Deep Learning for Dental Hyperspectral Image Analysis

Oleksandr Boiko[1] *, Joni Hyttinen*[1] *, Pauli Fält*[1] *, Heli Jäsberg*[2] *, Arash Mirhashemi*[1] *, Arja Kullaa*[2] *, and Markku Hauta-Kasari*[1]
[1] *School of Computing, University of Eastern Finland, Yliopistokatu 2, P.O. Box 111, 80101 Joensuu, Finland*
[2] *Institute of Dentistry, University of Eastern Finland, Yliopistonranta 1 C, P.O. Box 1627, 70211 Kuopio, Finland*

Abstract

The aim of this work is automatic and efficient detection of medically-relevant features from oral and dental hyperspectral images by applying up-to-date deep learning convolutional neural network techniques. This will help dentists to identify and classify unhealthy areas automatically and to prevent the progression of diseases. Hyperspectral imaging approach allows one to do so without exposing the patient to ionizing X-ray radiation. Spectral imaging provides information in the visible and near-infrared wavelength ranges. The dataset used in this paper contains 116 hyperspectral images from 18 patients taken from different viewing angles. Image annotation (ground truth) includes 38 classes in six different sub-groups assessed by dental experts. Mask region-based convolutional neural network (Mask R-CNN) is used as a deep learning model, for instance segmentation of areas. Preliminary results show high potential and accuracy for classification and segmentation of different classes.

Introduction

According to WHO statistics [1], oral and dental diseases are the most common noncommunicable diseases which affect peoples lives, causing them to suffer from discomfort, pain, and even death. For instance, the Global Burden of Disease Study [2] reports that dental caries of permanent teeth has affected half of the worlds population (3.58 billion people), being the most widely distributed condition assessed. Also, gum disease, which can lead to tooth loss was as the 11th most widespread disease in the world. It is essential to recognize signs of disease as early as possible as dental treatment is expensive and sometimes not affordable in low-and-middle-income countries. The goal of this study is to help dentists to automatically recognize early-stage signs of diseases from oral and dental spectral images. If diseases are treated in their early stages, patients' quality of life is improved, and more severe medical complications can be avoided.

Non-imaging examination techniques include visual observation with a periodontal probe and a mirror, and a chemical indicator for the demineralization process. Imaging methods include optical coherence tomography (OCT) [3][4], X-ray [5], spectral imaging [6][7][8], and dental photographs (Red/Green/Blue (RGB) imaging) to record patients medical history. Current methods are mainly used to detect caries or periodontitis.

Machine learning, deep learning, and hyperspectral imaging applied for computer vision tasks showed high potential in different fields over the last years [9]. Although, only a few recent neural network studies have concentrated on dental applications: Lee et al. [10] have developed an algorithm based on convolutional neural networks (CNN) for prediction and diagnosis of periodontally compromised teeth. Their dataset contains periapical radiographic datasets captured by applying X-Ray imaging. Ionizing radiation is not safe for patients and exposure should be limited. Furthermore, a radiographic image is a grayscale image with relatively limited amount of information. The system of Jae-Hong Lee et al. works as a black box solving only the classification problem, without showing the actual location of the disease. Another study of Jae-Hong Lee et al. [11] is related to detection and diagnosis of dental caries using a deep learning-based CNN algorithm. This study is similarly limited to periapical radiographic datasets. While their algorithm can detect and classify the image of teeth, it does not provide information on caries location, which would be very important for treatment. The authors of the next publication [12] applied deep learning for instance segmentation of the teeth from X-ray images, no disease classification or segmentation is done. The captured hyperspectral dataset used in this work is unique; there is only one publicly available dental dataset of X-ray images used for caries screening [13].

In this paper, we introduce a novel approach to oral and dental image segmentation, which is a combination of hyperspectral imaging and deep learning. Hyperspectral imaging applications are widely used in forestry [14], agriculture and food quality control [15], urban remote sensing [16], and ecology monitoring by remote sensing [17]. Hyperspectral sensors allow us to capture the reflected or emitted electromagnetic energy from an image scene over hundreds of narrow, contiguous spectral bands, from visible to infrared wavelengths [18]. This provides us a possibility to define the physical and chemical properties of scene objects and to detect or identify scene materials. The information is stored in a hyperspectral cube: each layer in the cube represents a spectral band of narrow wavelength range and each pixel contains a spectrum of a corresponding scene area. Thus, hyperspectral images contain significantly more information about the objects in the image scene than grayscale or RGB images.

Medical image feature extraction is a complicated process which requires a high-level specialist. In dentistry, accurate automatic segmentation and classification of different intraoral areas could enable identification of diseases with minimal human resources. Moreover, automated image segmentation can be used by non-experts, and also as an educational tool in dentistry.

Due to their outstanding predictive power, machine learning algorithms have become a key tool for modern hyperspectral image analysis [19]. Deep learning is a part of machine learning methods with deeper layer architecture inspired by neuroscience. Deep neural networks lead in the fields of computer vision, image processing, natural language processing, automatic speech recognition, and fraud detection, providing top performance with no need for manual feature extraction. Deep convolutional neural networks (see Fig. 1) inspired by the mammalian visual cortex system have rapidly become a state-of-the-art methodology for medical image analysis [20]. CNN can find patterns in the image: first few convolutional layers can identify low-level features like lines and corners passing them to further layers. Then, the network begins recognizing more complex patterns as shapes: this provides excellent performance in most modern image recognition challenges [19].

CNN usually contains convolutional layers, pooling layers, activation layers, and fully-connected layers. The main aim of
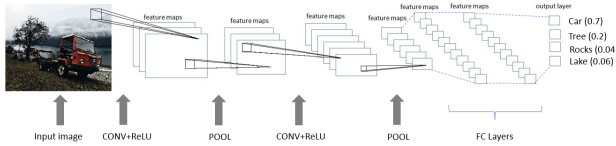
Figure 1. *CNN architecture. The initial input layer stores image pixels (left) and output stores class scores (only the top 4 presented). CONV+ReLU layers compute dot-product between weights and local regions from an input, then apply activation function max (0, x) element-wise. POOL layers downsample feature maps along the spatial dimensions. FC layers compute class scores.*

the convolutional layers is to perform filtering to find some patterns. Pooling layer is used to reduce dimensionality. There are two pooling layer types used: max-pooling to output maximum value from some region; average-pooling to output average value of some region. Activation layers change the input value using some function to change the range; for example, ReLU changes negative values to zero. Fully connected layer neurons are connected to all the neurons from another layer.

In this work, mask region-based convolutional neural network (Mask R-CNN) [21] is used, for instance segmentation of different dental areas like enamel, gum, and calculus. This network is an extension of Faster R-CNN [22], adding a binary mask classifier. Instance segmentation is relatively complicated task aimed to detect all objects in the scene while also precisely segmenting each instance. Mask R-CNN developed by Facebook AI Research (FAIR) outperformed all existing neural networks in three tracks of the COCO suite of challenges [24], including instance segmentation, bounding box object detection, person keypoint detection [21]. Implementation [23] of current network was used for training on our dataset. It is based on Feature Pyramid Network and a ResNet-101 [32] backbone. Examples of instance segmentation of Mask R-CNN based on MS COCO dataset are presented in Figure 2.

Figure 2. *Mask R-CNN results on the COCO test set [24]. The results are based on backbone ResNet-101. The bounding box, category prediction, and class score are presented.*

Practically, training of any CNN requires a relatively large annotated dataset [28]. We used pre-trained weights for MS COCO challenge [24]. Although MS COCO dataset does not contain dental images, it does contain many pictures of natural scenes, which help to recognize the most common patterns. So, we reused most parts of the pre-trained network, changed its output layer and retrained the network on our dataset to detect required classes. This method is called transfer learning.

Materials and methods

Spectral images in the dataset used in this work were captured at the Computational Spectral Imaging Laboratory of the University of Eastern Finland (Joensuu, Finland) and the Dental School Clinic of the University of Eastern Finland (Kuopio, Finland). The research ethical permission for the imaging was issued by the Hospital District of Northern Savo (Kuopio, Finland) and a fully informed written consent was acquired from each patient prior to the imaging. Overall, the dataset contains 116 hyperspectral images of dental lesions acquired from 18 individual participants. The spectral imaging setup used for dataset collection is shown in Figure 3.
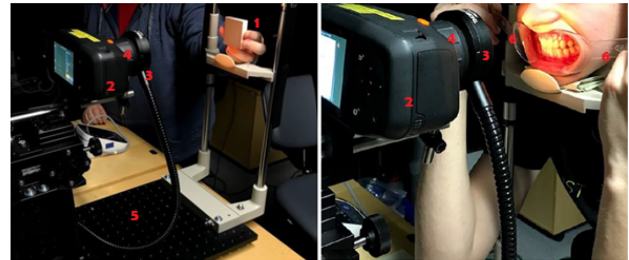

Figure 3. *The process of capturing hyperspectral images. Left: spectral image capture of a neutral gray reference sample. Right: spectral image capture of front teeth and gingiva. Images of the oral cavity and teeth were taken from different viewing angles. List of objects: 1 - neutral gray reference sample; 2 - hyperspectral camera; 3 - illumination ring; 4 - lens; 5 - construction; 6 - lip retractor.*

Dental experts assessed the hyperspectral images and provided ground truth annotations. Annotations include 38 classes in six different subgroups: technical issues, hard tissue and augmentations, soft tissue, hard tissue issues, soft tissue issues, and miscellaneous. Technical issues can be overlapped with other subgroups. An example of rendered RGB image overlayed with annotation masks can be found in Figure 4. The spectral cameras used were Specim IQ (Specim, Spectral Imaging Ltd., Finland) and CRi Nuance EX (PerkinElmer, Inc., USA). The former captures spectral images in the 400-1000 nm wavelength range and the latter in the 450-950 nm range.

Mask R-CNN extends Faster R-CNN by adding a branch for mask generation in parallel to classification and bounding box regression branches (Fig. 5). Prediction of segmentation masks is made by a small Fully Convolutional Network (FCN) on each Region of Interest (RoI), adding only a small computational overhead to Faster R-CNN. Faster R-CNN itself consists of two steps: a) Region Proposal Network is an end-to-end FCN which predicts rectangular object boundaries and prediction scores by sliding over the feature map from the last shared convolutional layer; b) parallel classification and bounding box extraction from features extracted by RoiAlign from previously predicted object boundaries 5.

The Mask R-CNN used in current research is implemented on Python3, Tensorflow 1.8, and Keras [23]. Keras is an opensource neural network library created for fast deployment of deep neural networks which uses TensorFlow as it is backend engine for tensor manipulations. ResNet-101 backbone was used as a

Figure 4. *The example of rendered RGB image overlayed with annotation masks.*
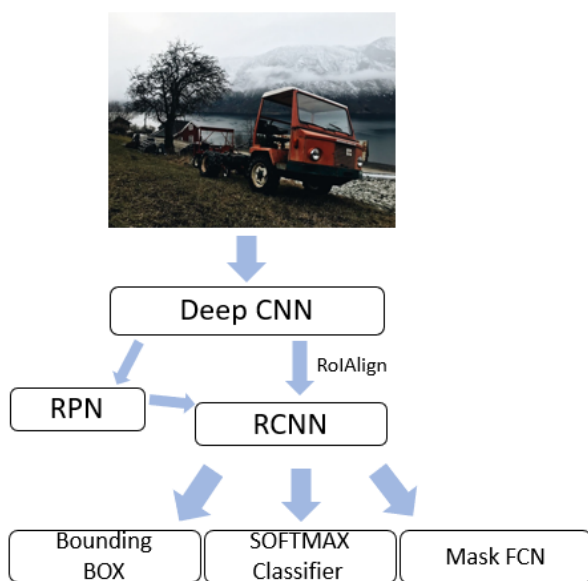

Figure 5. *Mask R-CNN architecture. Input is a 3-channel image. Region Proposal Network (RPN) is a fully convolutional network responsible for rectangular object boundaries prediction. Region-based CNN (R-CNN) generates candidate region proposals and feeds them into CNN. Then, parallel classification, bounding box detection, and mask generation are done.*

feature extractor for Mask R-CNN. Generally, CNN-like neural networks are rarely trained from scratch because it requires a relatively large dataset. We used a pre-trained model on MS COCO dataset. MS COCO includes 330k images (¿200k labeled), 1.5 million object instances, 80 object categories and 91 stuff categories used for object segmentation and recognition in context [24]. This gives us a considerable variety of different natural scenes and objects, which is essential to form filters for low-level and high-level feature extractions. Transfer learning technique is applied to the deep neural network to re-train the last output layer for segmentation and classification of different areas inside the mouth, while previous layers are responsible for extracting low-level features. Grouping of different low-level features leads to different class-representations.

Initial preprocessing of spectral images includes flat-field correction [33] and feature scaling to range [0,1]. Flat field correction is applied to exclude the effect of illuminant and camera by using the radiance of standard gray diffuse reference panel.

The equation is used to compute the sample's reflectance [25]:

$$R_s(x,y;\lambda) = \frac{DN_s(x,y;\lambda) - DN_d(x,y;\lambda)}{DN_{gr}(x,y;\lambda) - DN_d(x,y;\lambda)} \cdot R_{gr}(\lambda) \qquad (1)$$

where $R_s(x,y;\lambda)$ - the reflectance of imaged scene, $DN_s(x,y;\lambda)$ - the intensity values of imaged scene, $DN_{gr}(x,y;\lambda)$ - the intensity values of gray sample measurement data, $DN_d(x,y;\lambda)$- the intensity of dark current data, $R_{gr}(x,y;\lambda)$ is the spectral reflectance of the gray sample. Gaussian filtering is applied to eliminate the effect of uneven illumination. Poorly illuminated areas at the sides of the oral cavity is a common issue for dental imaging, and it is not yet solved. Due to relatively long capture time for Nuance EX (approximately 1.5 min - 2 min), generalized Dual Bootstrap Iterative Closest Point (GDBICP) [26] is used to to register spectral bands misalignment to eliminate the effect of movement.

To apply two-dimensional Mask R-CNN on hyperspectral images, we need to reduce the number of bands by choosing the ones containing the most meaningful information. The pre-trained network is trained on 3-channel RGB images. The image dataset used for transfer learning should be resized to the same spatial resolution, and all input images should have three channels to enable the use the pre-trained network.

Spatial resizing is performed by the system automatically to 1024 by 1024 px while preserving the aspect ratio. Zero-padding is added at the top-bottom or left-right for non-square images. The training process was done using 3.20 GHz Intel Core i7-8700 CPU, 16 GB RAM DDR4, GPU NVIDIA GeForce GTX 1080, 8 GB, on a Windows 10 operating system. Batch size is set to one image due to the limited amount of GPU memory. Learning rate is set to 0.001. Training of the full dataset takes approximately 9-13 minutes for one epoch, 100 steps per each epoch.

Results and discussion

Accurate, automatic instance segmentation would be a very useful tool for the prevention of oral and dental diseases. Hyperspectral imaging allows us to see medically-relevant features that can be challenging or even impossible to see using visual observation, or by using grayscale, RGB, or X-ray images. Combination of hyperspectral imaging and modern deep learning CNN techniques gives us a possibility to tackle the issue in a completely new way, providing promising and precise results. Moreover, open-source deep learning libraries help to perform the research in easier and faster way.

Mask R-CNN predicts boundary box, object class and builds the mask for each detected object. The class score represents class prediction accuracy. Bounding boxes are not shown on preliminary result images for simplicity and better visibility. The accuracy of predicted mask is evaluated via Intersection over Union (IoU) evaluation metric. IoU is a ratio between the area of overlap between ground-truth labeled area and predicted mask area by the area of union.

Spectral band selection with relevant features and Principal component analysis (PCA) were considered as dimensionality reduction methods. PCA did not show any good result; average IoU score was lower than 5% even for well-represented classes like enamel. The 3-channel images combined from PCA inner product images (calculated using the three most significant eigenvectors) are not similar to the MS COCO images, that could be the reason why the algorithm failed.

A previous study [27] identifies 450 nm, 500 nm, and 600 nm bands as the most important features for calculus, gingiva, erosion and caries detection. These bands were used to create 3-channel image dataset from captured hyperspectral images.

Enamel segmentation
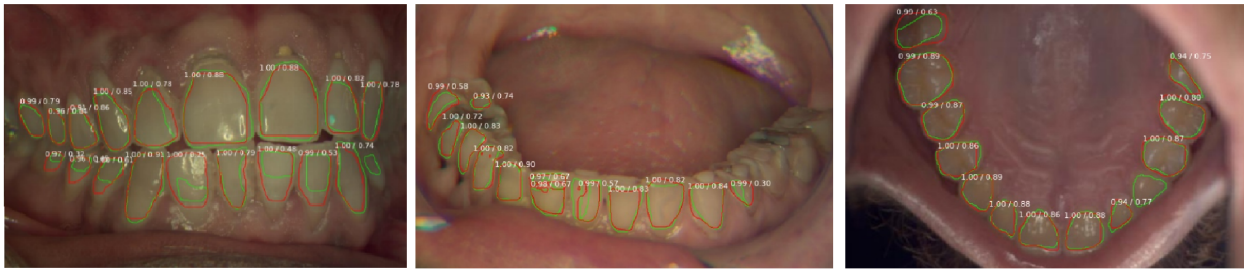
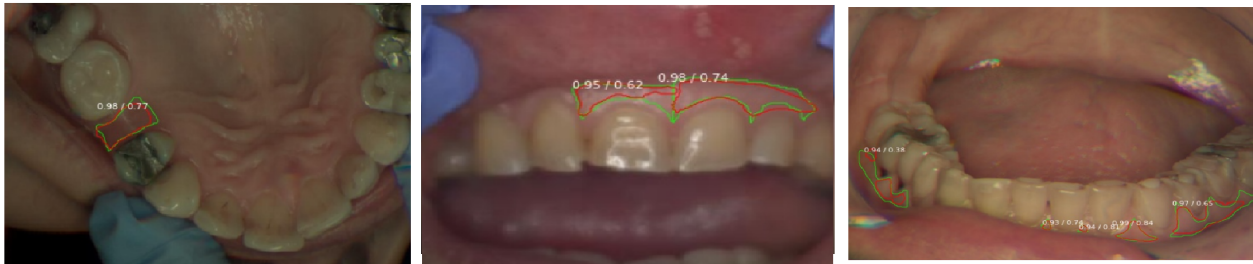# Attached gingiva segmentation

Figure 6. *Instance segmentation results of enamel and attached gingiva using Mask R-CNN. Ground truth annotation is marked as green, predictions marked as red. Caption next to each segmented region represents class prediction score / IoU score in range [0,1].*

The segmentation results for a few classes (enamel, attached gingiva) shows that the approach proposed in this paper is promising (Fig. 6). For enamel: average class prediction accuracy is 99%, and IoU is 77%. For attached gingiva: average class prediction accuracy is 96%, and IoU is 68%.

Further improvements can be done to increase the accuracy of prediction. First, the lack of training data leads to poor segmentation results. This issue could be solved by image augmentation. There are different types of augmentations that can be used to extend the size of the dataset: zooming (rescaling), rotation, flipping (vertically, horizontally, diagonally) and shifting.

Conclusions

Oral and dental diseases are affecting people's lives all over the world. Early-stage detection and prevention is an important issue not yet solved efficiently. Nowadays, periapical radiographic X-ray imaging is one of the most used technology to detect dental lesions, but this technique uses ionizing radiation which can be harmful to the patient and provides only a limited amount of information. Moreover, a highly trained dental expert is required for disease identification.

In this paper, hyperspectral imaging and deep learning convolutional neural networks were applied to develop a novel approach, for instance segmentation and classification of different areas-of-interest inside the mouth. The novelty of the proposed method lies in the usage of non-destructive and effective technology for dental and oral spectral imaging as well as automated image segmentation which can be used by non-experts as well as for educational purposes. Preliminary results of predicted classes (enamel, attached gingiva) proved the potential of the method providing relatively high segmentation accuracy (average IoU scores: 77% for enamel, 68% for attached gingiva). In the future, the accuracy of prediction can be improved by expanding of spectral image dataset both by capturing new spectral images and by using augmentation techniques. Also, additional deep learn-

ing methods will be tested to provide higher accuracy instance segmentation results.

References

[1] Petersen, Poul Erik, The World Oral Health Report 2003: continuous improvement of oral health in the 21st century–the approach of the WHO Global Oral Health Programme., Community dentistry and oral epidemiology, 31, pg. 3-23 (2003).

[2] GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016, The Lancet, 390, 10100, pg. 1211-1259 (2017).

[3] Amaechi, B. T., Higham, S. M., Podoleanu, A. G., Rogers, J. A., Jackson, D. A. 2001. Use of optical coherence tomography for assessment of dental caries: Quantitative procedure. Journal of Oral Rehabilitation.

[4] Le, M. H., Darling, C. L., Fried, D. 2010. Automated analysis of lesion depth and integrated reflectivity in PS-OCT scans of tooth demineralization. Lasers in Surgery and Medicine.

[5] Birdal, R. G., Gumus, E., Sertbas, A., Birdal, I. S. 2016. Automated lesion detection in panoramic dental radiographs. Oral Radiology.

[6] Prasanth, C. S., Betsy, J., Jayanthi, J. L., Nisha, U. G., Prasantila, J., Subhash, N. 2013. In vivo inflammation mapping of periodontal dis-

ease based on diffuse reflectance spectral imaging: a clinical study . Journal of Biomedical Optics.

[7] Zakian, C., Pretty, I., Ellwood, R., Hamlin, D. 2008. In vivo quantification of gingival inflammation using spectral imaging. Journal of Biomedical Optics.

[8] Usenik, P., Brmen, M., Fidler, A., Pernu, F., Likar, B. 2012. Automated classification and visualization of healthy and diseased hard dental tissues by near-infrared hyperspectral imaging. Applied Spectroscopy.

[9] Gewali, U. B., Monteiro, S. T., Saber, E. 2018. Machine learning based hyperspectral image analysis: A survey. ArXiv, abs/1802.08701.

[10] Jae-Hong Lee, Do-Hyung Kim, Seong-Nyum Jeong, Seong-Ho Choi, Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm, Journal of Dentistry, 77, pg. 106-111 (2018).

[11] Lee JH, Kim DH, Jeong SN, Choi SH, Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm., J Periodontal Implant Sci.,48(2), pg. 114123 (2018).

[12] Jader, G., Fontineli, J., Ruiz, M., Abdalla, K., Pithon, M., Oliveira, L. 2019. Deep Instance Segmentation of Teeth in Panoramic X-Ray Images. In Proceedings - 31st Conference on Graphics, Patterns and Images, SIBGRAPI 2018.

[13] Rad, A. E., Rahim, M. S. M., Rehman, A., Saba, T. 2016. Digital Dental X-ray Database for Caries Screening. 3D Research.

[14] Azadeh Ghiyamat and Helmi ZM Shafri, A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment, International Journal of Remote Sensing,31(7), pg. 1837-1856 (2010).

[15] Laura M Dale, Andre Thewis, Christelle Boudry, Ioan Rotar, Pierre Dardenne, Vincent Baeten,and Juan A Fernandez Pierna, Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review, Applied Spectroscopy Reviews, 48(2), pg. 142-159 (2013).

[16] Elizabeth A Wentz, Sharolyn Anderson, Michail Fragkias, Maik Netzband, Victor Mesev, Soe W Myint, Dale Quattrochi, Atiqur Rahman, and Karen C Seto, Supporting global environmental change research: A review of trends and knowledge gaps in urban remote sensing, Remote Sensing, 6(5), 3879-3905 (2014).

[17] Shukla, A., Kot, R, An Overview of Hyperspectral Remote Sensing and its applications in various Disciplines, IRA-International Journal of Applied Sciences (ISSN 2455-4499), 5(2), pg. 85-90 (2016).

[18] Michael Theodore Eismann, Hyperspectral remote sensing, (SPIE Press Bellingham, 2012).

[19] Gewali, Utsav B., Sildomar T. Monteiro and Eli Saber. Machine learning based hyperspectral image analysis: A survey., CoRR, abs/1802.08701 (2018).

[20] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, Clara I. Snchez, A survey on deep learning in medical image analysis, Medical Image Analysis, 42, pg. 60-88 (2017).

[21] He, Kaiming, Georgia Gkioxari, Piotr Dollr and Ross B. Girshick. Mask R-CNN., (ICCV, 2017), pg. 2980-2988.

[22] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Computer Vision and Pattern Recognition (CVPR) (2016).

[23] Waleed Abdulla, Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow, Github (2017).

[24] Lin, Tsung-Yi, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ra-

manan, Piotr Dollr and C. Lawrence Zitnick, Microsoft COCO: Common Objects in Context, (ECCV, 2014).

[25] Li, Q., He, X., Wang, Y., Liu, H., Xu, D., Guo, F. 2013. Review of spectral imaging technology in biomedical engineering: achievements and challenges. Journal of biomedical optics.

[26] Yang, G., Stewart, C. V., Sofka, M., Tsai, C.-L. 2007. Registration of challenging image pairs: Initialization, estimation, and decision. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[27] Hyttinen, J. Flt, P. Fauch, L. Riepponen, A. Kullaa, A. Hauta-Kasari, M., Contrast Enhancement of Dental Lesions by Light Source Optimization. Image and Signal Processing, 8th International Conference, ICISP Proceedings, 10884, pg. 499-507 (2018).

[28] ”CS231n Convolutional Neural Networks for Visual Recognition”. cs231n.github.io. (2018).

[29] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, pg. 91-99. (2015).

[30] Z. Zhong, J. Li, Z. Luo, and M. Chapman, Spectral-spatial residual network for hyperspectral image classification: A 3-d deep learning framework, IEEE Transactions on Geoscience and Remote Sensing, PP (99), pg. 1-12 (2017).

[31] Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi, Deep feature extraction and classification of hyperspectral images based on convolutional neural networks”, IEEE Transactions on Geoscience and Remote Sensing, 54(10), pg. 6232-6251 (2016).

[32] He, Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pg. 770-778 (2016).

[33] Li, Qingli, Xiaofu He, Yiting Wang, Hongying Liu, Dongrong Xu and Fangmin Guo, Review of spectral imaging technology in biomedical engineering: achievements and challenges, Journal of biomedical optics, 18 10, 100901 (2013).