RESEARCH ARTICLE OPEN ACCESS

Learning To See In the Dark Using Neural Network

Komal Mourya[1], Sharda Patil[2], Tabasum Nadaf[3], Divya Voccaligara[4], Prof.Harsha Chari[5], Dr.Shailendra Aswale[6]

[1,2,3,4,5,6]*(Department of Computer Engineering, Shree Rayeshwar Institute of Engineering and Information Technology, Goa, India)*

ABSTRACT

Machine learning has been recently applied to extreme low -light imaging which provides stirring results. As there are many limitations of using image capturing devices moreover because the presence of the non-ideal environment, the standard of digital image gets degraded at a good level. Despite of getting much advancement in digital imaging science, the captured images do not always fulfil the user's expectation for clear and soothing view of image. Also, we encounter different algorithms used for image processing; each of it's has its improvements and limitations. This paper provides a detail investigation of assorted neural networks and its architecture used for low -light image processing. Here we have proposed a reliable framework for the task of semantic segmentation for better performance result. Our framework consists of offbeat deep learning architecture i.e. Unet and ResUnet and an offbeat loss function which is predicted on Dice loss. We have addressed several open challenges including model overfitting, reducing number of parameters and handling test, train, prediction, investigating the effect because of network architecture on accuracy, dice-coefficient, iou. The proposed framework approach outperforms all the comparing entities, which demonstrates its superiority over various recently developed states of the arts.

***Keywords* -** Dice-coeff, FCN, iou**,** ResUnet, U-Net

Date of Submission: 06-08-2020 Date of Acceptance: 20-08-2020

I. INTRODUCTION

Several researches have been carried out on image processing techniques that are more superior over the old traditional techniques. Low light is a case that cannot be avoided in our everyday life. The amount of light present during night is comparatively low than the day light and the images captured during night have to suffer more problems as compared to the images captured during day time. This can be also due to the noise present in the environment. Noise is the most common issue that is faced in low light images and the presence of this makes the processing of images much more difficult. To avoid this problem many different approaches have been introduced and carried out.

For processing low-light images different approaches have been discussed in different papers, one of the most common procedures discussed are pre-processing and semantic segmentation which are carried on the images to improve the output. Preprocessing is the initial step that is carried out in image processing pipeline. Here we intensify the images to detect finer details of images and also noise is taken care here. Image enhancement methods can also be considered for noise elimination and smoothing [1]. And next comes the segmentation part, which is a technique that helps in simplifying and representation of an image that can be easier to analyze. Here, images are segmented to extract hidden details that are further used for feature extraction. The most common use of image segmentation is to locate objects and their boundaries such as lines and curves in an image.

Here in this paper, our main focus is on low light images and images with low contrast. The entire image can be constructed again using the raw format. Further image processing techniques also depends on the network which is used. Commonly used technique is training the deep neural networks [2]. The networks are trained in such a way that they adopt and learn the techniques, which include the denoising, demosaicking, color transitions, image enhancement. To avoid error, noise and other problems, the neural network is trained using end-to-end approach. Neural network is popular technique that is carried out in many image processing experiments [3]. Several network architectures have been designed for different purposes, each of which serves different improvements over other. We have also come across that the depth of the model which indicates the performance of the model [4]. Models such as deep convolutional neural network, super resolution CNN and ResNet uses deep architectures.

However, it is difficult to train deep architecture due to the problem of information loss [5]. We have also discussed about some network architecture, its uses, and application and how it can be further improved by using different parameters. Our main focus in this experiment is using two different model architectures using the same dataset for both and analyzing the final results. The rest of this paper is documented as follows: Section II provides a summary of the related work where different research works on pre-processing, segmentation is done and discussion on neural networks and types of neural networks in brief. Section III gives a short discussion on the proposed approach where the methodology used in the proposed model is discussed. Section IV includes the experimentation and performance analysis and Section V includes the conclusion to the paper.

II. INTRODUCTION

Low light images suffer from lack of information, image quality, due to the presence of noise thus reducing the performance of many techniques and algorithms. Here we discuss about some existing methods and short details about them.

Image Enhancement

The idea behind image enhancement is to enhance the perception or accountability of data within a picture for human viewers, or to produce `better' input for other automated image processing and analysis techniques. It includes various existing techniques like Retinex model, histogram processing, Gamma correction, LIME and LLNet. Histogram Equalization is carried-out by amending the intensity levels of a picture drew on input intensities of the Probability Density Function (PDF). The output of HE sometimes gets over enhanced and hence tending to lose some information, near some areas where quantum jumps occur within the CDF. Histogram processing serves a statistical relationship between its occurrence frequency and therefore the number of every gray level [6]. Its chief feature is that the probability to look within the uniform values throughout the scope of the same. Thus, it is also called as rectangular distribution or similar of probability distribution. AGC uses HSV (Hue, Saturation, Value) color space that assists in splitting the color and its brightness details of an image into hue (H), saturation (S), and value (V) channels. The main building block of LLNet is deep neural network classes specifically referred to as Stacked Sparse Denoising Autoencoder (SSDA) [7]. It is used to intensify and magnify low-light image. Retinex model doesn't give correct output for unnatural images. This model imposes regularization parameter by making use of L2-norm minimization and spatially adaptive weight map, which is generated by local variance map and Bright Channel Prior (BCP). LIME only takes into consideration illumination map [8]. Adaptive Gamma Correction (AGC) classifies an image into low and moderate or high contrast [9].To develop illumination map, one must find maximum intensity from each pixel in R, G, B, channels to approximate this we make use of constancy methods likely Max-RGB.

Image Denoising

Image denoising cites to the recuperation of a digital image that becomes contaminated by noise. The existence of noise in any image is inevitable. It is introduced during image recording, formation, or transmission phase. Additional processing of the image generally requires that the noise is removed or is at least reduced to some extent. Even if a smallest amount of noise is present can be harmful when a very high accuracy is needed. In-order to improve the flexibility nuclear norm minimization. A new method which is known as weighted nuclear norm minimization [10]. The main aim of WNNM is to clear image from a noisy image WNNM produces very artifacts and edge structures are preserved greatly than any other known competing methods. Birth of image denoising is when a corrupted image is added to white Gaussian noise, a common result of various acquisition channels. Image inpainting issue arises when certain pixel values are mis placed or when someone wants to eliminate more enlightened patterns, such as transient texts or any other objects, within an image. A Multi-Layer Protocol (MLP) consist of huge capacity of networks very large patch-size, and training sets are also large enough [11]. Thus, with all this we can implement MPL on GPUs which is ideally suited for the necessary computation of neural network and train. Various noises present in image denoising are Gaussian noise, which is distributed over the signal evenly next in salt and pepper noise also known as intensity spikes [12]. It occurs due to transmission error. It is an example of impulse noise. Fast burst denoising technique is compared with BM3D luck imaging, optimal flow and spatial temporal filtering from comparing BM3D last burst denoising and spatial temporal filtering it is concluded that it leaves behind certain noise inside flag region [13]. In fast burst denoising noise is removed completely and process is also fast [14]. Collaborative filtering combines BM3D grouping in filtering procedure. On performing noise attenuating, even the finest detail which is shared by the group patches is been revealed by collaborative filtering.

Neural Network

As discussed, former numerous techniques have been introduced for image denoising, concern to this, there is one more technique called U-net that has been introduced. U-net has the flexibility to beat the constraints of previous old methods. Noise is that the commonest and arbitrary variation present on color information in images that exploits the image. We also come across Denoising convolutional neural network (DNCNN) that is designed to assume residual image i.e. the difference between noisy observation and the clear image [15]. Deeper the network, higher is the performance. Deep residual U-net, is an architecture that take advantages of stability from both deep residual learning and U-net. The skip connections within the residual units and between the encoding and decoding paths of the network promotes information propagation which allows designing simple yet powerful neural network and shows cleaner result with less noise [16]. The Res-Unet has outstanding feature-extraction ability, that is, it accomplishes higher segmentation accuracy, also ensures that the gradient flows smoothly and never exit during the training of a network [17]. An end-to-end semantic-segmentation network of ResUnet is proposed; during which residual block is added to U-Net because the residual block has the flexibilty to extract higher level features and reduce the over-fitting problem. It also avoids the vanishing gradient problem. With Unet ultrasound segmentation to attain accurate result becomes too difficult. The models are optimized using algorithms like, Adam optimizer algorithm, an optimizer that is derived from adaptive estimates. It combines the benefits of two methods i.e. Adagard and RMSProp. Adam is taken into the account to be most popular algorithm within the field of machine learning because of its ability of fast results [18]. Also, for segmentation we come upon another concept called Fully Convolutional Network (FCN). Neural networks are built for segmentation, classification and detection. The network consists of layers which improve the architecture, such as using multi-resolution layers [19]. So as to extract higher-level abstraction features, CNN need not require artificial feature pattern for different scenes. CNN helps to enhance and increase the robustness for semantic segmentation [20]. In the field of biomedical segmentation SK-Unet [21] (Selective Kernel-Unet) is used to detect cardio-vascular diseases. The SK module makes the acquired characteristics maps more explanatory in both structural and channel-wise. Faster R-CNN is a type of neural network which supports images of variable sizes which forms the input data to sliding window and CNN, a proposal scanning that is helpful for the convolution neural network [22]. The speed of this network to detect objects is nearly as good as real time. Faster R-CNN uses region proposal for object detection. Training a deep neural network is to some extent hard because of slow conversion state. Fast residual learning and high learning rates are used to build a very deep network and to overcome the limits [23]. The deep CNN model is proficient in region proposals deformation approach, as well as in object detection present in pruned images [24]. For semantic-segmentation a method based on U-Net is proposed and to deal with the complex images and to construct the contraction part of the U-Net it uses the residual modules of Res-Unet [25]. And also adopt the two post-processing methods that is fully connected conditional random field (CRF) and a morphological operation which will help to produce a definite segmentation result [26]. The driving force behind FCN is fully convolutionl.i.e. FCN the layers are convolutional layers. Many present image segmentation models are based on FCN, such as U-Net [27]. A U-Net is improved FCN structure, which is used for image segmentation, U-Net consists of the left path called encoder and right path as the decoder. The encoder is convolutional neural network architecture, composed of convolutional layer and batch normalization layers. The decoder consists of convolutional layers and de-convolution layers [28]. When the decoder performs upsampling, the feature maps on the encoder side of the corresponding level are concatenated by matrix [29]. This action is useful to supplement the missing pixel position information during the convolution process, thereby improving the segmentation accuracy. Another concept is of light weight U-Net, this network is used to remove particle prediction from underdone particle images with composite backgrounds at the beginning of the improved algorithm.

III. PROPOSED METHODOLOGY

Dataset

Here, we make use of the Sony SID dataset from Github. The Sony SID dataset is a collection of low light images which are captured in dark environment. It is divided into two parts that is short exposed images and long exposed images where short exposed image consists of all total 231 images and the corresponding long exposed images for 231 short exposed images are 2685 images. Further, long exposed images are considered as Ground truth which refers to information collected on location. The dataset consists of raw short-exposure low light images, with corresponding long-exposure reference images. Here, in this paper we propose to use an end to end learning for image processing. For processing the entire image processing pipeline, we need to train a fully convolutional network (FCN).

According to recent research it is shown that FCN is used to represent image processing algorithm in more effective manner. The focused general structure of FCN that is used to form the core of proposed pipeline is a U-Net. The U-Net architecture is built in such a way that it yields better segmentation of images. Thanks to its symmetry, the network features a large feature maps within the up-sampling path, which allows transferring information [29]. We are also using another architecture called Residual Unet (ResUnet), which guarantees better performance over limited parameter. ResUnet makes use of an identity mapping instead of skip connection like in FCN. Using both architectures, we run our experiment and try to compare them with each other and record performances. After knowing the results, we end up finding the more suitable model. Both the models consist of same architecture but include some changes during processing. The architecture of both the models are divided in three parts i.e. encoding or contracting path, bottleneck or bridge and decoding or expanding path. Encoding and decoding process is done in order to ensure accurate mapping of information takes place. For pre-processing we convert the raw data into clean data using bayer raw as our first step where color channels in images is converted from RGB to RGGB and then using Gamma correction black level is eliminated and then in next step brightness is increased with an amplification factor of 300. Then, in the next step we built and trained our two models that are UNet and ResUnet where training of model is done using Adam optimization algorithm and loss function as binary cross-Entropy through which model will try to adopt the features or parameter required to clear a low light image. Finally, graph showing the different plots for models is built and analyzed.

IV. EXPERIMENTATION AND PERFORMANCE ANALYSIS

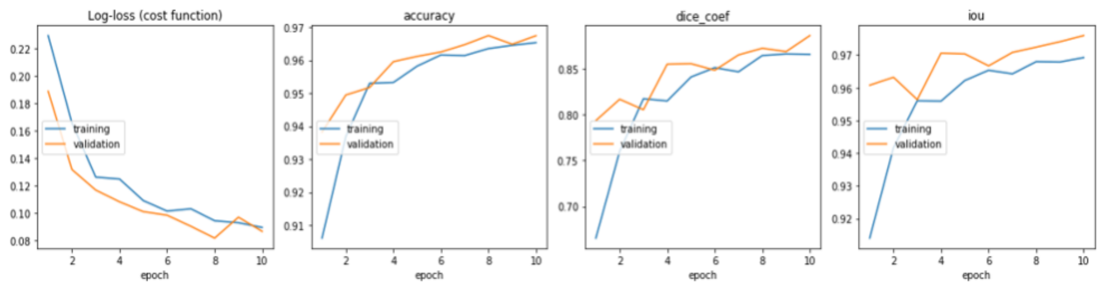
The experiment is implemented using python programming language on Anaconda, which is a free, open-source for distribution of the Python and R programming languages for scientific computing like machine learning application, data science, large-scale processing, predictive analytics and plenty of more. Thereunder we use Jupyter notebook which is an open-source web application that you just can use to make and share documents that contain live codes, visualizations, equations and text.

Preprocessing

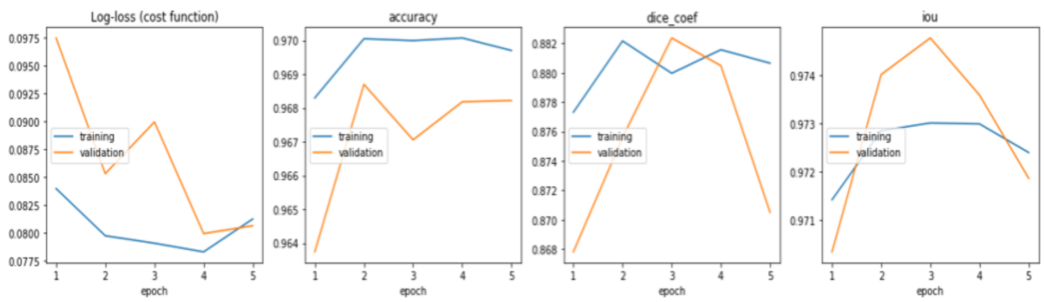
Pre-processing is used to transform the raw data into a clear data. The aim of preprocessing is detecting, correcting and improvements of images that are abolished due to the presence of noise. The steps used for preprocessing are transformation structure followed by elimination of black level followed by amplification of images in order to reduce the noise and for better processing of image.

Model Building U-net: at initial the input image is given which is resized to 512×512×3 where 512×512 is image width and height and 3 is the number of channels. Then we built model using 3×3 convolutional block or kernel. Each followed by a rectified linear unit (ReLu) and 2×2 max pooling operation with stride equals to 2 for contracting path. As the input values are of integer type, we need to convert them to floating point values because only floating points values are considered in convolutional layers. As mentioned earlier the architecture of U-net consists of contracting and expanding path. At each contracting path the feature channel doubles and at each expanding path concatenation is done with the corresponding feature channels at the contracting path. This helps in combining the low level and high-level information. Padding is set to the value same in order to add extra pixel to edges because to make the input image equal to the output image in dimensions. Thus, achieves better and promising performance. Residual U-net: the architecture is similar to that of u-net. It also uses 3×3 convolutional operator. Here, in this paper we make use of residual unet (ResUnet) and Unet to compare performance of both. We have 23 convolutional layers to be compared with 15 convolutional layers of ResUnet Residual unet has many advantages over Unet. ResUnet has advantage of both Unet and deep residual learning and its ease for training network makes it a good model for semantic segmentation and the skip connection in the network eases the information propagation, allowing us to form network with less parameter however improving performance. Further, we train our model using loss function. Here in this paper we use binary cross-entropy loss function. The loss function calculates how far aside from the accurate value the prediction for each of the classes and moderates these class-wise errors to attain the final loss. Validation is also performed on the model; validation refers to the process where trained model is evaluated with testing dataset. It is carried out after model training. Model training and model validation point to find an ideal model with the best performance. To test the generalization capacity of a trained model testing is done. We also come across the prediction values of training, testing and validation images. Prediction is used to predict how accurate our result is to the original value. After successful implementation, the outputs are recorded. The models are compared with respect to some parameter such as model loss, accuracy, dice coefficient and iou. Loss tells us the rate of loss occurred in model, accuracy is used to show how accurate the model is, dice coefficient and iou which are basically loss function which are used to decrease the loss rate when the performance rate is increasing. The Fig 1 below shows the performance of both the model in terms of accuracy and loss. It shows that the performance curve of training and validation for loss and accuracy function is improved for ResUnet. Hence, the achievement of Resunet is better than that of Unet. And, Table 1 below shows the comparison between two models

Loss and accuracy plot for U-Net



Loss and accuracy plot for Res-UNet



w.r.t the parameters used. Figure 1. Graph showing the plot for training and validation curves for loss and accuracy for Unet and ResUnet

Figure 1. Graph showing the plot for training and validation curves for loss and accuracy for Unet and ResUnet

Table 1 Comparison between Unet and ResUnet

Sr. no	Parameter	Unet	Resunet
1	Initial Loss	0.0856	0.2283
2	Final Loss	0.0811	0.0827
3	Initial Accuracy	0.9668	0.9114
4	Final Accuracy	0.9685	0.9686
5	Initial dice_coefficient	0.8751	0.6748
6	Final dice_coefficient	0.8812	0.8778
7	Initial iou	0.9706	0.9197
8	Final iou	0.9723	0.9719
9	Initial Validation loss	0.0845	1.3047
10	Final Validation loss	0.0759	0.0911
11	Initial Validation Accuracy	0.9677	0.9191
12	Final Validation Accuracy	0.9695	0.9659
13	Initial Validation dice_coefficient	0.8793	0.4330
14	Final Validation dice_coefficient	0.8820	0.8773
15	Initial Validation iou	0.9747	0.8127
16	Final Validation iou	0.9741	0.9726
17	Model Summary	• Total params: 1,941,105 • Trainable params: 1,941,105 Non-trainable params: 0	• Total params: 4,723,057 • Trainable params: 4,715,761 Non-trainable params: 7,296

V. CONCLUSION

In this paper, we used semantic segmentation network to tackle the problem of low-light images. In consideration of the features and ability, we used U-Net and Res-UNet for segmentation task. We conclude that models can be trained, tested and validated to improve the performance and increase the accuracy of results. As a result of comparison, ResUnet architecture offers many advantages over U-Net. The ability of ResUnet to give information on finer details and ease of training makes it more suitable for semantic segmentation. Further, we can assure that Resunet as well as Unet both can be used for different purposes in the field of medical and data analysis processes. FCN is more effective compared to the traditional method as the accuracy provided by these techniques was more compared to traditional methods or techniques. Later, when compared both the models we found that ResUNet obtains much better result and it can effectively extract the spatial and specific characteristics.

REFERENCES

[1]. Jinwen Yang,Weihe Zhong ,Zheng Miao.On the Image enchancement Histogram Processing.In IEEE,2016.
[2]. X. Guo, Y. Li, and H. Ling. LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2), 2017
[3]. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 26(7), 2017.
[4]. Zhengxin Zhang, Qingjie Liu (Member IEEE) and Yunhong Wang (Senior Member

IEEE). Road Extraction by Deep Residual U-Net. IEEE 2017.
[5]. Zhengquan Chu, TianTian∗, RuyiFeng, LizheWang.Sea –Land segmentation with Res-UNet and fully connected CRF. IEEE 2019 © 201XJETIRMonth201X, Volume X, Issue Xwww.jetir.org (ISSN-2349-5162) Paper id Journal of Emerging Technologies and Innovative Research (JETIR) www.jetir.org 30.
[6]. Sergei Yelmanov, Yuriy Romanyshyn. Image Enchancement Using a Modified Histogram Equation.In 2018.
[7]. K. G. Lore, A. Akintayo, and S. Sarkar. LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 61, 2017.
[8]. S. Park, S. Yu, B. Moon, S. Ko, and J. Paik. Low-light image enhancement using variational optimization-based Retinex model. IEEE Transactions on Consumer Electronics, 63(2), 2017.
[9]. Shanto Rahman, Md Mostafijur Rahman, M. Abdullah-Al-Wadud, Golam Dastegir Ai-Quaderi and Mohammad Shoyaib.An adaptive Gamma Correction for Image Enchancement.,2016.
[10]. S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In CVPR 2014
[11]. J. Zhang, J. Du, H. Liu, X. Hou, Y. Zhao, and M. Ding. LU-NET: An Improved U-Net for Ventricular segmentation. IEEE Access, 2019.
[12]. Q. Chen, J. Xu, and V. Koltun. Fast image processing with fullyconvolutional networks. In ICCV, 2017. 4, 7
[13]. H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising:Can plain neural networks compete with BM3D? InCVPR, 2012.
[14]. J. Xie, L. Xu, and E. Chen.Image denoising and in painting with deep neural networks.In NIPS, 2012 www.ijirset.com International Journal of Innovative Research in Science, Engineering and Technology Vol. 2, Issue 3, March 2013Copyright to IJIRSET www.ijirset.com 787 A Comparative Study of Image Denoising TechniquesJyotsna Patil1, Sunita Jadhav2 Assistant Professor, Department of EXTC, SCOE, Mumbai University, Maharashtra, India1 Assistant Professor, Department of EXTC, SCOE, Mumbai University, Maharashtra, India 2 ,2013
[15]. Z. Liu, L. Yuan, X. Tang, M. Uyttendaele, and J. Sun. Fast burst images denoising. ACM Transactions on Graphics, 33(6), 2014.1
[16]. J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In CVPR, 2015.
[17]. T. Ramez, O.Litany, R.Giryes, and A.M.Bronstein: Deep Convolutional Denoising of Low-Light Images. IEEE Transactions on Image Processing, 2017
[18]. D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In ICLR, 2015, 2014
[19]. Rina Komatsu and tad Gonsalves. Effectiveness fcn U-Net in denoising RGB images,2019
[20]. H. Zhao, O. Gallo, I. Frosio, and J. Kautz. Loss functions for image restoration with neural networks.IEEE Transactions on Computational Imaging, 3(1), 2017
[21]. Z. Wang, S. Yang, M. Tang, Y. Wei, L. He, J. Zhang, and X. Han. SK-Unet: An Improved U-net Model with Selective Kernel for the Segmentation of Multi-sequence Cardiac MR ,2019.
[22]. http://dx.doi.org/10.5201/ipol.2012.l-bm3d 2014/07/01 v0.5 IPOL article class analysis and implementation of the BM3D image denoising method marcLebrun11CMLA ENS CaChan,France(marc.lebrun@cmla.enscachan.fr), 2012.
[23]. B. Na, and G. C. Fox. Object Classification by a Super-resolution and a Convolutional Neural Networks. INTERNATIONAL JOURNAL OF DATA MINING SCIENCE, 2019 Recognition, 61, 2017.
[24]. B. Mildenhall, Jonathan T. Barron, JiawenChen, DillonSharlet, Ren Ng, Robert Carroll. Burst Denoising with Kernel Prediction Networks,2018
[25]. Gharbi, G. Chaurasia, S. Paris, and F. Durand. Deep joint demosaicking and denoising. ACM Transactions on Graphics, 35(6), 2016. 2, 7
[26]. Kim, J. K. Lee, and K. M. Lee. Accurate image superresolution using very deep convolutional networks. In CVPR, 2016. 4
[27]. Chen Chen,QifengChen,JiaXu,VldlenKoltun. Learning to see in the dark,CVPR 2018
[28]. O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, 2015. 4, 5, 7 [29] Rui Wang (Member IEEE), Hui Shen, Meng Zhou. Ultrasound Nerve Segmentation of Brachial Plexus Based on Optimized ResU-Net. IEEE 2019