Automatic Site-Specific Multiple Level Gum Disease Detection Based on Deep Neural Network

Guan-Hua Li
School of Information Engineering
Guangdong University of Technology
Guangzhou, People's Republic of China
2111903047@mail2.gdut.edu.cn

Tai-Chiu Hsung*
Department of Computer Science
Chu Hai College of Higher Education
Hong Kong
ORCID: 0000-0001-7699-1937

Wing-Kuen Ling
School of Information Engineering
Guangdong University of Technology
Guangzhou, People's Republic of China
ORCID: 0000-0002-0633-7224

Walter Yu-Hang Lam
Faculty of Dentistry
The University of Hong Kong
Hong Kong
ORCID: 0000-0001-5530-2645

George Pelekos
Faculty of Dentistry
The University of Hong Kong
Hong Kong
ORCID: 0000-0003-1917-1988

Colman McGrath
Faculty of Dentistry
The University of Hong Kong
Hong Kong
ORCID: 0000-0001-9379-0889

***Abstract*—The Gum diseases (gingivitis and periodontitis) is one of the most prevalent dental diseases which are initiated by dental plaque (bacterial biofilm). It has been strongly linked to the systemic diseases including cardiovascular (atherosclerosis, hypertension, stroke), respiratory (aspiration pneumonia), adverse pregnancy outcomes and even cancer via systemic routes with significant health implications. As the inflammation of gum is manifested as increased in redness (colour), increase in volume (oedema), and loss of surface characteristics (stippling; gum fibre attachment). These diseased sites are site-specific (i.e. subject can have healthy and disease sites in a mouth) can be identified by visual examination of dentists. Moreover, these inflammatory changes of gum can also be recognized by intraoral photography which has been clinical practice of regular dental check-up. The aim of this study is to train the computer to identify the inflamed disease sites in pixel level by deep learning approach. We collected 337 and 110 images for training and validation respectively from 110 patients' standard intraoral photographs and randomly. They are labeled into four health status levels (healthy, questionable healthy, questionable diseased and diseased) and verified by a dental specialist with more than 15 years clinical experience. The proposed semantic segmentation architecture is based on the DeepLabv3+ network with Xception and MobileNetV2 as the backbone. Experimental results show the effectiveness of the proposed system, which shows possible application on dental self check-up using mobile app particularly during the disease pandemic where visit to dentists are difficult or even impossible.**

***Keywords*—gingivitis, semantic segmentation, deep learning, network, Dentistry**

I. INTRODUCTION

The most prevalent dental diseases are tooth decay (caries) and gum diseases (gingivitis and periodontitis). It is evidence that these diseases are caused by dental plaque (bacterial biofilm) [1]-[3]. Although most patients brush their teeth every day, they cannot keep all their teeth clean. Areas in the mouth that are difficult to access, such as crowded areas, posterior teeth or interdental areas, are usually affected (site-specific) [4]. After thorough professional tooth cleaning, plaque will begin to accumulate on the surface of the teeth near the edge of the gum within few days. Clinical studies indicating that regular disruption to plaque are needed and can prevent and arrest gum disease [5]. However, dental diseases may take years to develop, the patient usually does not have any pain symptoms [6] unless the disease has progressed to the advanced stage. Significant amount of resources and clinical time have been used to motivate and instruct patients to keep their mouth clean and yet the results are not satisfactory. It is desirable to adopt an automated technique for monitoring oral health daily so we can seek for treatment when it is needed.

Patients' response to plaque accumulated at the gum margin is by inflammation which brings more blood cells to the site to fight against the bacterial invasion [7]. Inflammation of gum is manifested as increased in redness (colour), increase in volume (oedema), and loss of surface characteristics (stippling; gum fibre attachment) [8]. These diseased sites can be identified by visual examination of dentists. Moreover, these inflammatory changes of gum can also be recognized by intraoral photography or photograph mapped 3D models [9]-[11]. The objective of this research is to apply artificial intelligent (AI) techniques to detect gum inflammation from intraoral photographs. As the target inflammation site is at gum margin with varied shape and size, semantic segmentation at pixel level is needed.

The first attempt to automated segmentation of gingival diseases from intraoral images with deep learning approach is proposed in [12]. It adopts an autoencoder network architecture with deep convolution neural network. The dataset used comprises 405 color-augmented intraoral biomarker images from 150 individuals. Areas of gingival inflammation were labelled by dental professional and the trained network is able to predict the inflammation with area under the receiver operating characteristic curve (AUC) 0.746. The precision and recall values are 0.347 and 0.621, respectively. The network was trained with the labeling on diseased gum. Some calculus on teeth was also predicted as diseased gum as its yellowish color is technically close to that of diseased gum. Moreover, some parts of uninterested gingival area were also predicted as diseased gum.

In [13], a Multi-task deep learning framework was developed for locating regions with 5 common gum diseases (periodontal disease, dental caries, soft deposits, calculus and tooth discoloration). This data set contains 3182 oral photos annotated by dental experts. It achieves high positioning accuracy with an average detection sensitivity of 0.787.

In this study, we consider a pixelwise multiple level detection of gingival inflammation sing semantic segmentation network architecture. According to the nature of the application and the limitation of the data set, many network architectures for segmentation are proposed. In [14], Long et. al proposed an end-to-end, pixels-to-pixels fully

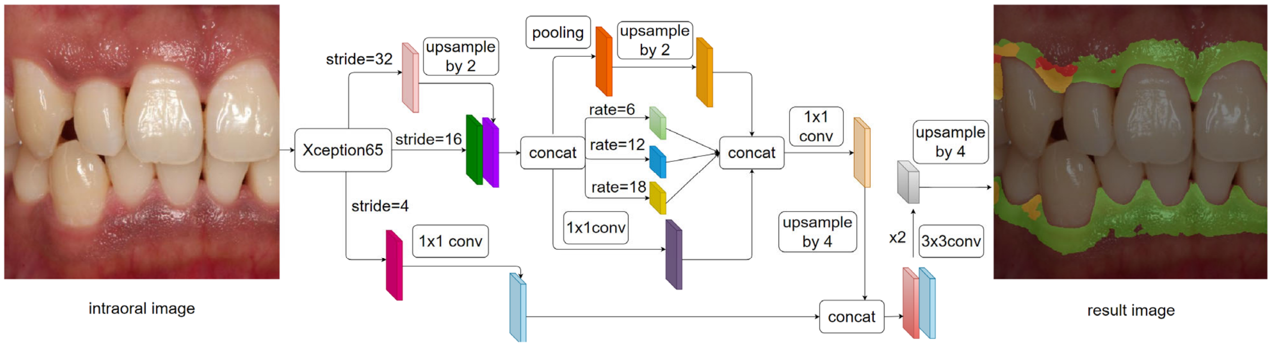


Fig 1. The proposed network architecture based on DeepLabv3+ with Xception65 as the backbone. The input intraoral image on the left is firstly fed into the encoder-decoder structure, follows with multi-scale ASPP modules and bilinear upsampling. Prediction result is overlaid on the input image on the right.

convolutional network for semantic segmentation. It adapts contemporary classification networks and transfers their learned representations to output segmentation tasks. It reaches 62.7% and 62.2% respectively on the PASCAL VOC 2011 and 2012 test sets in terms of mean intersection-over-union (mIOU). Zhao et.al [15] proposed pyramid scene parsing network (PSPNet) for scene parsing. It fuses multi-scale pooling layers features information from a pretrained ResNet by concatenation and convolution layer to obtain per-pixel prediction, reaching 41.68% and 80.04% on ADE20K dataset in terms of mIOU and pixel accuracy. Chen et. al [16][17] proposed adopting atrous convolution to spatial pyramid pooling module to extract multi-scale features. This improves the decoder with more effective segmentation results especially along object boundaries. It achieves mIOU 89.0% and 82.1% performance on the PASCAL VOC2012 and Cityscapes test sets, respectively. Sandler et. al [18] proposed a novel layer module: the inverted residual block with linear bottleneck. This block further improves the performance. In this module, the compressed features obtained from input images are expanded to high dimension and being filtered with lightweight depthwise convolution. Then project back to the same lower dimension as the input. It achieves mIOU 75.32% and only requires 2.75B multiply-adds (MAdds) on the PASCAL VOC 2012 validation set with the MobileNetV2 + DeepLabv3 [19] inference strategy. Chollet [20] proposed depthwise separable convolutions and created the novel Xception architecture. Xception shows small gains in classification performance on the ImageNet dataset and large gains on the JFT dataset, compared to similar parameter count Inception V3.

In [21], Ronneberger et. al proposed U-net for medical image (transmitted light microscopy images) segmentation. The network concatenates the feature maps from upsampling output and corresponding encoding layer output, to get more contextual information. It works excellent for semantically simple targets such as edges, lines, and etc. However, it is less effective for more complicated semantic segmentation dataset. Badrinarayanan et.al [22] proposed SegNet, which is an encoding-decoding network, for multi-class semantic segmentation. The decoder uses pooling indices computed in the max-pooling step of the encoder so avoid decoder learning for upsampling. This improves training time and reduces realization costs.

Among the existing network architectures, DeepLabv3+ [16] appears to be the most suitable for the current applications due to its high transferability. For gingival inflammation, the variability of image features is high in terms of shape, color and textures. DeepLabv3+ offers several pre-trained checkpoints on (1) PASCAL VOC 2012, (2) Cityscapes, and (3) ADE20K, which can be used for transfer learning to the gingival dataset. In this paper, the objective of the current study is to segment the multi-level health status of the gums, based on the intraoral image annotated by dental experts. DeepLabv3+ encoder-decoder network with state-of-the-art pre-trained backbone Xception and MobileNetV2 is proposed to perform pixelwise semantic segmentation of the gingival inflammation from the intraoral photographs. The photographs are indexed by a dental specialist with more than 15 years clinical experience to obtain the index category images for the network training. In the next section, the dataset and adopted network are described. Finally, it is evaluated in section III.

II. MATERIAL AND METHODS

A. Compliance with Ethical Standards

Standard intraoral photographs were taken using a single lens reflex (SLR) camera (EOS 500D Canon) with a macro lens (EF 100mm f/2.8, Canon) and a ring flash (Macro Ring Lite MR-14EX, Canon). The intraoral photographs of patients from the Faculty of Dentistry, The University of Hong Kong (HKU), which underwent periodontic treatment, were collected for the preliminary study. The study was approved by the Institutional review board of HKU (UW20-230).

B. Ground truth development

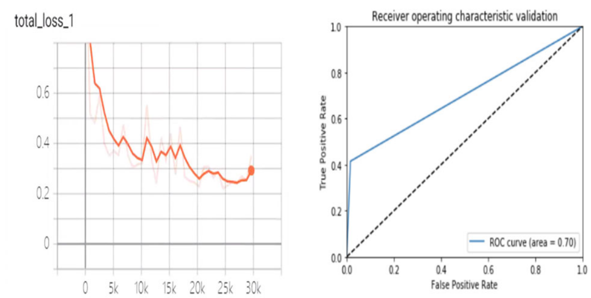
In this study, the gum conditions were marked on the printed intraoral photographs of the patients by an experienced dentist. The annotations were digitized to label maps using LabelMe [23]. The area of interest of gum (about 5mm from teeth gum margin) is labeled into 4 categories: 1. diseased, 2. healthy, 3. questionable diseased, and 4. questionable healthy. Other areas of the image were set to background corresponding label 0. The label images were further converted to single channel index image format.

C. Data preparation

Total 110 standard intraoral photographs with different resolutions were collected. They are manually cropped into different smaller images, and the target labels occupy the largest possible image, which is very beneficial for training. The size of the cropped image is unified to 512 × 512. The completed dataset is divided into two sets, respectively 337 images for training, and 110 images for validation. Considering that there are multiple images corresponding to one patient, so when dividing the dataset, the image of the same patient will not appear in the two divided datasets. The dataset will save as TFRecord file format.

D. Data augmentation

To enhance the network training, Random Crop, Random Rotation and Vertical Flip are applied to the original image and ground truth mask image as the data augmentation methods.



(a) (b)

Fig 2. Loss(a) and receiver operating characteristic curve(b)

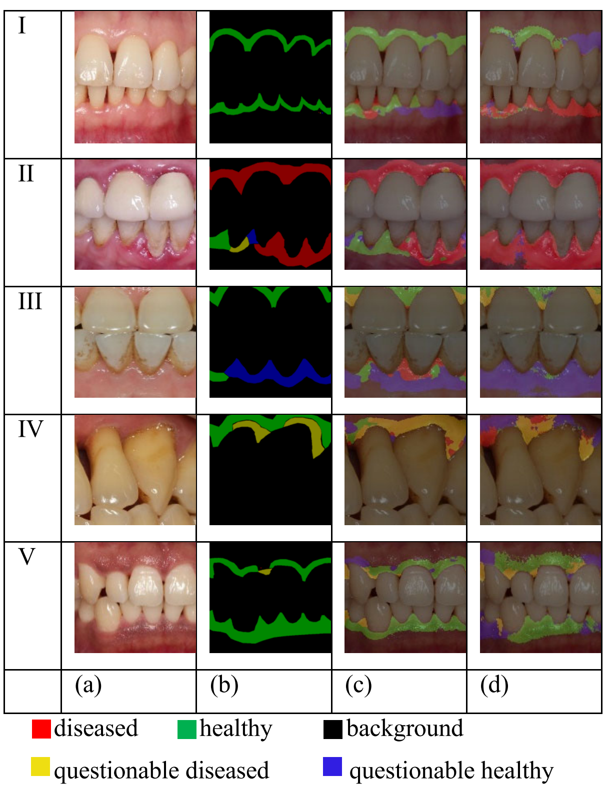


Fig 3. Selected prediction results on the validation dataset using the adopted semantic segmentation network. Rows (I)-(V) show the input images and prediction from different patients. (a) Input intraoral image. (b) Ground truth segmentation interest marked by expert. Prediction result overlay on input image with the backbone of (c) Xception65 and (d) MobileNetV2.

E. The adopted segmentation network

For the site-specific gum disease detection using neural network on mobile platforms, computational cost of prediction is crucial. The adopted network structure is based on DeepLabv3+ (as shown in Fig 1) implemented in TensorFlow [24], running on a Linux system with graphics card NVIDIA GeForce GTX 1080Ti. There are two available backbones for transfer learning: MobileNetV2 and Xception65 with model size of 14 MB and 88 MB respectively. The batch size was set to 4, the number of training iterations was 30,000. Performance metric for multi-class segmentation was mean intersection-over-Union (mIOU). Performance metric for two-class segmentation were mIOU, AUC and precision and recall.

III. RESULT AND DISCUSSION

In this section, we would like to see performance of the proposed system. Firstly, the two-class segmentation suggested in [12] is implemented with our dataset. Then, the performance of the Five-class segmentation of gingival condition model is analyzed. Finally, the performance of the Four-class segmentation of gingival condition model is analyzed. The adopted DeepLabv3+ model with the backbone of Xception65 employs atrous spatial pyramid pooling (ASPP) of 6, 12, 18 and the decoder modules output strides of 4, and the backbone of MobileNetV2 not employs ASPP and decoder modules for fast computation.

A. Two-class segmentation of gingival condition

The ground truth mask image is modified to contain two classes: keeping label 1 (diseased) and the rest set 0 (background). The adopted segmentation model was trained and tested with intraoral images from the training and validation dataset. Some of the prediction results are shown in Fig 2. The area under the receiver operating characteristic curve (AUC) is 0.7, and the precision and recall values are 0.606 and 0.415, respectively. It reaches the mIOU 0.650.

B. Five-class segmentation of gingival condition

The adopted segmentation model was tested with intraoral images from the validation dataset. The ground truth mask images are set to have 5 classes: 0. background, 1. diseased, 2. healthy, 3. questionable diseased, and 4. questionable healthy. In Fig 3, it shows that the visual performance of the system with (c) backbone Xception65 with output stride = 16 and ASPP module with atrous rates 6,12,18, reaching the mIOU 0.379, and (d) the backbone of MobileNetV2 not employed ASPP and decoder modules reaching the mIOU 0.355.

C. Four-class segmentation of gingival condition

In additions to Five-class semantic segmentation, it is also interesting to see how Four-class segmentation performs. The questionable diseased and questionable healthy were combined into one category, where the index of these two categories are encoded as the same index before feeding to the model in training. In Fig 4, it shows the visual performance of the backbone Xception65 with output stride = 16 and ASPP module with atrous rates 6,12,18, reaching the mIOU 0.485, and the backbone of MobileNetV2 not employed ASPP and decoder modules reaching the mIOU 0.449.

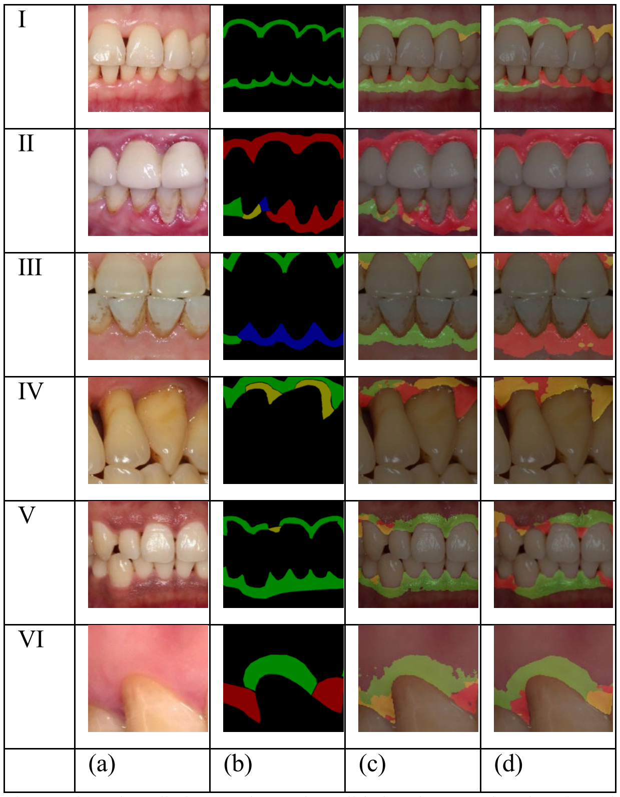


Fig 4. Selected prediction results on the validation dataset using the adopted semantic segmentation network. Rows (I)-(VI) show the input images and prediction from different patients. (a) Input intraoral image. (b) Ground truth segmentation interest marked by expert. (c)-(d) Prediction results, with yellow refer to questionable healthy and questionable diseased, using DeepLabV3+ model with the backbone of Xception65 and MobileNetV2.

IV. RESULT AND DISCUSSION

In this paper, a semantic segmentation network for identifying gum condition was presented. MobileNetV2 and Xception65 are adopted as the backbone of the DeepLabv3+ network. The proposed network model can predict the contour of the site-specific gingival area. Experiment results show that the proposed segmentation model can accurately divide most of the gum inflammation area into five or four categories. The mIOU with Xception65 as the backbone network are 0.379 and 0.485, respectively, and the mIOU with MobileNetV2 as the backbone network are 0.355 and 0.449, respectively.

From [25], it was found that MobileNetV2 is 20 times faster than using Xception at inference time in DeepLabv3 system running on Samsung Galaxy S8 without any additional computing device (e.g. laptop, cloud server), achieving 17.96B FLOPs and 5452 ms inference time with 513 × 513 crop size and 2.18M Params. Considering the real-time performance and memory footprint, MobileNetV2 is the backbone choice in DeepLabv3+ network for mobile applications. To further improve the gingival disease detection, one possible direction is to expand the dataset and optimize the network structure. It is because big data [26] can avoid overfitting the deep learning models while data augmentation cannot overcome all biases present in small dataset.

ACKNOWLEDGMENT

This work is supported by the Guangdong Provincial Key Laboratory of Intellectual Property and Big Data under Grant 2018B030322016.

REFERENCES

[1] J.K. Clarke, "On the bacterial factor in the Ætiology of dental caries," *British Journal of Experimental Pathology*, vol. 5, no. 3, pp. 141-147, 1924.

[2] S.S. Socransky and A.D. Haffajee, "The bacterial etiology of destructive periodontal disease: current concepts," *Journal of periodontology*, vol. 63, pp. 322-331, 1992.

[3] W.F. Liljemark and C. Bloomquist, "Human oral microbial ecology and dental caries and periodontal diseases," *Critical Reviews in Oral Biology & Medicine*, vol. 7, no. 2, pp. 180-198, 1996.

[4] H.J. Breen, N.W. Johnson and P.A. Rogers, "Site-specific attachment level change detected by physical probing in untreated chronic adult periodontitis: review of studies 1982-1997," *Journal of periodontology*, vol. 70, no. 3, pp. 312-328, 1999.

[5] H. Löe, "Oral hygiene in the prevention of caries and periodontal disease," *International dental journal*, vol. 50, no. 3, pp. 129-139, 2000.

[6] L. Croxson, "Periodontal awareness: the key to periodontal health," *International dental journal*, vol. 43, (2 Suppl 1), pp. 167-177, 1993.

[7] R.J. Genco and J. Slots, "Host responses in periodontal diseases," *Journal of dental research*, vol. 63, no. 3, pp. 441-451, 1984.

[8] G.C. Armitage, "Clinical evaluation of periodontal diseases," *Periodontology 2000*, vol. 7, pp. 39-53, 1995.

[9] Y.H. Lam, T.C. Hsung, Y.Y. Cheng and H.N. Pow, "Mapping intraoral photographs on virtual teeth model," *Journal of Dentistry*, vol. 79, pp. 107-110, Dec. 2018.

[10] T.C. Hsung, Y.H. Lam and H.N. Pow, "Image to Geometry Registration for Virtual Dental Models," *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, Shanghai, China, pp. 1-4, 2018.

[11] H. Tang, T.C. Hsung, Y.H. Lam, Y.Y. Cheng and H.N. Pow, "On 2D-3D Image Feature Detections for Image-To-Geometry Registration in Virtual Dental Model," *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP),* Macau, China, pp. 140-143, 2020.

[12] A. Rana, G. Yauney, L.C. Wong, O. Gupta, A. Muftu and P. Shah, "Automated segmentation of gingival diseases from oral images," *2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*, Bethesda, MD, USA, pp. 144-147, 2017.

[13] Y. Liang, et al., "OralCam: Enabling Self-Examination and Awareness of Oral Health Using a Smartphone Camera," *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA, pp. 1-13, 2020.

[14] E. Shelhamer, J. Long and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640-651, 1 April 2017.

[15] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 6230-6239, 2017.

[16] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Proceedings of the European conference on computer vision (ECCV)*, pp. 833-851, 2018.

[17] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A.L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834-848, 1 Apr. 2018.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 4510-4520, 2018.

[19] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *CoRR*, abs/1706.05587, 2017.

[20] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 1800-1807, 2017.

[21] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

[22] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, Dec. 2017.

[23] B.C. Russell, A. Torralba, K.P. Murphy and W.T. Freeman, "LabelMe: A Database and Web-Based Tool for Image Annotation," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 157-173, May 2008.

[24] M. Abadi, et al., "Tensorflow: Large scale machine learning on heterogeneous distributed systems," *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, USENIX Association, pp. 265-283, 2016.

[25] W. Kim and J. Seok, "Indoor semantic segmentation for robot navigating on mobile," *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN),* Prague, Czech Republic, pp. 22-25, 2018.

[26] C. Shorten and T.M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, article no. 60, 2019.