Motion Artifact Detection in Confocal Laser Endomicroscopy Images

Maike Stoeve[1], Marc Aubreville[1], Nicolai Oetter[2,3], Christian Knipfer[4,3], Helmut Neumann[5,3], Florian Stelzle[2,3], Andreas Maier[1,3]

[1] Pattern Recognition Lab, Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2] Department of Oral and Maxillofacial Surgery, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg
[3] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander-Universität Erlangen-Nürnberg
[4] Department of Oral and Maxillofacial Surgery, University Medical Center Hamburg-Eppendorf
[5] First Department of Internal Medicine, University Hospital Mainz, Johannes Gutenberg-Universität Mainz

maike.stoeve@fau.de

Abstract. Confocal Laser Endomicroscopy (CLE), an optical imaging technique allowing non-invasive examination of the mucosa on a (sub)-cellular level, has proven to be a valuable diagnostic tool in gastroenterology and shows promising results in various anatomical regions including the oral cavity. Recently, the feasibility of automatic carcinoma detection for CLE images of sufficient quality was shown. However, in real world data sets a high amount of CLE images is corrupted by artifacts. Amongst the most prevalent artifact types are motion-induced image deteriorations. In the scope of this work, algorithmic approaches for the automatic detection of motion artifact-tainted image regions were developed. Hence, this work provides an important step towards clinical applicability of automatic carcinoma detection. Both, conventional machine learning and novel, deep learning-based approaches were assessed. The deep learning-based approach outperforms the conventional approaches, attaining an AUC of 0.90.

1 Introduction

With over 500,000 diagnosed cases each year, head and neck squamous cell carcinoma (HNSCC) is considered the sixth most common cancer type worldwide [1]. For diagnosis of HNSCC, invasive biopsy and subsequent histopathological examination is applied as gold standard method [2]. As alternative, non-invasive optical imaging technologies as narrow-band imaging, autofluorescence imaging and confocal laser endomicroscopy (CLE) are gaining interest in research [3]. Among those technologies, CLE has already proven to be a valuable diagnostic tool in gastroenterology and has been successfully applied to examine lesions in the oral cavity [4]. Both, the interpretation of CLE images and the histopathological examination of tissue samples require experience and proficiency and entail a subjective component [5]. Thus, promising approaches using machine learning techniques for automatic carcinoma detection based on CLE images were developed [6].

As the accuracy of these algorithms is highly affected by the occurrence of artifacts, artifact-tainted images were excluded manually prior to training in a time-consuming manual labeling step in all known approaches. In the scope of this work, an automatic motion artifact detection pipeline was developed and evaluated. Hence, this work provides the basis to integrate a fully automatic motion artifact detection into existing carcinoma detection frameworks, an important step towards clinical applicability.

1.1 Motion artifacts

A frequent cause of image impairment are motion artifacts. They can either be caused by movements of the investigated anatomical structures or motion induced by the physician. The proportion of motion artifact images compared to good quality images is highly dependent on the experience of the physician [7]. As shown in Fig. 1, two different manifestations can be observed. The first manifestation is characterized by streaky patterns originating from the repeated acquisition of the same, shifted image line. This pattern can be observed, when the sum of the velocities of probe, organ and scanning is approximately equal to zero. This requires an organ movement in the same direction as the sampling pattern or a probe motion in the correspondingly opposite direction. If a significant relative motion results from organ or probe motion, the cells are stretched or blurred. This second manifestation is hard to detect, as distinction between elongated cells is difficult without the observation of adjacent frames. Finally, motion artifacts can impair the whole image leading to a total loss of information or only influence parts of the image still allowing a diagnosis based on untainted regions [7]. Due to the meander-shaped optical sampling pattern, only whole image rows are affected by motion artifacts.

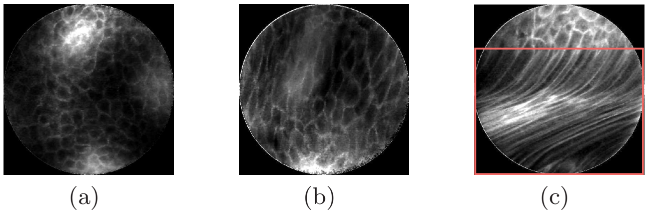

(a) (b) (c)

Fig. 1. CLE images containing (a) no artifacts or motion artifacts manifesting in (b) stretched cells or (c) stripe patterns (region marked in red).

2 Materials and methods

For the present work, 116 CLE sequences comprising 11,234 images from 12 patients and 4 sites within the oral cavity, namely the upper alveolar ridge, the lower inner labium, the palatal region and the lesion site itself were utilized. The CLE sequences were recorded at the department of Oral and Maxillofacial Surgery of the University Hospital Erlangen by a standalone probe-based CLE system (Cellvizio, Mauna Kea Technologies, Paris, France). The obtained images are approximately of size 576×578 pixels. The overall image quality was assessed by an expert and artifact regions were annotated manually within each image. Details on the data can be found in [6].

In the scope of this work, two different methodical approaches were established for the detection of motion artifact-tainted CLE image regions. The first approach uses conventional pattern recognition methods extracting characteristic image features, whereas the second approach applies deep learning strategies. For both approaches, images with low signal-to-noise ratio are excluded. The remaining CLE images are converted to 8-bit integer values after a quantile-based dynamic range compression following Aubreville et al. [6] was performed.

2.1 Feature-based motion artifact detection

The feature-based approach consists of three steps: pre-processing, feature extraction and classification. The pre-processing step is required due to the unusual round shape of CLE images complicating the feature extraction process. Jaremenko et al. circumvented this problem by dividing the image in overlapping, square patches and concatenating the information of all patches for the classification of the whole image [8]. As motion artifacts always cover the whole width of an image, slices with a width of the maximum extent of the CLE image in x-direction, a fixed height of 128 pixels and an overlap of 30 % (HOG) or 50 % (corrAngle) are extracted. To allow the detection of motion artifact-tainted slices of an image and differentiate them from untainted slices of the same image, either Histogram of Oriented Gradient (HOG) or angle of maximum correlation (corrAngle) features are extracted for each slice. For classification, a random forest classifier (RF) and a support vector machine (SVM) classifier were used. Undersampling of the majority class is applied to deal with the class imbalance. Prior to training of the linear SVM classifier, the feature vector is standardized. For evaluation purposes of both classifiers, 5-fold cross-validation is used. The different slices of one image are all assigned to the same fold.

Histogram of oriented gradients (HOG) The Histogram of Oriented Gradient (HOG) feature descriptor of Dalal and Triggs is frequently used in the field of object recognition [9]. Basically, the occurrence of gradient directions is computed in a local image region characterizing the local shape of an object. In contrast to the original pipeline, no gamma and color normalization is used. A cell size of 32×32 and block size of 64×64 resulted in the best classification performance. Due to the varying length of CLE image slices, the obtained HOG feature vectors are of varying length. Hence, mean, standard deviation, skewness and kurtosis are computed over all 9 bins of the feature vector individually. The statistical properties for all bins are subsequently concatenated to form a feature vector of a fixed length of 36.

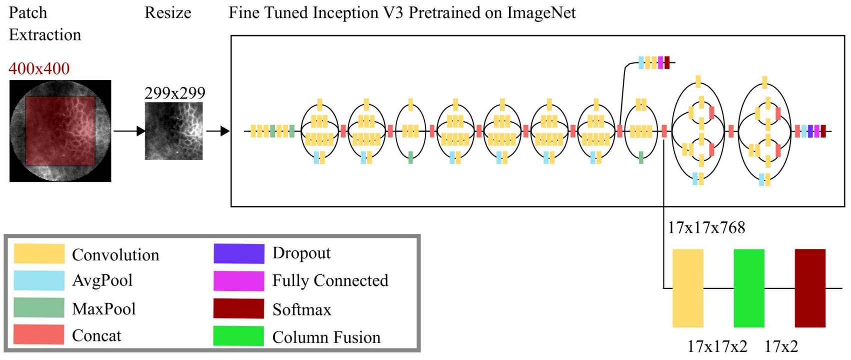
Angle of Maximum Correlation (corrAngle) The angle of maximum correlation feature (corrAngle) is designed under consideration of the origin of the streaky patterns visible in most motion artifact-tainted images. The direction of the relative motion accounts for the angle θ characterizing the direction of the stripe pattern. To create a feature vector describing the presence or absence of motion artifact patterns in an image slice with length L, a centered reference segment of row i of the CLE image slice is extracted. For a set of angles equally distributed between $\frac{\pi}{8}$ and $\frac{7}{8}\pi$ and a fixed radius R, comparative segments are extracted with center at $(\frac{L}{2}+R\sin(\theta), j+R+R\cos(\theta))$ using bilinear interpolation. Then, the correlation coefficient of the reference segment and each comparative segment is computed as measurement of similarity. The angle responsible for the highest correlation coefficient is saved. This procedure is repeated for each row despite the first and last R rows. Finally, all approximated motion angles are concatenated to form the final feature vector. For a motion artifact-tainted image slice, the approximated angle is constant (stripe pattern). In contrast, CLE image slices without motion artifacts show high variations of the approximated motion angle over image rows.

2.2 Deep learning-based motion artifact detection (artiNet)

Building on the Inception v3 architecture of Szegedy [10] pretrained on ImageNet, the deep learning-based detector of motion artifact-tainted CLE image areas (artiNet) is build by inserting a 2d 1×1 convolutional layer after the eighth inception block followed by a column fusion and a softmax layer as depicted in Fig. 2. Thus, the input representation for the new layers still entails spatial information. The convolutional layer is used to map the 17×17×768 dimensional input to the two output classes motion artifact and good quality. The result is a 17×17 grid of predictions. To obtain predictions of motion artifact regions covering the whole image width comparable to the slices of the conventional approach, a column fusion layer extracts the maximum over the image width resulting in a 17×2 representation fed to the final softmax layer.

Prior to training the grayscale images are transformed to RGB color representation. Then, a centered image patch of 400×400 pixels is extracted and resized to 299×299 pixels. Due to the slice-wise detection of motion artifacts and available corresponding labels effectively representing a fully convolutional approach with a network capacity according to one with smaller patches, data augmentation seems less important. Within the TensorFlow framework, the Inception network was fine-tuned in 2000 training steps using the Adam optimizer with an initial learning rate of $5 \cdot 10^{-6}$. For the training of the new layers, a learning rate of $5 \cdot 10^{-5}$ was used. In each training step, a minibatch consisting of 25

Fig. 2. Visualization of the artiNet for motion artifact detection in CLE images based on an Inception v3 network [10] pre-trained on ImageNet.



randomly selected instances of both classes is processed to deal with the class imbalance of the data set. For evaluation, a leave-one-patient-out cross-validation was performed.

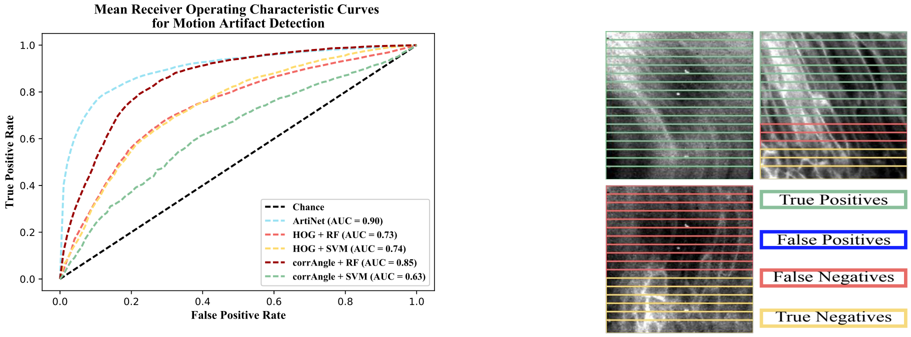
3 Results

The top result is obtained by the artiNet, where an AUC of 0.90 is achieved. The best performance following the conventional pattern recognition pipeline is achieved by the corrAngle feature and RF classification. For this combination, an AUC of 0.85 is reached. The corrAngle-based motion detection with RF performs significantly better than the RF and SVM predictors trained on HOG, where approximately equal ROC curves with AUC values of 0.73 and 0.74 are achieved. In contrast to the RF classification performance, the results of the SVM classifier trained on the corrAngle features are poor. For additional insight into the performance of the artiNet, the comparison of predictions for single image slices and respective labels is visualized in Fig. 3.

4 Discussion

Due to the included patch extraction step of the artiNet, the image borders are discarded. For motion artifact detection, the center of the slice is representative as a motion induced deterioration covers the whole extent in x-direction. Still, the information at the top and bottom of the image is removed. Thus, motion artifacts only deteriorating the rejected areas can not be detected. An improved performance is to be expected if the whole image is used. As the manifestation of stretched cells is underrepresented in the data set, a performance gap between the two possible manifestations of motion artifacts, stripe patterns and stretched cells might occur. Moreover, the performance of the proposed methods was only assessed on data of one clinical team. Hence, additional training instances are required to obtain a robust artifact detector.

Fig. 3. Results of motion artifact detection in CLE images.



(a) ROC curves visualizing the average cross-validation performance of the motion artifact detection.

(b) Example images showing slicewise predictions of the artiNet compared to class labels.

References

1. Blaszczak W, Barczak W, Wegner A, et al. Clinical value of monoclonal antibodies and tyrosine kinase inhibitors in the treatment of head and neck squamous cell carcinoma. Med Oncol. 2017;34(4):60.
2. Oetter N, Knipfer C, Rohde M, et al. Development and validation of a classification and scoring system for the diagnosis of oral squamous cell carcinomas through confocal laser endomicroscopy. J Transl Med. 2016;14(1):159.
3. Nathan CAO, Kaskas NM, Ma X, et al. Confocal laser endomicroscopy in the detection of head and neck precancerous lesions. J Otolaryngol Head Neck Surg. 2014;151(1):73–80.
4. Thong PS, Tandjung SS, Movania MM, et al. Toward real-time virtual biopsy of oral lesions using confocal laser endomicroscopy interfaced with embedded computing. J Biomed Opt. 2012;17(5):0560091–05600910.
5. Neumann H, Vieth M, Atreya R, et al. Prospective evaluation of the learning curve of confocal laser endomicroscopy in patients with IBD. Histol Histopathol. 2011;26(7):867.
6. Aubreville M, Knipfer C, Oetter N, et al. Automatic classification of cancerous tissue in laserendomicroscopy images of the oral cavity using deep learning. Sci Rep. 2017;20(7):11979.
7. Sharman M, Bacci B, Whittem T, et al. In vivo confocal endomicroscopy of small intestinal mucosal morphology in dogs. J Vet Intern Med. 2013;27(6):1372–1378.
8. Jaremenko C, Maier A, Steidl S, et al. Classification of confocal laser endomicroscopic images of the oral cavity to distinguish pathological from healthy tissue. Proc BVM. 2015; p. 479–485.
9. Dalal N, Triggs B, et al. Histograms of oriented gradients for human detection. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. vol. 1 of CVPR '05. IEEE; 2005. p. 886–893.
10. Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit; 2016. p. 2818–2826.