Segmentation of Diabetic Retinopathy Lesions by Deep Learning: Achievements and Limitations

Pedro Furtado[a]
CISUC, Universidade de Coimbra, Polo II, Coimbra, Portugal

Keywords: Medical Imaging, Deep Learning, Segmentation, EFI.

Abstract: Analysis of Eye Fundus Images (EFI) allows early diagnosis and grading of Diabetic Retinopathy (DR), detecting micro-aneurisms, exudates, haemorrhages, neo-vascularizations and other signs. Automated detection of individual lesions helps visualizing, characterizing and determining degree of DR. Today modified deep convolution neural networks (DCNNs) are state-of-the-art in most segmentation tasks. But the task of segmenting lesions in EFI is challenging due to sizes, varying shapes, similarity and lack of contrast with other parts of the EFI, so that the results are ambiguous. In this paper we test two DCNNs to do a preliminary evaluation of the strengths and limitations using publicly available data. We already conclude that the accuracies are good but the segmentations still have relevant deficiencies. Based on this, we identify the need for further assessment and suggest future work to improve segmentation approaches.

1 INTRODUCTION

Diabetic Retinopathy (DR) is a fast-progressing disease, often resulting in blindness, early diagnosis is crucial to prevent further damage. Analysis of Eye Fundus Images (EFI) allows detection of lesions and the degree of DR. In earliest stages, a few micro-aneurisms can be seen in the EFI (enlarged capillaries resembling small red dots, e.g. less than 5) (Wilkinson et al., 2003). Later stages may include exudates (which are yellow deposits corresponding to proteins and lipids) (Oliveira, 2012) and haemorrhages. The number of micro-aneurisms may also have increased. In later (Proliferative) Diabetic retinopathy there is neo-vascularization (Jaafar et al., 2011) and related lesions.

Automated detection allows the medical doctor to visualize the lesions, characterize them and conclude regarding the degree of DR (Wilkinson et al., 2003). Deep convolution neural networks (DCNN) are state-of-the-art in segmentation of medical images. Totally automated lesion detection can be based in those segmentation DCNNs. A DCNN is built and trained based on error back-propagation on groundtruth images and corresponding segmentation masks. A DCNN architecture designed for segmentation has two main stages, encoding and decoding. The encoding stage is similar to a deep convolution neural network (DCNN), with convolution layers successively compressing the image into smaller feature maps. The fully-connected layer is replaced by additional convolution layers followed by up-sampling or deconvolution layers that create outputs with larger sizes than the inputs, this way successively restoring the original image size. Training consists in giving images as inputs and the backpropagation learning algorithm iteratively backpropagates the error between the correct segmentation masks given as groundtruth and the output of the encode+decode network, thus effectively learning how to segment images for a specific purpose.

DCNNs are most often cited as achieving very high accuracies in either classification or segmentation, therefore we wanted to test the quality of segmentation in lesion detection. There are several difficult challenges in the task, in particular the small sizes of many lesions, microaneurisms and others, varied lesion morphologies and also some similarity of colour and texture between lesions and other structures, such as parts of the vascular tree. In order to train and experiment with segmentation DCNNs, both eye fundus images and corresponding lesion mask groundtruths are needed. The Indian Diabetic Retinopathy Image Dataset (IDRiD) (Porwal and Meriaudeau, 2019) is such a dataset, prepared for experimentation with identification, localization and segmentation of lesions and structures in the EFI. Ac-

[a] https://orcid.org/0000-0001-6054-637X

Furtado, P.
Segmentation of Diabetic Retinopathy Lesions by Deep Learning: Achievements and Limitations.
DOI: 10.5220/0008881100950101
In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020) - Volume 2: BIOIMAGING*, pages 95-101
ISBN: 978-989-758-398-8; ISSN: 2184-4305

cording to its authors, IDRID dataset is the only one having pixel-level annotations of diabetic retinopathy lesions and of other retinal structures. This dataset provides information on the disease severity of diabetic retinopathy, and diabetic macular edema for each image. This makes it perfect for development and evaluation of image analysis algorithms for early detection of diabetic retinopathy". The first sub-challenge of IDRID, of especial interest to our current work, is segmentation of retinal lesions associated with diabetic retinopathy, microaneurysms, haemorrhages, hard exudates and soft exudates.

In this work we do some preliminary testing with two DCNNs to provide evidence for the research questions: What is the accuracy segmenting DR lesions, and are there significant limitations? We provide preliminary evidence, and suggest future work to evaluate and improve the solutions.

The paper is structured as follows: section 2 reviews related work. Section 3 discusses materials and methods, in effect summarily introducing the two architectures tested, the dataset and the experimental setup. Results are shown and analyzed in detail in section 4, section 5 concluding the paper.

We end the current section by illustrating the segmentation context using an example image. The problem of segmenting diabetic retinopathy lesions equates to finding, identifying and outlining microaneurysms (MA), soft exudates (SE), hard exudates (EX) and hemorrhages (HE) in EFI images. Figure 1 shows an example EFI from IDRID dataset, and Figure 2 shows the corresponding groundtruth mask detecting the lesions and the optic disk structure as well. Put very simply, the optic disk is a big rounded yellowish region, haemorrhages are blood-coloured regions, microaneurisms are very small red dots and exudates are small yellowish plus larger yellowish regions. Figure 1 is the EFI image, Figure 2 is the pixelmap groundtruth with actual lesions and optic disk locations.

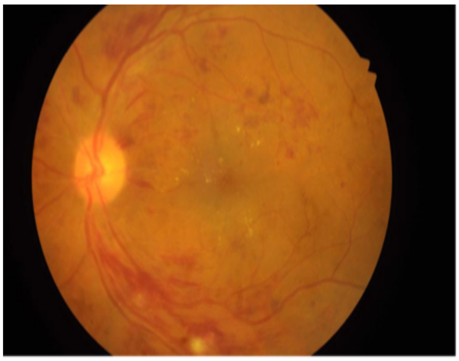

Figure 1: Example EFI.

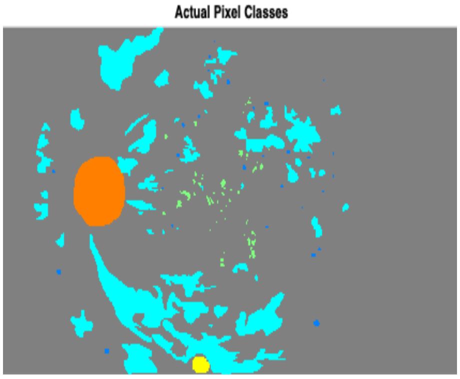

Figure 2: Example EFI GT pixelmap.

2 RELATED WORK

The Indian Diabetic Retinopathy Image challenge (IDRiD) (Porwal and Meriaudeau, 2019) is a challenge for segmentation of EFI images that supplies a dataset with pixel-level annotations of diabetic retinopathy lesions and of other retinal structures. The aim of the challenge, posed as part of the organization of the "Diabetic Retinopathy: Segmentation and Grading Challenge" workshop at IEEE International Symposium on Biomedical Imaging (ISBI-2018), is to evaluate algorithms for automated detection and grading of diabetic retinopathy and diabetic macular edema using retinal fundus images, and in particular sub-challenge 1 involves segmentation of retinal lesions associated with diabetic retinopathy. The IDRiD leaderboard for sub-challenge 1 shows a set of results ranked by score of segmentation of microaneursms (MA), hard exudates (HE), soft exudates (SE) and haemorrhages (EX) score. The scores are very far from perfection (whch would be a value of 1). Consequently, it is important to evaluate the quality of these deep learning approaches on the task.

In this paper we test two DCNN architectures and do a preliminary analysis of the results to draw conclusions regarding achievements and limitations of the approaches. Segmentation of biological structures using deep learning has been the focus of much research in latest years, and (Menze et al.,) (BRATS, 2014) already featured works applying deep convolution neural networks to segment brain tumours and structures (Davy, 2014)(Urban et al., 2014)(ZikicD et al., 2014), with for instance (ZikicD et al., 2014) reporting 83.7+-9.4 accuracy on brain tumor tissues versus 76.3+-12.4 for non-deep learning randomized forests. Since then segmentation deep learning networks based on DCNNs became the standard in segmentation tasks.

There has been some prior evidence in related works that deep segmentation networks can have some difficulties with variability and size of segmented objects. For instance, in (Badrinarayanan et al., 2017) the authors evaluate and compare approaches on a SUN RGB-D dataset (Song et al., 2015) (a very challenging and large dataset of indoor scenes with 5,285 training and 5,050 testing images). The results have shown that all the deep architectures share low Intersect over Union and boundary metrics, where larger classes have reasonable accuracy and smaller classes have lower accuracies.

Next, we briefly review some of the milestones in the evolution of deep learning segmentation networks, from the first ones to DeepLabV3 and alike. The Fully Convolutional Network (FCN) for image segmentation was proposed in [14]. It modified well-known architectures, such as VGG16 [15], replacing all the fully connected layers by convolutional layers with large receptive fields and adding up-sampling layers based on simple interpolation filters. Only the convolutions part of the network was fine-tuned to learn deconvolution indirectly. The authors achieved more than 62% on the Intersect over Union (IoU) metric over the 2012 PASCAL VOC segmentation challenge using pretrained models on the 2012 ImageNet dataset. The authors in (Noh et al., 2015) proposed an improved semantic segmentation algorithm, by learning a deconvolution network. The convolutional layers are also adapted from VGG16, while the deconvolution network is composed of deconvolution and unpooling layers, which identify pixel-wise class labels and predict segmentation masks. The proposed approach reached 72.5% IoU on the same PASCAL VOC 2012 dataset. (Ronneberger et al., 2015) proposed the U-Net, a DCNN specially designed for segmentation of biomedical images. The authors trained the network end-to-end from very few images and outperformed the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks. The contracting part of the U-Net computes features, while the deconvolution part localizes patterns spatially in the image. The contracting part has an FCN-like architecture, extracting features with 3x3 convolutions, while the expanding part uses deconvolutions to reduce the number of feature maps while increasing the size of the images. Cropped feature maps from the contracting part are also copied into the expanding part to avoid losing pattern information. At the end, a 1x1 convolution processes the feature maps to generate a segmentation map assigning a label to each pixel of the input image. DeepLab (Chen et al., 2017) proposed three main innovations. Convolutions with upsampled filters, or 'atrous convolution', explicitly controls the resolution at which feature responses are computed and enlarges the field-of-view of filters to incorporate larger contexts without increasing the number of parameters or the amount of computation. Atrous convolution is also known as dilated convolution, consisting of filters targeting sparse pixels with a fixed rate. Atrous spatial pyramid pooling (ASPP) segments objects at multiple scales, by probing incoming convolutional feature layers with filters at multiple sampling rates and effective fields-of-views, thus capturing objects as well as image context at multiple scales. Finally, localization of object boundaries is improved by combining methods from deep convolution neural networks (DCNNs) and probabilistic graphical models. This is done by combining the responses at the final DCNN layer with a fully connected Conditional Random Field (CRF), which improves localization both qualitatively and quantitatively. "DeepLab" achieved 79.7% IoU on PASCAL VOC-2012 semantic image segmentation task, and improvement over (Long et al., 2015) and (Noh et al., 2015).

3 MATERIALS AND METHOD

For this work we apply one of the most recent and best performing DCNN segmentation network architectures, DeepLabV3 (Chen et al., 2017), and also Segnet (Badrinarayanan et al., 2017), for comparison purposes. In this section we review the two architectures briefly and the IDRiD dataset that was used in our experimental analysis. Finally, we discuss training setup, timings and results.

3.1 DCNN Architectures

The DeepLabV3 (Chen et al., 2017) architecture in this work uses imageNet's pretrained Resnet-18 network, with atrous convolutions as its main feature extractor. DeepLabV3 introduces a set of innovations. Figure 3 shows a plot of the overall architecture of DeepLabV3 we used. First of all, it uses multiscale processing, by passing multiple rescaled versions of original images to parallel CNN branches (Image pyramid) and by using multiple parallel atrous convolutional layers with different sampling rates (ASPP). In the modified ResNet model, the last ResNet block uses atrous convolutions with different dilation rates, and Atrous Spatial Pyramid Pooling and bilinear upsampling are used in the decoder module on top of the modified ResNet block. Additionally, structured prediction is done by fully connected Conditional Random Field (CRF). CRF is a postprocessing step used to improve segmentation results, a graphical model which ‘smooths’ segmentation based on the underlying image intensities. CRF works based on the observation that similar intensity pixels tend to be labeled as the same class. CRFs can typically boost scores by 1-2%.

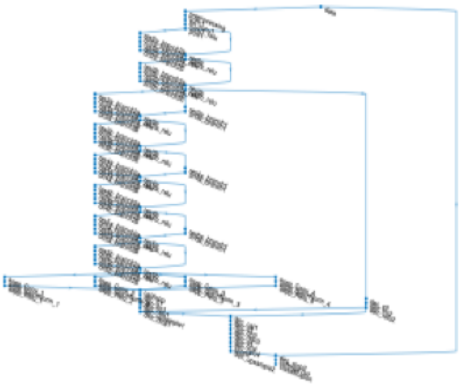

Figure 3: Matlab2019 plot of DeepLabV3 net.

Segnet is another deep convolutional encoder-decoder architecture for image segmentation proposed in (Badrinarayanan et al., 2017) and shown in Figure 3, with 5 encoder and 5 decoder “stages”, plus the central encoder-decoder stage, resulting in a total of 87 layers and corresponding connections. At the encoder, convolutions and max pooling are performed. There are 13 convolutional layers from VGG-16. Each encoder stage is made of two successive conv+bn+relu layers (bn is batch normalization), plus 2x2 max pooling, with the corresponding max pooling indices (locations) used as forward connecting links, to perform non-linear up-sampling. Each decoder stage does unpool (using the pooling indices), followed by two deconv+bn+relu. The last two layers are softmax and the pixel classification layer, with weight balancing.

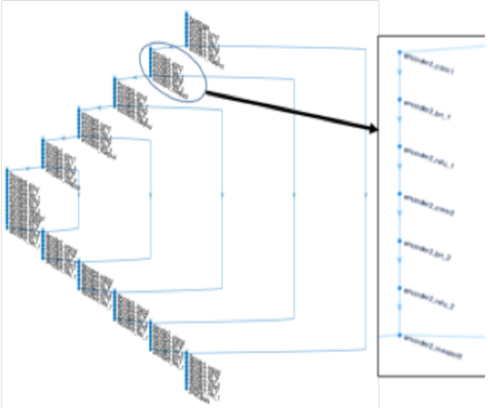

Figure 4: Matlab2019 plot of Segnet.

3.2 The IDRID Dataset

The challenge of the Indian Diabetic Retinopathy Image Dataset (IDRiD) (Porwal and Meriaudeau, 2019) is to evaluate algorithms for automated detection and grading of diabetic retinopathy and diabetic macular edema using retinal fundus images. According to the organizers of the challenge, “IDRiD is the only dataset constituting typical diabetic retinopathy lesions and normal retinal structures annotated at a pixel level. The dataset provides information on the disease severity of diabetic retinopathy, and diabetic macular edema for each image. This makes it perfect for development and evaluation of image analysis algorithms for early detection of diabetic retinopathy. In particular, the lesion segmentation task of the challenge aims at segmenting retinal lesions associated with diabetic retinopathy, which can be microaneurysms, hemorrhages, hard exudates and soft exudates. The task also includes identifying and segmenting the optic disc correctly”. The fundus images in IDRiD were captured by a retinal specialist at an Eye Clinic located in Nanded, Maharashtra, India. From the thousands of examinations available, 516 images were extracted to form the dataset (from which 81 were selected for the lesion segmentation sub-challenge). Experts verified that all images are of adequate quality, clinically relevant, that no image is duplicated and that a reasonable mixture of disease stratification representative of diabetic retinopathy (DR) and diabetic macular edema (DME) is present. The medical experts graded the full set of 516 images with a variety of pathological conditions of DR and DME.

The images were acquired using a Kowa VX-10 alpha digital fundus camera with 50-degree field of view (FOV), and all are centered near to the macula. The images have a resolution of 4288x2848 pixels and are stored in jpg file format. The size of each image is about 800 KB. This dataset for the lesion segmentation sub-challenge consists of 81 colour fundus images with signs of DR. Precise pixel level annotation of abnormalities associated with DR like microaneurysms (MA), soft exudates (SE), hard exudates (EX) and hemorrhages (HE) is provided as a binary mask for performance evaluation of individual lesion segmentation techniques. It includes color fundus images (.jpg files) and binary masks made of lesions (.tif files). Number of images (some images contain multiple lesions) with binary masks available for particular lesion is given as follows: MA – 81, EX – 81, HE – 80, SE – 40. In addition to all the abnormalities, binary masks for the optic disc region are provided for all 81 images.

3.3 Experimental Setup

The original IDRiD training dataset was divided randomly into 5 folds with 80%/20% combinations used by choosing one of the folds as test data and the remaining dataset as train data. Our experiments involved evaluating DeepLabV3 and Segnet in the task of segmenting the IDRiD dataset. The networks and experimental setup were implemented in Matlab2018, and the networks were modified to balance class weights. The following initial training options were used, the validation patience was set to infinite and the number of training epochs was set to 500, but an interactive training progress view and manual stopping option allowed us to stop when visual inspection of the training curve showed that the training progress converged to a final steady state. MR classNames=[”BackGround”, ”MA”, ”HE”, ”SE”, ”EX”]; ’LearnRateSchedule’ = ’piecewise’, ’LearnRateDropPeriod’ = 10, ’LearnRateDropFactor’=0.8, ’Momentum’, 0.9, ’InitialLearnRate’, 0.001; ’MaxEpochs’=500, ’MiniBatchSize’=8, ’Shuffle’=’every-epoch’, ’Plots’=’training-progress’, ’ValidationPatience’=Inf; Average training time of DeepLabV3 was 19 mins, Segnet was 362 minutes. Figure 5 is a depiction of deepLabV3 training accuracy evolution along iterations.

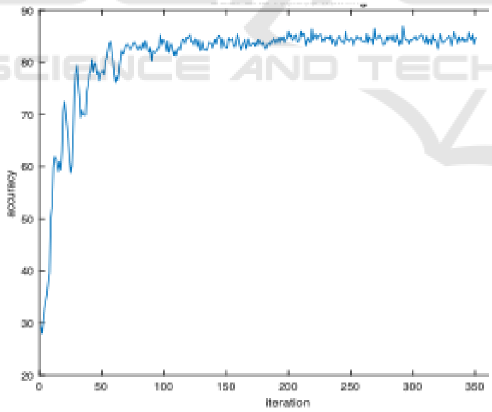

Figure 5: Evolution of DeepLabV3 training accuracy.

3.4 Analysis Metrics

The metrics used to analyse the results are some of the most commonly available in DCNN toolboxes, including accuracy and intersect-over-the-union (IoU) or Jaccard Index. The IoU is one of the most commonly used metrics for evaluating segmentation (sometimes the Dice coefficient is used instead, however the two are highly positively correlated). In practice, all these metrics are useful in the evaluation of segmentation outcomes, each one returning an important interpretation of what is observed. Importantly, we analyse the results not only overall (global accuracy, mean accuracy, mean IoU, weighted IoU), but considering each class (lesion) separately (per-class accuracy, IoU). The use of these metrics was fundamental to allow us to reach relevant conclusions regarding the strengths and limitations.

4 RESULTS

After training the networks with IDRiD we proceeded to analyse and interpret the results. Section 4.1. visualizes sample images and corresponding results. This gives an initial impression of the quality of segmentation, although still only specific cases. In section 4.2. we report numerical results using the defined metrics, analyze and interpret those results. This allows us to conclude regarding the quality, strengths and limitations of the approaches, together with suggestion of more evaluation and future work.

4.1 Visualizing Sample Images

Visual inspection helps verify the segmentation result on samples, a preliminary way to test the quality of segmentation. Figure 6 and Figure 7 show the groundtruth (left) and segmentation (right) results for two sample EFI using DeepLabV3 and Segnet. Figure 8 and Figure 9 shows the results for another image. DeepLabV3 was able to segment the optic disk almost perfectly in both samples, and it was also able to detect and match many of the lesions, verified by the similar lesion patterns in the groundtruth and the segmentation itself. However, many background pixels were also identified as lesions. The same phenomena is seen in Segnet, but there the false positives are much more prevalent, together with more wrong pixel classifications.

The main conclusion from inspection of these images is that DeepLabV3 seems to segment better, but both DCNNs confuse parts of the background as lesions.

Figure 10 and Table 1 shows the evaluation of the tested approaches based on global metrics (those that evaluate over all pixels). Table 2 shows the metric Intersection-over-the-union for each lesion.

4.2 Results and Analysis

Global accuracy metrics (Figure 10 and Table 1): the results of DeepLabV3 reveal that global accuracy, mean accuracy and weighted IoU are good (80 to

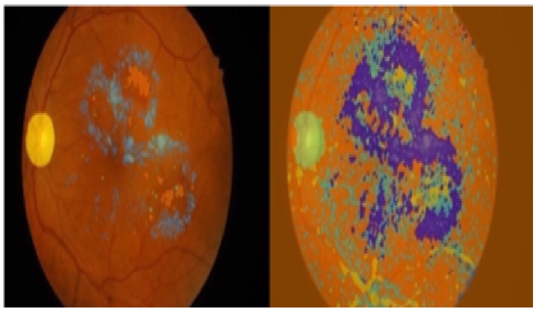

Figure 6: Segmentation of EFI image 1 (DeepLabV3).

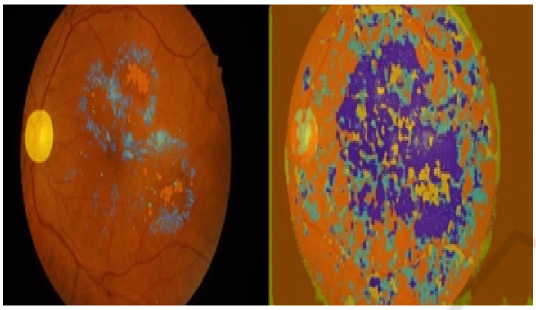

Figure 7: Segmentation of EFI image 1 (Segnet).

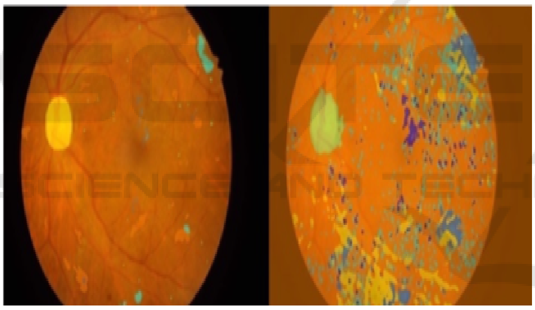

Figure 8: Segmentation of EFI image 2 (DeepLabV3).

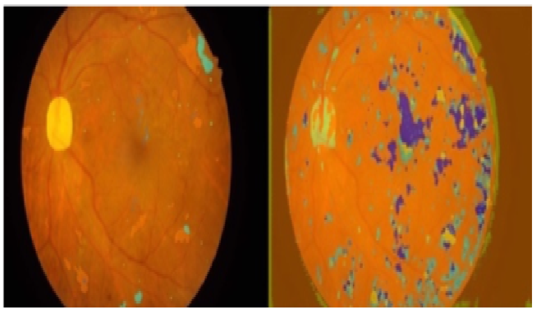

Figure 9: Segmentation of EFI image 2 (Segnet).

87%). Segnet has much worse results in all those metrics (47% to 57%), confirming that DeepLabV3 segmentation results are much better than those of Segnet.

The IoU of each lesion, however, reveals that the degree of matching of result segments with groundtruth regions is only good for the background and the optic disk in both approaches, much worse always in the case of Segnet. Good accuracy but bad IoU of DeepLabV3 means the approach is good identifying lesion pixels but at the expense of wrongly classifying many background pixels as lesions. These results provide an indication that there are definite deficiencies in the deep learning approaches applied to segmentation of this kind of images. Further study, evaluation and analysis of deep learning approaches applied to this kind of problem is necessary, as well as identification of the main limitations and proposals of improvement avenues.

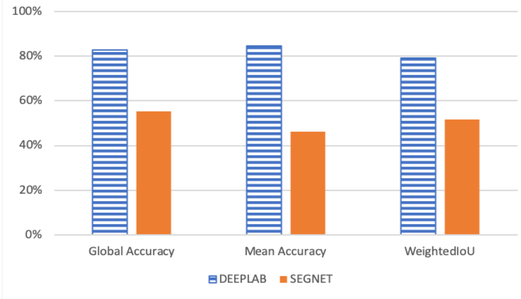

Figure 10: Accuracy comparison.

Table 1: Comparison of accuracy (technique and data setup).

Method	Global Accuracy	Mean Accuracy	weighted IoU
DeppLabV3	88%	82%	80%
Segnet	57%	47%	54%

Table 2: Comparison of IoU (technique and lesion).

Method	Backg	OpticD	SoftEx
DeepLab	84%	71%	15%
Segnet	52%	18%	2%
Method	HardEx	Haemo	MAneu
DeepLab	17%	21%	2%
Segnet	3%	15%	2%

5 CONCLUSIONS AND FUTURE WORK

Segmentation of medical images is a hard task in the presence of difficulties such as lack of contrast, confusion between structures and plasticity of shapes and textures, among others. In this work we compared two DCNN segmentation architectures in the task of segmentation of lesions in Eye Fundus Images (EFI). We defined the two typical DCNN segmentation architectures and used a public dataset to experiment with training and then testing segmentation of lesions on EFI. By analysis of the results we concluded that the best performing approach (DeepLabV3) was competent segmenting lesions, but we also found that there is a low degree of matching of segments to groundtruth regions, which means that current state-of-the-art still needs significant improvement. This was a preliminary study, we propose as future work a more complete evaluation and analysis of the approaches, plus proposal of possible improvements and solutions to the problem.

ACKNOWLEDGMENTS

We use the IDRiD challenge dataset for this work (Porwal and Meriaudeau, 2019). We would like therefore to thank the IDRiD challenge organizers for sharing the dataset and making this work possible.

REFERENCES

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. In *In IEEE transactions on pattern analysis and machine intelligence,*, pages 2481–2495.

BRATS (2014). Brain tumour segmentation challenge. *[URL Accessed 8/2019]. URL: https://sites.google.com/site/miccaibrats2014/*.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs". In *In IEEE transactions on pattern analysis and machine intelligence,*, pages 834–848.

Davy, A. (2014). Brain tumor segmentation with deep neural networks. In *Proceedings of Multimodal Brain Tumour Segmentation Challenge*.

Jaafar, H., Nandi, A., and Al-Nuaimy (2011). Automated detection and grading of hard exudates from retinal fundus images. In *in European Signal Processing Conference,*, pages 66–70.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., and Kirby, J. The multimodal brain tumor image segmentation benchmark (brats)&quot;,. *IEEE Transactions on Medical Imaging*, pages 1993–2024.

Noh, H., Hong, S., and Han, B. (2015). Learning deconvolution network for semantic segmentation". In *In Proceedings of the IEEE international conference on computer vision*, pages 1520–1528.

Oliveira, J. (2012). Estudo e desenvolvimento de tecnicas de processamento de imagem para identificacao de patologias em imagem de fundo do olho. *Biomedical Engineering Thesis, U. Minho*.

Porwal, Prasanna, S. P. R. K. M. K. G. D. V. S. and Meriaudeau, F. (2019). Indian diabetic retinopathy image dataset (idrid). *IEEE Dataport.*

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation". In *In International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer,, Cham.

Song, S., Lichtenberg, S. P., Xiao, J., and "SUN (2015). Rgb-d: A rgb-d scene understanding benchmark suite. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 567–576.

Urban, G., Bendszus, M., Hamprecht, F., and Kleesiek1, J. (2014). Multi-modal brain tumor segmentation using deep convolutional neural networks. In *Proceedings of Multimodal Brain Tumour Segmentation Challenge*.

Wilkinson, C., Ferris III, F. L., Klein, R. E., Lee, P. P., Agardh, C. D., Davis, M., Dills, D., Kampik, A., Pararajasegaram, R., Verdaguer, J. T., et al. (2003). Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*, 110(9):1677–1682.

ZikicD, I. Y., Brown, M., and A., C. (2014). Segmentation of brain tumor tissues with convolutional neural networks. In *Proceedings of Multimodal Brain Tumour Segmentation Challenge*.