Machine learning framework for maxillofacial preoperative planning

Jakob Botvidsson (BME–19)

***Abstract*—Artificial intelligence in biomedical image processing is approaching human performance at object localization while saving immense amounts of time for the physicians. These AI algorithms have the potential to automatically segment anatomical structures for preoperative planning. However, there are currently no tools such tools on the market. This study propose a framework of generating effective machine learning algorithms, applicable on different anatomical structures, to be used to increase automation in virtual surgical planning software. In this study a limited data set consisting of 34 CT image volumes was used to generate labelled training data to a Convolutional Neural Network (CNN) called Unet. The networks were evaluated with metric evaluation as well as visually evaluated. The framework produced two networks for automatic segmentation, one for the orbital bone and one for the mandibular bone. The orbital automation made useful segmentations ready for 3D printing while the mandible automation needs more work to be able to make printable segmentations. In conclusion this framework provides a viable approach of generating anatomical models for virtual surgical planning.**

I. Introduction

ARTIFICIAL INTELLIGENCE (AI), is a term used widely to explain many different programs and systems. The OECD:s (Organisation for Economic Co-operation and Development) definition is ”an AI system is a machine-based system that is capable of making recommendations, predictions or decisions for a given set of objectives”[1]. The strength of AI is the capability of finding patterns in large volumes of data. When used in applicable environments it can contribute towards productivity and accuracy[1]. A fitting environment being healthcare, where time-consuming manual task can be replaced with automation and relieve doctors in their work[2].

Worth noting is that when provided with biased input the AI also makes biased recommendations, predictions and decisions. One bias being perception bias, which is when AI provided with data that over-/under-represents a population would result in that AI operating better for that population[1]. Another issue with AI is the black box machine learning method, where the black box refers to the underlying incomprehensible network algorithm. The research concerning explainability of networks is growing, it is referred to as Explainable Artificial Intelligence (XAI). XAI seeks to justify decisions made by the AI and provide transparency and traceability of the black box algorithm[3]. The key is to make the network learn the right way.

Inlämnat den June 5, 2022
Emejladress:{ja1416bo-s@student.lu.se}
Teknisk och klinisk handledare: Einar Heiberg, Department of Clinical Sciences, Lund

A. Machine Learning

AI and machine learning is closely linked. Machine learning is when a system adapts and learns from the data that it is processing. The process of machine learning consists of three major steps called data exploratory phase, training phase and validation phase. The purpose of the data exploratory phase is to find patterns in the data. These patterns will then be used to construct a hypothesis to be used in the following steps. The training phase creates a model, ground truth, based on the hypothesis in hopes that when later tested, in the validation phase, it produces results towards the hypothesis[4].

There are many approaches to machine learning, one is the supervised approach. Supervised learning uses labels in the training data to indicate a predetermined output. The objective for the system is to find patterns and create rules for prediction[5].

B. Neural networks

The artificial neural network (ANN) evolved from inspiration taken from our biological neural network in the brain. Its a widely used method of machine learning that is versatile and fast. Like the biological neural network, ANN uses interconnected nodes. Each neural network has specific way of processing signals with weighing between the different nodes and a specific organization and interconnection between nodes. An example can be seen in Figure 1[6].

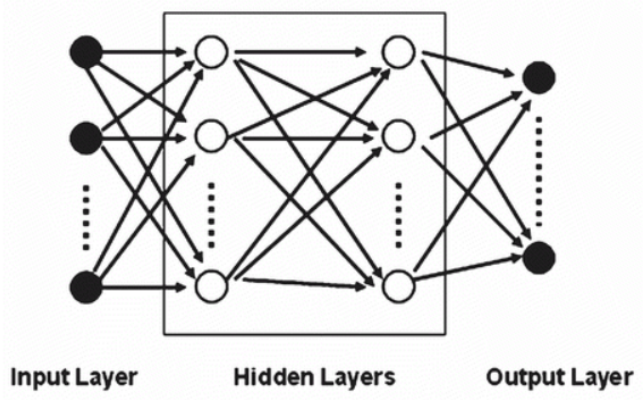

Fig. 1. Feedforward ANN. Networks can contain different amounts of hidden layers.

The network ”learns” through adjusting the weights between nodes. During supervised training the weights are set to minimize amount of faulty prediction of outputs compared to

the provided labeled outputs. The mathematical function used to minimize this error in prediction is called loss function. The power of the network increases with the amount of nodes. But too large networks may risk loosing generalization of unseen data. The network can be extended by adding hidden layers, making it deeper and referred to as deep learning. Emperically, the network is then capable of more generalization and can learn from less data[7].

C. Convolutional neural networks

Convolutional neural networks (CNN) are used for image recognition and object localization. An easy explanation of CNN is ”to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction”[8]. The CNN is similar to a feed-forward network but here the layers of the network has three dimensions. The input could be an image in which case the dimensions would be width, height and depth. Depth refers to number of features, specific patterns in the different small rectangular areas of the image, and width and height refers to dimensions. There are three types of layers in the CNN: convolution, pooling and Rectified Linear Unit (ReLU) layer, each with different operations. The convolutional operation places filters covering the entire image. The dimensions of the filter and the amount of filters determine the dimensions of the next hidden layer, illustrated in Figure 2. Each filter tries to find a feature thus often resulting in a large amount of filters to find a large amount of features. When filtering at the edges of an image information is lost. This is solved with padding, adding pixels with value 0 around the border of the image. Following a convolution is either a pooling or ReLU layer[7].

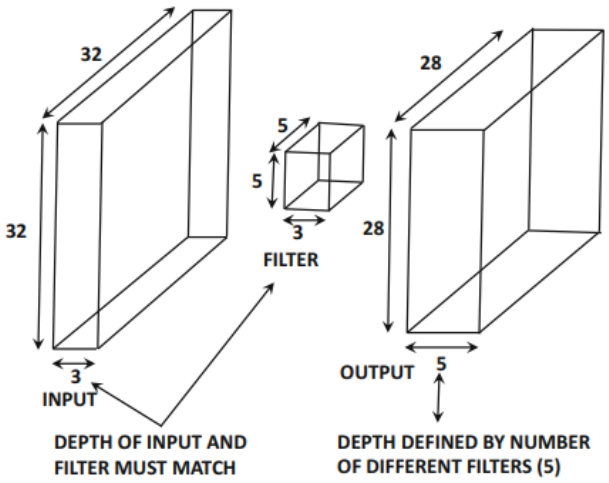

Fig. 2. A convolutional operation between an input of size 32x32x3 and output of dimension 28x28 when applying a filter of size 5x5x3. The change of depth is caused by the amount filters applied to the input.

The pooling operation is used to reduce the size of the image while keeping the information in the image. Max pooling is a pooling operation that takes a small grid region of each layer of the image and outputs the maximum value of that grid, keeping the same depth[7].

A ReLU layer does not change the dimensions of the image. It uses an activation function which puts negative values to zero. This layer increases processing speed, allowing use of deeper models and generating more accurate networks. It is so commonly used in CNN that it is often not shown in network architecture illustrations[7].

1) Unet: There is always a need for an effective neural network in biomedical image processing, to localize important features in the output. Effective networks are built on large amounts of training images, which is difficult if not impossible to attain within the biomedical field. Unet is a network that uses data augmentation to solve this issue and make best use of the available images. Data augmentation is a process that helps stretching and enlarging the data pool through reconstructing the original images with different property changes, without including more patients. The network also uses an expanding, upsampling, path increasing the resolution while applying information from the convolutional, contracting, path in order to localize features in the high resolution image[9]. The high performance of a finely tuned Unet is underscored in the paper *nnu-net: Self-adapting framework for u-net-based medical image segmentation*[10], where the results show ”nnU-net” surpassing or perform on par with specialised deep learning networks in multiple competitions making it a state of the art segmentation tool. nnUnet uses rotation, scaling, noise and many more techniques of augmentation in training[10].

When using Unet to segment a 3D structure a GPU-effective way is to apply the network to 2D images in all three directions, returning 3 outputs. The resulting output is generated through 3D-voting. Here each output ”votes” for the pixel to be of a certain class, the class with most ”votes” win and the pixel is classified and segmented, see Figure 3[11].

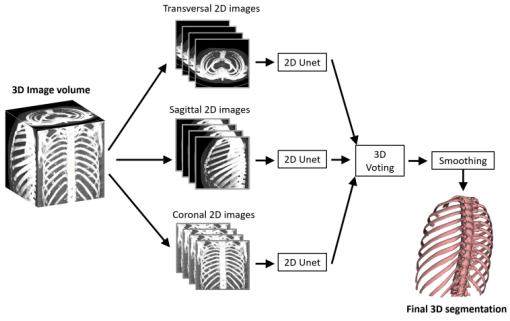

Fig. 3. Illustration of the 3D-voting in a network.

D. Automated segmentation in surgical planning

Mandible and orbital segmentations are two common segmentation tasks in maxillofacial preoperative planning. Therefore, these two task was chosen as perfomance test of the proposed segmentation framework. Studies show successful segmentation of both mandibular and orbital bones[12][13].

The technology is able to perform similarly to manual segmentations by clinicians, with 30 min saved for each segmentation. These images can be used as a diagnostic tool for orbital assessment and for creating physical models for pre-bending orbital implants in patients with fractures[13]. Companies can be found applying these algorithms to their Virtual Surgical planning (VSP) software[14]. But no company markets a software that automatically segment the mandibular bone, which advertises the need for such a software. Especially a comprehensive software containing specialized tools for many different task.

E. Segment 3DPrint

Segment 3DPrint is a software created by Medviso in Lund, Sweden[15]. It uses standardized medical dicom-files to produce printable 3D-models as well as using 3D models to help plan surgery and reducing operation time. The software has multiple tools for delineations of structures, adding and removing of pixels and automatic segmentation of the skull, fetuses and structures of the heart.

Integrating automatic segmentation in VSP could potentially reduce the workload for clinicians and their time spent on manual segmentation tasks. The potential time gained could be applied to other sections of surgical planning or reduce overall stress for clinicians, leading to more successful surgeries. In this study we present a framework for generating machine learning algorithms, that can automatically segment anatomical models, and be added to virtual surgical planning software. This report will present a method of creating machine learning algorithms segmenting the mandibular and orbital bone. The purpose was to evaluate the segmented models made by these algorithms, thus assessing the framework.

II. Method

A. Dataset

We managed to gather a dataset consisting of 3D CT scans centered around the skull and used for 3D-printing of organs or diagnosis of diverse types of lesions. Images were provided by 3D Centre at Skåne University Hospital. Images were completely anonymised except for age and sex. Informed consent was waived by National Ethical Review Board (Dnr 2019-02264). 34 images were acquired.

1) Orbita: Lesions affecting the training data was patients with orbital fractures. All lesions were included in the training as it diversifies the networks. For this dataset, 20 CT scans were generated to be ground truth, 20 were used to evaluate segmentation visually and 13 were either deemed inadequate or unfinished for training due to lack of time.

2) Mandible: Lesions affecting the training data was patients with jaw implants, braces, dentures and no teeth. All lesions were included in the training as it diversifies the networks. For this dataset, 15 CT scans were generated to be ground truth, 1 CT scan was used to evaluate segmentation statistics and to evaluate segmentation visually and 17 were deemed inadequate or unfinished for training due to lack of time.

B. Generating ground truth

This process used 3D CT images as input and the output is, through segmentation, a labeled image were the target anatomical structures are delineated and classified, in the form of .mat-files. Segment 3DPrint and its functions was used for this. Segment 3DPrint and its tools was taught by supervisors and experienced personnel along with the anatomical structures of interest to be outlined manually. Therefore the delineation and estimation of completeness of the images could be done by the author. The earliest images were also inspected by supervisors to ensure the proper quality and precision. It was also assured that only one person did the delineation to allow for a similar method and therefore homogeneous training data.

Images with bad quality, such as varying intensity and contrast, lots of artifacts and low resolution, are put aside in order to work on higher quality images.

The manual moments were carefully executed to ensure precise training data. Tools in Segment 3DPrint were used to achieve this and also remove unnecessary manual tasks. The skull segmentation-function was used on all images to attain correct threshold intensity and label everything over threshold as Skull and below threshold as Background. This also provided a stable base method of segmenting the skull. The images were cropped into the sought after 3D-region-of-interest (ROI) in order to minimize computational demand and the 2D slices were similarly aligned in the transversal, sagittal and coronal view to allow for a consistent delineation approach.

Medical images always vary in intensity and contrast. In order to achieve similar characteristics in the different image segmentations, different tools and methods of Segment 3DPrint were utilized. Grainy images were refined using smoothing. Artifacts derived from metal implantations were removed by adjusting threshold intensity, isolating and subtracting objects and erasing pixels. After this the delineation process differs between the ground truth generation for the two networks.

1) Mandible: The mandible was separated and isolated from the skull and the two different objects were classified as mandible and bone. The separation was done through erasing the pixels interconnected between the skull and the mandible at the temporomandibular joint and, if the patient was biting down, at the teeth, with the sagittal and coronal point of view. This allowed for creating a new object from the now isolated mandible and classifying it. By subtracting the volume of the mandible-object from the skull-object, the bone-object was created and classified, giving images the classes: Mandible, Bones and Background. It was determined that this was the

simplest delineation process and that this would result in a easier ground truth for the network to train on.

2) Orbita: For the orbita training data, pixels were added in the areas where the bone was too thin to be segmented using skull segmentation. Pixel adding was done by drawing lines in each image slice between visual cues in the underlying CT-image and some individually segmented pixels. This was done with the coronal point of view. All pixels were classified as bones since the orbital bones together with the skull is a solid mass with no edge in between. Classes in the training data was Orbita (skull) and Background.

C. Hyper parameters

The tuning of the hyperparameters of the model were in reference to a previous study by Amanda Nilsson at Lund University[11]. The hyperparameters were encoder depth, learning rate, loss function as well as some augmentation parameters not included in Table I.

The different augmentation settings can be seen in Table I together with parameter settings for Image volume and Patch creation.

TABLE I
DIFFERING PARAMETERS FOR THE TRAINING THE TWO NETWORKS.

Parameters	Orbita	Mandible
Image volume		
IMinV	-200	-500
IMaxV	1000	2000
Patch creation		
Pdirections	[1 2 3]	[2 3]
PBlackProbability	0.01	0
PGlobalWriteFactor	0.2	0.3
Augmentation		
APhantomWriteFactor	0	0.5
AXScaleRange	[0.8 1.3]	[0.7 1.3]
ATranslate	0.2	0
ATranslateXRange	[-0.25 0.25]	[-0.5 0.5]
ATranslateYRange	[-0.25 0.25]	[-0.5 0.5]
ARotateRange	[-15 15]	[30 30]

The qualifying and subsequently final parameters were established by the author and supervisor through rational analysis of the anatomy and segmentation result. Although it is impossible to assure that these were the superior settings.

D. Segmentation evaluation

1) Orbita: The images to test the segmentation on, needs to be unseen by the network. Unseen meaning that the images was not used as training data, to validate future use of the network on new patients. We picked 20 CT scans for testing the segmentation, consisting of patients evenly split between female and male and representing ages between 23-102yo.

After some testing it was clear that aligning the transversal and coronal plane as seen in Figure 4, resulted in a better

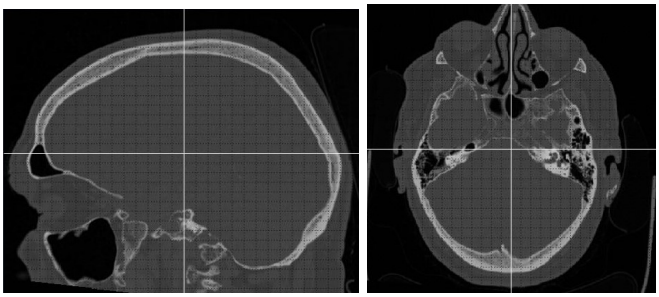

Fig. 4. Optimal alignment of the transversal and coronal plane for automatic segmentation.

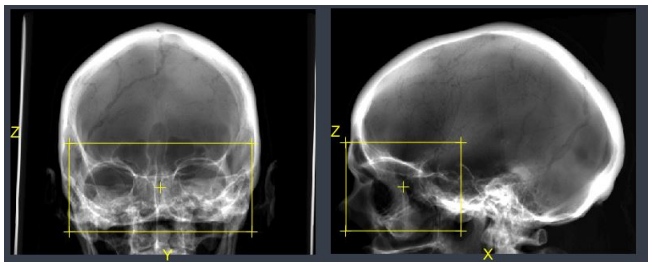

Fig. 5. Optimal ROI, within the yellow borders, for automatic segmentation.

segmentation of the orbital bone. The automatic segmentation-ROI was chosen to contain the orbital bone together with some overlap into the surrounding skull, see Figure 5.

Because of the nature of the issue that the automatical orbital segmentation tries to solve, the best method of evaluation is a visual assessment. It is difficult to decide, based on qualitative measures such as Dice Score, if there are any anatomically incorrect holes in the orbital region. In order to have a statistical, universal result we created a grading of the different automatic orbital segmentations. The different grades are further disclosed in Table II.

A segmentation assigned the grade 3 can be seen in Figure 6, Figure 7, Figure 8. In this example the holes of Figure 6 and Figure 8 could easily be filled in and made good enough for printing within a reasonable amount of time, which should result in a grade 4. The grade 3 is due to the hole in Figure 7. It is not fixable in an acceptable time-frame using simple Segment 3DPrint-tools. Comparing this to the

TABLE II
GRADING CRITERIA. 5 IS THE BEST AND 1 IS THE WORST.

Grade	Criteria
5	Excellent quality, can be printed directly
4	Some holes, acceptable, needs some clicks to prepare for printing
3	Some holes, acceptable, needs work to prepare for printing
2	Some holes, acceptable, needs extensive work to prepare for printing
1	Too many flaws to print, needs extensive work to fix

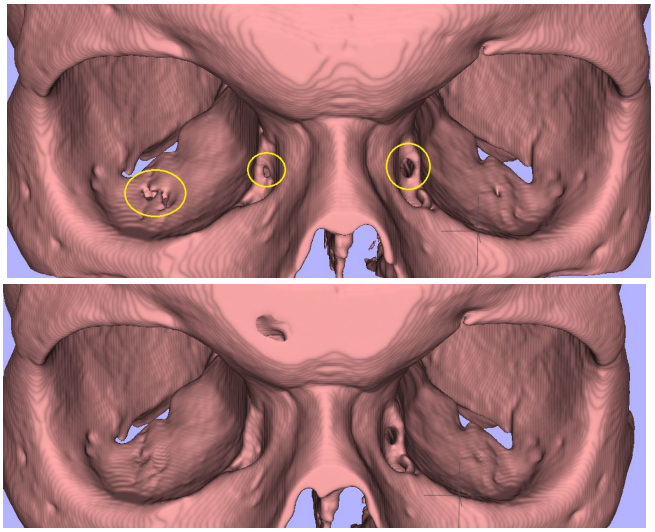

Fig. 6. Top image, finished segmentation of orbital bone with some holes marked with yellow. Bottom image after ∼30s of clicking to fill holes.

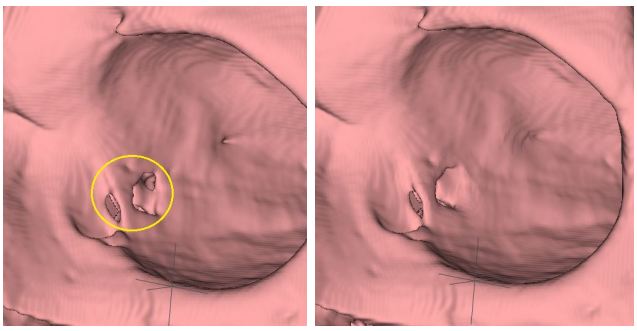

Fig. 7. Proximal view of left orbital bone. Left image, finished segmentation of orbital bone with hole marked with yellow. Right image after ∼30s of clicking to fill holes.

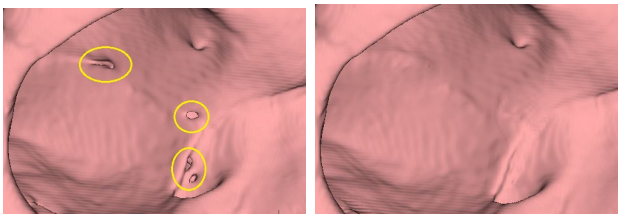

Fig. 8. Proximal view of right orbital bone. Left image, finished segmentation of orbital bone with hole marked with yellow. Right image after ∼30s of clicking to fill holes.

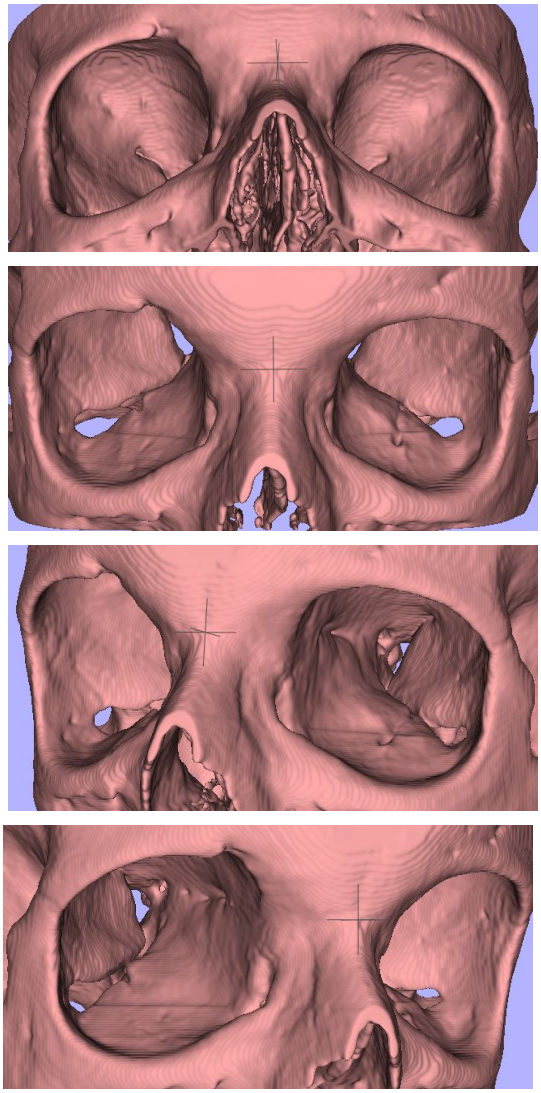

Fig. 9. An overview of an orbital segmentation of excellent quality.

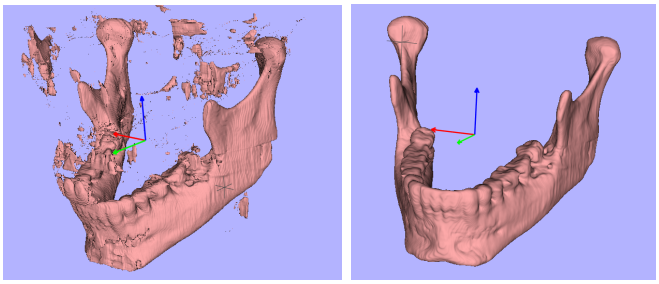

Fig. 10. Left image is an automatic segmentation of the mandible. Right image is the same segmentation after some work.

grade 5 example, Figure 9, we can see that no work is needed to cover any holes and the segmentation is ready to print immediately. All holes and cavities still seen in Figure 9 is anatomically correct and do not alter the grade.

2) Mandible: The images to test the segmentation on, needs to be unseen by the network, unseen meaning that the images was not used as training data, to validate future use of the network on new patients. We picked 2 CT scans to validate segmentation visually and 1 CT scan to validate with evaluation metrics of segmentation accuracy. The small amount of test data is a consequence of wanting to use as much classified data as possible for training. The visual assessment uses the same grading as the orbital evaluation, see Table II. A visual example is illustrated below in Figure 10. The automatic segmentation-ROI was chosen to contain the mandibular bone together with some overlap into the surrounding skull and background. Dice and 95%-percentile were used as evaluation metrics. The Dice coefficient formula is two times the area of overlap divided by total number of pixels in both images, resulting in a value 0 - 1 where 1 is the top score.

A segmentation assigned the grade 2 can be seen in Figure 10. This example was made ready for printing within a considerate amount time using different Segment 3dprint-tools. The segmentation is still acceptable since the segmentation is useful. If the automation produces a segmentation similar to the right image of Figure 10 it would need no work and be assigned grade 5.

III. Results

TABLE III
ORBITAL ACCURACY EVALUATIONS IN TERMS OF GRADING OF SEGMENTATION IN 20 PATIENTS DIVIDED IN 10 FEMALES AND 10 MALES EACH REPRESENTING ALL AGE GROUPS EXCEPT ADOLESCENTS. SORTED BY AGE.

Age	Gender	Grade	Age	Gender	Grade
102	Female	5	62	Male	5
101	Female	5	61	Female	5
98	Male	5	52	Female	4
98	Male	5	51	Male	2
92	Female	2	41	Female	4
92	Male	4	41	Male	3
82	Female	4	33	Female	3
82	Male	4	32	Male	5
71	Female	5	24	Male	4
71	Male	5	23	Female	5

A. Orbita

Table III is sorted by age and no correlation can be found in the grading. More test data is needed to make any assumptions. The average grade across all images is 4,2 and for the genders, the female average grade is 4,4 and the male average grade is 4. Each segmentation took around 15 seconds.

B. Mandible

Test results show that the network creates an average segmentation. The accuracy metrics were a dice score of 85.8%. The visual evaluation of the segmentation is a grade 2. Each segmentation took around 30 seconds.

All segmentations were performed on NVIDIA GeForce RTX 2070.

IV. Discussion

Algorithms for automatic mandible and orbital segmentation has been developed using a shared general segmentation framework. The results were varying, with the orbital segmentation being integrated in software directly and the mandible segmentation needing more work or a bigger dataset to train on.

A. Orbital segmentation

The visual grading may be regarded as harsh, since many segmentations would be usable in preoperative planning. The grading was instead determined in regards to aesthetics and comparison to ground truth. Perhaps grading could be determined more based on preoperative usability since this is what the segmentations will be used for. This would lead to an easier time motivating the use of this framework in creating algorithms. Only evaluating the orbital segmentations visually was more in-line with use in preoperative planning, but there was also a rationale behind this decision. Evaluation metrics would not have targeted the orbital bone but the whole segmented volume, which the orbital bone only was a part of, resulting in misleading evaluation of the segmentation.

A logical reason for the success of this network is the simple localization of the sinuses below the orbital bone. The reason is that since these cavities are air-filled there is no contrast, making it easier for the network to localize where the contrast appears, which is where the orbital bone starts.

B. Mandible segmentation

There can be many factors that influence the result of the mandible segmentation. One plausible factor could be the unavoidable complexity of the problem of differentiating between the mandible and the rest of the bone, given that the mandible is a very unique structure between individuals. This, as mentioned, is of course unavoidable but can be compensated as describe below.
In this case the algorithm could perhaps suffer from it's limited training data. A machine learning process start with gathering data and that data is the solid ground on which algorithms are built upon. This being said there can never be too much training data, but there can be too little. The beginning, exploratory phase, of this study should have made a further use of its resources, the supervisors, in order to have a more solid base.
The limited data was not the only issue, as there seemed to be an issue with how the network either trained on or applied itself onto images. As seen in Table I, the mandible algorithm trained on images in the sagittal and coronal planes and, in the code, discrepancies could be found regarding the reference to different planes. Because of the structure of the parameters for flipping, mirroring and transposing would be zero but there was a suspicion of discrepancies here as well.
Dealing with these mentioned issues could potentially yield a better result.

C. Ethical machine learning

In order for the algorithms to be ethical there need to be diversity and transparency of the AI. There is always a risk of generating a biased and non-diversified algorithm, and it might not be intentional. But with the limited amount of images in healthcare the algorithms are forcefully biased, leading to mistakes made by the algorithm if its applied to difficult and odd images.
The advantage of segmentation is that it is visual and do only give suggestions to the clinicians in their preoperative planning. This is therefore less of an ethical dilemma.

D. Sustainability

Machine learning involves a heavy use of powerful computers to make sure training does not take too long. This consumes a lot of energy since training each network took at least 12 hours, with the specified GPU. There is no assurance that the network produced is even usable. If it is not it requires a new training.
But there is a turning point where the automated segmentations undercuts manual segmentations in consumed energy, since the automation take much less time.

The algorithms of automating segmentation will simplify the preoperative process of 3D-printing, resulting in more material use of 3D-printing materials. But this use of materials allow for better preoperative planning and thus a better prepared surgeon, hopefully resulting in more successful surgeries and a better recovery for the patient. Healthier patients mean less strain on healthcare and perhaps less use of resources.

V. Conclusions

The framework shows promise with one generated algorithm yielding great results and can be integrated in software directly. Applied to more difficult anatomical structures the generated algorithm stumbles, requiring more tuning and more nuanced training data. Concluded the framework is a feasible way forward in automatic biomedical segmentation.

References

[1] OECD. Scoping the oecd ai principles. (291), 2019. doi: https://doi.org/https://doi.org/10.1787/d62f618a-en. URL https://www.oecd-ilibrary.org/content/paper/d62f618a-en.

[2] Jiangchang Xu, Jiannan Liu, Dingzhong Zhang, Zijie Zhou, Xiaoyi Jiang, Chenping Zhang, and Xiaojun Chen. Automatic mandible segmentation from CT image using 3D fully convolutional neural network based on DenseASPP and attention gates. *International Journal of Computer Assisted Radiology and Surgery*, 16(10):1785–1794, October 2021.

[3] *Explainable AI in healthcare and medicine : building a culture of transparency and accountability.* Studies in computational intelligence: 914. Springer, 2021. ISBN 9783030533519. URL http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=cat07147a&AN=lub.6682424&site=eds-live&scope=site.

[4] Sumeet Dua and service) SpringerLink (Online. *Machine Learning in Healthcare Informatics. [Elektronisk resurs].* Intelligent Systems Reference Library: 56. Springer Berlin Heidelberg, 2014. ISBN 9783642400162. URL http://ludwig.lub.lu.se/login?url=https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=cat07147a&AN=lub.5715638&site=eds-live&scope=site.

[5] Mohamed Alloghani, Dhiya Al-Jumeily Obe, Jamila Mustafina, Abir Hussain, and Ahmed Aljaaf. *A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science*, pages 3–21. 01 2020. ISBN 978-3-030-22474-5. doi: 10.1007/978-3-030-22475-2_1.

[6] Jinming Zou, Yi Han, and Sung-Sau So. Overview of artificial neural networks. *Methods in molecular biology (Clifton, N.J.)*, 458:15—23, 2008. ISSN 1064-3745. doi: 10.1007/978-1-60327-101-1_2. URL https://doi.org/10.1007/978-1-60327-101-1_2.

[7] Charu C. Aggarwal. *Neural Networks and Deep Learning - A Textbook*. Springer, 2018. ISBN 978-3-319-94462-3. doi: 10.1007/978-3-319-94463-0. URL https://doi.org/10.1007/978-3-319-94463-0.

[8] S. Saha. A comprehensive guide to convolutional neural networks-the eli5 way. 2018. URL https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-

[9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[10] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, and Klaus H. Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018. URL https://arxiv.org/abs/1809.10486.

[11] Nilsson, Amanda. Quantification of Fetal Volume in Magnetic Resonance Images Using Deep Learning, 2021. ISSN 1404-6342. Student Paper.

[12] Bingjiang Qiu, Jiapan Guo, Joep Kraeima, Haye H Glas, Ronald J H Borra, Max J H Witjes, and Peter M A van Ooijen. Automatic segmentation of the mandible from computed tomography scans for 3d virtual surgical planning using the convolutional neural network. *Physics in Medicine &amp Biology*, 64(17):175020, sep 2019. doi: 10.1088/1361-6560/ab2c95. URL https://doi.org/10.1088/1361-6560/ab2c95.

[13] Jared Hamwood, Beat Schmutz, Michael J Collins, Mark C Allenby, and David Alonso-Caneiro. A deep learning method for automatic segmentation of the bony orbit in MRI and CT images. *Scientific Reports*, 11(1):13693, July 2021.

[14] Brainlab. Craniomaxillofacial planning, unique software tools for unique cases, 2022. URL https://www.brainlab.com/surgery-products/digital-cmf-surgery/cmf-planning/.

[15] Medviso. Software solution for medical 3d printing, 2022. URL http://segment.heiberg.se.