Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 828–835

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Segmentation of Sella Turcica in X-ray Image based on U-Net Architecture

Kaushlesh Singh Shakya[a,b], Priti K[a,b], Manojkumar Jaiswal[c,*], Amit Laddi[a,b,*]

[a]*Academy of Scientific & Innovative Research (AcSIR), Ghaziabad 201002, India*
[b]*CSIR-Central Scientific Instruments Organisation, Chandigarh 160030, India*
[c]*Oral Health Sciences (Dental) Centre, PGIMER, Chandigarh 160012, India*

Abstract

The computer-aided evaluation of medical imaging and orthodontic examinations requires the segmentation of Sella Turcica for feature extraction using cephalometric X-ray images, which is a novel and complex task. However, x-ray image segmentation is more complicated than CT and MR images due to less association of dense thin white tissue of Sella Turcica that is difficult to distinguish from cephalometric radiographs. This paper presents deep learning methods for completely automated Sella Turcica segmentation from cephalometric X-ray images using a fully connected neural network ("U-Net"). The experiments were performed and analyzed using secondary datasets provided by PGIMER, Chandigarh. The different pre-trained U-Net models (VGG16 and 19, ResNet34, 50, and 152) are used as an encoder to improve the prediction accuracy of the Sella structure. The author compared the IoU score of all the considered pre-trained models and found that VGG19 and ResNet34 have high IoU scores compared to other models, but out of all the considered models, ResNet34 gives good prediction results.

Keywords: Sella Turcica; X-ray image; Segmentation; U-net architecture; Deep learning

1. Introduction

Sella Turcica (ST) has a significant role in orthodontics. ST is a crucial anatomical structure in the measurement of lateral cephalometry. While scrutinizing dentofacial cephalometry and neurocranial morphology, the Sella point, a central landmark located in the middle of ST, is a predominant reference used for evaluating craniofacial morphology, the relationship of the intermaxillary to the cranium, and the cephalometric serial superimposition [1-7].

* Corresponding author. Tel.: +91-788820664, +91-9878495624.
E-mail address: drmanojjaiswal@yahoo.in, amitladdi@csio.res.in

1877-0509
10.1016/j.procs.2023.01.063

ST is a depression of saddle-shape placed on the intracranial surface of the sphenoid bone, which is the base of the skull, housing the pituitary gland. Any variation in Sella shape could manifest disturbance in endocrine hormones release such as prolactin, growth hormone, gonadotrophins hormones, and other associated chemical secretions [3]. In addition, the literature suggests that the Sella Turcica region is crucial for neural crest cells migrating towards the anterior pituitary gland, maxillary, and teeth developments during human embryological development [8].

On lateral radiographs, orthodontists usually deal with ST. Clinicians need to be familiarized with normal Sella anatomy to recognize morphological irregularity, which will detect pathological conditions [5]. Sella Turcica Bridging (STB) is a prominent structure variation of the Sella Turcica, characterized by merging anteriorly and posteriorly through two clinoid processes [4]. STB prevalence varies between 1.1% and 13%; however, it is higher in patients with severe craniofacial abnormality [1, 2, 9-12]. The ST and teeth require cells of the neural crest for their growth and development, and thought anatomic differences in the ST could be linked to dental changes [13, 14].

Therefore, many studies have investigated association between ST and dental abnormalities and also have interested many researchers in the course of the overview of the previously published works in the related field. The findings communicate that the various analytical studies related to ST on many subjects enumerate their results using statistical methods like Regression, Pearson coefficient, and Student t-test. However, it mainly configs lack efficient statical analysis and risk of bias problems related to the above analytical studies because methodological flaws can result in bias.

The study aims to fill the gap of flaws in methodology and risk of biases in results, the need to consider segmentation techniques likewise other biomedical disciplines are doing.

There are six sections to elaborate on this study. Section 2 demonstrates the considered U-Net network architecture in the study, which consists of two parts down-scaling and up-scaling of image resolution. Section 3 represents the network training and data augmentation process. Section 4 articulates the experiments, composed of different models with their variants, used hyperparameters, and evaluation metrics IoU. Section 5 provides empirical results of different methods, and lastly, section 6 demonstrates the conclusion, which consists of the study's contribution, limitation, and possible future scope.

2. Architecture of used network (U-Net)

The network architecture depicted in Fig. 1 represents two portions: the first comprises contracting (on the left side) known as an encoder, and the second is an expanding part (on the right side) called a decoder. The encoder part of convolutional architecture consists of two 3×3 convolutions; an optimizer rectified linear unit (ReLU) and a 2×2 max pooling function with two strides for upscaling.

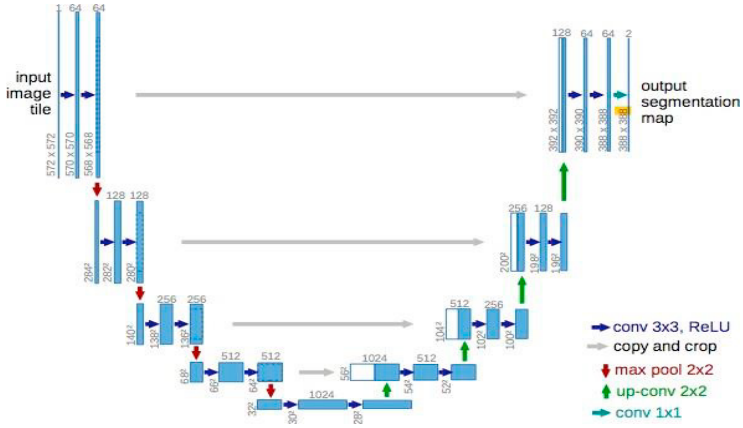


Fig. 1. U-Net Architecture work from work by Olaf Ronneberger et.al, 2015 [15].

With each down-sampling step in the encoding part of the architecture, the feature channels are increased by two. The expansivepath comprises feature map up sampling, is accompanied by a 2×2 up convolution, which reduces the feature channel by half, consists of a concatenation path with the proportionally cropped feature map through the contracting path, each step followed by an optimizer called ReLU. Cropping is needed in each convolutional layer to avoid the loss of boundary pixels. The final 1×1 convolution layer converts 64 feature vector components to get specified, multiple classes. A total of 23 convolutional layers are presented in the network. In order to apply all 2×2 max-polling functions to the layer in an even x and y-dimension, it is crucial to choose the input tile size appropriately [15-18].

3. Network training

The cephalometric X-ray input images and corresponding segmentation maps are involved in the network training process (seeFig. 2). Adam is employed in the optimization process [19, 20]. The input images are employed with a constant border width that is larger than the output to avoid the loss of border pixels in each convolution. This study preferred large input tiles over huge batch sizes to save overhead, make the greatest use of GPU memory, and condense the batch to a single image. The author set a high momentum (0.99) to compensate for uncertain gradients, so the current optimization step is dictated by many previously encountered training samples.

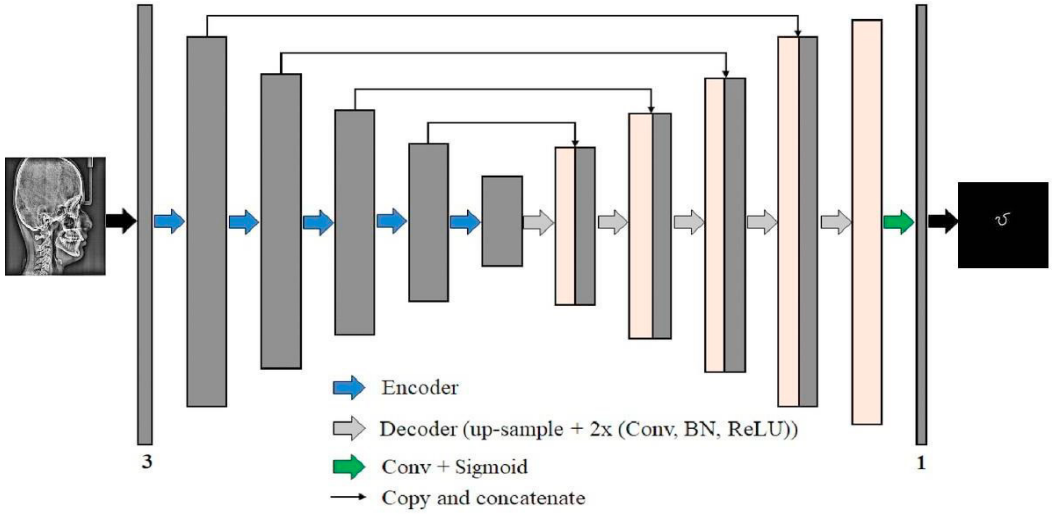


Fig. 2. Training of the semantic segmentation network for Sella Turcica segmentation.

This study put two different training methods to the test. The first VGG network has a relatively small receptive field (3×3, thesmallest size that captures both left and right and up and down). There are additional 1x1 convolution filters that perform a linear change of the input before passing it through a ReLU unit. The convolution stride is set at 1px to preserve spatial resolution following convolution [17]. The author applied the ResNet network to different layers, solving the famous vanishing gradient problem in the second training method. In CNN network an encoder (down-sampling path) extracts feature information from the input image data and the resulting feature maps are expanded using a decoder part to yield a pixel-wise segmentation mask.

In deep networks with numerous convolutional layers and diverse pathways, proper weight initialization is critical. Otherwise, some sections of the network might be active all of the time while others aren't. In an ideal situation, the initial weights are required to be adjusted so that while training the CNN algorithm, the variance for every feature map generated is around one unit. Our network can achieve this variance by setting weights from a common distribution function called Gaussian with S.D (standard deviation) of $\sqrt{2/N}$, where N is the count of approaching nodes per unit neuron [21]. For example, N = 9*32 = 288 if the previous layer contained a 3×3 convolution and 32 feature maps.

3.1. Data augmentation

When the data size is small for training a deep learning model, the data augmentation can play a crucial role in increasing the data size (see Fig. 3). The data augmentation is generated using different methods like translating, rotating, or flipping training images. Data augmentation is performed at runtime when data is retrieved for training and hence requires no additional storage capacity. As a result, it increases the data size and variability in training images.

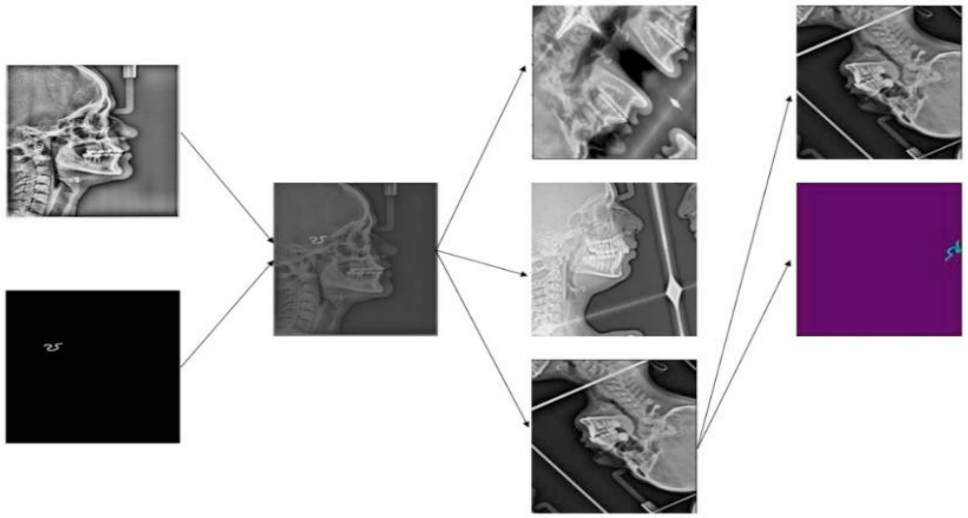


Fig. 3. The whole process of data augmentation.

4. Experiments

The study used standard U-Net convolutional networks to train models on lateral cephalometric datasets. A random sample of 60% of each dataset was used for training, while the remaining 40% was used for validation. Table 1 lists the optimizers and hyperparameters that were used for training.

Table 1. Used hyper-parameters.

	VGG16 (256×256)	**VGG19** (256×256)	**ResNet34** (256×256)	**ResNet50** (256×256)	**ResNet152** (256×256)
Optimizer			Adam		
Learning Rate	0.001	0.001	0.001	0.001	0.001
Data Size (Images)	351	351	351	351	351
Batch Size	4	4	4	16	16
Epochs	300	300	300	300	300

The various networks were trained (see Table 1) using data from the PGIMER in Chandigarh (351 annotated X-ray images from 351 patients). As these images were scanned at different scales, therefore, such images were resized to a standard size of 640×480px and normalized to grey value range [0, 1]. Then, using the above augmentations, the study used 1500 training image tiles with 256×256px and the correspondingly transformed segmentation maps. All networks were trained using Adam, with learning rates ranging between 0.01 and 0.001, which were reduced twice by a factor of 10 after 1500 iterations. The entire training procedure of a single network took about 24 hours on an Nvidia GeForce GTX 1080 8GB GPU. An evaluation matrix known as Intersection Over Union (IoU) is used in the study to assess the quality of semantic segmentation network models. IoU is calculated as the ratio of the number of pixels present between the area of intersection of referenced and predicted labels to the number of

pixels present in both labels. The mathematical formula for calculating IoU is presented (equation 1) below:

$$IoU = \frac{Reference \cap Prediction}{Reference \cup Prediction} \quad (1)$$

Average IoU and loss assessment results of 256×256 U-Net architectures are shown in the Table 2.

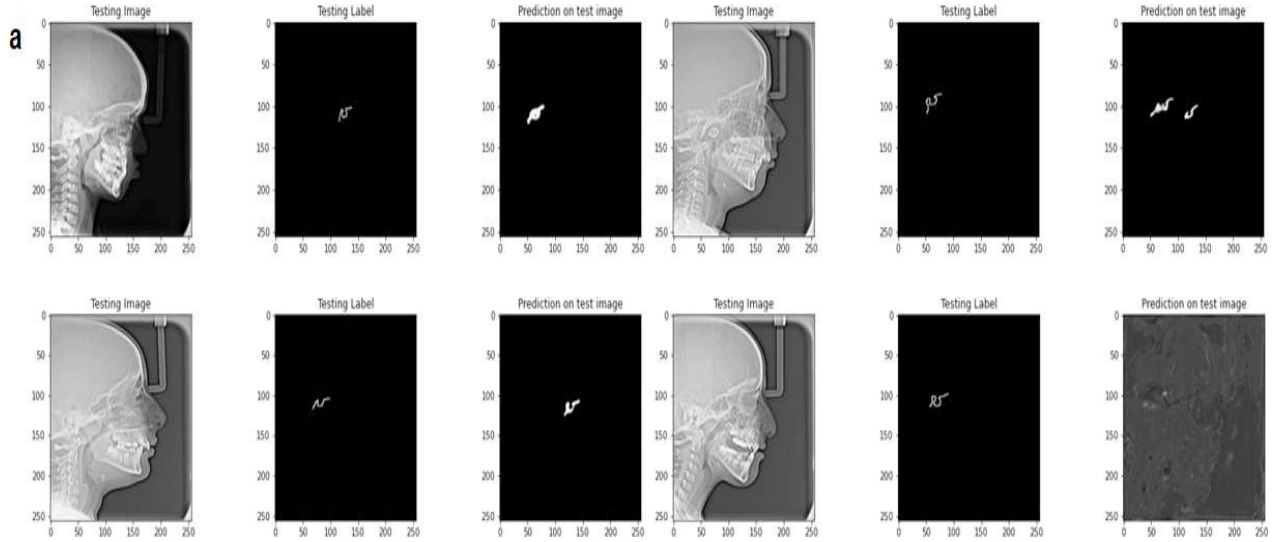
Table 2. Different model architectures providing memory usage, computation time and average IoU with loss scores.

Model	Data Size	Epochs	Memory Usage (mb)	Average IoU	Loss	Computation Time
VGG16	351	300	528	0.42	0.23	15 h 47 min
VGG19	351	300	549	**0.47**	**0.21**	17 h 23 min
ResNet34	351	300	102	**0.41**	**0.25**	15 h 53 min
ResNet50	351	300	102	0.37	0.33	21 h 07 min
ResNet152	351	300	230	0.39	0.30	23 h 13 min

Table 2 provides the results of multiple segmentation model based on pre-trained architecture networks. The obtained average IoU score of optimized VGG16, ResNet50, and ResNet152 is less compared to rest of other employed architectures like VGG19, ResNet34. Thus, both the latter models accomplish improved semantic segmentation results, with average IoU scores of 0.47% and 0.41%, respectively.

5. Results

This result section performed all ResNet experiments with different layers, data sizes, and epochs. The results for training, validation, and loss IoU graphs are quite satisfactory, but when comparing VGG and ResNet model results, the accuracy of VGG architecture is higher than the ResNet architecture for the 256×256 dataset. But the ResNet networks are able to predict the Sella Turcica shape.



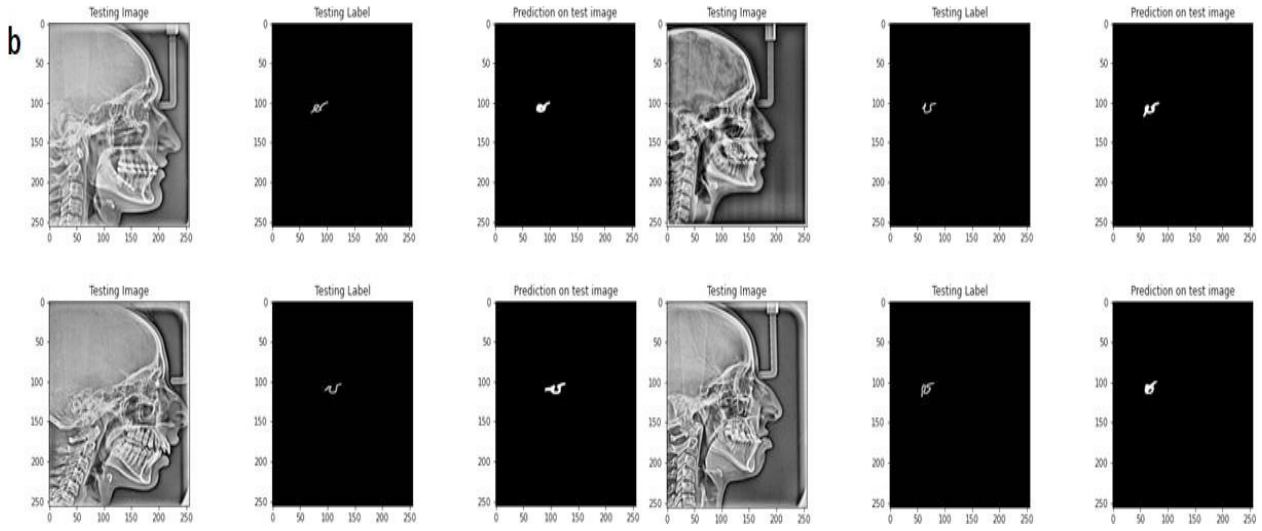
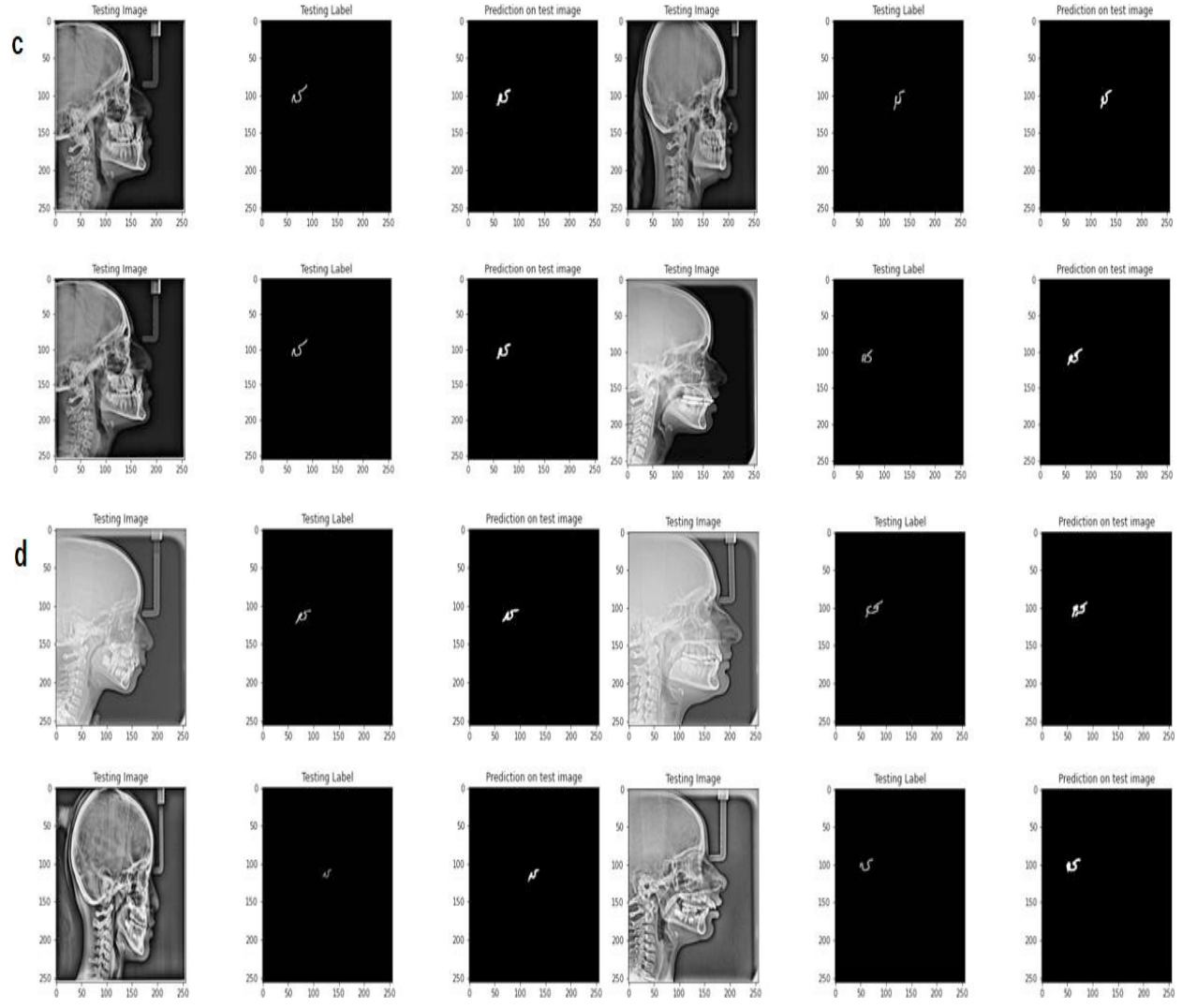


Fig. 4. Segmentation results of (a)VGG16 and (b) VGG19 U-Net architecture (256×256) with 300 epochs.



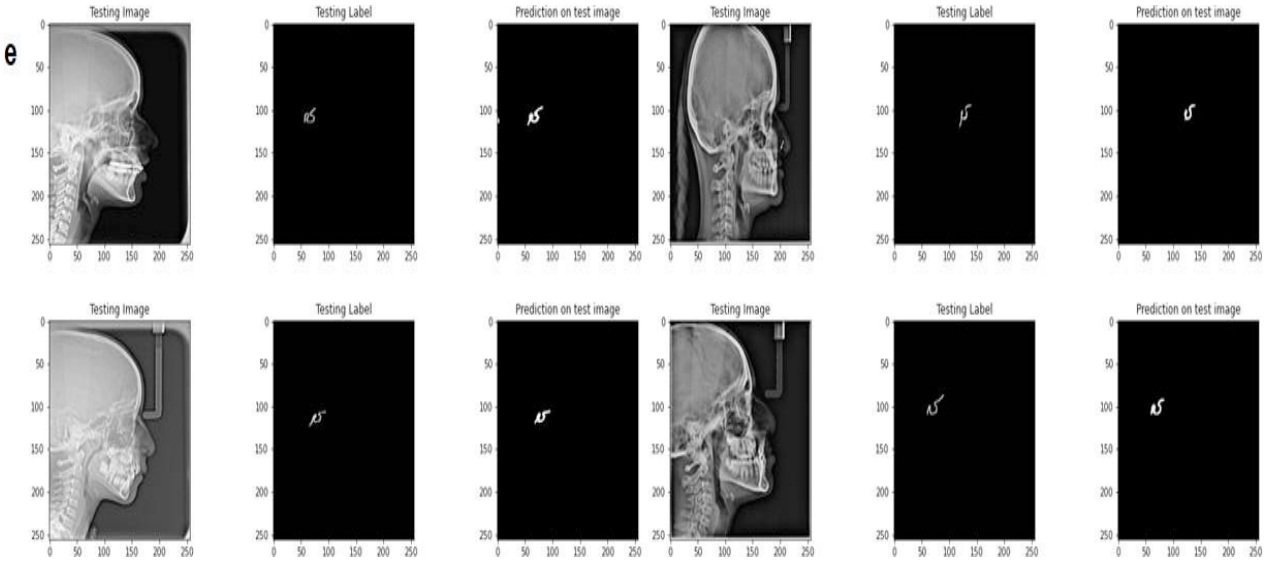


Fig. 5. Segmentation results of (c) ResNet34, (d) ResNet50 and (e) ResNet152 U-Net architecture (256×256) with 300 epochs.

The figure shows the testing results of segmented images generated by the models. White pixels in prediction on test images represent a segmented shape, while black pixels represent the background.

6. Conclusion

The present study employed a computer-aided segmentation technique, convolutional neural network-based U-Net models (VGG16, and 19, ResNet34, 50 and 152 variants) for Sella Turcica segmentation, which is a novel for lateral cephalometry. The experiment's result shows that the ResNet network has high prediction results with less training accuracy. Through this initial stage of the study, we can infer that for non-linear features detection of Sella, such segmentation techniques can work efficiently as existing studies include the statistical evaluation of morphological characteristics of Sella Turcica. This proposed work has taken a step toward automation through deep learning convolutional neural network-based U-Net models for Sella Turcica segmentation. These feature extraction techniques will help to detect different shapes of Sella Turcica. The future scope of the study is towards the classification of different morphological characteristics of Sella Turcica and further correlates these characteristics with facial deformities. The limitation of the study is that Sella, a non-linear structure segmentation technique, is enabled to extract Sella features more accurately. However, for further improvement in accuracy, we need to work on its ROI.

Acknowledgements

The authors thankful to PGIMER Chandigarh for providing data for this study under IEC-09/2021-2119 approval no.: PGI/IEC/2021/001488.

References

[1] Becktor, J.P., S. Einersen, and I. Kjær, 2000. A sella turcica bridge in subjects with severe craniofacial deviations. The European Journal of Orthodontics, **22(1)**: 69-74.

[2] Axelsson, S., K. Storhaug, and I. Kjær, 2004. Post-natal size and morphology of the sella turcica. Longitudinal cephalometric standards for Norwegians between 6 and 21 years of age. The European Journal of Orthodontics, **26(6)**: 597-604.

[3] Jones, R., et al., 2005. Bridging and dimensions of sella turcica in subjects treated by surgical-orthodontic means or orthodontics only. The Angle Orthodontist, **75(5)**: 714-718.

[4] Meyer-Marcotty, P., T. Reuther, and A. Stellzig-Eisenhauer, 2010. Bridging of the sella turcica in skeletal Class III subjects. The European Journal of Orthodontics, **32(2)**: 148-153.

[5] Sathyanarayana, H.P., V. Kailasam, and A.B. Chitharanjan, 2013. Sella turcica-Its importance in orthodontics andcraniofacial morphology. Dental Research Journal, **10(5)**: 571.

[6] Ali, B., A. Shaikh, and M. Fida, 2014. Association between sella turcica bridging and palatal canine impaction. AmericanJournal of Orthodontics and Dentofacial Orthopedics, **146(4)**: 437-441.

[7] Valizadeh, S., et al., 2015. Correlation of shape and size of sella turcica with the type of facial skeletal class in an Iraniangroup. Iranian Journal of Radiology, **12(3)**.

[8] Kjær, I., 2015. Sella turcica morphology and the pituitary gland—a new contribution to craniofacial diagnostics based onhistology and neuroradiology. European journal of orthodontics, **37(1)**: 28-36.

[9] Alkofide, E.A., 2007. The shape and size of the sella turcica in skeletal Class I, Class II, and Class III Saudi subjects. TheEuropean Journal of Orthodontics, **29(5)**: 457-463.

[10] Meyer‐Marcotty, P., et al., 2008. Morphology of the sella turcica in Axenfeld–Rieger syndrome with PITX2 mutation.Journal of oral pathology & medicine, **37(8)**: 504-510.

[11] Korayem, M. and E. AlKofide, 2015. Size and shape of the sella turcica in subjects with D own syndrome. Orthodontics &craniofacial research, **18(1)**: 43-50.

[12] Lata, J., N. Verma, and A. Kaur, 2015. Gorlin–Goltz syndrome: A case series of 5 patients in North Indian population withcomparative analysis of literature. Contemporary clinical dentistry, 6(Suppl 1), S192.

[13] Leonardi, R., et al., 2006. A sella turcica bridge in subjects with dental anomalies. The European Journal of Orthodontics, **28(6)**: 580-585.

[14] Leonardi, R., M. Farella, and M.T. Cobourne, 2011. An association between sella turcica bridging and dental transposition.The European Journal of Orthodontics, **33(4)**: 461-465.

[15] Ronneberger, O., P. Fischer, and T. Brox, 2015. U-net: Convolutional networks for biomedical image segmentation.International Conference on Medical image computing and computer-assisted intervention, 234-241.

[16] Shakya, K.S., Laddi, A. and Jaiswal, M., 2022. Automated methods for sella turcica segmentation on cephalometric radiographic data using deep learning (CNN) techniques. Oral Radiology, pp.1-18.

[17] Rastogi, P., K. Khanna, and V. Singh, 2022. Gland segmentation in colorectal cancer histopathological images using U-netinspired convolutional network. Neural Computing and Applications, **34(7)**: 5383-5395.

[18] Shakya, K.S., Jaiswal, M., Priti, K., Alavi, A., Kumar, V., Li, M. and Laddi, A., 2022. A novel SM-Net model to assess the morphological types of Sella Turcica using Lateral Cephalogram.

[19] Jia, Y., et al., 2014. Caffe: Convolutional architecture for fast feature embedding. Proceedings of the 22nd ACMinternational conference on Multimedia, 675-678.

[20] Rastogi, P., K. Khanna, and V. Singh, 2022. LeuFeatx: Deep learning–based feature extractor for the diagnosis of acuteleukemia from microscopic images of peripheral blood smear. Computers in Biology and Medicine, 142, 105236.

[21] He, K., et al., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification.Proceedings of the IEEE international conference on computer vision, 1026-1034.