




ORIGINAL ARTICLE

Fully convolutional networks in multimodal nonlinear microscopy images for automated detection of head and neck carcinoma: A pilot study

Erik Rodner PhD^{1,2†} | Thomas Bocklitz PhD^{3,4†} | Ferdinand von Eggeling PhD^{3,4,5} |
Günther Ernst MD⁵ | Olga Chernavskaia PhD⁴ | Jürgen Popp PhD^{3,4} | Joachim Denzler PhD¹ |
Orlando Guntinas-Lichius MD⁵ 

¹Department of Computer Science, Friedrich Schiller University, Jena, Germany

²Corporate Research and Technology, Carl Zeiss AG, Jena, Germany

³Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich Schiller University, Jena, Germany

⁴Leibniz Institute of Photonic Technology, Jena, Germany

⁵Department of Otorhinolaryngology, Jena University Hospital, Jena, Germany

Correspondence

Orlando Guntinas-Lichius, MD, Department of Otorhinolaryngology, Jena University Hospital, Am Klinikum 1, D-07747 Jena, Germany.
Email: orlando.guntinas@med.uni-jena.de

Funding information

German Research Foundation, Grant/Award Numbers: PO 563/29-1, EG 102/9-1, PO 563/30-1, BO 4700/1-1, RO 5093/1-1, DE 735/10-1; Leibniz ScienceCampus InfectoOptics

Abstract

Background: A fully convolutional neural networks (FCN)-based automated image analysis algorithm to discriminate between head and neck cancer and non-cancerous epithelium based on nonlinear microscopic images was developed.

Methods: Head and neck cancer sections were used for standard histopathology and co-registered with multimodal images from the same sections using the combination of coherent anti-Stokes Raman scattering, two-photon excited fluorescence, and second harmonic generation microscopy. The images analyzed with semantic segmentation using a FCN for four classes: cancer, normal epithelium, background, and other tissue types.

Results: A total of 114 images of 12 patients were analyzed. Using a patch score aggregation, the average recognition rate and an overall recognition rate of the four classes were 88.9% and 86.7%, respectively. A total of 113 seconds were needed to process a whole-slice image in the dataset.

Conclusion: Multimodal nonlinear microscopy in combination with automated image analysis using FCN seems to be a promising technique for objective differentiation between head and neck cancer and noncancerous epithelium.

KEYWORDS

coherent anti-stokes Raman scattering, convolutional neural networks, diagnostics, digital pathology, head and neck cancer, image analysis, second-harmonic generation, semantic segmentation, spectral histopathology, two-photon excited fluorescence

1 | INTRODUCTION

New optical diagnostic methods have the potential to become an important adjunct to the histopathological assessment of head and neck cancer tissue specimens as the gold standard for cancer diagnosis.^{1,2} Classical histopathological analysis needs tissue staining and specific molecular

pathology is time-consuming and expensive. Raman microspectroscopy principally is one of best new optical imaging techniques nondestructively visualizing the molecular composition of head and neck cancer tissue and of the tumor environment without the need for any external staining. This enables noninvasive high-resolution in vitro and even in vivo imaging.³ Raman spectroscopy has the disadvantage of long acquisition times, because of the intrinsic weak spontaneous Raman scattering.⁴ Coherent anti-Stokes Raman

[†]These authors contributed equally to this work.

scattering (CARS) microscopy is a fast nonlinear variant of Raman scattering allowing real-time in-vivo tissue visualization.^{5,6} If CARS is applied together with other nonlinear techniques such as two-photon-excited fluorescence (TPEF) and second-harmonic generation (SHG), a powerful tool for label-free discrimination between cancerous and noncancerous tissue in patients with head and neck cancer is formed, which is often called multimodal imaging.^{5,7,8} General problems hindering a clinical implementation for head and neck cancer diagnostics so far are that information extraction from the spectroscopic data and time-consuming and complex multivariate data analysis. Objective computerized evaluation and tissue classification systems are needed to establish multimodal nonlinear microscopy in clinical routine.⁵ Such a real-time automated image analysis would lead to a prediction of cancerous and noncancerous head and neck tissue, which is independent of the surgeon and pathologist.

Based on our previous pilot study to develop a comprehensive image analysis system to predict the diagnosis of head and neck cancer in imaged tissue sections,⁵ we now present a new semantic segmentation approach, that is, a pixel-wise classification of the imaged tissue sections, based on fully convolutional neural networks (FCN) allowing a fast and reliable classification of head and neck cancer specimens. FCN are a specific deep learning technique that has been the key technology to enable a significant improvement in segmentation in the last years and have not been applied and adapted to multimodal imaging so far. The key idea behind FCN and deep learning is to not only finding the right classification decision, but also finding the right representation of the input data at the same time. As a consequence, subjective and manual design of features from the images is avoided, which can easily result in a bias of the later classification step. Training means to find those parameter, that is, weights, of the neural network that minimizes the difference between expected and predicted output of the classifier. In deep learning the number of weights is in the range of 10-50 millions of standard network architectures.

2 | PATIENTS AND METHODS

2.1 | Study design and setting

This prospective observational pilot study was performed at the Department of Otorhinolaryngology, Jena University Hospital, Jena, Germany. Approval for the study was obtained through the local Institutional Review Board (No. 3008-12/10) and informed consent was obtained from all study participants.

2.2 | Patients

The study cohort consisted of 12 patients with primary head and neck squamous cell carcinoma (&1x cavity of the

mouth; 5x oropharynx; 4x larynx; 2x hypopharynx) treated in 2013. Primary tumor classification varied from pT2 (3x), pT3 (4x), to pT4 (5x). Inclusion criteria were age ≥ 18 years, diagnosis confirmed by standard histopathology, and written informed consent. As per patient, at least one section of a tumor sample and one section of a normal mucosa sample were analyzed. Images of different areas of the sections were taken. Overall, 114 images of 12 patients were analyzed. A ground truth existed for 30 of these images. This image subset was utilized for validation. The same study cohort as in the previous pilot study was used.⁵ The previous pilot study used a linear discriminant analysis approach to analyze the data. Using the same study cohort and exactly the same samples allowed a direct comparison of the new FCN approach with the linear discriminant analysis approach.

2.3 | Tissue processing and multimodal nonlinear microscopy imaging

Details on the tissue processing and the multimodal nonlinear microscopic imaging were reported previously.⁵ Briefly, the tissue samples were frozen in liquid nitrogen and sliced into 20 μm thin sections. Samples from the tumor were used. The sections included the primary tumor, its margin, and sometimes surrounding tissue. In addition, control biopsies from normal appearing mucosa of the same patients were taken. For each section under investigation, several adjacent parallel section of 5 μm thickness were obtained for various standard histopathology staining protocols, including H&E-staining and immunohistochemistry. Various tissue types (squamous cell cancer, epithelial tissue, inflamed area, glandular tissue, fat tissue, vessels, respectively) were labeled with distinct colors in H&E stained images by a trained pathologist (Günther Ernst). These labeled images were later co-registered with the multimodal images. The 20 μm thick sections were investigated by nonlinear multimodal microscopy without further washing or staining steps. The herein used multimodal imaging approach used the combination of CARS, TPEF, and SHG microscopy. The setup used for nonlinear multimodal microscopy was described in detail elsewhere.⁹

2.4 | Segmentation of the images with FCN

The preprocessed and aligned images were analyzed with semantic segmentation using a FCN. We applied our own FCN implementation, CN24, which is open source and publicly available (<https://www.github.com/cvjena/cn24>). The network architecture is depicted in Supporting Information Figure S1. Due to lack of a sufficient number of annotated data, the initialization of the weights was done by pretraining on the ILSVRC2012 dataset, which is common practice and has shown to provide a sufficient initial solution for a wide range of visual recognition tasks. The *ILSVRC2012 dataset*, sometimes summarized as ImageNet, is a standard benchmark

dataset of annotated images that show 1000 different everyday object categories, for example, furniture, animals, humans, cars, and so forth. *Pretraining* means that an initial set of weights of the FCN are found by training the task of classifying the images from the ILSVRC2012 dataset.

The task was to distinguish between four classes: head and neck cancer, epithelial tissue, background, and other tissue classes. The given images had a dimension between 2048×2048 and 8192×8192 pixels. Due to memory limitations of graphics cards, there was currently no way to process the given images with large FCNs directly, neither for training nor for testing. Therefore, we tiled the images into 384×384 regions and applied semantic segmentation individually on these tiles. It is worth noting that this does not correspond to down-sampling the original images, but to process it at its original resolution in a tile-wise manner. This allowed for tractable training but at the same time has the disadvantage of boundary artifacts that might appear but could be compensated for using postprocessing. Boundary artifacts might occur at the borders of the tiles, because two neighboring tiles are processed independently at their borders without considering the image data of the respective other tile. For testing, we currently performed a prediction on nonoverlapping tiles. However, reducing boundary artifacts could be done using overlapping tiles and averaging the predictions accordingly.

2.5 | Classification of whole-slide images

Whereas the segmentation result can also directly lead to a classification decision for each slide, we also tested an alternative approach for classification bypassing segmentation. The question that drove this approach was how much of the whole slide needs to be seen by an algorithm to make a reliable decision. We therefore sampled 100 patches of size 224×224 feeding them to a VGG19 network. The *VGG19 network* is a publicly available FCN trained on the ImageNet dataset as well. We then trained an SVM on the conv5_3 layer outputs. The *conv5_3 layer outputs* consist of activations of the neurons from the last convolutional layer of the VGG19 neural network. It is good practice to use those outputs as feature representation of the input images. During test, or each of the 100 patches, we get a classification score for each of the four classes, which we call *patch scores*. Final classification is then done by averaging all SVM patch scores.

2.6 | Statistical analysis

The main questions that drove our evaluation and statistical analysis were: can pixels of head and neck cancer, epithelial tissue, background, and other tissue classes be differentiated using multimodal images only? Each pixel of the image is decided into one of the four classes using the FCN architecture described before. For evaluation, we used a leave-one-

patient-out technique, which is common for robustly evaluating machine learning algorithms.¹⁰ The technique *leave-one-patient-out* means that we used annotated pixels for all 12 patients except the one for testing. During learning, a weighted loss-function in the FCN is optimized measuring the difference in the decision for each pixel between ground truth annotation and output of the FCN. The weight for each pixel of the training example was computed by the inverse of the class frequency of the ground truth class of that pixel. This step was necessary to tackle a common issue in semantic segmentation tasks that large differences in class frequencies will lead to a model focusing only on the background class, because it is comprised by the largest percentage of training pixels. It is worth noting that this weighted loss-function did not cause any bias on the final decision during testing.

The diagnostic accuracy of the presented algorithm, that is, the differentiation between pixels of head and neck cancer, epithelial tissue, background, and other tissue classes was evaluated by the overall and average recognition rate. The latter is computed by calculating the mean recognition rate of all four classes to reduce a bias of the performance with respect to the most frequent class. In addition, we evaluate the weighted-loss of the FCN during training indicating the success and stability of the training procedure.

3 | RESULTS

3.1 | Training analysis

As can be seen in Supporting Information Figure S2, the training converged to a reasonable performance on the training set. For further epochs, no change of the performance on the training set was seen. For this analysis, all labeled images from all patients except for the ones of the test patient were included.

3.2 | Leave-one-patient-out analysis

Quantitative results for the leave-one-patient-out scheme resulted in an overall recognition score of 83.39%. By calculating the mean recognition rate of all four classes to reduce a bias of the performance with respect to the most frequent class, an average recognition rate of 75.26% was reached. This decrease in accuracy indicated that the classifier still was not unbiased toward the class frequency. In the optimal case, both scores—overall and mean recognition rate—shall have almost equal values.

3.3 | Inference and segmentation time

On average 113 seconds to process a whole-slide image in the dataset were needed. However, multiple instances of our convolutional neural network framework were running concurrently on the machines equipped with different graphics

processing units. Therefore, the best machine was able to process the whole-slice images in 55 seconds. This was the total processing time excluding preprocessing steps, because the convolutional neural networks performed feature extraction and classification in one single step.

3.4 | Aggregating patch scores

Examples of the results of this approach are shown in Figure 1 and in high magnification in Figure 2. The patch score aggregation to discriminate cancer tissue from all other tissue resulted in an average recognition rate and an overall recognition rate of 88.9% and 86.7%, respectively. In contrast to the other method, the accuracy was not just higher, but also showed no bias toward class frequency.

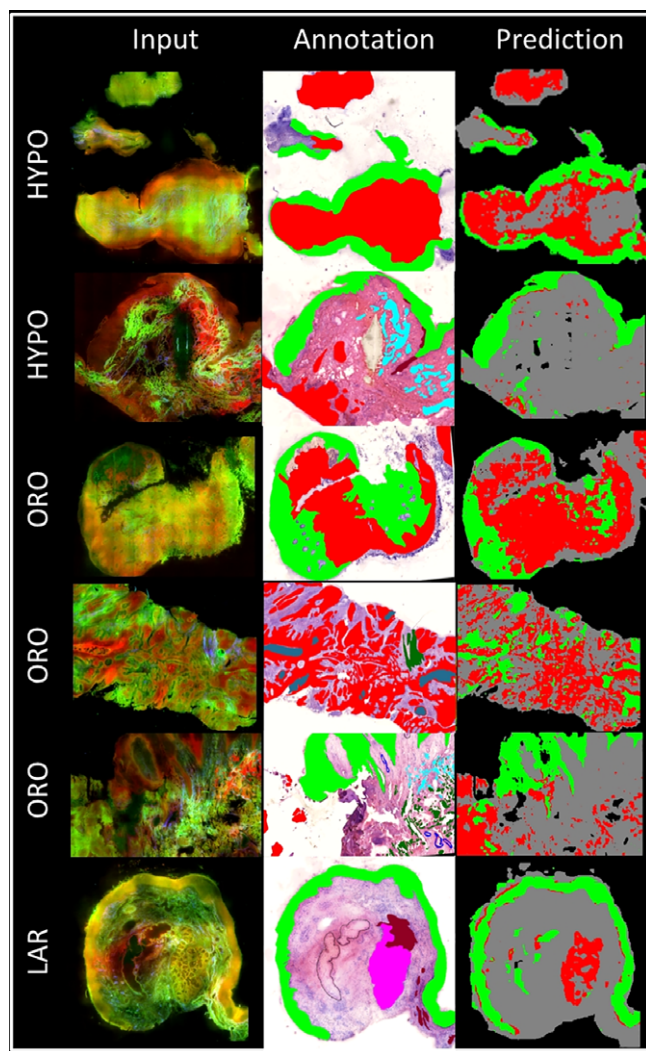


FIGURE 1 Examples for qualitative results using the leave-one-patient-out approach. The two colors show the classification of the cancer area (red) and noncancer area (green), respectively. Input, Input multimodal nonlinear microscopic image. Annotation, Tissue type annotation performed by a trained pathologist. Prediction, Result of the fully convolutional neural network (FCN) semantic segmentation approach. HYPO, hypopharyngeal cancer; ORO, oropharyngeal cancer; LAR, laryngeal cancer [Color figure can be viewed at wileyonlinelibrary.com]

4 | DISCUSSION

The presented approach for semantic segmentation showed a promising performance for the classification of head and neck squamous cell carcinoma based on multimodal nonlinear microscopic images combining data from CARS, TPEF, and SHG microscopy. When we recently introduced this multimodal nonlinear microscopy setting for head and neck cancer diagnosis, we used a linear discriminant analysis.⁵ The semantic segmentation (pixel-wise classification) of the four tissue classes of the presented FCN approach resulted in an average recognition rate and in an overall recognition rate of 75% and 83%, respectively. In comparison, the linear discriminant analysis approach resulted in recognition rates of 74% and 73%, respectively.⁵ Alternatively, a direct estimation approach with aggregating patch scores was used to provide a global dualistic score to help pathologists to separate in each image only cancerous from noncancerous tissue, an average recognition rate and in an overall recognition rate of 89% and 88%, respectively, was reached. In comparison, the linear discriminant analysis approach resulted in dualistic recognition rates of 79% and 80%, respectively.⁵ An example directly comparing the FCN with the linear discriminant analysis approach is shown in Figure 2. Hence, the FCN approach showed a better performance. Further postprocessing or data augmentation could be applied and might even increase the performance.

Recently, the combination of the three label-free nonlinear imaging modalities CARS, TPEF, and SHG was used to realize a fast, automated real-time prediction of the disease activity from tissue samples of patients with inflammatory bowel disease, that is, the approach seems to be helpful also for the prediction of nontumorous inflammatory diseases.¹¹ The same multimodal imaging approach was also used to generate computational pseudo hematoxylin and eosin (HE) images by multivariate statistics of mouse colon sections.⁶ These images allowed the valid identification of regions with colon cancer. The latter were analyzed further by targeted Raman-spectroscopy retrieving the tissue's molecular fingerprint. The CARS, TPEF, and SHG combination has the potential for a fast screening method. Meanwhile, the accuracy of CARS as a fast label-free technique for the evaluation of head and neck cancer has been confirmed by the second working group.¹² Raman spectroscopy is so far not suitable as screening tool due to the long acquisition time. But Raman spectroscopy is able to establish a detailed spectral histopathology analysis of cancerous and dysplastic tissue after screening with multimodal nonlinear microscopy.³ Furthermore, also another working group has already established a linear discriminant analysis-based approach to classify Raman spectra from tissue samples of patients tongue cancer into cancer vs nontumorous tissue.¹³

In general, any kind of modern optical diagnostics or optical biopsy is a complex process. Of course, the oncologic

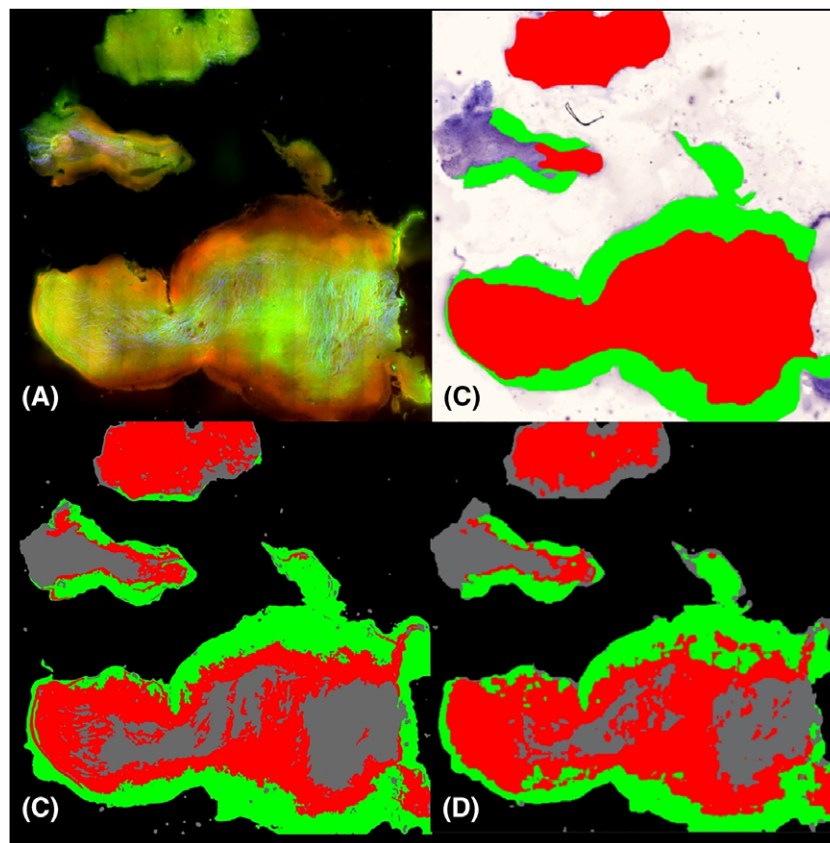


FIGURE 2 Example of images of a hypopharyngeal cancer for qualitative results using patient 1 in our leave-one-patient-out approach in comparison to our previously presented analysis using a linear discriminant analysis. The two colors show the classification of the cancer area (red) and noncancer area (green), respectively. Morphological postprocessing has been applied on the result images. A, Input image. B, Tissue type annotation performed by a trained pathologist. C, Result of linear discriminant analysis presented elsewhere.⁵ D, Result of the herein presented FCN semantic segmentation approach [Color figure can be viewed at wileyonlinelibrary.com]

surgeon should know the fundamental principles of optical imaging to understand which conclusions can be drawn from the images.¹⁴ Nevertheless, the head and neck surgeon is not trained to interpret the histology-like images of tissues *in vivo*.¹⁵ The pathologist is the natural partner to interpret multimodal nonlinear microscopy images. Automated classification methods such as FCN will not replace the pathologist but will help accelerate and improve the accuracy of intraoperative cancer diagnostics. It is conceivable that intraoperative tumor biopsies (in addition to or perspective as a substitute for frozen sections) are used for immediate multimodal nonlinear microscopy and are automatically classified. The pathologist could then focus on the tumor borders or areas of uncertainty. This would help the pathologist to process more tissue biopsies at the same time, too.

The next step is a multicenter validation of the setting by implementation of multimodal nonlinear microscopy into the clinical routine pathways in parallel to frozen sections. A big step will be to apply nonlinear imaging *in vivo*. The development of CARS imaging fiber probe has been reported recently.¹⁶ The *in vivo* application in combination with automated imaging analysis would make it possible in the future to display the results into a postprocessed image guiding the surgeon during surgery

in real time. Such an application has already been introduced for hyperspectral imaging using flexible endoscopy for laryngeal cancer detection.¹⁷

5 | CONCLUSIONS

Label-free multimodal nonlinear microscopy combined with a FCN approach for automated classification is feasible. It allows a reliable and objective offline image analysis for the differentiation of head and neck cancer tissue and other non-neoplastic tissue types. Next step is the demonstration of its usefulness by implementation of multimodal nonlinear microscopy in the routine setting of intraoperative pathological diagnostics. It has to be shown that the approach accelerates intraoperative diagnostics, reduces the pathologist's workload per biopsy, and finally increases the overall quality of intraoperative diagnostics.

ACKNOWLEDGMENTS

This research was partially supported by grant DE 735/10-1, RO 5093/1-1, BO 4700/1-1, and PO 563/30-1 of the German Research Foundation (DFG) and Leibniz ScienceCampus

InfectoOptics (BLOODi) and by the DFG-funded (EG 102/9-1 and PO 563/29-1) Core Facility Jena Biophotonic and Imaging Laboratory (JBIL)

FINANCIAL DISCLOSURE

The authors have no financial interest to declare in relation to the content of this article.

AUTHOR CONTRIBUTIONS

Each author listed on the manuscript has seen and approved the submission of this version of the manuscript and takes full responsibility for the manuscript.

Writing the first draft of the manuscript: Guntinas-Lichius

DISCLOSURE OF INTERESTS

No honorarium, grant, or other form of payment was given to anyone to produce the manuscript.

ORCID

Orlando Guntinas-Lichius  <https://orcid.org/0000-0001-9671-0784>

REFERENCES

- Green B, Cobb AR, Brennan PA, Hopper C. Optical diagnostic techniques for use in lesions of the head and neck: review of the latest developments. *Br J Oral Maxillofac Surg*. 2014;52(8):675-680.
- Singh SP, Ibrahim O, Byrne HJ, et al. Recent advances in optical diagnosis of oral cancers: review and future perspectives. *Head Neck*. 2016;38(Suppl 1):E2403-E2411.
- Bocklitz T, Brautigam K, Urbanek A, et al. Novel workflow for combining Raman spectroscopy and MALDI-MSI for tissue based studies. *Anal Bioanal Chem*. 2015;407(26):7865-7873.
- Downes A, Elfick A. Raman spectroscopy and related techniques in biomedicine. *Sensors (Basel)*. 2010;10(3):1871-1889.
- Heuke S, Chernavskaya O, Bocklitz T, et al. Multimodal nonlinear microscopy of head and neck carcinoma - toward surgery assisting frozen section analysis. *Head Neck*. 2016;38(10):1545-1552.
- Bocklitz TW, Salah FS, Vogler N, et al. Pseudo-HE images derived from CARS/TPEF/SHG multimodal imaging in combination with

- Raman-spectroscopy as a pathological screening tool. *BMC Cancer*. 2016;16:534.
- Meyer T, Baumgartl M, Gottschall T, et al. A compact microscope setup for multimodal nonlinear imaging in clinics and its application to disease diagnostics. *Analyst*. 2013;138(14):4048-4057.
- Meyer T, Guntinas-Lichius O, von Eggeling F, et al. Multimodal nonlinear microscopic investigations on head and neck squamous cell carcinoma: toward intraoperative imaging. *Head Neck*. 2013;35(9):E280-E287.
- Heuke S, Vogler N, Meyer T, et al. Multimodal mapping of human skin. *Br J Dermatol*. 2013;169(4):794-803.
- Duan L, Marvdashti T, Lee A, Tang JY, Ellerbee AK. Automated identification of basal cell carcinoma by polarization-sensitive optical coherence tomography. *Biomed Opt Exp*. 2014;5(10):3717-3729.
- Chernavskaya O, Heuke S, Vieth M, et al. Beyond endoscopic assessment in inflammatory bowel disease: real-time histology of disease activity by nonlinear multimodal imaging. *Sci Rep*. 2016;6:29239.
- Hoesli RC, Orringer DA, McHugh JB, Spector ME. Coherent Raman scattering microscopy for evaluation of head and neck carcinoma. *Otolaryngol Head Neck Surg*. 2017;157(3):448-453.
- Cals FL, Koljenovic S, Hardillo JA, Baatenburg de Jong RJ, Bakker Schut TC, Puppels GJ. Development and validation of Raman spectroscopic classification models to discriminate tongue squamous cell carcinoma from non-tumorous tissue. *Oral Oncol*. 2016;60:41-47.
- Keereweer S, Van Driel PB, Snoeks TJ, et al. Optical image-guided cancer surgery: challenges and limitations. *Clin Cancer Res*. 2013;19(14):3745-3754.
- Dittbner A, Rodner E, Ortmann W, et al. Automated analysis of confocal laser endomicroscopy images to detect head and neck cancer. *Head Neck*. 2016;38(Suppl 1):E1419-E1426.
- Lukic A, Dochow S, Chernavskaya O, et al. Fiber probe for nonlinear imaging applications. *J Biophotonics*. 2016;9(1-2):138-143.
- Regeling B, Thies B, Gerstner AO, et al. Hyperspectral imaging using flexible endoscopy for laryngeal cancer detection. *Sensors (Basel)*. 2016;16(8):pii: E1288. <https://doi.org/10.3390/s1608128>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Rodner E, Bocklitz T, von Eggeling F, et al. Fully convolutional networks in multimodal nonlinear microscopy images for automated detection of head and neck carcinoma: A pilot study. *Head & Neck*. 2018;1-6. <https://doi.org/10.1002/hed.25489>