Evaluation of a remote sensing segmentation network on medical imaging. Final Project Report

Nicolas Chahine[1] and Kevin Mercier[2]

[1] Telecom ParisTech
kevin.mercier@telecom-paristech.fr
[2] Telecom ParisTech
nicolas.chahine@telecom-paris.fr

Abstract. Segmentation in medical imaging is an important task, widely used on a variety of images. Most of the segmentation tasks are done nowadays using deep neural networks. However, in the medical field, we have many image types, and finding a good network architecture can be challenging. These images are usually more complicated then in other fields, and demand a different perspective. In this project, we train MAP-Net, a neural network developed in the remote sensing field, on three medical datasets. We compare the results with the state of the art models, for example, BCDU-Net [5]. Our work is a combination of implementation, training and testing.

Keywords: Segmentation · Remote Sensing · Medical Imaging.

1 Introduction

Segmentation in medical imaging is done on a lot of different media, from MRI to dermoscopic images, and aims to segment various object, from veins to tumors. Manual segmentation is tedious task and the field of medical imaging requires professionals to get accurate segmentation. Finding a network architecture that is able to generalize to the most of them would be great. It could be used as a reprocessing task other algorithm. It could also be used for triage ans streamlining the work of overbook medical professional or in regions where they are not many of those professionals. Satellite imaging is another field that focuses on the segmentation of different kind of objects with different sizes and in different environments. We evaluate a recent remote sensing network, MAP-Net [1], against more specific medical segmentation networks, to see its generalization ability on multiple medical imaging datasets: Retinal Blood Vessel Segmentation, Skin Lesion Segmentation and Lung Segmentation.

2 Problem Definition

Our segmentation problem can be seen as the binary classification of each pixel. We only have two classes, one for the detection and one for the background but it can easily be extended to a multi-class binary classification.

2.1 Loss function

A matching loss function to optimize is the cross-entropy loss. Since the model doesn't have a sigmoid at the end to force the output to be between 0 and 1, the sigmoid cross-entropy loss is used:

$$Loss_{ij} = -\left[y_{ij} * \ln p_{ij} + (1 - y_{ij}) \ln (1 - p_{ij})\right]$$
$$p_{ij} = \text{Sigmoid}\left(\hat{y}_{ij}\right) = \frac{1}{1 + e^{-\hat{y}_{ij}}}$$

Where (i, j) are the pixel indices, $y \in 0, 1$ the ground truth and $\hat{y}$ the output prediction. The average over all images is used has the loss function to optimize in the network.

2.2 Evaluation metrics

Multiple metrics are used to evaluate our experimental results, accuracy (AC), precision (PC), recall (RC), specificity (SF), F1-score (F1), Jaccard Similarity or Intersection Over Union (IOU) and the area under the curve (AOC) of the ROC curve, the plot of the true positive rate (TPR) against the false positive rate (FPR). Tey are defined as:

$$AC = \frac{TP + TN}{TP + TN + FP + FN}$$
$$PC = \frac{TP}{TP + FP}$$
$$SF = \frac{TN}{TN + FP}$$
$$RC = \frac{TP}{TP + FN}$$
$$F1 = \frac{2 * PC * RC}{PC + RC}$$
$$IOU(JS) = \frac{TP}{TP + FN + FP}$$

With, TP, TN, FN, FP, respectively, true positive, true negative, false negative and false positive.

3 Related Work

One of the most important and challenging tasks in medical imaging is semantic segmentation. Before the revolution of deep learning in computer vision, traditional handcrafted features were exploited for semantic segmentation. During the last few years, deep learning-based approaches have outstandingly improved the performance of classical image segmentation strategies. Based on the exploited deep architecture, these approaches can be divided into three groups.

3.1 Convolutional Neural Network (CNN)

Cui et al. [15] exploited CNN for automatic segmentation of brain MRI images. The authors first divided the input images into some patches and then utilized these patches for training CNN.Roth et al. [16] proposed a multi-level deep convolutional networks for pancreas segmentation in abdominal CT scans as a probabilistic bottom-up approach.

3.2 Fully Convolutional Network (FCN)

A problem of the CNN models for segmentation is that the spatial information of the image is lost when the convolutional features are fed into the fully connected layers. To overcome this problem the FCN was proposed in [17]. This network is trained end-to-end and pixels-to-pixels, in which all fc layers of the CNN architecture are replaced with convolutional and deconvolutional to keep the original spatial resolutions. Zhou et al. [18] exploited FCN for segmentation of anatomical structures on 3D CT images. An FCN with convolution and de-convolution parts is trained end-to-end, performing voxelwise multiple-class classification to map each voxel in a CT image to an anatomical label. Roth et al. [19] proposed to employ 3D FCN in a cascaded fashion for segmentation of the organs and vessels in CT images.

U-Net, [2], is one of the most popular FCNs for medical image segmentation. It has some advantages over the other segmentation-based networks. It works well with few training samples and the network is able to utilize the global location and context information at the same time. Milletari et al. 3D U-Net [20] is proposed for processing 3D volumes instead of 2D images. In which, all 2D operations of U-Net are replaced with their 3D counterparts. Li et al. proposed High-Res3DNet [21], which is a high-resolution, compact convolutional network for volumetric image segmentation.

3.3 Recurrent Neural Network (RNN)

Pinheiro et al. [22] proposed an end-to-end feed forward deep network consisting of an RNN that can take into account long range label dependencies in the scenes while limiting the capacity of the model. Visin et al. [23] proposed ReSeg for semantic segmentation. In that network, the input images are processed with

a pre-trained VGG-16 model and its resulting feature maps are then fed into one or more ReNet layers. Alom et al. [24] proposed Recurrent Convolutional Neural Network (RCNN) and Recurrent Residual Convolutional Neural Network (R2CNN) based on U-Net models for medical image segmentation. Bai et al. [25] combined an FCN with an RNN for medical image sequence segmentation, which is able to incorporate both spatial and temporal information for MR images.

Finally, BCDU-Net [5] was proposed as an extension of U-Net, showing better performance than state-of-the-art alternatives for the segmentation task.

4 Methodology

Medical segmentation is usually based on U-Net architectures so we tried to find other types of base architectures in other domains. We selected the DenseNet based network [7], MAPNet used in satellite imagery. The three medical datasets from [4] are quite different from each other and will be a good test to see if it can generalise well to medical problems.

4.1 MAP-Net network

The MAP-Net paper starts from the observation that repeated pooling layers and stride convolution lose spatial localization during the feature extraction procedure. U-Net based architectures often used for segmentation purposes, recovers localization using skip connections, but we observe in practice that the first feature maps are noisy and blurry, most likely due to the very architecture of the U-Net. Hence information is lost in the encoding-decoding process. To tackle this issue, they introduce a neural network architecture inspired of DenseNet where the features of high spatial dimension are well processed, and aggregated with features of lower dimension across all layers to preserve the resolution and calculating interesting multi-scale features. They notably use an attention mechanism to process together the local and global information and to aggregate feature maps.

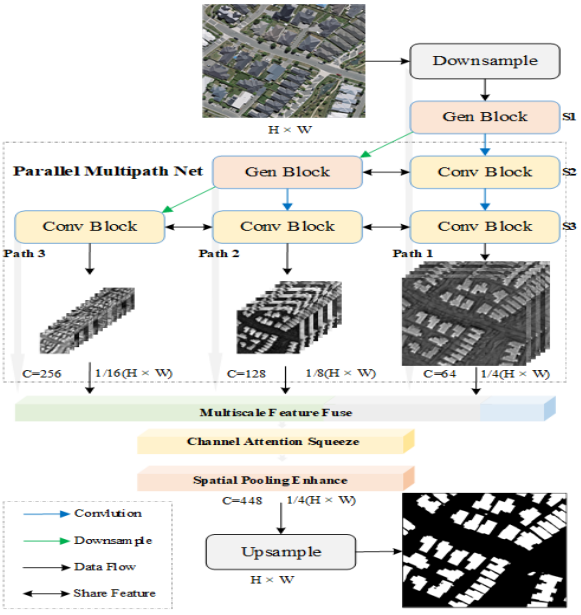


Fig. 1: MAP-Net network overview.

Localization-Preserved Multi-path Network The first stage is composed of two parallel paths, the next stage add a more downsampled path with double the number of channels. Feature maps extracted in each path maintain spatial resolution important for pixel-accuracy segmentation. Multi-scale features maps are created at the end of each path and re-scaled. The structure of MAP-Net is composed of three parallel paths for extracting multi-scale features. The spatial resolutions of feature maps are 1/4, 1/8 and 1/16 of the original image, with corresponding numbers of channels of 64, 128 and 256, respectively.

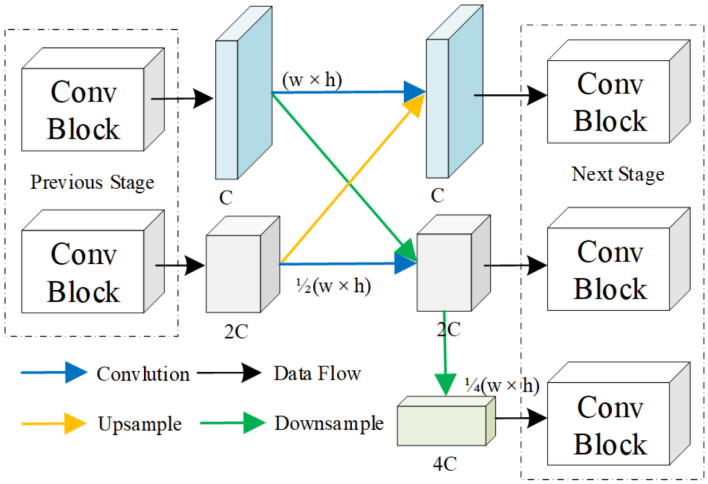


Fig. 2: Preserved multipath network.

Attention-Based Feature Squeeze The higher-resolution feature maps from the multiple paths have more accurate localization information and higher level semantic information. The lower-resolution ones have better global information. The different resolutions are harmonized with bi-linear interpolation to 1/4 of the input resolution and concatenated to a single feature vector. The channel attention squeeze module is a fully connected layer that measure the importance of each channel by learning the weight for each channel and automatically reconstructing the feature maps for optimal representation. The spatial pooling enhancement module is there to capture more global feature with a greater receptive field, it is especially useful when dealing with big objects that don't have a lot of textures.

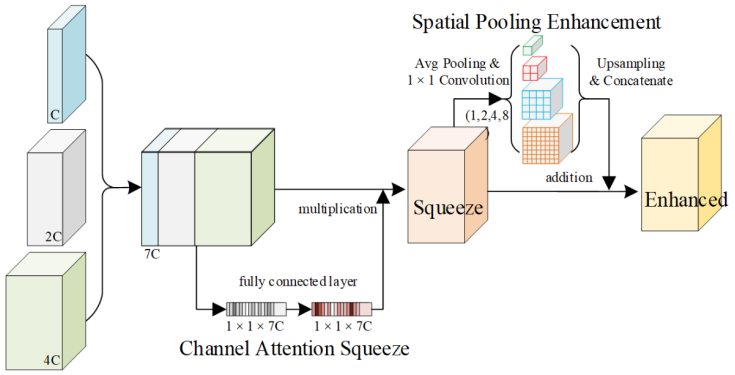


Fig. 3: Feature semantic enhancement module.

Basic block The main drawback of have a multi-path network is that it is similar to having multiple network which is computationally demanding. To reduce this complexity, the input of the multi-path is it passed through a Downsampling block, Subfig (a), the output is 64 features maps with 1/4 of the input spatial resolution. Convolution block Subfig (b) are made of Residual blocks Subfig (c), 1x1 convolution are used to reduces and get back the dimensions of features and 3x3 convolution are used to extract them. The Upsampling block in Subfig (d) take as an input the multi-scale features and generates the output segmentation mask. The upsampling is done with bi-linear interpolation and the convolution layers reduce the number of channels to 1.

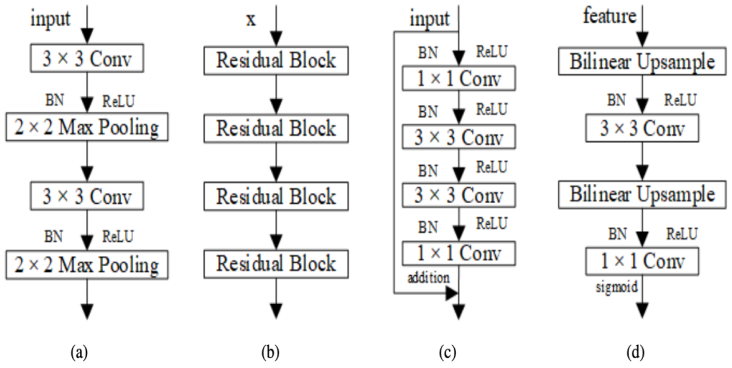


Fig. 4: Basic blocks: (a) Downsampling (b) Convolution (c) Residual (d) Upsampling

5 Evaluation

5.1 Datasets

Blood Vessel Segmentation The DRIVE dataset [8] contains 40 retina images and their corresponding blood vessel segmentation masks of size 565x584 pixels. We used 18 images for the training set, 2 for the validation test and 20 for test set. With such a low number of images we followed the reprocessing proposed by the R2U-Net paper [4]. We randomly select 171000 patches of size 64x64 for the training set and 18000 patches for the validation. Given the multi-scale nature of our network working with such small patches might lead to worse result. The original images are colored, but to be able to compare our results with other papers the images are converted to gray scale.

Image RGB Vessel Mask

Fig. 5: Example of the Drive dataset.

Skin Lesion Segmentation The ISIC dataset [9] contains 2594 dermoscopic images their corresponding melanoma segmentation masks. It was published by the International Skin Imaging Collaboration (ISIC) for cancer detection, in our case both cancerous and benign melanomas are segmented. Following [4] the 700x900 images are resized to 256x256 as shown in Fig. 6, and the training set is made of 1815 images, 259 in the validation set and 520 in the testing set.

Lung Segmentation The Lung Segmentation dataset from Kaggle [14] is composed of 1021 manually segmented 3d CT images. The dataset is then separated into 70% (714 images for the training) and 30 % (307 images) for the test. We don't use an evaluation set, since the dataset is already too small. The original images are on the gray scale level and are resized to 512x512. To prepare the data, the original mask is given as the segmented lungs. However, for training, we also save the reversed image, hence the outside mask as shown in Fig. 7

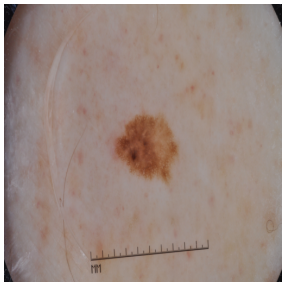
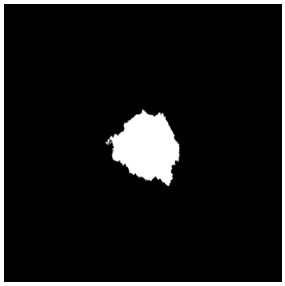
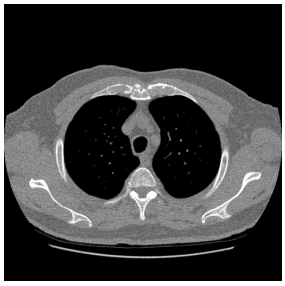


Image RGB

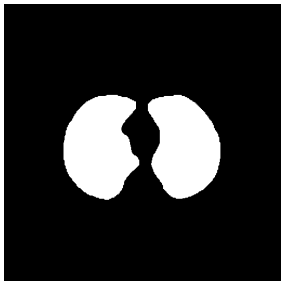


Lesion Mask

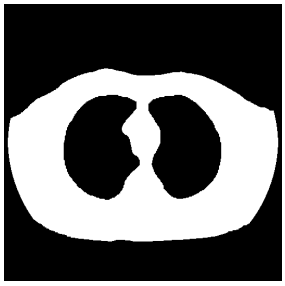
Fig. 6: Example of the ISIC18 dataset.



CT Image



Original Mask



Training Mask

Fig. 7: Example of the Lung dataset.

5.2 Training

Our code is based on the official MAP-Net Tensorflow 1.15 implementation. Their code was quite bare bone and specific to their workflow. We heavily modified it to be more generic and to work with any image size and any number of channels. The pre-processing and evaluation code is based on the official BCDU-Net implementation described in [4]. The code is quite messy and the dependencies it needs are quite out-of-date so we cleaned it a bit and updated it. The models were trained on Google Colab. It randomly provides a GPU, to use larger batch sizes the P100 with 16GB of VRAM is preferred.

Data augmentation The model data augmentation is the same for all dataset since they should all be invariant to horizontal flipping, vertical flipping and 90* rotations.

Blood Vessel Segmentation We trained it on the training set defined previously with the Adam optimizer for 50 epochs with a learning rate of 0.03 and 20

epochs with a learning rate of 0.001. The vessel in this dataset are quite small and their class is under represented in the images. To combat this unbalanced dataset, the loss is switched to his weighted version, from where weight of 2 is added when dealing with the under represented class.

Skin Lesion Segmentation The model was trained using Adam optimizer with a fixed learning rat of 0.0001. We take a small batch size of 4 since the GPU in google Colab cannot support a higher batch size (it crashed most of the time). That said, the training extended to almost 100 epochs were the peak IOU (Jaccard index) was reached at 60 epochs. The training took about 3 hours on a K80 GPU.

Lung Segmentation For this task, we use the same parameters for the training as the Skin Lesion task. However, the input images here are in gray scale and has a larger size (512x512). The training lasted also about 3 hours. Finally, The IOU reached its peak at almost 50 epochs.

5.3 Results

Blood Vessel Segmentation As expected, training with 64x64 images leads to poor performances, the multi-path and multi-resolution aspect of MAPNet have a hard time capturing the finest vessels.

Method	Year	F1-score	Sensitivity	Specificity	Accuracy	AUC
Hybrid Features [10]	2014	-	0.7252	0.9798	0.9474	0.9648
Three Stage Filtering [11]	2016	-	0.7250	**0.9830**	0.9520	0.9620
U-net [2]	2015	0.8142	0.7537	0.9820	0.9531	0.9755
Attention U-net [3]	2018	0.8155	0.7751	0.9816	0.9556	0.9782
R2U-Net [4]	2018	0.8171	0.7792	0.9813	0.9556	0.9784
BCDU-Net [5]	2019	**0.8222**	**0.8012**	0.9784	**0.9559**	**0.9788**
MAPNet [1]	2019	0.8083	0.7263	0.9805	0.9521	0.9593

Table 1: Results for the Retina DRIVE dataset.

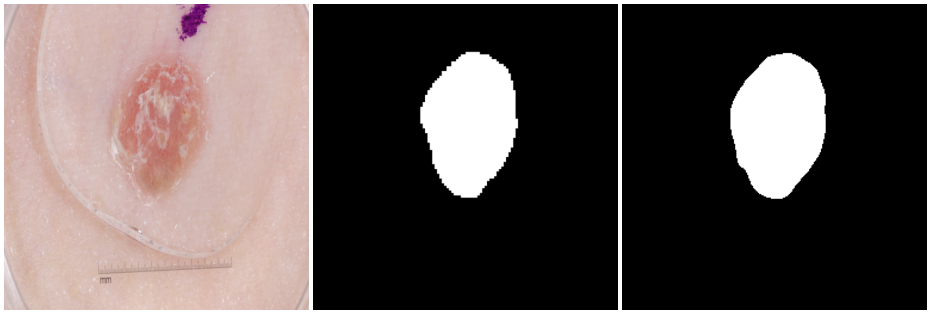
Image RGB Vessel Mask Output Mask

Fig. 8: Example of a bad detection, small vessels not detected.

Skin Lesion Segmentation The skin lesion test results surpassed the BCDU-Net, which was the state of the art before MCGU-Net [13](2020), in all the evaluation metrics. It also surpassed the MCGU-Net in most of the metrics with an impressing 96.6 % jaccard index and an F1-score of 88.2%, as shown in table 2. Moreover, Fig. 9 shows an example of a good segmentation, where the resulting mask gets very similar to the original mask.

Method	Year	F1-score	Sensitivity	Specificity	Accuracy	PC	JS
U-net [2]	2015	0.647	0.708	0.964	0.890	0.779	0.549
Recurrent Residual U-net [12]	2018	0.679	0.792	0.928	0.880	0.741	0.581
Attention U-net [3]	2018	0.665	0.717	0.967	0.897	0.787	0.566
R2U-Net [4]	2018	0.691	0.726	0.971	0.904	0.822	0.592
BCDU-Net [5]	2019	0.847	0.783	0.980	0.936	0.922	0.936
MCGU-Net [13]	2020	**0.895**	0.848	**0.986**	0.955	**0.947**	0.955
MAPNet [1]	2019	0.882	**0.893**	0.978	**0.966**	0.935	**0.966**

Table 2: Skin Lesion Segmentation results.



Original Image True Mask Result Mask

Fig. 9: Lesion result example.

Lung Segmentation The Lung segmentation task was evaluated using the inverse masks, i.e the region of the lungs instead of the outer region of the body. In this case, we can remove avoid the noise from the inside of the lungs for better evaluation. Looking at the results, we see that even though the Map-Net did not surpass the other models in most of the metrics, it gets pretty close to them as shown in table 3. That said, a deeper training using an adaptive learning rate and more data augmentation, may help to surpass the other models. For the quantitative results, Fig 10 shows the top 5 segmentation results.

Method	Year	F1-score	Sensitivity	Specificity	Accuracy	AUC	JS
U-net [2]	2015	0.9658	0.9696	0.9872	0.9872	0.9784	0.9858
Recurrent Residual U-net [12]	2018	0.9638	0.9734	0.9866	0.9836	0.9800	0.9836
R2U-Net [4]	2018	0.9832	**0.9944**	0.9832	0.9918	0.9889	0.9918
BCDU-Net [5]	2019	**0.9904**	0.9910	**0.9982**	**0.9972**	**0.9946**	**0.9972**
MAPNet [1]	2019	0.9652	0.9705	0.9904	0.9866	0.9805	0.9866

Table 3: Lung Segmentation results.

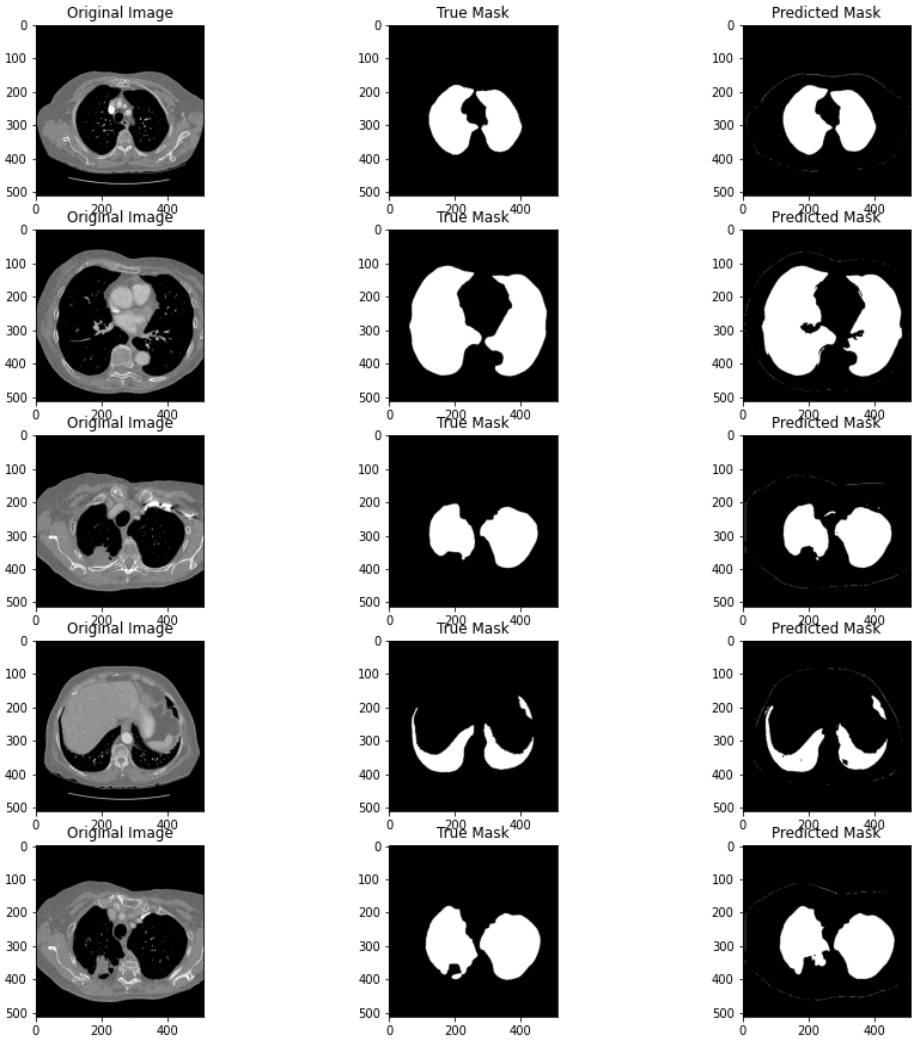


Fig. 10: Top 5 Lung segmentation results using MAP-Net.

6 Discussion

In this paper, we applied MAP-Net [1], a neural network dedicated for remote sensing segmentation, on different medical imaging segmentation tasks: Blood Vessel Segmentation, Skin Lesion Segmentation and Lung Segmentation. These datasets describes the variety of medical imaging, where each has its own problems, from very small datasets (as in the blood vessel dataset) to the badly conditioned imaging environment (as in skin lesion dataset). For each dataset, we re-implement the data extraction to adapt to the MAP-Net architecture while preserving the same data as in the training of the reference models. We also re-implement the MAP-Net training, testing process and train 3 models on Google Colab. The results are promising when we compare with the state-of-the-art models which are dedicated for medical imaging, as BCDU-Net [5] and MCGU-Net [13]. However, in the case where we have small to relatively small datasets and low input resolution, the model couldn't outperform the best model. However, with enough data, as in the case of the skin lesion, the model outperformed even the last existing model in 2020. That said, our training environment was not perfect since we were limited by what Google Colab offered, like limited GPU usage and random GPU types. Therefore, with sufficient resources, one can go further by tuning the hyperparameters of the model to get even better results and keep the full input resolution to take full advantage of the multi-path aspect of MAPNet. Moreover, we can do more data augmentation, as changing the hue of the image especially in RGB images like the blood vessels and skin lesion datasets.

References

1. Zhu, Qing and Liao, Cheng and Hu, Han and Mei, Xiaoming and Li, Haifeng, MAP-Net: Multi Attending Path Neural Network for Building Footprint Extraction from Remote Sensed Imagery, arXiv:1910.12060, 2019
2. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.
3. Oktay, Ozan, et al. "Attention u-net: Learning where to look for the pancreas." arXiv preprint arXiv:1804.03999 (2018).
4. Alom, Md Zahangir, et al. "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation." arXiv preprint arXiv:1802.06955 (2018).
5. Azad, Reza, et al. "Bi-Directional ConvLSTM U-Net with Densley Connected Convolutions." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2019.
6. Deep High-Resolution Representation Learning for Visual Recognition. Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, Bin Xiao.
7. Huang, Gao, Zhuang Liu and Kilian Q. Weinberger. "Densely Connected Convolutional Networks." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 2261-2269.
8. J. Staal, M. D. Abr'amoff, M. Niemeijer, M. A. Viergever,and B. Van Ginneken. Ridge-based vessel segmentation in color images of the retina. IEEE transactions on medical imaging, 23(4):501–509, 2004.
9. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A.Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra,H. Kittler, et al. Skin lesion analysis toward melanoma de-tection: A challenge at the 2017 international symposium onbiomedical imaging (isbi), hosted by the international skinimaging collaboration (isic). InISBI 2018, pages 168–172.IEEE, 2018.
10. Cheng, Erkang, Liang Du, Yi Wu, Ying J. Zhu, Vasileios Megalooikonomou and Haibin Ling. "Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features." Machine Vision and Applications 25 (2014): 1779-1792.
11. Roychowdhury, Sohini, Dara Koozekanani and Keshab K. Parhi. "Blood Vessel Segmentation of Fundus Images by Major Vessel Extraction and Subimage Classification." IEEE Journal of Biomedical and Health Informatics 19 (2015): 1118-1128.
12. Alom MZ, Hasan M, Yakopcic C, Taha TM, Asari VK. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955. 2018 Feb 20.
13. Asadi-Aghbolaghi M, Azad R, Fathy M, Escalera S. Multi-level Context Gating of Embedded Collective Knowledge for Medical Image Segmentation. arXiv preprint arXiv:2003.05056. 2020 Mar 10.
14. Finding and Measuring Lungs in CT Data.A collection of CT images, manually segmented lungs and measurements in 2/3D. Kaggle, https://www.kaggle.com/kmader/finding-lungs-in-ct-data/data.
15. Z. Cui, J. Yang, and Y. Qiao. Brain mri segmentation with patch-based cnn approach. In 2016 35th Chinese Control Conference (CCC), pages 7026–7031. IEEE, 2016.

16. H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers. Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In International conference on medical image computing and computer-assisted intervention, pages 556–564. Springer, 2015.
17. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in CVPR, 2015, pp. 3431–3440.
18. X. Zhou, T. Ito, R. Takayama, S. Wang, T. Hara, and H. Fujita, "Threedimensional ct image segmentation by combining 2d fully convolutional network with 3d majority voting," in Deep Learning and Data Labeling for Medical Applications. Springer, 2016, pp. 111–120.
19. H. R. Roth et al., "An application of cascaded 3d fully convolutional networks for medical image segmentation," Computerized Medical Imaging and Graphics, vol. 66, pp. 90–99, 2018.
20. O. C¸ ic¸ek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in International conference on MICCAI. Springer, 2016, pp. 424–432.
21. W. Li, G. Wang, L. Fidon, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "On the compactness, efficiency, and representation of 3d convolutional networks: brain parcellation as a pretext task," in International Conference on Information Processing in Medical Imaging. Springer, 2017, pp. 348–360.
22. P. O. Pinheiro and R. Collobert. Recurrent convolutional neural networks for scene labeling. Technical report, 2014.
23. F. Visin, M. Ciccone, A. Romero, K. Kastner, K. Cho, Y. Bengio, M. Matteucci, and A. Courville. Reseg: A recurrent neural network-based model for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 41–48, 2016.
24. M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari. Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955, 2018
25. W. Bai, H. Suzuki, C. Qin, G. Tarroni, O. Oktay, P. M. Matthews, and D. Rueckert. Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In International Conference on Medical Image Computing and Comp