to 24 minutes using i7-7700K, 16 GB RAM and 2 RTX 3080 Ti, and the test time from ~ 60 to 40 seconds using i5-2500K, 16 GB RAM and intel HD Graphics 3000.



(a) exposure time = 57.5 ms. (b) exposure time = 77.5 ms.

Fig. 3: Low-complexity image and heat map of channel group No.$1-24$ with ground truth of malignant lesion determined with red line marked by expert.

ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

5 Conclusion

In this contribution, we have shown that HSI on raw clinical data is feasible for laryngeal cancer detection with a U-Net architecture. The medical hypothesis that bands with large wavelengths should be favorable could be supported. Additionally, we were able to reduce the number of HSI channels and speed up the prediction and training process without reducing the performance. Furthermore, the results show clearly how the data complexity, as well as the exposure time, play significant roles in tumor detection and segmentation.

There are obvious shortcomings that need to be addressed in further work. All results have been obtained on a still small data set with an overall high variance, and we are well aware that labeling for the ground truth can be improved. The complexity analysis shows that potentially conditions need to be controlled somewhat more. Note that, however, it is not necessary to classify all pixels perfectly to reach our medium-term goal of providing the physician with an in-vivo online signal whereat to prevent overlooking critical lesions. This mediates the effect of noise on the pixel level but calls for more significant measures than plain pixel-wise accuracy or AUC to be employed. In summary, we believe that overall the results are very encouraging and show that beyond the visible spectrum, HSI can contribute strongly to cancer detection if the technical difficulties specifically in image recording and generation can be mastered.

References

[1] A. O. H. Gerstner. Früherkennung von Kopf-Hals-Tumoren. Entwicklung, aktueller Stand und Perspektiven. *Laryngo-Rhino-Otologie*, 87 Suppl 1(S 1):S1–20, 2008.

[2] A. O. H. Gerstner et al. Endoskopie des Larynx mit Hyperspectral Imaging. *HNO*, 60(12):1047–1052, 2012.

[3] K. Munir et al. Cancer diagnosis using deep learning: A bibliographic review. *Cancers*, 11(9), 2019.

[4] O. Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.

[5] D. Eggert et al. In vivo detection of head and neck tumors by hyperspectral imaging combined with deep learning methods. *Journal of Biophotonics*, 15(3):e202100167, 2022.

[6] C. Arens et al. Fortschritte der endoskopischen Diagnostik von Dysplasien und Karzinomen des Larynx. *HNO*, 60(1):44–52, 2012.

[7] T. D. Wang et al. Optical biopsy: a new frontier in endoscopic detection and diagnosis. *Clin. Gastroenterol. Hepatol.: the official clinical practice journal of the American Gastroenterological Association*, 2(9):744–753, 2004.

[8] M. Halicek et al. Cancer detection using hyperspectral imaging and evaluation of the superficial tumor margin variance with depth. *Medical Imaging 2019*, 10951:329 – 339.

[9] F. Meyer-Veit et al. Hyperspectral Endoscopy using Deep Learning for Laryngeal Cancer Segmentation. In *ICANN*, 2022.

[10] H. Iqbal. Harisiqbal88/plotneuralnet. 2018.

[11] S. Jadon. A survey of loss functions for semantic segmentation. In *CIBCB*, pages 1–7. IEEE, 2020.

[12] J A Hanley et al. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982. PMID: 7063747.

[13] A. Ng. Deep Learning Specialization. *Deep Learning AI*, 2018.