Performance analysis of lightweight CNN models to segment infectious lung tissues of COVID-19 cases from Tomographic images

Tharun J Iyer [1], **Alex Noel Joseph Raj** [2], **Sushil Ghildiyal** [1], **Ruban Nersisson** [Corresp. 1]

[1] School of Electrical Engineering, Vellore Institute of Technology University, Vellore, Tamil Nadu, India

[2] Department of Electronic Engineering, Shantou University (□□□□□□□), Shantou, Guangdong, China

Corresponding Author: Ruban Nersisson
Email address: nruban@vit.ac.in

The pandemic of Coronavirus Disease-19 (COVID-19) has spread around the world causing an existential health crisis. Automated detection of COVID-19 infections in the lungs from Computed Tomography (CT) images offers huge potential in tackling the problem of slow detection and augments the conventional diagnostic procedures. However, segmenting COVID-19 from CT Scans is problematic, due to high variations in the types of infections and low contrast between healthy and infected tissues. While segmenting Lung CT Scans for COVID-19, fast and accurate results are required and furthermore, due to the pandemic, most of the research community has opted for various cloud based servers such as Google Colabs etc to develop their algorithms. High accuracy can be achieved using Deep Networks but the prediction time would vary as the resources are shared amongst many thus requiring the need to compare different lightweight segmentation model. To address this issue, we aim to analyse the segmentation of COVID-19 using four Convolutional Neural Networks (CNN). The images in our dataset are preprocessed where the motion artifacts are removed. The four networks are UNet, Segmentation Network (Seg Net), High-Resolution Network (HR Net) and VGG UNet. Trained on our dataset of more than 3000 images, HR Net was found to be the best performing network achieving an accuracy of 96.24% and a Dice score of 0.9127. The analysis shows that lightweight CNN models perform better than other neural net models when to segment infectious tissue due to COVID-19 from CT slices.

Performance Analysis of Lightweight CNN Models to Segment Infectious Lung Tissues of COVID-19 cases from Tomographic Images

Tharun J Iyer[1], Alex Noel Joseph Raj[2], Sushil Ghildiyal[1], Nersisson Ruban[1]

[1] School of Electronics and Electrical Engineering, Vellore Institute of Technology Vellore, Tamil Nadu, India
[2] Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Department of Electronic Engineering, Shantou University, Shantou, Guangdong, China

Corresponding Author:
Nersisson Ruban[1]
(Associate Professor, School of Electrical Engineering, Vellore Institute of Technology, Vellore -632014, Tamilnadu, India)
Email Address: nruban@vit.ac.in

Abstract

The pandemic of Coronavirus Disease-19 (COVID-19) has spread around the world causing an existential health crisis. Automated detection of COVID-19 infections in the lungs from Computed Tomography (CT) images offers huge potential in tackling the problem of slow detection and augments the conventional diagnostic procedures. However, segmenting COVID-19 from CT Scans is problematic, due to high variations in the types of infections and low contrast between healthy and infected tissues. While segmenting Lung CT Scans for COVID-19, fast and accurate results are required and furthermore, due to the pandemic, most of the research community has opted for various cloud based servers such as Google Colabs etc to develop their algorithms. High accuracy can be achieved using Deep Networks but the prediction time would vary as the resources are shared amongst many thus requiring the need to compare different lightweight segmentation model. To address this issue, we aim to analyse the segmentation of COVID-19 using four Convolutional Neural Networks (CNN). The images in our dataset are preprocessed where the motion artifacts are removed. The four networks are UNet, Segmentation Network (Seg Net), High-Resolution Network (HR Net) and VGG UNet. Trained on our dataset of more than 3000 images, HR Net was found to be the best performing network achieving an accuracy of 96.24% and a Dice score of 0.9127. The analysis shows that lightweight CNN models perform better than other neural net models when to segment infectious tissue due to COVID-19 from CT slices.

Introduction

During the winter months December 2019, a highly contagious disease out broke in Wuhan, china [1,4]. High grade fever and other flu like symptoms were noticed and most of the patients developed Pneumonia. The pathogen causing the disease was identified as corona virus, and named as Severe Acute Respiratory Syndrome Corona Virus -2 (SARS-CoV-2) [2]. The disease caused by the virus is named by World health Organization (WHO) as Corona Virus Disease (COVID-19). WHO also declared COVID 19 spread as Global public health emergency [2]. As of 9th May 2020, more than 200 countries around the world are affected by COVID-19. There are around 4 million people affected by the disease worldwide with a mortality rate of 6% (around 2.75 million people lost their lives). Developed countries like US, Europe and most of the developing countries are suffering a lot from the outbreak. The scientific community is largely involved in devising an antidrug and vaccines for the device. But unfortunately there are no positive results till date and more over it is reported that, due to mutations, characteristics are changing which makes the vaccine development even more challenging. Taking the situation into account there are very few ways we can control the virus; like staying isolated from the world and breaking the spreading chain of the virus, maintaining the personal hygiene, early detection of the symptoms and taking necessary precautions are few of them.

The successful control of the outbreak depends on the rapid and accurate detection and identification of the symptoms isolating the patient from the community, so that the spread of the disease can be stopped. Currently, the method used for the detection is Real-time reverse transcriptase polymerase chain reaction (RT-PCR) [5]. It is the standard procedure used by many hospitals and clinics for testing COVID-19 cases. Even though this method remains the reference standard, there are many reported false negative cases using this RT-PCR [8], which is an alarming fact on the situation. It is also time consuming and the limited supply of RT-PCR kits for the rural areas make the testing more difficult [6]. Since the COVID-19 patients develop breathe related discomfort and Pneumonia as the outcome of the disease progress, Radiological studies can play a vital role in diagnosing the lung infections caused by this episode [9]. The CT chest scan can be used to identify the early stages of lung infections and related problems. The chest CT reveals the initial pulmonary abnormalities for COVID-19 patients for whom RT-PCR gave negative results [3].

The common manifestations of SARS-CoV-2 in chest CT scan are:- ground glass opacities, consolidation, crazy paving, dilation of vessel width in some cases and round shape lesions in few cases [7,9]. The effectiveness of the chest CT scan based COVID-19 management depends on the efficient automatic detection and segmentation of regions in the scan. So in that context the recent developments in the imaging technologies come handy. There are plenty of imaging tools which give very high and accurate quantification of abnormal conditions. This procedure of image based diagnosis system involves capturing the image, analysing the image by a trained, experienced radiologist and annotation is made for the ground truth segments. The current

scenario slows down the annotation of the images, labelling and getting the ground truth processes due to the increasing number of patients day by day, lack of radiologist and the over duty burden of existing radiologists. So automatically detecting the infected regions from the chest CT scan using computer based algorithms are the current trends in research that gives wonderful results and aids in medical diagnostics.

The main objective of the research work is comparing the segmentation performance of computationally non-intensive models deep learning model when subjected to Lung CT Scans for that are affected by COVID-19 a. The models utilized for the research belong to the U-Net variants models which are the most popular models of choice for segmentation of Medical Images. Here we compare the traditional U-Net model as proposed by Ronneberger with other variants such as Seg Net, U-Net based on VGG16 and HR-Net (High Resolution Net) and present both qualitative and quantitative results.

Materials & Methods

The Block diagram describing the entire methodology is shown below in Figure 1. The description of each method is described below.

Dataset Considered:

The dataset used consists of 3770 images and their corresponding ground truths. 3020 training images are used and 750 testing images are used. The CT scans of 50 patients were taken from mosmed.ai [19] were openly accessible Neuroimaging Informatics Technology Initiative (NIFTI) images were provided. The data was collected from the Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department. The CT Scans were obtained between 1st March 2020 and 25th April 2020. Each NIFTI file was decompressed to PNG images and used for the study. The CT scans of another 20 patients were taken from zenodo.org [20] where the NIFTI files of 20 patients were provided. The images were annotated by two radiologists and verified by an experienced radiologist. For both datasets, MATLAB was used to extract the PNG images.

Pre-Processing:

The images in the dataset are riddled with motion artefacts and noise. Motion artefacts are caused due to improper imaging techniques and are a specific kind of noise relevant to CT Scans. Therefore, removing this noise is important or else it will cause the algorithms to learn improperly. MATLAB is used to remove the noise and motion artefacts. The original image is converted to grayscale from RGB and then, the image properties are extracted. Area and Solidity are used and then, the image is thresholded after selecting the max area and highest solidity. Once a mask is ready, the mask is multiplied with the original image to get the pre-processed image. A comparison is given below in Figure 2. As can be seen, motion artefacts are removed and the image has more clarity.

HR Net:

HRNet is developed at Microsoft and has signified state of art presentation in the areas of semantic segmentation, image classification, facial detection, object detection and pose estimation [10]. Its attention is on training High Resolution (HR) representation. The existing techniques recuperate representation of high resolution from representation of low resolution formed by high to low resolution network. In HRNet, from first stage commencement high-resolution network, progressively augment high to low resolution networks successively to arrange more steps and associate the multi-resolution network in parallel.

HRNet is able of uphold high-resolution representation throughout the process as repeated multi-scale combinations are conducted by switching the information through the multi-resolution parallel subnetworks repeatedly throughout the process [11]. The architecture of resulting network is displayed in Figure 3. This network has advantages in contrast to existing networks like Segnet, UNET, Hourglass etc. These existing networks lose a lot of essential information in the progression of recovering high-resolution from low-resolution representation. HRNet links high to low resolution networks in parallel instead of series and this gives high-resolution representation throughout the process, correspondingly the estimated heatmap is much accurate, spatially much precise.

Multi-resolution sequential subnetwork

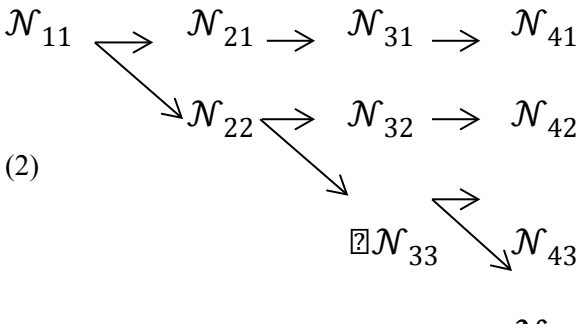
Existing models works by linking high to low resolution convolutions subnetwork in series, where each individual subnetwork form a platform, collection of an arrangement of convolutions furthermore, there is a down sample layer through end-to-end subnetworks to split the resolution into halves.

Let $\mathcal{N}_{sr}$ be the subnet in the stage s^{th} and resolution index r. First subnet resolution is given by $\frac{1}{2^r - 1}$. The high-to-low system with S phases/stages (I.e.: 4) can be indicated as:

$$\mathcal{N}_{11} \rightarrow \mathcal{N}_{22} \rightarrow \mathcal{N}_{33} \rightarrow \mathcal{N}_{44} \qquad (1)$$

Multi-resolution parallel subnetwork

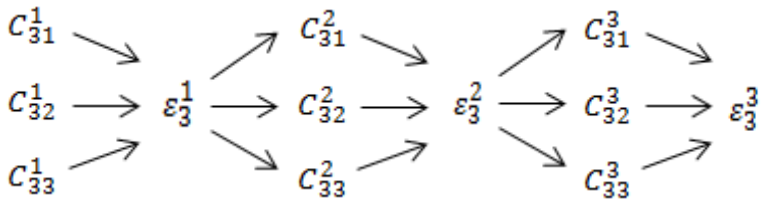
Starting from first phase/stage begin with high resolution subnet, slowly enhance high to low resolution subnet, generating new phases/stages, and associate multi-resolution subnet in parallel. Eventually, the parallel subnet resolution of a later phase/stage comprises of the resolution from an earlier stage and below one stage. The network shown below contains 4 parallel subnets.



(2)

Multi-scale repeated fusion

In this network exchange units were introduced throughout parallel subnet in such a way that an individual subnet continuously collects information from parallel subnets. How information is exchanged lets understand this process through an example here third stage is subdivided into multiple exchange blocks and every block consists of three parallel convolution modules, having exchange units followed by parallel units which is shown below,



(3)

Where:

C_{sr}^{b} – Convolution module,

ε_{s}^{b} – Exchange Unit,

and s is the stage, r is the resolution and b is the block

Explanation of exchange units is show in Figure 4. The input mapping is given by : $\{X_1, X_2, X_3, \ldots\ldots, X_s\}$ and the output mapping was given by: $\{Y_1, Y_2, Y_3, \ldots\ldots, Y_s\}$. The width and resolution of the output is same as input. Every output is a sum of input mapping i.e $Y_K = \sum_{i=1}^{s} a(X_i, K)$. Assume of 3×3 stride was done for down sampling and for up sampling 1×1 convolution (nearest neighbor).

HRNet experimental results (when tested with different datasets) show remarkable results for the applications like facial detection, semantic segmentation, and object detection.

Seg Net:

At the university of Cambridge, UK, team of the robotics group researched and developed that SegNet is a deep encoder decoder architecture for multiclass pixel-wise segmentation [12]. The framework comprises order of non-linear processing layers which is called encoders and a similar set of decoders afterward a pixel wise classifier. Generally, encoder have made up of a ReLU non-linearity and one or more convolutional layers with batch normalization, subsequently non-overlapping maxpooling and subsampling. Using Max-pooling indices in encoding sequence, for up sampling the sparse encoding in consequence the pooling process to the decoder. Use of max-pooling indices in the decoders is the one important feature of the SegNet to execute the sampling of low resolution maps. For segmented images the tendency to retain high frequency details and capable enough to decrease the number of parameters in the decoder needed for training are some advantages of SegNet. Using stochastic gradient descent this framework can be trained end-to-end.

SegNet is composed of encoder and decoder after a last pixel-wise classification layer. The architecture is shown in Figure 5. The encoder in SegNet is composed of convolution layers which are 13 in number, and these layer matches with the 13 starting layers of VGG16, considered for classifying the objects [13].

Figure 6 illustrates the decoding method utilized by SegNet in which there is no learning engaged with the up-sampling stage. The upsampling of decoder network's feature map (input) is done by learned maxpooling indices from the equivalent encoder feature map. Dense feature maps are generated by combining feature maps and trainable decoder channel.

SegNet a deep network was used for semantic segmentation. Basically, It was designed because the motivation behindhand was to propose an architecture for roads, outdoor and indoor sites which is proficient together in terms of computational time and memory. Feature map's maxpooling indices are only stored in SegNet and to attain better performance it uses them in its decoder network.

UNet:

The UNet design is based upon the fully convolution network and adjusted such that it produces better segmentation results in medical imaging. UNet consists of two paths named as contracting and and expansive. In the contracting path it captures the context whereas in expansive path it enables exact localization. While contracting path is a classical architecture of UNet. It includes two 3x3 convolutions, max pooling operation with repeating application. The Figure 7 illustrates the architecture of UNet, which is U in shape that itself gives the name 'UNet'. The main philosophy behind this network is, it replace pooling operation by using upsampling operators [14]. So, ultimately the resolution will increase layer by layer. The main feature of UNet is the large number of channels which lead to higher resolution. Moreover, in every downsampling it doubles the feature channels.

Each stage in the expansive path involves upsampling of the feature channel followed by (2x2) convolution that splits the number of feature channels into halves. In contracting path, it crops the feature map because of loss in border pixel in each convolution. Final layer is mapped by 1x1 convolutions which is used to map all 64 units feature vector. The network contains total 23 convolutional layers. UNet performs well on image segmentation [15].

While training the UNet model, the cross- entropy loss function united with the last feature map and by applying a pixel- wise softmax over it, the softmax is denoted as:

$$p_k = \frac{e^{(a_k(x))}}{\sum_{k=1}^{K} e^{(a_k x)}} \quad (4)$$

In addition, the energy function is calculated by:

$$E = \sum_{x \in \Omega} w(x) log\ (p_{l(x)}(x)) \quad (5)$$

Where:

a_k : Represents the activation in feature map k

p_k : Represents estimated maximum function

K: No. of class

$x \in \Omega$: Pixel position

$p_{l(x)}$: Deviation

In the training data set, to counterbalance the diverse frequency of pixels from a specific class the weight map is pre calculated for ground- truth segmentation, and enforcing the network to study the minor separation borders amid touching cells introduce by us.

The morphological operation used to calculate separation borders, the weight map calculated using:

$$w(x) = w_c(x) + w_0 . e^{(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2})} \quad (6)$$

Where:

w : denotes the weight map

d_1: distance upto border of nearest first cell

d_2: distance upto border of nearest second cell

VGG UNet:

Image segmentation, which is performed pixel wise is most preferable task in the field of computer vision. Encoders- decoders when combined they form UNET architectures, which are very famous for image segmentation in medical imaging and satellite images etc. The weights of the pre-trained models (like ImageNet) are used to initialize the weights of the neural network (i.e trained on large dataset) as it gives better performance major than those models, which are trained on small dataset from scratch. Models accuracy is very important in some applications like traffic safety and medicine pre-trained encoder can enhance the architecture and performance of UNET. Applications like Object detection, image classification and scene

understanding have improved their performance after the introduction of convolutional neural network (CNN). Nowadays, CNN has outperformed in several fields over human experts.

Image segmentation plays vital role in the field of medical imaging to enhance the diagnostic capabilities. Fully connected network (FCN) is amongst the most popular state- of-the-art machine learning technique [16]. Segmentation accuracy attained by some advancement in

FCN as compared to PASCAL VOC [17] common approach on standard datasets
UNet consists of two paths named as contracting and and expansive. In the contracting path it captures the context whereas in expansive path it enables exact localization. The contracting path sticks with the design of a convolutional network with pooling operations, alternating convolution, gradually down sample feature channels and expanding many feature maps\ layer simultaneously, each stage in the expansive path composed of an up-sampling of the feature channel along with a convolution. The VGGUnet architecture is illustrated in Figure 8. The encoder for UNET model is composed of 11 successive (series) layers VGG family and denoted by VGG-11 [3]. VGG-11 consist of 7 convolution layers each using rectified linear unit (ReLu) activation function, 5 maxpooling operations each reduces feature channel by 2 and the kernels size 3x3 is used for every convolutional layer [18].
Common loss function i.e binary cross entropy can be used for classification problem where $\hat{y}_i$ denotes the prediction, y_i denotes the true value and m denotes the no. of samples

$$H = -\frac{1}{m}\sum_{i=1}^{m}(y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)) \qquad (7)$$

Performance Validation:

To validate the performance of the models presented above, Sensitivity, Specificity, Jaccard Index, Dice Coefficient, Accuracy and Precision are used. To measure the accuracy of the segmented image, accuracy and precision are used and to measure the quality of segmentation, sensitivity and specificity are used. The various performance measures are described below:
Accuracy and Precision are used to calculate the accuracy of the segmentation model itself. Accuracy is as the ratio of correct predictions to the total number of predictions and Precision is defined as the ratio of correctly predicted positive observations to the total number of correctly predicted observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (8)$$

$$Precision = \frac{TP}{TP + FP} \qquad (9)$$

In the case of segmentation, accuracy and precision are used to measure the binary segmentation of each pixel of the image by the model. Although precision and accuracy may seem to be enough to describe the performance of the model, other factors are also important to describe the quality of segmentation.

Sensitivity and Specificity are used to measure the quality of segmentation between the classes. In this case, the models are performing binary segmentation. So, Sensitivity, or the True Positive Rate, measures the quality of segmentation of one class and Specificity, or the True Negative Rate, measures the quality of segmentation of the other class. Sensitivity and Specificity can be defined as:

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (10)$$

$$Specificity = \frac{TN}{(TN + FP)} \quad (11)$$

With Sensitivity and Specificity, having a high value for each is good as it shows that the model is able to segment the pixels correctly without any errors.

The Jaccard Index and Dice Coefficient are used to quantify the similarity between the original image and the segmented image. Jaccard Index and Dice Coefficient are similar to the Intersection over Union (IoU) used to evaluate Object detection models. Jaccard Index and Dice Coefficient ranges from 0 to 1 where 0 means no overlap and 1 mean full similarity. Jaccard Index and Dice Coefficient can be defined as:

$$Dice\ Coefficient = \frac{2TP}{2TP + FP + FN} \quad (12)$$

$$Jaccard\ Index = \frac{TP}{TP + FP + FN} \quad (13)$$

While the Dice coefficient and the Jaccard Index are quite similar, ideally, the two measures have to be equal. So, to measure the quality of similarity of the model, a similarity between the Dice Coefficient and Jaccard Index can be viewed to measure the quality of segmentation.

Results and Discussion

The experiments were conducted in the Google Colab platform. As shown in Table 1, HR Net is shown to have the highest performance as compared to the other models. The second best model is the classical UNet, the third best model is the VGG UNet and the model with the worst performance is the Seg Net. The reason for the high performance of the HR Net is the fact that the HR Net extracts high resolution information and retains in throughout the segmentation process. This is due to the parallel networks that are able to maintain essential information. HR Net indicates a high accuracy of segmentation with an Accuracy of 0.9624 and a Specificity of 0.9930.

Figure 9. shows the various outputs obtained from the models which segmented the test image. HR Net shows the best segmentation performance while UNet shows good performance too. UNet is able to obtain a proper boundary similar to the test image. Seg Net has performed poorly to segment the image. Neither has it obtained a boundary nor has it segmented the finer details properly. The VGG UNet has segmented the image properly but not to the extent of HR Net or UNet. We can see that HR Net has the best performance. With decreased area, the performance

decreases which means that UNet is the second best performance, VGG UNet is the third best and Seg Net has the worst performance amongst the models.

From Figure 10, we can infer that the performance of the HR Net is the best as it has the greatest area. The 6 points are the performance measures namely, Sensitivity, Specificity, Precision, Accuracy, Jaccard Index and Dice Coefficient. The glyph plot is a good way to directly compare the performance of various models through the use of polygons.

As can be seen from Table 2, HR-Net has the highest number of layers at 1043 with other models having less than 100 layers. But, HR-Net has similar parameters to VGG UNet and an even lesser number of parameters than SegNet. This is due to the architecture of the HR-Net. It is able to extract deeper features than the other models while maintaining the overall file size and number of parameters. The reason for HR-Net having the highest performance is its architecture which makes it the best model to be used for fast inference. Comparing the SegNet, VGG UNet and UNet, SegNet has the poorest inference speed at 84 ms with the largest model size. It is therefore inferred that SegNet is the worst model to use for the Segmentation of COVID-19 based on our study. VGG UNet and UNet have different metrics due to the fact that VGG UNet is trained on the VGG-16 weights. It, therefore, takes far higher time to load on and produce an inference than UNet. If HR-Net cannot be used, VGG UNet is the next best network for segmenting COVID-19.

Conclusion

In this paper, we analyzed four models for segmenting COVID-19 from Lung CT Images. With the growing number of cases worldwide, quick and accurate testing is needed. To solve this problem, we approached the problem by reviewing four lightweight models that do not take a long time for training or testing. First, we remove motion artefacts from the tomographic images through Thresholding and use the pre-processed images for training the models. The four models trained are Seg Net, UNet, VGG UNet and HR-Net. We evaluate the models on their performance using accuracy, dice, Jaccard index and precision. We also used Specificity and Sensitivity as secondary evaluation characteristics. The results obtained demonstrate that lightweight convolutional networks have high latent ability to segment COVID-19 from CT images with HRNet being the best network out of the four models analysed. In the future, we will increase the number of CT images by collaborating with hospitals and multi-centre collaborations.

Acknowledgements

We would like to thank VIT University for the nurturing environment as well as the exposure it provided for us to complete this project. We also would like to thank MosMedData for the Covid 19 dataset which is used for this research work.

References

1. Zhu N, Zhang D, Wang W et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. The New England journal of medicine 2020; 382: 727–733.

2. World health Organization: WHO Director General's remarks at the media briefing on 2019-n CoV on 11th February 2020. At: https://www.who.int/dg/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020. Published February 11, 2020.

3. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z. and Xia, L., 2020. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*, p.200642.
4. Liu, W., Zhang, Q., Chen, J., Xiang, R., Song, H., Shu, S., Chen, L., Liang, L., Zhou, J., You, L. and Wu, P., 2020. Detection of Covid-19 in children in early January 2020 in Wuhan, China. *New England Journal of Medicine*, *382*(14), pp.1370-1371.
5. Li, Y. and Xia, L., 2020. Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. *American Journal of Roentgenology*, pp.1-7.
6. Chen, X., Yao, L. and Zhang, Y., 2020. Residual Attention U-Net for Automated Multi-Class Segmentation of COVID-19 Chest CT Images. *arXiv preprint arXiv:2004.05645*.
7. Hamer, O.W., Salzberger, B., Gebauer, J., Stroszczynski, C. and Pfeifer, M., 2020, March. CT morphology of COVID-19: Case report and review of literature. In *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*. © Georg Thieme Verlag KG.
8. Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. 2020 Jan 24. pii: S0140-6736(20)30154-9. doi: 10.1016/S0140-6736(20)30154-9. [Epub ahead of print]
9. Zu, Z.Y., Jiang, M.D., Xu, P.P., Chen, W., Ni, Q.Q., Lu, G.M. and Zhang, L.J., 2020. Coronavirus disease 2019 (COVID-19): a perspective from China. *Radiology*, p.200490.
10. Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep high-resolution representation learning for human pose estimation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019-June, 5686–5696.
11. Sun, K., Zhao, Y., Jiang, B., Cheng, T., Xiao, B., Liu, D., … Wang, J. (2019). High-Resolution Representations for Labeling Pixels and Regions.
12. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495.
13. Mannem, R., Ca, V., & Ghosh, P. K. (2019). A SegNet based image enhancement technique for air-tissue boundary segmentation in real-time magnetic resonance imaging video. 25th National Conference on Communications, NCC 2019.

14. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9351, 234–241.
15. Livne, M., Rieger, J., Aydin, O. U., Taha, A. A., Akay, E. M., Kossen, T., Madai, V. I. (2019). A U-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. Frontiers in Neuroscience, 13(FEB), 1–13. https://doi.org/10.3389/fnins.2019.00097
16. J. Long, E. Shelhamer and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
17. M. Everingham, et al. The Pascal Visual Object Classes Challenge: A Retrospective. International Journal of Computer Vision 111, 98-136, 2015.
18. Iglovikov, V., & Shvets, A. (2018). TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. Retrieved from http://arxiv.org/abs/1801.05746.
19. Morozov, S.P., Andreychenko, A.E., Pavlov, N.A., Vladzymyrskyy, A.V., Ledikhova, N.V., Gombolevskiy, V.A., Blokhin, I.A., Gelezhe, P.B., Gonchar, A.V. and Chernina, V.Y., 2020. MosMedData: Chest CT Scans with COVID-19 Related Findings Dataset. *arXiv preprint arXiv:2005.06465*.
20. Ma Jun, Ge Cheng, Wang Yixin, An Xingle, Gao Jiantao, Yu Ziqi, He Jian. (2020). COVID-19 CT Lung and Infection Segmentation Dataset (Version Verson 1.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3757476

Figure 1

Comparison between original image and pre-processed image

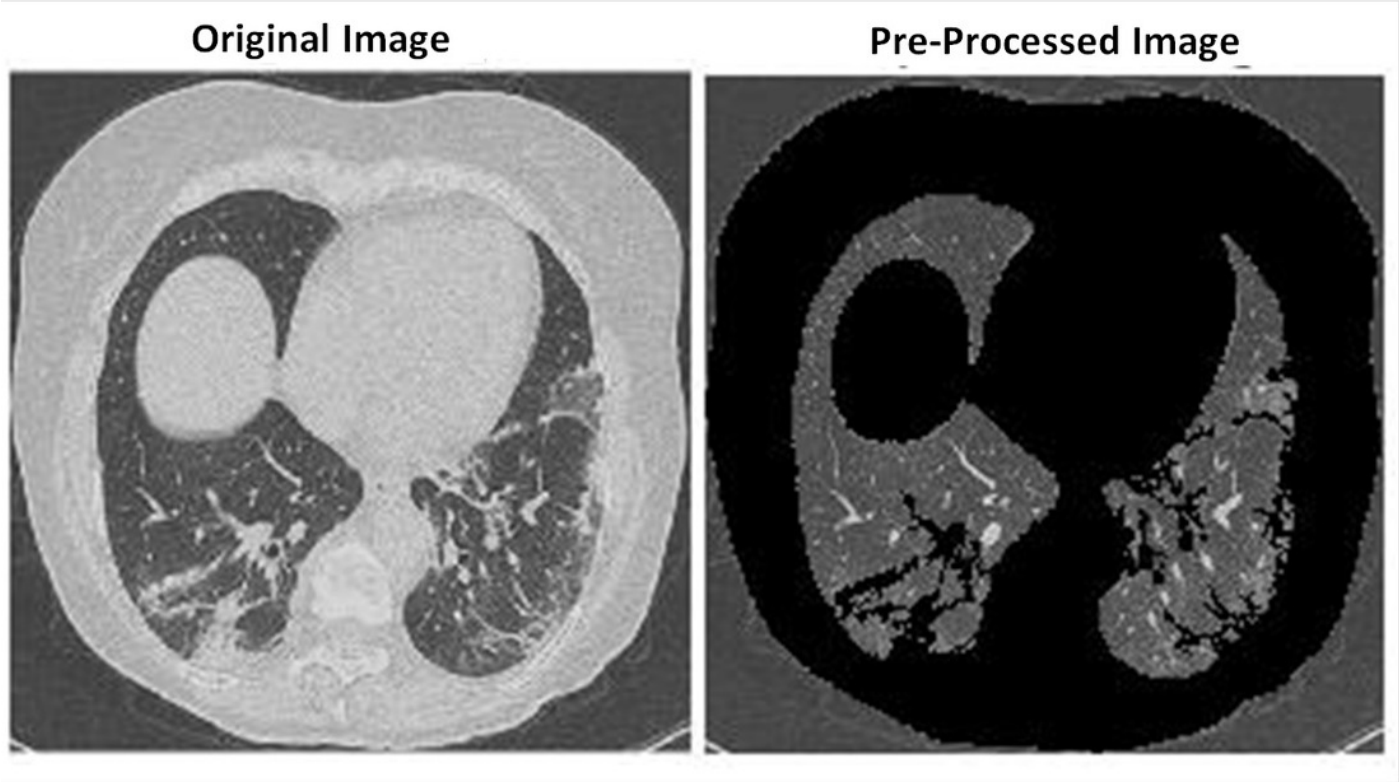


Figure 2

Block Diagram of proposed method

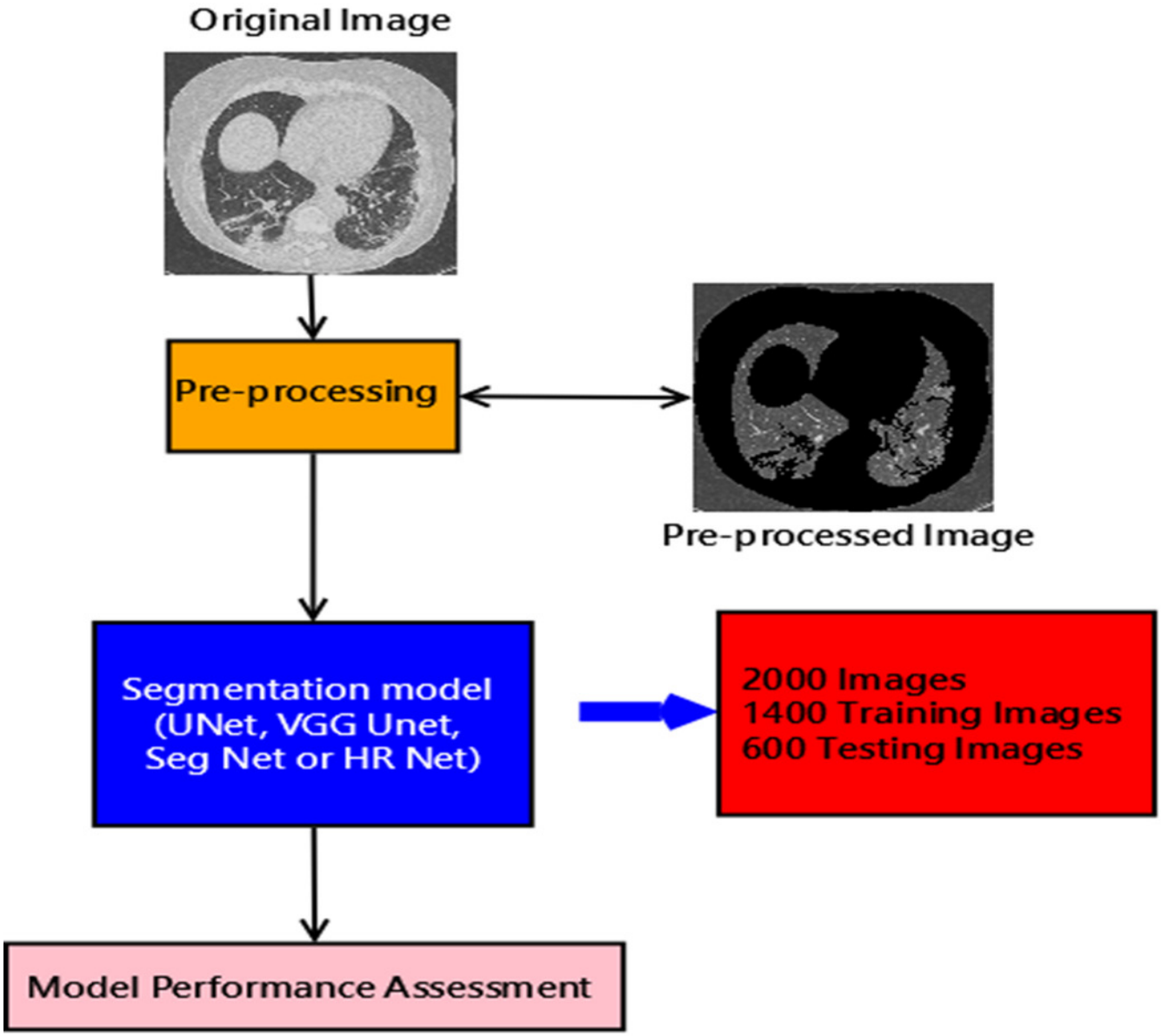


Figure 3

Architecture of HRNet

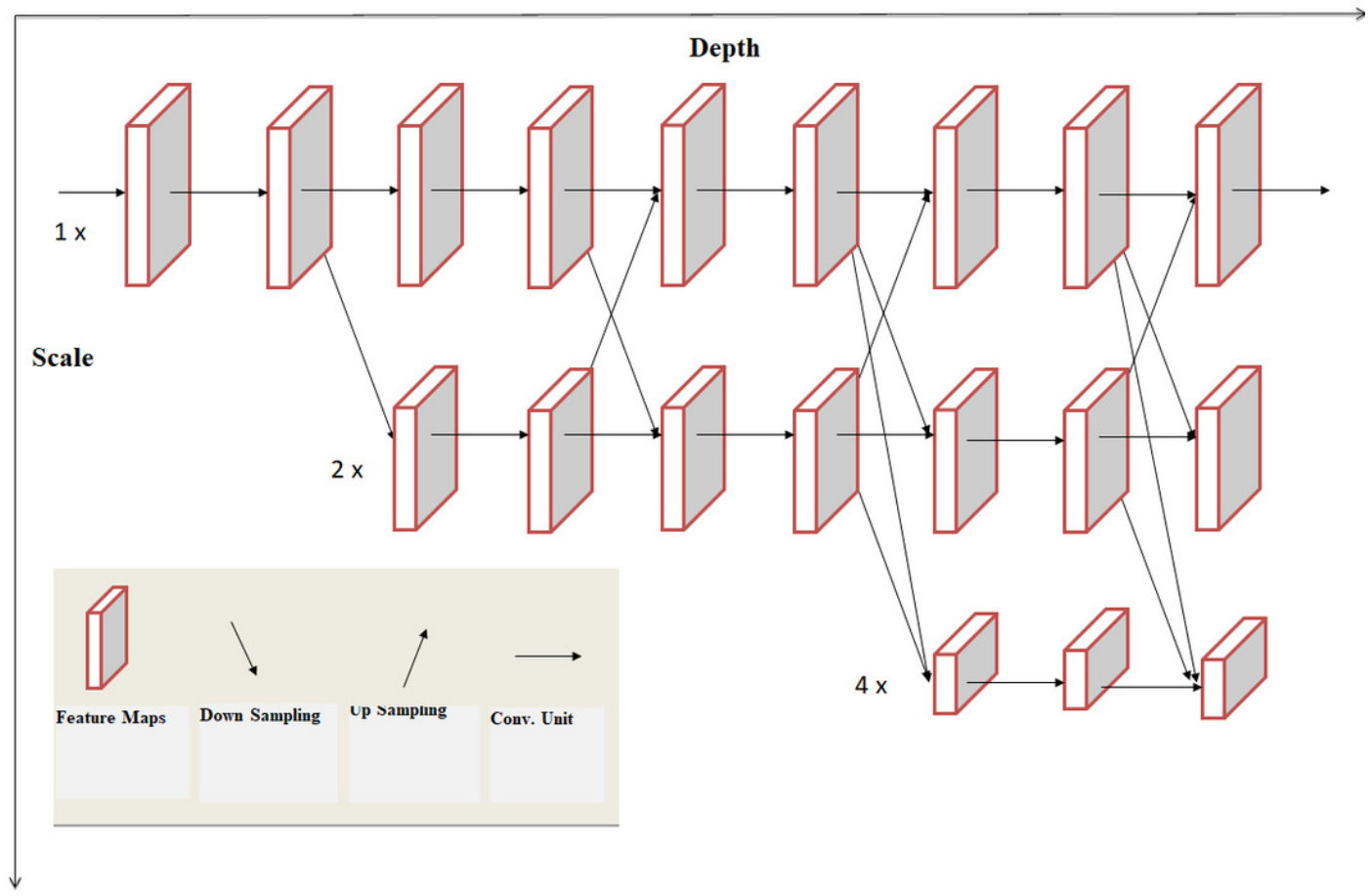


Figure 4

Low, Medium and High resolution (left to right)

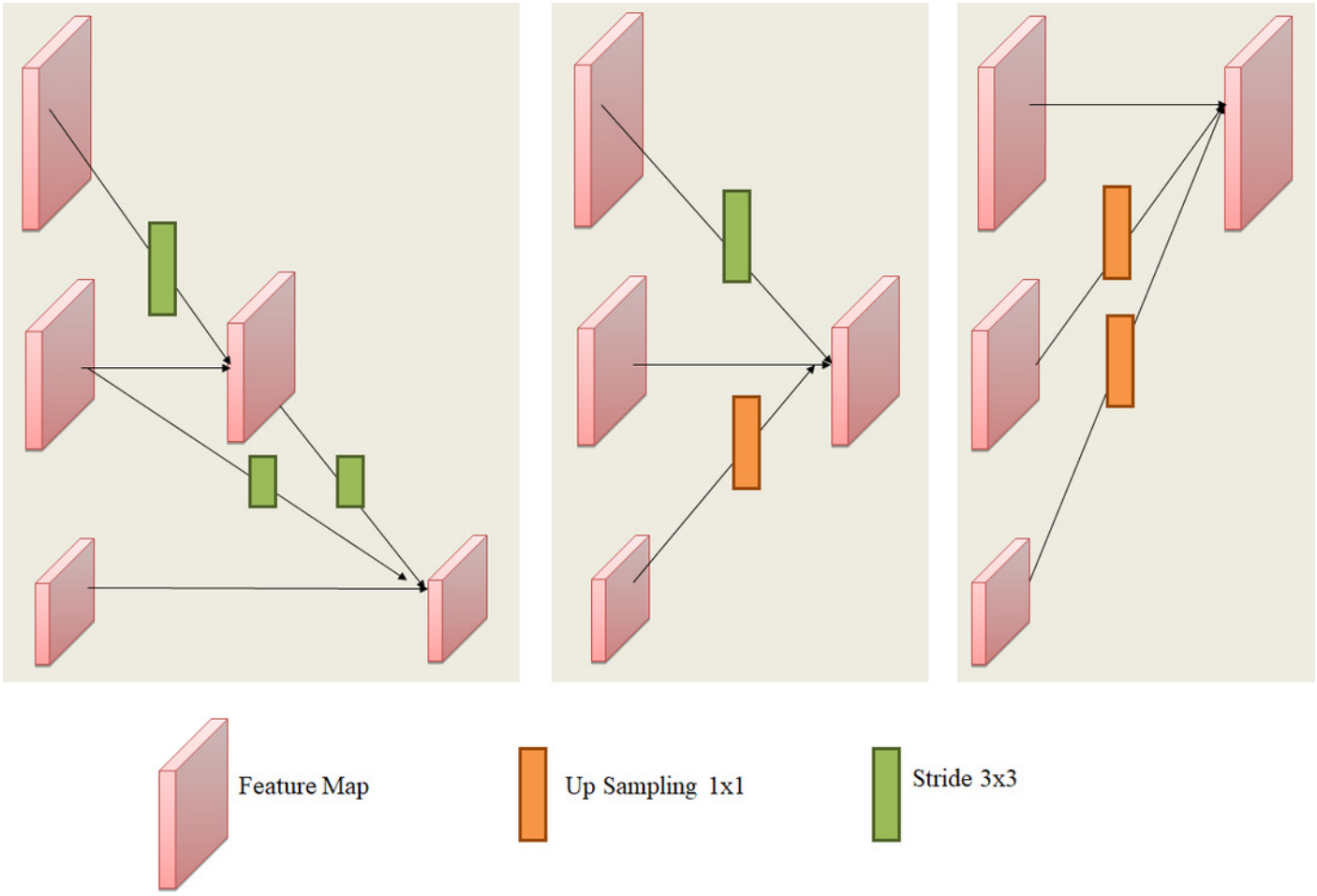


Figure 5

SegNet Architecture

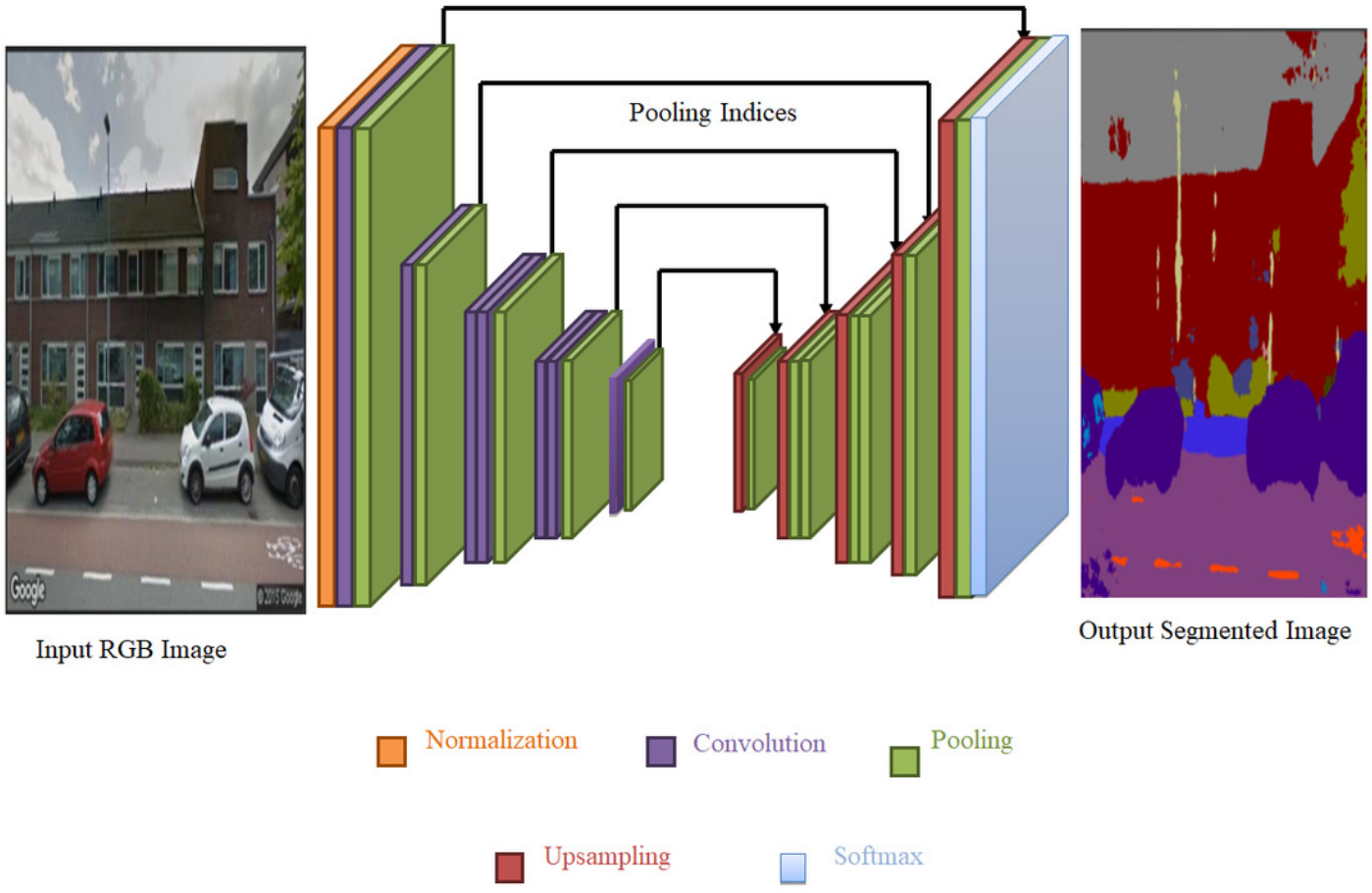


Figure 6

Decoding Techniques used by SegNet

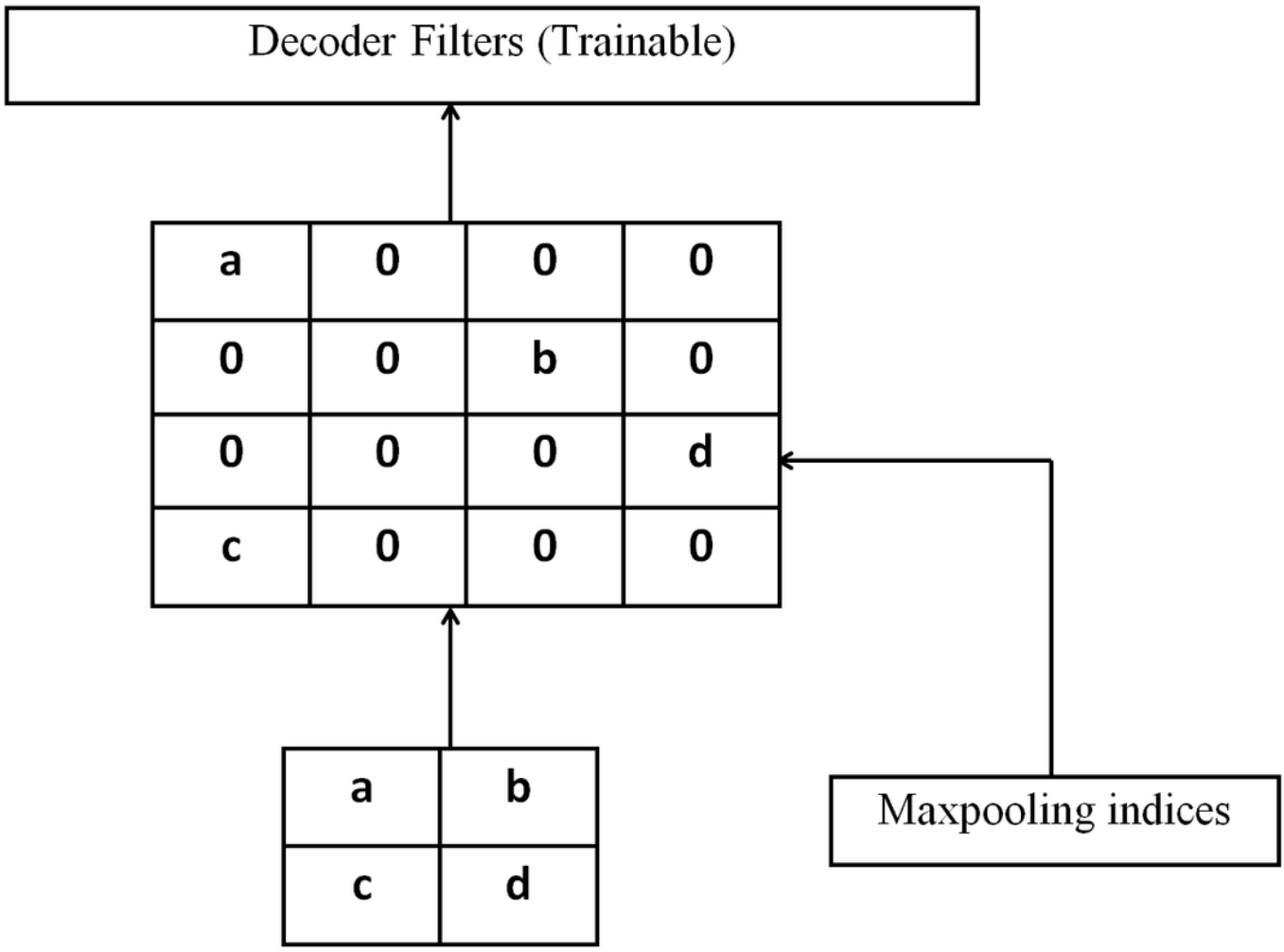


Figure 7

Illustrating the architecture of U-Net

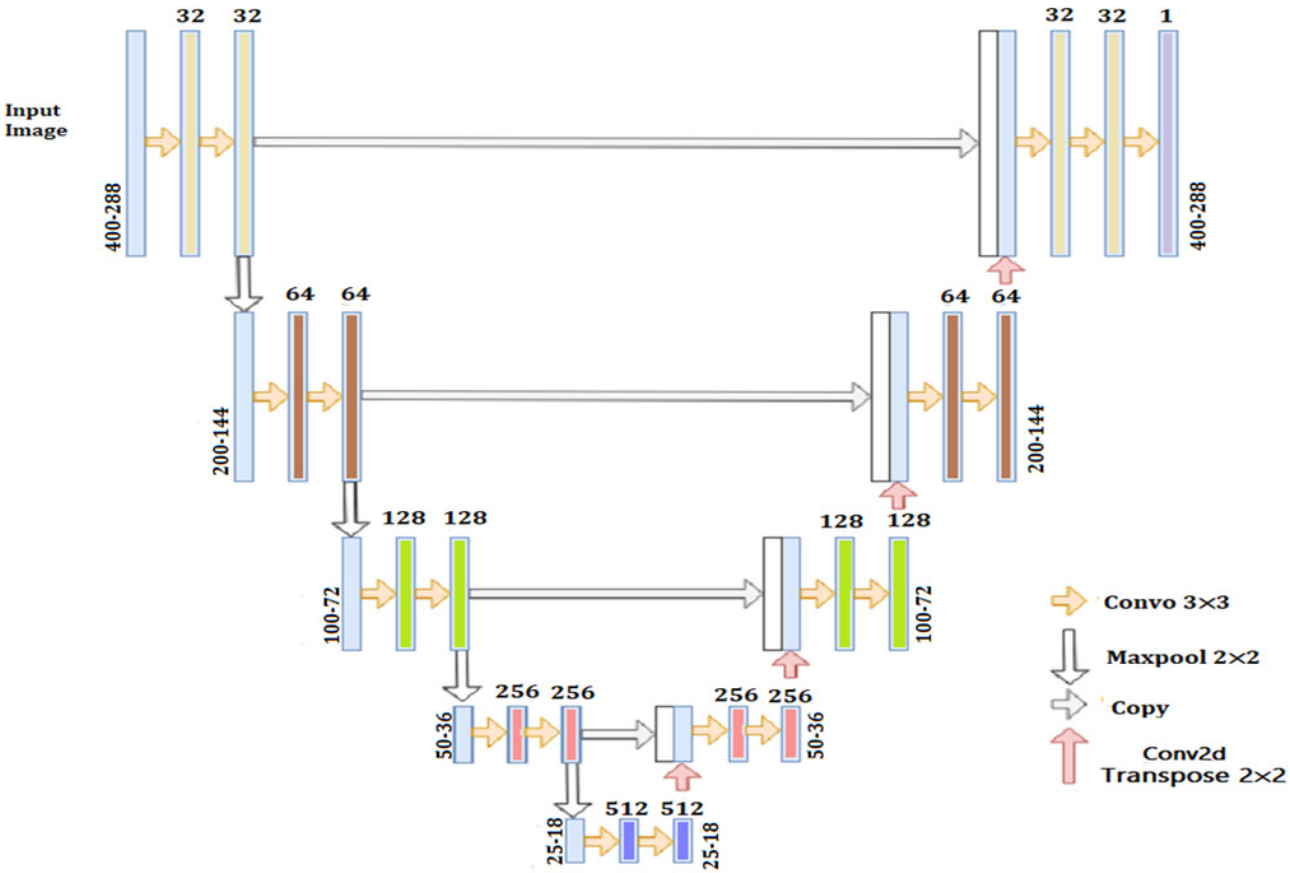


Figure 8

Illustrating encoder decoder architecture also known as VGGUNet

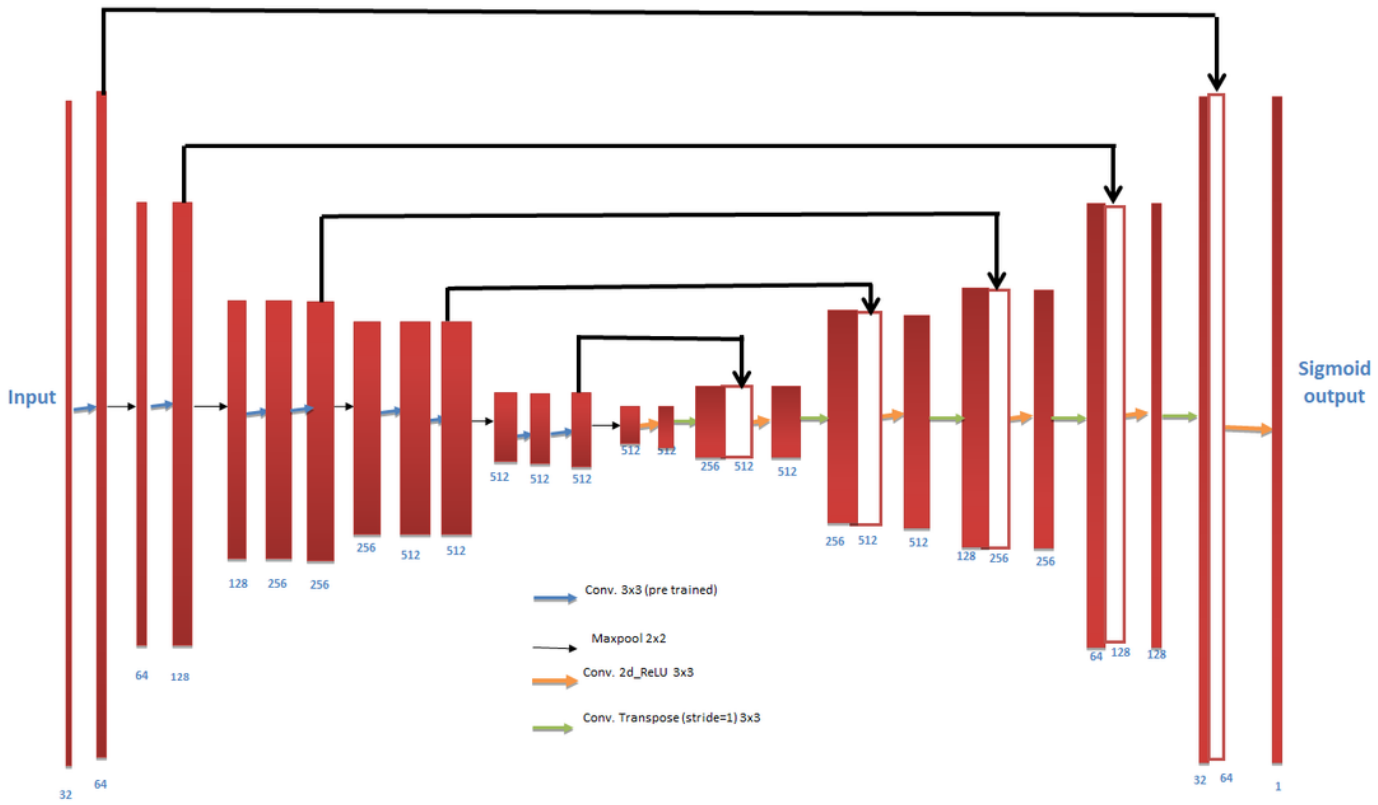


Figure 9

Sample test image and corresponding results obtained with the proposed method: (a) Original test Image after the artifact removal; (b) Binary image; (c) Region of Interest; (d)-(g) Segmentation results obtained with HR-Net, Seg Net, UNet, and VGG Net resp

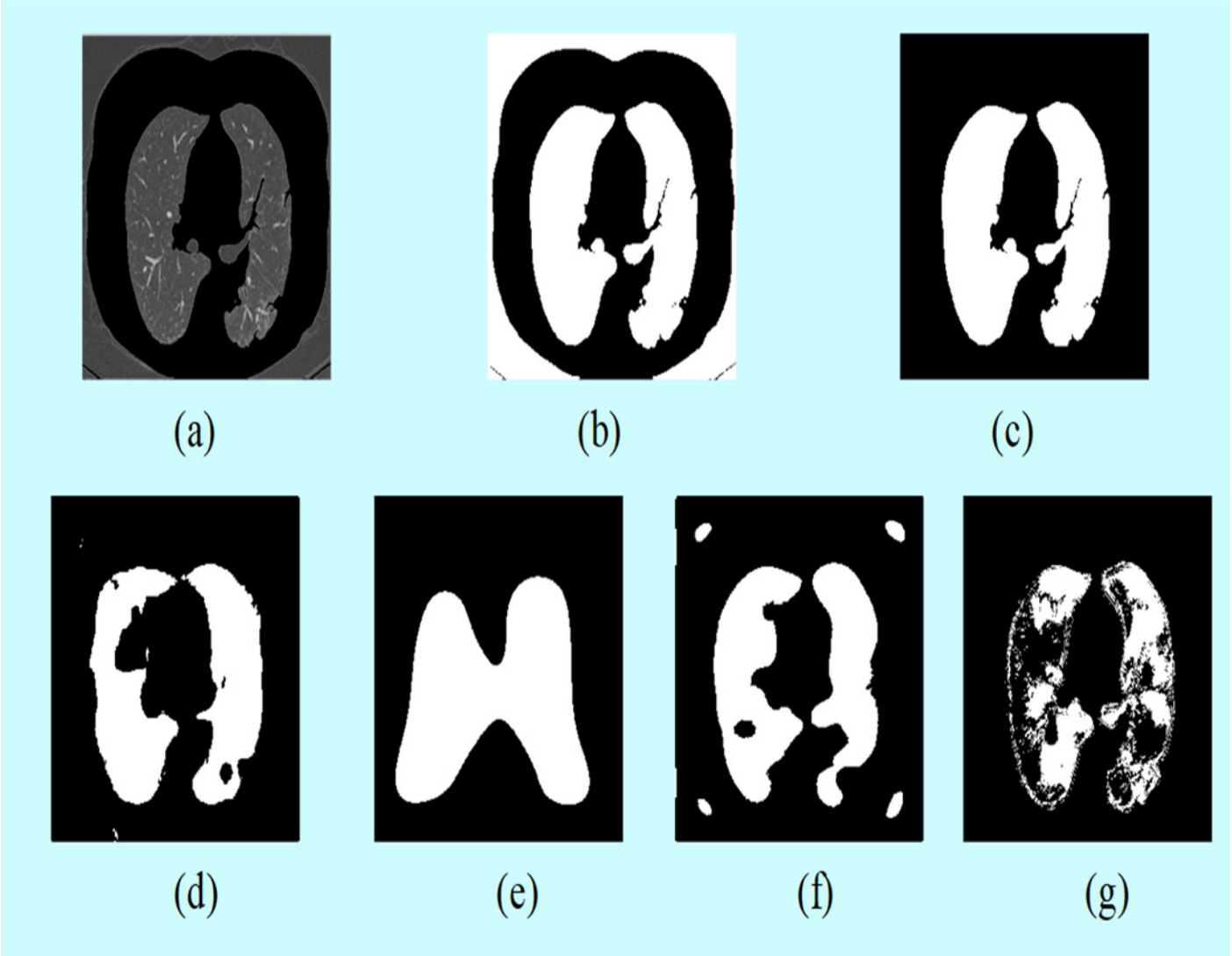


Figure 10

Glyph plot representing the performance of the models over the performance measures

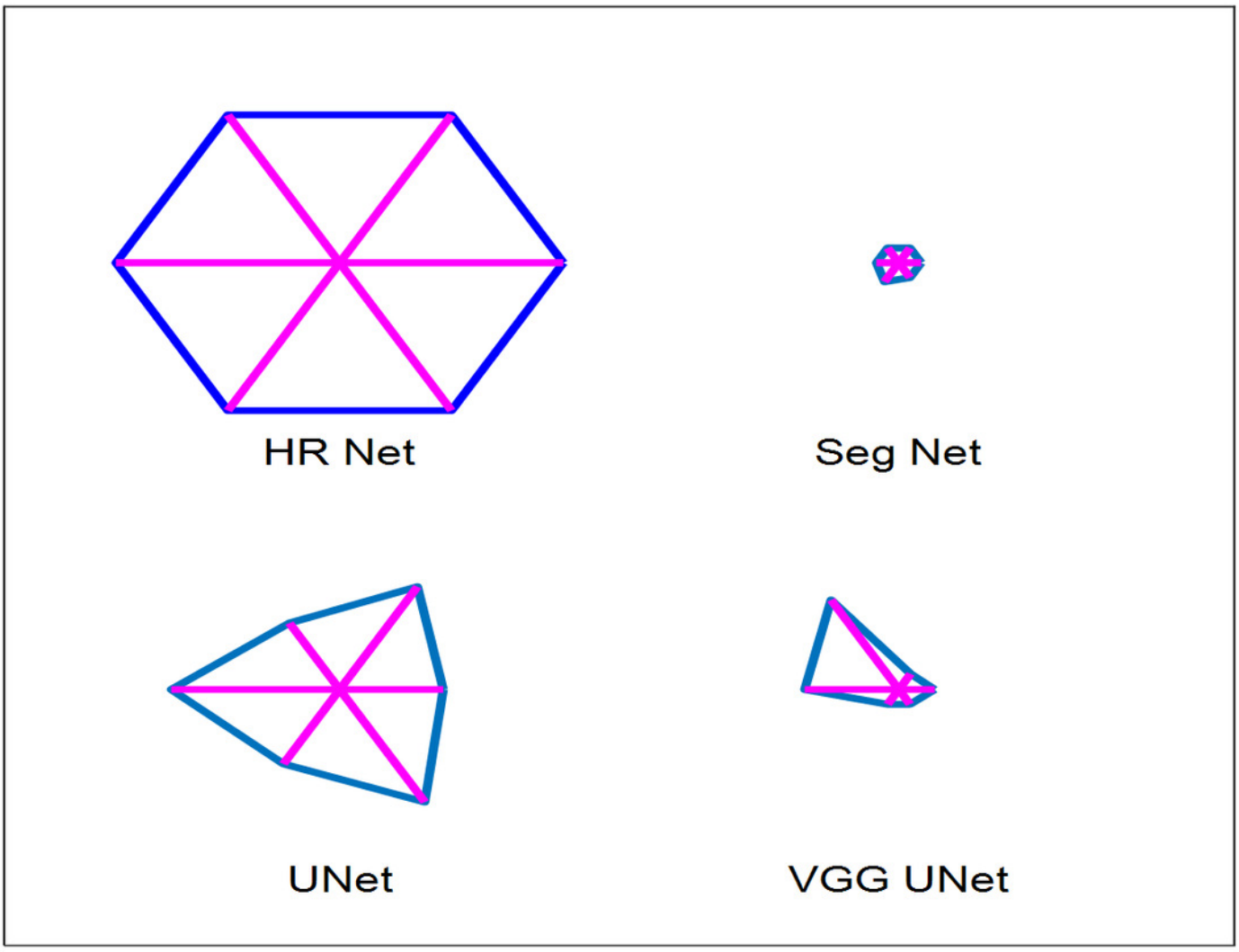


Table 1(on next page)

Performance measures of segmentation performed on various models mentioned

1 Table 1. Performance measures of segmentation performed on various models mentioned

Method	TPrate= Sensitivity	TNrate= Specificity	Jaccard	Dice	Accuracy	Precision
HR_Net	**0.8862**	**0.9930**	**0.8428**	**0.9147**	**0.9624**	**0.9593**
Seg_Net	0.7859	0.7952	0.7962	0.8014	0.8816	0.8416
UNet	0.8215	0.9195	0.8143	0.8836	0.9105	0.9281
VGG_UNet	0.7918	0.7964	0.8224	0.8418	0.8794	0.8416

2

Table 2(on next page)

Inference speed of the four models along with other parameters to affecting inference speed

Table 2: Inference speed of the four models along with other parameters to affecting inference speed

Name of Model	Inference Time (ms)	Number of Layers	Number of Params	Model Size (MB)
UNet	42	32	7,759,521	30
SegNet	84	64	31,819,649	122
VGG UNet	65	85	25,882,433	99
HR Net	140	1043	28,607,456	112