2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

U-net for auricular elements segmentation: a proof-of-concept study

Michaela Servi*, Elisa Mussi, Roberto Magherini, Monica Carfagni, Rocco Furferi, Yary Volpe

Abstract— **Convolutional neural networks are increasingly used in the medical field for the automatic segmentation of several anatomical regions on diagnostic and non-diagnostic images. Such automatic algorithms allow to speed up time-consuming processes and to avoid the presence of expert personnel, reducing time and costs. The present work proposes the use of a convolutional neural network, the U-net architecture, for the segmentation of ear elements. The auricular elements segmentation process is a crucial step of a wider procedure, already automated by the authors, that has as final goal the realization of surgical guides designed to assist surgeons in the reconstruction of the external ear. The segmentation, performed on depth map images of 3D ear models, aims to define of the contour of the helix, antihelix, tragus-antitragus and concha. A dataset of 131 ear depth map was created;70% of the data are used as the training set, 15% composes the validation set, and the remaining 15% is used as testing set. The network showed excellent performance, achieving 97% accuracy on the validation test.**

I. Introduction

Artificial Intelligence (AI) improves almost every field it touches, and the world of healthcare makes no exception. In the medical field, AI can be used in a wide spectrum of medical specialties, such as cardiology, orthopedics, gastroenterology, etc., to help physicians to faster recognize and diagnose diseases and provide much more effective treatments. Currently, research is mainly focused on modeling human body parts [1] and recognizing conditions from various medical imaging sources [2–5] (e.g., cardiograms, CT scans, ultrasound scans, etc.). Among AI-based techniques, Artificial Neural Networks (ANNs) are increasingly used in the field of biomedical systems [6,7]. In fact, ANNs are exploited in several medical applications, such as biochemical analysis [8], for example to track blood glucose or attempt to calculate blood ion levels, or image analysis for tumor detection or classification of tissues and vessels [9–11]. The strength of these approaches lies in the fact that it is not necessary to provide a specific algorithm on how to identify a disease, since the neural networks learn through examples.

One of the application fields that has attracted most ANN researchers is the segmentation of medical images, which plays an important role in their analysis. In medicine, the task of image segmentation is often required on diagnostic images, to define, for example, the contours of a tumor, or on microscope images, to isolate cells for biological analysis. For this purpose, manual approaches are generally adopted, but in many cases the segmentation process is slow and often tedious. In this scenario, there is a growing demand for computer algorithms that can perform segmentation quickly and accurately without human interaction. An efficient automatic processing without human involvement can avoid human error and reduce time and costs. Compared to classical segmentation methods based on image processing techniques (e.g. thresholding based methods, region and clustering based methods, etc.), the usage of algorithms based on neural networks is considerably growing. In the last years, Deep Learning has demonstrated to improve the performances of ANNs, becoming very popular. Not by chance, Convolutional Neural Networks (CNN) play today an important role in the field of medical image segmentation [12–14]. Typically, CNNs are used for classification tasks where, for each image, the output is a single label. However, in many applications, such as biomedical image processing, the desired output includes localization, i.e. it requires that a class label is assigned to each pixel, in other words, the desired output is the delineation of the contours of an object. Among the many types of CNN available, the U-Net architecture [15] is one of the most famous fully convolutional network architectures suitable for medical semantic segmentation tasks. The U-Net architecture demonstrated to achieve excellent performance on very different biomedical segmentation applications: identification of pulmonary nodules [16], skin lesions [17], brain tumors [18], carotid artery [19], etc. Thanks to data enhancement with elastic deformations, U-net architecture needs only very few annotated images and has an excellent training time. This constitutes a great advantage in the medical field, where most of the time a shortage of labeled data is observed as labeling the dataset requires an expert in this field which is expensive, and it requires a lot of effort and time [20]. In light of the widely demonstrated effectiveness of the U-net architecture in semantic segmentation, the authors intend to exploit this tool for automatic segmentation of auricular anatomical elements. Such segmentation is crucial in the context of autologous ear reconstruction (AER) surgery, i.e. the reconstruction of the outer ear from the patient's costal cartilage tissue. This surgical technique is used in cases where the patient presents total or partial absence of the ear (due, for example, to congenital malformations such

M. Servi, E. Mussi, R. Magherini, M. Carfagni, R. Furferi, Y. Volpe are with Department of Industrial Engineering, University of Florence, Florence, Italy. E-mail: {michaela.servi, elisa.mussi, roberto.magherini, monica.carfagni, rocco.furferi, yary.volpe}@unifi.it

*corresponding author: (phone: +39 0552758703, email: michaela.servi@unifi.it)

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

as microtia) and offers many advantages when compared to other types of treatment.

A. Autologous Ear Reconstruction: a new way of treatment

The AER surgical procedure involves the realization of a 3D structure of the ear, obtained by carving and sculpting the patient's costal cartilage, and its implant in a subcutaneous pocket located in the auricular region. According to the technique proposed in [21], the auricular elements to be reconstructed are helix, antihelix, tragus-antitragus (colored elements in Figure 1), plus a support base.

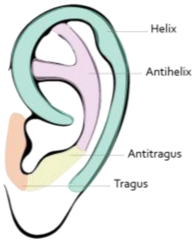


Figure 1. Anatomical elements of the ear.

The result of the surgery, however, strongly depends on the surgeon's manual skills in modeling auricular 3D geometries. The surgery aims at achieving a result that ensures symmetry of the face with the contralateral ear, but the operation is a real challenge for the surgeon since the geometry of the ear is actually very difficult to reproduce. In order to help the surgeon in this procedure the authors have studied and realized 3D printed surgical guides that provide a simplified representation of the patient's ear anatomy and guide the surgeon in cutting the different anatomical elements. An example of the cutting guides is shown in Figure 2.

The realization of the patient-specific medical devices involves the acquisition of the 3D anatomy of the healthy ear (with optical scanning techniques or from CT scan), a mirroring operation to obtain the target anatomy to be reconstructed, and the CAD modeling with appropriate modeling tools.

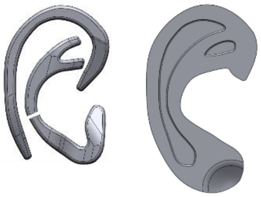


Figure 2. Example of CAD models of the surgical guides created by simplifying the original anatomy.

In detail, the CAD procedure is performed on the correctly oriented 3D model, i.e. the ear must be oriented in such a way that all the elements involved in the reconstruction are visible to the user view point (coincident with one of the global reference system planes e.g. the XY-plane), then the procedure takes place on the 2D sketch created on such plane (example in Figure 3). On the so defined sketch, through well-known CAD operations, the printable models of the surgical guides are created.

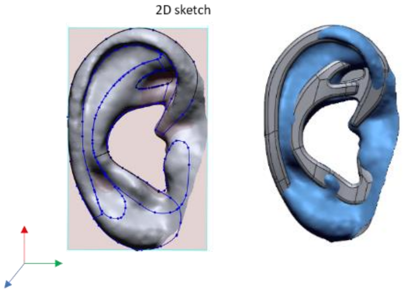


Figure 3. Main phases of the manual modelling procedure of the surgical guides. a) initial anatomy orientation; b) result of the cad modelling procedure.

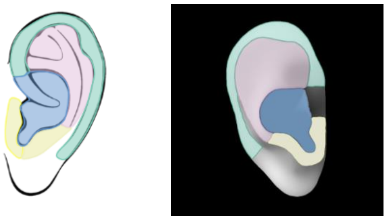
With the aim of making the physician autonomous in the realization of patient-specific cutting guides for AER surgery, the research vision is to develop software tools for designing the semi-automatic design of the medical devices' CAD models. The development of tools easily manageable at hospital level would allow to realize a streamlined and fast production process, to be used within the common clinical practice. To this end, through a software routine, the authors automated the CAD modeling of the surgical guides [22]; the algorithm requires the contours of each anatomical element as input. For the complete automation of the process, it is thus necessary to segment the ear in the involved anatomical components. For this purpose, it was chosen to implement a 2D segmentation algorithm starting from the depth map of the ear, obtained from the correctly oriented 3D model. As well-known the depth map is an image that contains information relating to the distance of the surfaces of scene objects from a viewpoint. Therefore, the proposed algorithm exploits the 2D characteristics of the depth map, which contains, in its definition, depth information defined by the 3D model.

In this perspective, a first segmentation software was proposed by the authors in [23], where state-of-the-art segmentation algorithms based on image processing techniques were analyzed, without being able to find a suitable algorithm to perform segmentation of ear elements. As a result, a very accurate ad-hoc algorithm based on image processing techniques was developed, which however requires setting some initial parameters. To overcome this shortfall, the feasibility of using Deep Learning techniques, specifically the U-Net architecture, for the ear segmentation is evaluated in this study.

II. Materials and Methods

A. Dataset – image annotation

The dataset for training and testing the network consists of 131 ear depth maps. To create the dataset were used 62 computerized tomographies of the head (CTs) and 18 ear scans (the number of retrieved scans exceeds 131 since not all CT scans allow both left and right ear anatomies to be used due to the patient's position during the scan or to congenital malformations of ear). In detail, the anonymized CT scans were retrieved from the CQ500 dataset [24] and further processed with the Mimics Materialise software package [25] to obtain the 3D model of the ears. The remaining 18 ear models were already in 3D form as they were obtained with a professional 3D scanner. The dataset can be considered heterogeneous since the ear constitutes a biometric element whose shape and size are independent of age, gender, and ethnic group [23]. Starting from such three-dimensional models of the ear, to create the depth maps of the dataset was used the algorithm implemented in [23] able to properly orient the ear and create the depth map. The depth maps were then manually annotated in collaboration with a physician. For this feasibility study, it was decided to combine in a single element i) the tragus with the antitragus elements and ii) the antihelix with the triangular fossa and the root of the helix. Moreover, to implement a more comprehensive algorithm to be used in a variety of applications, the concha element (blue in Figure 4) is also considered. Figure 4 shows the target segmentation of the anatomical elements and an example of manual annotation.



(a) (b)

Figure 4. a) ear elements definition and b) example ear manual segmentation.

B. *Model architecture*

The neural network model chosen for the ear segmentation is based on the U-Net architecture [15].

A U-Net consists of an encoder (downsampler) and a decoder (upsampler); in fact the original architecture of the network consists of a contraction path and an expansion path. As for the contraction it follows the typical architecture of convolutional neural networks, i.e. repeated application of two 3x3 convolutions followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. At this stage, each downsampling step doubles the number of channels. The expansion path consists of an upsampling of feature maps followed by a 2x2 convolution which halves the number of channels, a concatenation with the corresponding feature map of the contracting path, and two 3x3 convolutions followed by a ReLU. As a final layer, a 1x1 convolution is used to map each component feature vector to the desired number of classes. The U-Net also provides skip connections in the encoder decoder architecture, this way fine-grained details can be retrieved in the prediction.

In this work, a modified version of the standard U-Net was used. As said, in the first half of the network the characteristics of the input images are extracted using the encoder. Since the task of the encoder is to extract generic characteristics, the initial learning phase based on random input parameters can be replaced with a pre-trained model in order to learn robust features and reduce the number of trainable parameters. More precisely, the intermediate layers outputs of a pre-trained MobileNetV2 model [26] are used as the encoder. As for the decoder, it follows the general structure of the original U-Net. The final transposed convolution has six output channels, since there are six possible labels for each pixel, corresponding to the four anatomical elements (see Figure 1), the rest of the ear anatomy and the background. The network architecture is shown in Figure 5. As far as training is concerned, taking advantage of the use of already pre-trained levels, it is necessary to train only the decoder and the final classifier levels. Moreover, a data augmentation process was applied on the dataset by mirroring all training images and changing the image brightness. The authors did not test for orientation variations since the segmentation task is embedded within a workflow that provides for robust automatic orientation of the 3D model prior to creating the depth map, thus resulting in the standardization of the orientation of the images to be segmented. Taking advantage of transfer learning and data augmentation, it is possible to obtain good results by using a reduced number of input data.

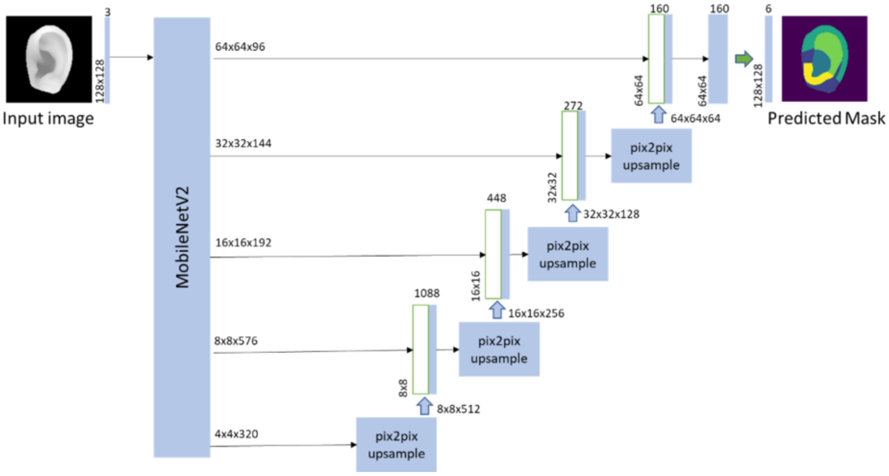
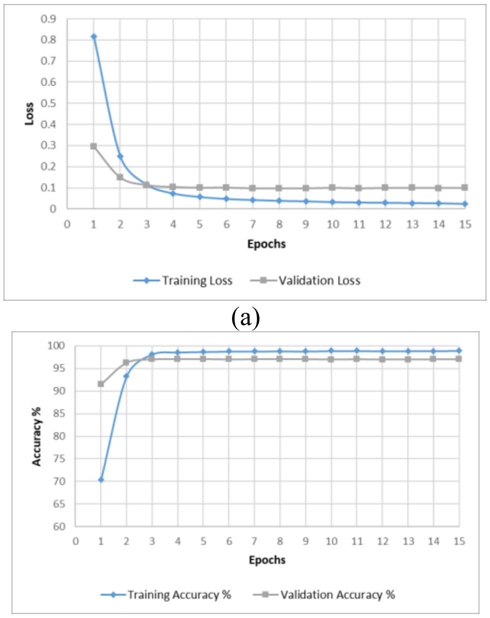


Figure 5. Architecture of the implemented network.

III. Results and Discussion

The neural network was implemented with Tensorflow [27] using the high-level API Keras. The Adam optimizer was used and as loss the sparse categorical crossentropy was evaluated, since there are more than two classes as network output. To train the network the dataset was divided as follows: 70% of the data are used as the training set, 15% composes the validation set, and the remaining 15% is used as testing set. The number of epochs was set to 15, considering that subsequently the accuracy of the network remains stable and the loss increases as shown in Figure 6a. In particular, in Figure 6 is possible to observe the loss trend (Figure 6a) and the accuracy trend (Figure 6b) obtained during the 15 epochs of the training phase, both for the training set and the validation set. As can be seen in the graphs the network reaches an accuracy over 95% after few epochs on both sets, reaching finally an accuracy of about 99% on the training set and about 97% on the validation set. In Figure 6a it can be seen how the loss on the training set has a decreasing trend as the training epochs increase and how this does not happen in the validation set: for this reason, and to avoid overfitting, it is not necessary to carry out a higher number of training epochs.



(a)

(b)

Figure 6. a) loss and b) accuracy graphs for the training epochs.

The network was developed using Colab (a service provided by Google) as it offers many advantages including free access to GPUs so as not to overload personal machines, but with all the disadvantage of not being able to choose and therefore not having a fixed configuration of processing machine. For this reason with the aim of providing reference times, both in terms of training time and prediction time, a fixed configuration was also used. In particular, the machine has a Nvidia GeForce MX150 GPU. Figure 7 shows the average prediction time, calculated on all the dataset images, and the average time per epoch, calculated on 1000 epochs.

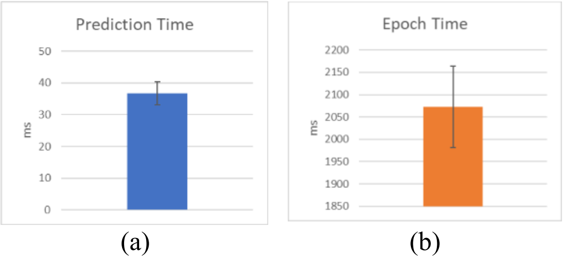


Figure 7. Average prediction time a) and average time for epoch b) obtain with Nvidia GeForce MX150 GPU.

Figure 8 shows some of the segmentations predicted by the network (predicted mask) compared with the corresponding ground truth (true mask).

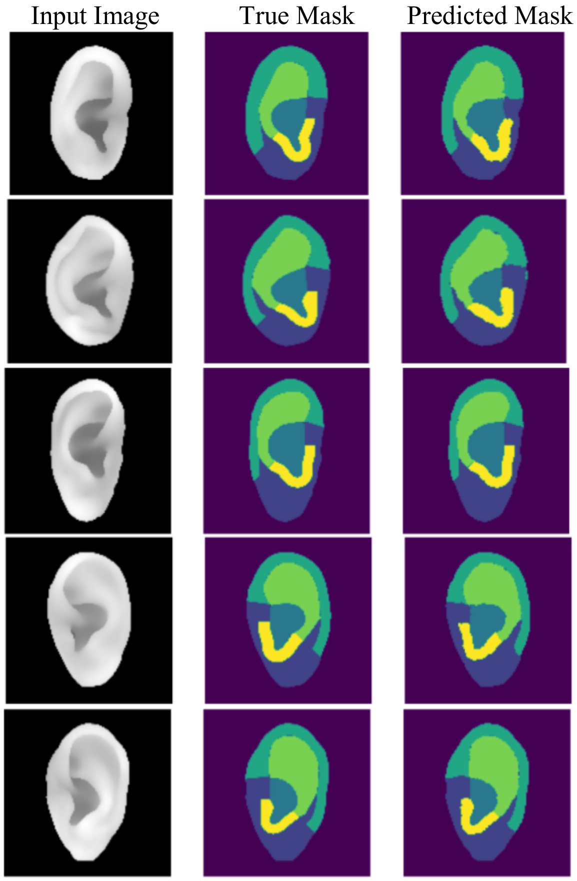


Figure 8. Subset of network experimental results.

The network is able to achieve excellent performance, as can be seen both from the accuracy values in Figure 6b and qualitatively in Figure 8. These results are in any case strongly related to the ear's position within the image: the correct segmentation is in fact strongly dependent on the correct orientation of the ear and, at the moment, the network is not able to produce correct segmentations with input scans not correctly oriented, both on the frontal plane and in space before creating the depth map.

IV. Conclusions

The increasing use of Artificial Intelligence techniques in the medical field has allowed the development of fast, accurate and reliable systems in support of common clinical practice. One of the areas most revolutionized by the introduction of artificial neural networks is the segmentation of medical images. In this context, this work proposes the use of a convolutional neural network, the U-Net, to address the problem of ear segmentation to recognize its anatomical elements with the idea of integrating it into the process of creating custom surgical guides to assist autologous ear reconstruction surgery. The realized network takes as input ear depth map images created after the correct orientation of the 3D model and provides, as output, the ear segmentation in the main anatomical elements. The net was trained using 131 images (plus the ones obtained by mirroring and changing the image brightness) manually annotated and appropriately divided between train test and validation set. The net was trained over 15 epochs reaching an accuracy of 97% on the validation set and, in general, the results obtained are very satisfactory especially considering the reduced number of images used in the training process. Moreover, the average prediction time is ~36 ms on a machine mounting a Nvidia GeForce MX150 GPU, and the average training time for one epoch is ~2 s. The study demonstrated the potential of using this type of network to address the task of ear segmentation. Future developments will concern the analysis of a network for the segmentation of the ear that is applicable to ears not properly oriented on the frontal plane, and therefore on the extension of the dataset to these cases.

References

[1] H.R. Roth, C.T. Lee, H.C. Shin, A. Seff, L. Kim, J. Yao, L. Lu, R.M. Summers, Anatomy-specific classification of medical images using deep convolutional nets, in: Proc. - Int. Symp. Biomed. Imaging, IEEE Computer Society, 2015: pp. 101–104. https://doi.org/10.1109/ISBI.2015.7163826.

[2] F.F. Ting, Y.J. Tan, K.S. Sim, Convolutional neural network improvement for breast cancer classification, Expert Syst. Appl. 120 (2019) 103–115. https://doi.org/10.1016/j.eswa.2018.11.008.

[3] M. Zihlmann, D. Perekrestenko, M. Tschannen, Convolutional Recurrent Neural Networks for Electrocardiogram Classification, 2018. https://github.com/yruffiner/ecg-classification (accessed December 28, 2020).

[4] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A.Y. Ng, Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks, n.d. https://stanfordmlgroup. (accessed December 28, 2020).

[5] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, M. Eramian, Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network, J. Digit. Imaging. 30 (2017) 477–486. https://doi.org/10.1007/s10278-017-9997-y.

[6] E. Anwaitu Fraser, Okonkwo, Obikwelu R, Artificial Neural Networks for Medical Diagnosis: A Review of Recent Trends, Int. J. Comput. Sci. Eng. Surv. 11 (2020) 1–11. https://doi.org/10.5121/ijcses.2020.11301.

[7] S.K. Pandey, R.R. Janghel, Recent Deep Learning Techniques, Challenges and Its Applications for Medical Healthcare System: A Review, Neural Process. Lett. 50 (2019) 1907–1935. https://doi.org/10.1007/s11063-018-09976-2.

[8] A.S. Vickram, A.R. Kamini, R. Das, M.R. Pathy, R. Parameswari, K. Archana, T.B. Sridharan, Validation of artificial neural network models for predicting biochemical markers associated with male infertility, Syst. Biol. Reprod. Med. 62 (2016) 258–265. https://doi.org/10.1080/19396368.2016.1185654.

[9] F. Taher, R. Sammouda, Lung cancer detection by using artificial neural network and fuzzy clustering methods, in: 2011 IEEE GCC Conf. Exhib. GCC 2011, 2011: pp. 295–298. https://doi.org/10.1109/IEEEGCC.2011.5752535.

[10] D. Roffman, G. Hart, M. Girardi, C.J. Ko, J. Deng, Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network, Sci. Rep. 8 (2018). https://doi.org/10.1038/s41598-018-19907-9.

[11] V. Bevilacqua, N. Pietroleonardo, V. Triggiani, A. Brunetti, A.M. Di Palma, M. Rossini, L. Gesualdo, An innovative neural network framework to classify blood vessels and tubules based on Haralick features evaluated in histological images of kidney biopsy, Neurocomputing. 228 (2017) 143–153. https://doi.org/10.1016/j.neucom.2016.09.091.

[12] L. Cai, J. Gao, D. Zhao, A review of the application of deep learning in medical image classification and segmentation, Ann. Transl. Med. 8 (2020) 713–713. https://doi.org/10.21037/atm.2020.02.44.

[13] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, I. Ben Ayed, HyperDense-Net: A hyper-densely connected CNN for multi-modal image segmentation, IEEE Trans. Med. Imaging. 38 (2018) 1116–1126. http://arxiv.org/abs/1804.02967 (accessed December 28, 2020).

[14] M. Bakator, D. Radosav, Deep Learning and Medical Diagnosis: A Review of Literature, Multimodal Technol. Interact. 2 (2018) 47. https://doi.org/10.3390/mti2030047.

[15] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, n.d. http://lmb.informatik.uni-freiburg.de/ (accessed November 6, 2020).

[16] G. Tong, Y. Li, H. Chen, Q. Zhang, H. Jiang, Improved U-NET network for pulmonary nodules segmentation, Optik (Stuttg). 174 (2018) 460–469. https://doi.org/10.1016/j.ijleo.2018.08.086.

[17] B.S. Lin, K. Michael, S. Kalra, H.R. Tizhoosh, Skin Lesion Segmentation: U-Nets versus Clustering, (2017). http://arxiv.org/abs/1710.01248 (accessed December 28, 2020).

[18] H. Dong, G. Yang, F. Liu, Y. Mo, Y. Guo, Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks, Commun. Comput. Inf. Sci. 723 (2017) 506–517. http://arxiv.org/abs/1705.03820 (accessed December 28, 2020).

[19] T. Zhou, T. Tan, X. Pan, H. Tang, J. Li, Fully automatic deep learning trained on limited data for carotid artery segmentation from large image volumes, Quant. Imaging Med. Surg. 11 (2021) 67–83. https://doi.org/10.21037/QIMS-20-286.

[20] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation, ArXiv. (2018). http://arxiv.org/abs/1802.06955 (accessed December 28, 2020).

[21] S. Nagata, A new method of total reconstruction of the auricle for microtia, Plast. Reconstr. Surg. 92 (1993) 187–201. https://doi.org/10.1097/00006534-199308000-00001.

[22] E. Mussi, M. Servi, F. Facchini, M. Carfagni, Y. Volpe, A computer-aided strategy for preoperative simulation of autologous ear reconstruction procedure, Int. J. Interact. Des. Manuf. (2020) 1–4. https://doi.org/10.1007/s12008-020-00723-3.

[23] E. Mussi, M. Servi, F. Facchini, R. Furferi, L. Governi, Y. Volpe, A novel ear elements segmentation algorithm on depth map images, Comput. Biol. Med. 129 (2021) 104157. https://doi.org/10.1016/j.compbiomed.2020.104157.

[24] Download Head CT CQ500 dataset. http://headctstudy.qure.ai/dataset (accessed September 28, 2020).

[25] Materialise Mimics | 3D Medical Image Processing Software. https://www.materialise.com/en/medical/mimics-innovation-suite/mimics (accessed September 28, 2020).

[26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, MobileNetV2: Inverted Residuals and Linear Bottlenecks, n.d.

[27] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D.G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, TensorFlow: A system for large-scale machine learning, Proc. 12th USENIX Symp. Oper. Syst. Des. Implementation, OSDI 2016. (2016) 265–283. http://arxiv.org/abs/1605.08695 (accessed December 28, 2020).