T-Net: A Resource-Constrained Tiny Convolutional Neural Network for Medical Image Segmentation

Tariq M. Khan[1] Antonio Robles-Kelly[1] Syed S. Naqvi[2]
[1] School of IT, Faculty of Sci. Eng. & Built Env., Deakin University, Waurn Ponds, VIC 3216, Australia
tariq045@gmail.com, antonio.robles-kelly@deakin.edu.au
[2] Dept. of Electrical and Computer Eng., COMSATS University Islamabad, Islamabad, Pakistan
saud_naqvi@comsats.edu.pk

Abstract

In this paper, we present T-Net, a fully convolutional network particularly well suited for resource constrained and mobile devices, which cannot cater for the computational resources often required by much larger networks. T-NET's design allows for dual-stream information flow both inside as well as outside of the encoder-decoder pair. Here, we use group convolutions to increase the width of the network and, in doing so, learn a larger number of low and intermediate level features. We have also employed skip connections in order to keep spatial information loss to a minimum. T-Net uses a dice loss for pixel-wise classification which alleviates the effect of class imbalance. We have performed experiments with three different applications, retinal vessel segmentation, skin lesion segmentation and digestive tract polyp segmentation. In our experiments, T-Net is quite competitive, outperforming alternatives with two or even three orders of magnitude more trainable parameters.

1. Introduction

The correct segmentation of anatomical structures in the medical image field is an important success factor in diagnosis and eventual treatment [1]. Medical image segmentation can be challenging even to seasoned experts with ample experience [2]. This is due to the structural boundary ambiguity, heterogeneous texture, segmented area uncertainties, intensity inhomogeneity and large contrast variations often found in medical imagery.

Existing methods for the segmentation of medical images can be divided into unsupervised and supervised methods. Unsupervised methods employ low-level features and ad-hoc rules that are manually designed and, therefore, often show poor generalisation properties. Supervised methods use human annotated training images and generally have greater segmentation accuracy than unsupervised approaches. Of these supervised methods, deep learning has been particularly effective, allowing for end-to-end segmentation with high accuracy and better generalisation properties than other alternatives [3].

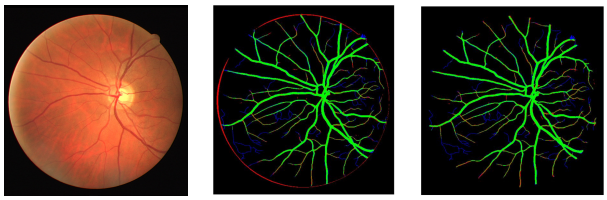

Figure 1: From left-to-right: A sample retinal image from the DRIVE drive data set, the vessel segmentation yielded by U-Net [14] and that delivered by T-Net. Despite being two orders of magnitude smaller, our network provides a margin of improvement.

Thus, many convolutional neural networks (CNNs) have been proposed for medical image segmentation [4–9]. Fully Convolutional Networks (FCN) for semantic segmentation with skip-layers to preserve spatial localization information were proposed in [4]. Inspired by FCNs, U-Net was proposed in [5]. Guo *et al* [10] have proposed a CNN model based upon a reinforcement strategy. To improve segmentation results, Li *et al* [11] proposed a connection-sensitive U-Net. In a related development, a pre-trained CNN model has been used in [12] for retinal image segmentation. Noting that, in [12], the segmentation task can become cumbersome if images are corrupted by noise, Yan *et al* [13] propose the use of segmentation-level as well as a pixel-level joint loss.

Despite these methods generate supervised segmentation results of a quality well beyond their unsupervised counterparts, training and testing these networks can be computationally expensive. This is compounded by the limited amounts of densely annotated data for a wide variety of conditions and imaging modalities. Moreover, in medical

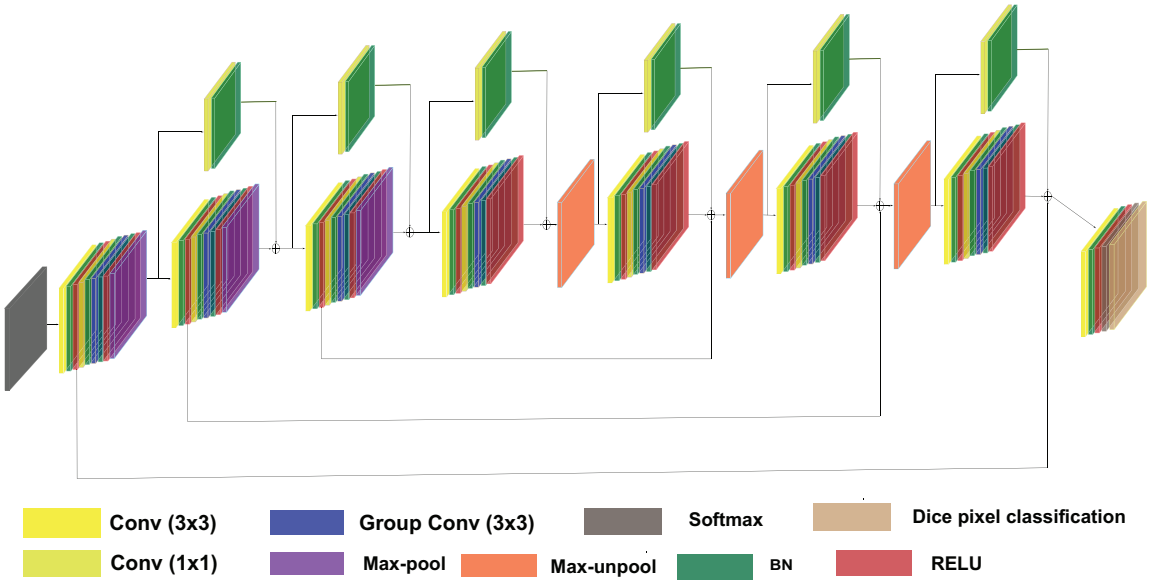

Figure 2: Diagram showing the structure of T-Net.

imaging, usually the image size is large, further increasing computational complexity. As a result, most of the CNN architectures above fail to train on high-resolution data sets on a single mid-range GPU. For such approaches, either multiple GPU's with much larger memory are required or the data set under consideration must be down-sampled before training. This is particularly undesirable in medical imaging, where down-sampling of images is discouraged as critical features may be downgraded to the point of affecting the diagnosis. An alternative, in terms of complexity, is shallow alternatives with a reduced number of layers [15, 16]. In these lightweight networks, the number of filters per layer is often been reduced by search, stemming from hardware needs [15].

This paper presents an effective tiny neural network architecture which we call T-Net. T-Net is a general architecture specifically tailored for resource-constrained environments and mobile platforms with limited memory and computational resources. This is due to the small number of parameters in T-Net, which requires a much smaller memory and GPU footprint for testing as compared to alternatives with much larger number of parameters.

2. T-Net

T-Net aims at preserving boundary information and detail by using the least possible number of pooling layers. This is since these often reduce the dimension of the feature maps and can also cause the loss of spatial information. Further, we have also employed skip connections and, so as to keep the complexity low, we have used a small number of convolutional layers. We have also reduced the overall number of convolutional filters within each layer and used grouped convolutions to further reduce the complexity of the network. Finally, but not least, we have addressed the problem of class imbalance by using a Dice loss in the pixel classification layer. To illustrate the utility of our method for purposes of medical image segmentation, we have performed experiments in three different medical applications. These are retinal vessel segmentation, intestinal polyp detection and skin lesion segmentation.

In Figure 2 we show the structure of T-Net. Note that our network has six convolutional blocks, where the first block of these is the input one, followed by two down-sampling convolutional blocks. There is an intermediate convolutional block which bridges between the down and up-sampling blocks. There are two up-sampling convolutional blocks followed by the final convolutional output block which is equipped with the essential final layers responsible of creating the pixel-wise segmentation map. All the encoder blocks generate the respective collection of features making use of convolutions between the input feature maps and the filter banks. Following [17] we have performed batch normalisation on these features followed by the application of a ReLU. The resulting feature maps are then passed on to the max-pooling or unpooling layers, depending upon whether the block under consideration is a down-sampling or up-sampling one. All max-pooling and unpooling layers have set to be a size of 2×2, non-

overlapping with a stride of size 2. In our network, we have used only two max-pooling and two unpooling layers. All the convolutional filters in our network are of size 3×3.

2.1. Preserving Boundary Information

In our network, we preserve the boundary structure of the foreground regions making use of two different strategies. Firstly, boundary information at the convolutional block level is preserved through residual skip connections comprised of a 1×1 convolutions and a batch normalisation operation. For the structural information preservation, we have also employed identity skip connections between the encoder and the corresponding decoder blocks. Our motivation to use identity skip connections as an alternative to dense skip pathways also stems from the notion that feature preservation within each convolutional block can help bridging the semantic gap between the encoder and decoder while helping to maintain computational overhead under check. Recall that grouped convolution is an important factor in reducing CNN complexity and thus facilitating training of larger neural networks. Here we use grouped convolutional layers for channel-wise separable (also known as depth-wise separable) convolution. The layers are convolved with the input for each group by moving the filters vertically and horizontally along the input, computing the dot product of the weights and the input, and then adding a bias term.

2.2. Dice Coefficient Loss

Note that it is not uncommon in medical image segmentation for the anatomy of interest to occupy only a small region of the scan. This frequently causes the learning process to become trapped in local minima of the loss function, resulting in a network with predictions that are heavily biased towards the background. As a result, the foreground region is frequently missing or detected only partially. In this paper, we employ an L-2 loss on the Dice coefficient so as to attain the advantage of avoiding the need to assign weights to balance the contribution to the loss of pixels arising from different classes to balance foreground and background.

Thus, we use the loss given by:

$$\mathcal{L} = \sum_{I \in \mathcal{I}} (1 - D_I)^2 \tag{1}$$

where D_I is the Dice coefficient for the image I in the dataset $\mathcal{I}$ under consideration.

Recall the Dice coefficient is a scalar ranging from 0 to 1 which can be written as:

$$D_I = \frac{2 \sum_{i \in I} y_i \hat{y}_i}{\sum_{i \in I} {y_i}^2 + \sum_{i \in I} \hat{y}_i^2}$$

where y_j represents the predicted binary segmentation label for to the i^{th} pixel in the image $I \in \mathcal{I}$ and $\hat{y}_i$ denotes the corresponding ground truth.

With these ingredients, its straightforward to compute the gradient of the loss in Equation 1 with respect to the predicted segmentation label, which is given by

$$\nabla_{y_i} \mathcal{L} = 2(1 - D_I) \frac{\partial D_I}{\partial y_i}$$

where, as before, the pixel indexed i belongs to the image I. In the equation above, the partial derivative of the Dice coefficient can be directly computed using the following expression

$$\frac{\partial D_I}{\partial y_i} = \left[\frac{\hat{y}_i \left(\sum_{i \in I} y_i^2 + \sum_{i \in I} \hat{y}_i^2 \right) - 2 y_i \left(\sum_{i \in I} y_i \hat{y}_i \right)}{\left(\sum_{i \in I} y_i^2 + \sum_{i \in I} \hat{y}_i^2 \right)^2} \right]$$

3. Experimental Setup

3.1. Datasets

We have evaluated our network using six publicly available databases across three medical imaging segmentation applications. These are retinal vessel and skin lesion segmentation and digestive tract polyp detection. For the retinal vessel segmentation, we have evaluated our network using four publicly available image data sets: CHASE–DB1 [18][1], DRIVE [19][2], STARE [20][3] and e high-resolution fundus (HRF) [4]. The DRIVE data set originated from a diabetic retinopathy screening program held in the Netherlands. It covers a wide age range of diabetic patients and consists of 20 color images for training and 20 color images for testing that are saved in JPEG format with an image size of 584×565 pixels. Among these 40 images, only seven images have small signs of mild early diabetic retinopathy. A binary field of view (FOV) mask for each image is available. Both training and test images are equipped with manual vessel segmentation as ground truth that has been annotated by experts.

The STARE data set is a collection of 20 color retinal fundus images captured at 35° FOV with an image size of 700×605 pixels. Out of these 20 images, 10 images contain pathologies. Two different manual segmentation maps are available for each of these images. Here we employ the first expert segmentation as ground truth. The CHASE data set consists of 28 color images of 14 school children in England. A 30° FOV centered at optic disc is used to capture each image with an image resolution of 999×960 pixels. Two different manual segmentation maps are available as

[1]The data set can found at `https://blogs.kingston.ac.uk/retinal/chasedb1/`

[2]The data set is widely available at `https://drive.grand-challenge.org/`

[3]More information regarding the STARE project can be found at `https://cecas.clemson.edu/~ahoover/stare/`

[4]The dataset can be found at `https://www5.cs.fau.de/research/data/fundus-images/`

ground truth. Again, here we employ the first experts segmentation for our experiments. The CHASE dataset doesn't contain any dedicated training or testing sets. Here we have used the first 20 images for training and the last 8 images for testing.

HRF consists of 45 high resolution images of 3504×2336 pixels (15 images of glaucomatous patients, 15 images of patients with diabetic retinopathy and 15 images of healthy patients) [21]. For a fair comparison, we constructed the same training set comprising the first 5 images of each subset and tested on all remaining images, as reported in [21, 22].

For the skin lesion segmentation, we have used the PH2 [23] and the ISBI 2016 Skin Lesion Challenge data set [24]. These are public data sets [5]. The ISBI 2016 dataset corresponds to the "Skin Lesion Analysis Towards Melanoma Detection" challenge. This dataset contains 900 images of different sizes. The dataset of PH2 includes 200 dermoscopic images acquired at the Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal.

Finally, for the polyp detection, we have used the CVC-ClinicDB dataset [25] [6]. The CVC-ClinicDB is a database of frames extracted from 29 colorectal colonoscopy videos, all of which have at least one polyp. In addition to the frames, ground truth is provided. The dataset hence contains 612 images in tiff format of size 384×288 pixels with its corresponding label maps.

3.2. Implementation and Training

All our experiments have been effected on an Intel(R) Xeon(R) W-2133 3.6 GHz CPU with 96GB RAM and a GeForce GTX2080TI GPU. Our implementation of T-Net used stochastic gradient descent with a fixed learning rate.

For all our experiments, a weighted cross-entropy loss is used as an objective function for training. This choice stems from the observation that, in vessel segmentation, the "non-vessel" pixels in each retinal image heavily outweigh the "vessel" pixels. For the assignment of the loss weights, different methods can be used. Here, we calculate class association weights by using median frequency balancing [26].

Note that there is no dedicated test set available for STARE or CHASE data sets. For STARE, in the literature, typically a "leave-one-out" approach is used [27]. Here, we have used both "leave-one-out" and a 50%-50% data split, i.e. 10 images for training and 10 for testing. For the CHASE data set we have used the first 20 images for training and the last 8 images for testing.

[5]The data sets are accessible at `https://challenge.kitware.com/#phase/566744dccad3a56fac786787` and `https://www.fc.up.pt/addi/ph2database.html`, respectively

[6]The dataset is available at `https://polyp.grand-challenge.org/CVCClinicDB/`

Also, since the retinal vessel segmentation data sets used here are quite small in nature, we have used data augmentation to generate enough data for training. For the data augmentation, we have used rotation and contrast enhancement. For the rotations, each training image is rotated by 1 degree. The contrast enhancement has been done by randomly increasing and decreasing the image brightness. This yields 7600 images for the DRIVE and CHASE_DB data sets and 7000 images for each of the leave-one-out trails of the STARE data.

Here we have followed the experimental protocols in [28] for the PH2+ISBI 2016 data set, whereby the 900 images from the ISBI 2016 data set are employed for training and, for testing, the 200 images from the PH2 data set are used. For the CVC-ClinicDB data set [25] we have used a similar training and testing strategy as that adopted by [29, 30]. Thus, we have selected 80% images randomly for training, 10% are used for validation and 10% for testing.

3.3. Evaluation Criteria

Recall that vessel segmentation maps are binary, whereby a pixel is marked as corresponding to a vessel or the background. The "ground truth" provided with publicly available datasets are manually marked by expert clinician. Thus, in each image, each pixel is classified into an area of interest (retinal vessels, skin lesions, intestinal polyps, etc.) if present. Note that, for each output image, there are four results: pixels which are correctly predicted as an area of interest (true positive (TP)), pixels which are correctly predicted as non-interesting (true negative (TN)), non-interesting pixels incorrectly predicted as such (false positive (FP)) and, finally, interesting pixels incorrectly predicted as such (false negative (FN). Making use of this ingredients, in the literature, four common parameters (Sensitivity, Specificity, Accuracy and F1) are often used to compare methods with one another.

In our experiments hereafter we denote the accuracy, showing the ratio of the pixels segmented correctly to the total number of pixels in the expertly annotated mask, as Acc. The Se and Sp reflect the sensitivity and specificity that show how the vessel and non-vessel pixels are identified accurately in the model. In our vessel segmentation results we also show the area under the curve (AUC) for the receiver operating characteristic (ROC). We have done this since these data sets have an unbalanced distribution of positives and negatives [31], where the AUC-ROC is often considered to be a good indicator of how the model can separate positive and negative classes in segmentation problems. Also, note that, for segmentation, and viewing it as binary classification task the Sørensen–Dice (Dice) coefficient is equivalent to the F1 score.

Finally, in our polyp segmentation experiments, we fol-

low [29], where the weighted Dice metric F_β^ω is used to amend Dice's "Eual-importance flaw". The mean absolute error (MAE) is used to evaluate the pixel-level accuracy and the S_α the similarity between the predicted segmentation and the ground-truth [32]. We evaluate the pixel-level and global level similarity making use of a recently proposed measure called enhanced-alignment metric ($E_\varphi^{\max}$) [29] [7].

4. Ablation Study

We commence by presenting an ablation study for our network. To do this, we tested different hyper-parameters, the number of filters used in the convolutional layers and the influence of the skip connections. We have done this to assess both their impact on the performance as well as on the number of trainable parameters. Table 1 shows the impact of number of filters across different convolutional layers. As shown in Figure 2, T-Net has 21 convolutional layers (12 in the encoder and 9 in the decoder) distributed across 7 blocks (4 in the encoder and 3 in the decoder). Each of these blocks is comprised by two single and one group convolution.

Also, note that the blocks corresponding to the encoder-decoder pairs have the same number of filters. Following this structure, in Table 1 we show the results yielded by T-Net for different filter configurations when applied to the DRIVE dataset. In the table, each of the variants of our network has been denoted using the triplet $\{b_1, b_2, b_3\}$, where b_i, $i \in \{1, 2, 3\}$ corresponds to the number of filters in each block, whereby the fourth block on the encoder has the same number of filters as the third block of the encoder and the decoder, i.e. b_3 filters. From the table, we can conclude that reducing the number of filters in the convolutional layers does not overly affect the performance but it significantly reduces the overall size of the network. Moreover, the often used filter triplet size of $\{64, 128, 256\}$ yields a number of trainable parameters of more then 2 million. However, if this is reduced to $\{8, 16, 24\}$, then the number of trainable parameters is reduced nearly two orders of magnitude while the overall performance remains quite comparable. Moreover, the removal of the skip connections does have a detrimental effect on the performance. Further, skip connections only add 1186 trainable parameters to the $\{8, 16, 24\}$ configuration.

Following our results, we have employed the $\{8, 16, 24\}$ configuration in all our experiments. We have also tested this configuration with different initial learning rates (ILRs). As shown in Table 2, an ILR of $1e^{-03}$ gives the best results and, hence, we have used this value in all our experiments. Its also worth noting, however, that our network is quite robust to different values of ILR, whereby the variations in performance in the table are quite minor across the range of learning rates we tested.

[7] We have used the implementation available at `https://github.com/DengPingFan/PraNet`. For our evaluation results we used the default parameters in the implementation by the authors.

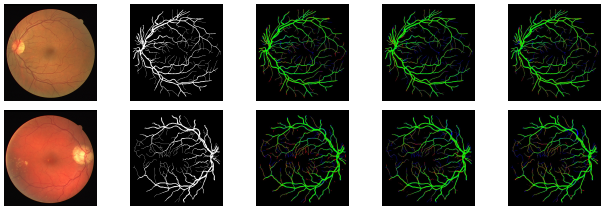

Figure 3: Sample retinal vessel segmentation results on the DRIVE dataset. From left-to-right, we show the input images, the ground truth vessel map manually annotated by an expert and the results yielded by T-Net, SegNet [33] and U-Net [14].

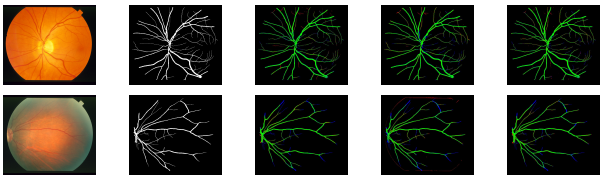

Figure 4: Sample retinal vessel segmentation results on the STARE dataset. From left-to-right, we show the input images, the ground truth vessel map manually annotated by an expert and the results yielded by T-Net, SegNet [33] and U-Net [5].

5. Results and Comparison

5.1. Medical Image Segmentation

We now present results obtained on the five data sets under consideration. We commence by presenting qualitative retinal vessel segmentation results on the DRIVE dataset in Figure 3. In the figure, we show, from left-to-right, the input images, the expertly annotated vessel map (ground truth) and the segmentation results yielded by our method, SegNet [33] and U-Net [14]. In the figure, and in all our qualitative results hereafter, we use green and black for the correctly predicted vessel pixels and red and blue for the false positives and false negatives, respectively. In Table 3, we show the corresponding qualitative results, now also including alternatives such as Image BTS-DSN [37] and VesselNet [38]. Note that, our network delivers a marginal performance improvement over all the alternatives even when its much smaller in terms of its number of parameters. This is consistent with our qualitative results, which show the three networks deliver segmentation maps which are quite comparable.

We now turn our attention to the STARE data set. In Fig-

Filters	Se	Sp	Acc	F1-score	Layers	Parameters
T-Net $\{64, 128, 256\}$	0.8258	0.9834	0.9695	0.8257	81	2228870
T-Net $\{32, 64, 128\}$	0.8284	0.9828	0.9693	0.8250	81	596482
T-Net $\{16, 32, 64\}$	0.8281	0.9825	0.9689	0.8233	81	150278
T-Net $\{8, 16, 32\}$	0.8251	0.9816	0.9678	0.8180	81	35270
T-Net $\{8, 16, 24\}$	**0.8262**	**0.9862**	**0.9697**	**0.8269**	**81**	**25910**
T-Net $\{8, 16, 24\}$ without skip connections	0.8270	0.9822	0.9689	0.8232	66	23886

Table 1: Results yielded by T-Net for different filter configurations when applied to the DRIVE dataset

ure 4 we show sample results as delivered by our network, SegNet [33] and U-Net [14]. We show quantitative results in Table 4. Again, note that our method performs quite competitively against all the alternatives for both, leave-one-out and 50%-50% training strategies while being much more economical in terms of size as compared to all the other methods under consideration. For instance, the second best F1 score after T-Net on leave-one-out is Patch BTS-DSN [37], whose trainable parameters are almost 200 times as many as those in our network. U-Net [14], which has the best accuracy in leave-one-out has over 3.4 million parameters. T-Net comes second with more than 100 times less parameters.

In Figure 5 we show qualitative results on the CHASE data set. In the figure we have followed the same sequence

ILR	Se	Sp	Acc	F1 score
$3e^{-04}$	0.8263	0.9828	0.9690	0.8235
$4e^{-04}$	0.8307	0.9825	0.9691	0.8250
$5e^{-04}$	0.8283	0.9828	0.9693	0.8251
$6e^{-04}$	0.8254	0.9826	0.9689	0.8225
$7e^{-04}$	0.8277	0.9825	0.9689	0.8233
$8e^{-04}$	0.8313	0.9821	0.9688	0.8237
$\mathbf{1}e^{-03}$	**0.8262**	**0.9862**	**0.9697**	**0.8269**
$2e^{-03}$	0.8281	0.9826	0.9690	0.8238

Table 2: Results for different initial learning rates (ILR) on the DRIVE dataset

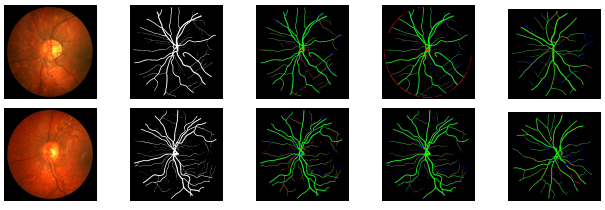

Figure 5: Sample retinal vessel segmentation results on the CHASE dataset. From left-to-right, we show the input images, the ground truth vessel map manually annotated by an expert and the results yielded by T-Net, SegNet [33] and U-Net [14].

of presentation as that in Figures 3 and 4. The qualitative results for the CHASE data set are presented in Table 5. T-Net is the best performer by all measures except for the specificity (Sp), where its outperformed by MS-NFN [39]. Nonetheless, T-Net is still quite competitive regardless of having approximately 14 times less parameters.

The qualitative results of T-Net on the HRF dataset are presented in Figure 6. It is worth noting that, for our experiments, we have used then original image size (3504×2336). This is possible due to the low overhead of T-Net. This contrasts with the common practice of downsampling the images and labels for training and testing by a factor of 4 [21]. It is also noteworthy that T-Net can capture tiny vessel information on high-resolution images. Table 6 summarises the quantitative results of the proposed T-Net with existing state-of-the-art methods. T-Net achieves far better accuracy and specificity with comparable sensitivity as compared to methods using Pixel-wise loss [21], Joint losses [21], the random field in [22] and M2U-Net [42].

In Figure 7 we show sample segmentation maps yielded by our method when applied to skin lesion segmentation using the PH2+ISBI 2016 challenge data set. Note that our method can cope well with a wide variety of lesion size, shape, colour and texture. Table 7 presents the quantitative comparison of our method against other five state-of-the-art segmentation methods. As per the table, our method achieves the overall best performance on both, the F1-score (DICE) and Jaccard coefficient. Moreover, overall, our approach achieves much better performance than all the other methods under consideration in Table 7 despite being, by far, the smaller in size.

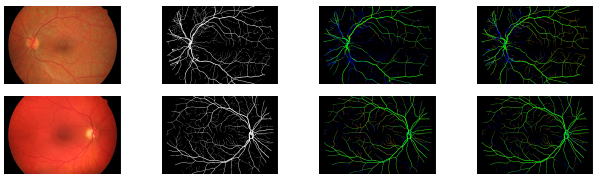

Figure 6: Retinal vessel segmentation results on the HRF dataset. From left-to-right: input images, ground truth, result obtained by M2-UNet [42] and that yielded by T-Net.

Method	Se	Sp	Acc	AUC	F1 score	Params(M)
SegNet [33]	0.7949	0.9738	0.9579	0.9720	0.8182	28.4
Three-stage FCN [34]	0.7631	0.982	0.9538	0.9750	N.A	20.4
VessNet [35]	0.8022	0.9810	0.9655	0.9820	N.A	9
DRIU [36]	0.7855	0.9799	0.9552	0.9793	0.8220	7.8
Image BTS-DSN [37]	0.78	0.9806	0.9551	0.9796	0.8208	7.8
U-Net [14]	0.7849	0.9802	0.9554	0.9761	0.8175	3.4
Vessel-Net [38]	0.8038	0.9802	0.9578	0.9821	N.A	1.7
MS-NFN [39]	0.7844	0.9819	0.9567	0.9807	N.A	0.4
FCN [40]	0.8039	0.9804	0.9576	0.9821	N.A	0.2
T-Net	**0.8262**	**0.9862**	**0.9697**	**0.9867**	**0.8269**	**0.026**

Table 3: Comparison results on the DRIVE dataset

Method	Se	Sp	Acc	AUC	F1 score	Dataset split
DRIU [36]	0.8036	0.9845	0.9658	**0.9970**	0.831	50%/50% (train/test)
Patch BTS-DSN [37]	0.8212	0.9843	0.9674	0.9859	**0.8421**	50%/50% (train/test)
Image BTS-DSN [37]	0.8201	0.9828	0.966	0.9872	0.8362	50%/50% (train/test)
SegNet [33]	0.8118	0.9738	0.9543	0.9728	0.8162	50%/50% (train/test)
T-Net	**0.8249**	**0.9875**	**0.9754**	0.9890	0.8333	50%/50% (train/test)
U-Net [14]	0.764	**0.9867**	0.9637	0.9789	0.8133	leave-one-out
VessNet [35]	**0.8526**	0.9791	0.9697	**0.9883**	N.A	leave-one-out
Three-stage FCN [34]	0.7735	0.9857	0.9638	0.9833	N.A	leave-one-out
T-Net	0.8319	0.9871	**0.9741**	0.9875	**0.8311**	leave-one-out

Table 4: Results on the STARE dataset. Best results are in bold.

Method	Se	Sp	Acc	AUC	F1 score
Three-stage FCN [34]	0.7641	0.9806	0.9607	0.9776	N.A
MS-NFN [39]	0.7538	**0.9847**	0.9637	0.9825	N.A
Vessel-Net [38]	0.8132	0.9814	0.9661	0.9860	N.A
BTS-DSN [37]	0.7888	0.9801	0.9627	0.9840	0.7983
DEU-Net [41]	0.8074	0.9821	0.9661	0.9812	0.8037
SegNet [33]	0.8190	0.9735	0.9638	0.9780	0.7981
U-Net [14]	**0.8355**	0.9698	0.9578	0.9784	0.7792
Proposed	0.8323	0.9844	**0.9739**	**0.9889**	**0.8143**

Table 5: Comparison results on the CHASE dataset. Best results are in bold.

Method	Se	Sp	Acc	Pr
Pixel-wise loss [21],	**0.8084**	0.9417	0.9298	0.5930
Joint losses [21]	0.7881	0.9592	0.9437	0.6647
Orlando [22]	0.7874	0.9584	–	0.6630
Laibacher [42]	–	–	0.9635	–
T-Net	0.8024	**0.9822**	**0.9685**	**0.7936**

Table 6: Comparison results on the HRF dataset. Best results are in bold.

Method	F1 score	Jacc	Params(M)
JCLMM [43]	82.85	-	NA
MSCA [44]	81.57	72.33	NA
SCDRR [45]	86	76	NA
Multistage FCN [28]	90.66	83.99	10
FCN+BPB+SBE [46]	91.84	84.3	8
T-Net	**0.9282**	**0.8696**	**0.026**

Table 7: F1 score and Jaccard coefficient comparison on PH2+ISBI 2016 challenge dataset. Best results are in bold. The first three alternatives are not based upon neural networks and, hence, the number of parameters does not apply (NA)

Finally, we present example intestinal polyp results on the CVC-ClinicDB data set [25] in Figure 8. Table 8 presents the quantitative comparison of our method with five state-of-the-art segmentation methods. Table 8 shows that the performance of T-Net is the best in terms of Mean F1-score and $E_{\varphi}^{\max}$. For F_{β}^{ω} and S_{α} and accuracy (Acc) PraNet [29] performs better while our network comes second best. SFA [48] yields the best mean absolute error (MAE) with our method being again the second. These

Methods	Mean F1 score	S_α	F_β^ω	Acc	$E_\varphi^{\max}$	MAE	Params(M)
U-Net++ [47]	0.794	0.729	0.785	0.873	0.931	0.022	9
SFA [48]	0.701	0.607	0.647	0.793	0.885	**0.042**	19.8
PraNet [29]	0.899	**0.849**	**0.896**	**0.936**	0.979	0.009	32.5
T-Net	**0.931**	0.852	0.901	0.9781	**0.996**	0.010	**0.026**

Table 8: Quantitative results on the CVC-ClinicDB data set. Best results are in bold.

Method	DRIVE			CHASE		
	Dice	Acc	AUC	Dice	Acc	AUC
M2U-Net [42]	0.8091	0.9630	0.9714	0.8006	0.9703	0.9666
ERFNet [49]	0.7652	0.9598	0.9633	0.7872	0.9716	0.9785
MobileNet-V3-Small [15]	0.6575	0.9371	0.9376	0.6837	0.9571	0.9673
T-Net	**0.8269**	**0.9697**	**0.9867**	**0.8143**	**0.9739**	**0.9889**

Table 9: Comparison with recent light-weight networks in terms of quantitative performance on DRIVE and CHASE datasets. Best results are highlighted in bold font.

Model	Params	Size	Platform
MobileNet-V3-Small [15]	2.5M	11.0MB	NVIDIA GTX 1080Ti
ERFNet [49]	2.06M	8.0MB	Tegra TX1
M2U-Net [42]	0.55M	2.2MB	NVIDIA GTX 1080Ti
T-Net	**0.026M**	**0.11MB**	NVIDIA GTX 1080Ti

Table 10: Comparison of computational requirements of recent light-weight networks and T-Net. Lower values are preferred.

results are consistent with those presented throughout the section, where our network is quite competitive and often outperforms the alternatives.

5.2. Comparison with Light-weight Architectures

We now turn our attention to the comparison of T-Net with recent light-weight networks on retinal vessel segmentation. In Table 9, we present the Dice, Acc and AUC delivered by a number of these networks as compared to T-Net when applied to the DRIVE and CHASE datasets.

It can be seen from Table 9 that T-Net outperforms the alternatives in terms of all three quantitative measures as

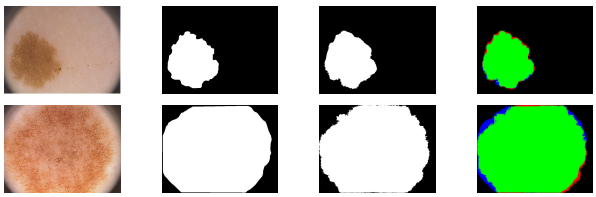

Figure 7: Qualitative results on images from the PH2+ISBI 2016 challenge data set. From left-to-right: input images, ground truth and segmentation map delivered by T-Net and accuracy map showing correctly segmented pixels (black and green), false positives (red) and false negatives (blue).

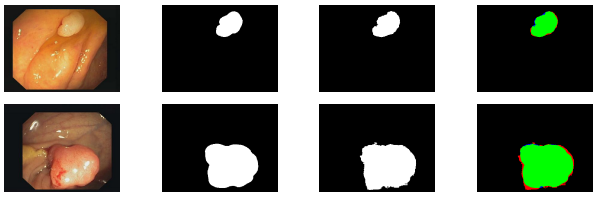

Figure 8: Example results on images from the CVC-ClinicDB data set. From left-to-right: input images, ground truth and segmentation map delivered by T-Net and accuracy map showing correctly segmented pixels (black and green), false positives (red) and false negatives (blue).

compared with the state-of-the-art light-weight networks in the table. This is consistent with the results presented earlier in the section. A potential reason for the low performance of the MobileNet-V3-Small [15] can be attributed to its light-weight segmentation head, which features large pooling kernel and a large stride [15], thus, failing to capture information at finer scales. Moreover, current state-of-the-art networks are not well-suited to high resolution medical images due to their large memory footprint during training. For instance, ERFNet [49] can not be trained on the HRF dataset using our training setup, *i.e.* a GPU with 11GB of RAM. This also applies to M2U-Net [42].

6. Conclusions

In this paper, we have presented T-Net, a convolutional neural network for medical image segmentation which is quite small as compared to alternatives elsewhere in the literature. Our network is particularly well suited for resource constrained and mobile devices, which cannot accommodate much larger networks. Here, we have kept pooling operations to a minimum and integrated skip connections into the network so as to preserve spatial information. Our network also employs a small number of kernels per convolutional layer. We have illustrated the utility of T-Net in retinal vessel segmentation and skin lesion segmentation and intestinal polyp recognition. In our experiments, T-Net is quite competitive, outperforming a number of alternatives that are much larger in terms of trainable parameters.

References

[1] D. Nie, L. Wang, L. Xiang, S. Zhou, and E. Adeli, "Difficulty-aware attention network with confidence learning for medical image segmentation," in *AAAI Conference on Artificial Intelligence*, vol. 33, 07 2019, pp. 1085–1092.

[2] H. Park, H. J. Lee, H. G. Kim, Y. M. Ro, D. Shin, S. R. Lee, S. H. Kim, and M. Kong, "Endometrium segmentation on transvaginal ultrasound image using key-point discriminator," *Medical Physics*, vol. 46, no. 9, pp. 3974–3984, 2019.

[3] Y. Lv, H. Ma, J. Li, and S. Liu, "Attention guided U-Net with atrous convolution for accurate retinal vessels segmentation," *IEEE Access*, vol. 8, pp. 32 826–32 839, 2020.

[4] L. Jonathan, S. Evan, and D. Trevor, "Fully convolutional networks for semantic segmentation," in *IEEE conference on computer vision and pattern recognition*, 2015, p. 3431–3440.

[5] P. Ronneberger, O..and Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[6] Z. Xiaomei, Y. Wu, G. Song, L. Zhenye, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Medical Image Analysis*, vol. 43, 02 2017.

[7] T. M. Khan, A. Robles-Kelly, S. S. Naqvi, and A. Muhammad, "Residual multiscale full convolutional network (rmfcn) for high resolution semantic segmentation of retinal vasculature," in *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops, S+SSPR 2020, Padua, Italy, January 21–22, 2021, Proceedings*. Springer Nature, 2021, p. 324.

[8] R. Imtiaz, T. M. Khan, S. S. Naqvi, M. Arsalan, and S. J. Nawaz, "Screening of glaucoma disease from retinal vessel images using semantic segmentation," *Computers & Electrical Engineering*, vol. 91, p. 107036, 2021.

[9] T. M. Khan, A. Robles-Kelly, and S. S. Naqvi, "A semantically flexible feature fusion network for retinal vessel segmentation," in *International Conference on Neural Information Processing*. Springer, Cham, 2020, pp. 159–167.

[10] Y. Guo, Ü. Budak, L. J. Vespa, E. Khorasani, and A. Şengür, "A retinal vessel detection approach using convolution neural network with reinforcement sample learning strategy," *Measurement*, vol. 125, pp. 586–591, 2018.

[11] R. Li, M. Li, and J. Li, "Connection sensitive attention u-net for accurate retinal vessel segmentation," *arXiv preprint arXiv:1903.05558*, 2019.

[12] Z. Jiang, H. Zhang, Y. Wang, and S. Ko, "Retinal blood vessel segmentation using fully convolutional network with transfer learning," *Computerized Medical Imaging and Graphics*, vol. 68, pp. 1–15, 2018.

[13] Z. Yan, X. Yang, and K. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1912–1923, 2018.

[14] G. Song, "DPN: Detail-preserving network with high resolution representation for efficient segmentation of retinal vessels," 2020.

[15] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[16] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[17] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning - Volume 37*, 2015, pp. 448–456.

[18] M. M. Fraz, S. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, A. R. Rudnicka, and C. G. Owen, "An approach to localize the retinal blood vessels using bit planes and centerline detection," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 600 – 616, 2012c.

[19] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[20] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, 2000.

[21] Z. Yan, X. Yang, and K. T. Cheng, "Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1912–1923, 2018.

[22] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 16–27, 2017.

[23] T. Mendonça, P. Ferreira, J. Marques, A. Marçal, and J. Rozeira, "PH2 - a dermoscopic image database for research and benchmarking," *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, vol. 2013, pp. 5437–5440, 07 2013.

[24] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *2018 IEEE 15th International Symposium on Biomedical Imaging*, 2018, pp. 168–172.

[25] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs.

saliency maps from physicians." *Computerized medical imaging and graphics*, vol. 43, pp. 99–111, jul 2015.

[26] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[27] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1214–1222, 2006.

[28] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, "Dermoscopic image segmentation via multistage fully convolutional networks," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2065–2074, 2017.

[29] D. Fan, G. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Pranet: Parallel reverse attention network for polyp segmentation," in *Medical Image Computing and Computer Assisted Intervention*, 2020, pp. 263–273.

[30] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen, "ResUNet++: An advanced architecture for medical image segmentation," in *2019 IEEE International Symposium on Multimedia (ISM)*, 2019, pp. 225–2255.

[31] Q. Li, B. Feng, L. Xie, P. Liang, H. Zhang, and T. Wang, "A cross-modality learning approach for vessel segmentation in retinal images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 109–118, 2016.

[32] D. Fan, M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *2017 IEEE International Conference on Computer Vision*, 2017, pp. 4558–4567.

[33] K. Tariq M., A. Musaed, A. Khursheed, A. Muhammad, N. Syed S., and N. S. Junaid, "Residual connection based encoder decoder network (RCED-Net) for retinal vessel segmentation," *IEEE Access*, 2020.

[34] Z. Yan, X. Yang, and K. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1427–1436, 2019.

[35] A. Muhammad, O. Muhamamd, M. Tahir, C. Se Woon, and P. Kang Ryoung, "Aiding the diagnosis of diabetic and hypertensive retinopathy using artificial intelligence-based semantic segmentation," *Journal of Clinical Medicine*, vol. 8, no. 9, Sep. 2019.

[36] K.-K. Maninis, J. Pont-Tuset, P. Arbeláez, and L. Van Gool, "Deep retinal image understanding," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells, Eds. Cham: Springer International Publishing, 2016, pp. 140–148.

[37] G. Song, W. Kai, K. Hong, Z. Yujun, G. Yingqi, and L. Tao, "Bts-dsn: Deeply supervised neural network with short connections for retinal vessel segmentation," *International Journal of Medical Informatics*, vol. 126, pp. 105 – 113, 2019.

[38] Y. Wu, Y. Xia, Y. Song, D. Zhang, D. Liu, C. Zhang, and W. Cai, "Vessel-net: Retinal vessel segmentation under multi-path supervision," in *Medical Image Computing and Computer Assisted Intervention*, 2019, pp. 264–272.

[39] Y. Wu, Y. Xia, Y. Song, Y. Zhang, and W. Cai, "Multiscale network followed network model for retinal vessel segmentation," in *Medical Image Computing and Computer Assisted Intervention*, 2018, pp. 119–126.

[40] O. Américo, P. Sérgio, and A. S. Carlos, "Retinal vessel segmentation based on fully convolutional neural networks," *Expert Systems with Applications*, vol. 112, pp. 229 – 242, 2018.

[41] B. Wang, S. Qiu, and H. He, "Dual encoding U-net for retinal vessel segmentation," in *Medical Image Computing and Computer Assisted Intervention*, 2019, pp. 84–92.

[42] T. Laibacher, T. Weyde, and S. Jalali, "M2u-net: Effective and efficient retinal vessel segmentation for real-world applications," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019, Long Beach, CA, USA, June 16-20, 2019*, 2019, pp. 115–124.

[43] R. Anandarup, P. Anabik, and G. Utpal, "JCLMM: A finite mixture model for clustering of circular-linear data and its application to psoriatic plaque segmentation," *Pattern Recognition*, vol. 66, pp. 160 – 173, 2017.

[44] L. Bi, J. Kim, E. Ahn, D. Feng, and M. Fulham, "Automated skin lesion segmentation via image-wise supervised learning and multi-scale superpixel based cellular automata," in *International Symposium on Biomedical Imaging*, 2016, pp. 1059–1062.

[45] B. Bozorgtabar, M. Abedini, and R. Garnavi, "Sparse coding based skin lesion segmentation using dynamic rule-based refinement," in *Machine Learning in Medical Imaging*, 2016, pp. 254–261.

[46] H. J. Lee, J. U. Kim, S. Lee, H. G. Kim, and Y. M. Ro, "Structure boundary preserving segmentation for medical image with ambiguous boundary," in *Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4816–4825.

[47] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018, pp. 3–11.

[48] Y. Fang, C. Chen, Y. Yuan, and K. Tong, "Selective feature aggregation network with area-boundary constraints for polyp segmentation," in *Medical Image Computing and Computer Assisted Intervention*, 2019, pp. 302–310.

[49] E. Romera, J. M. Álvarez, L. M. Bergasa, and R. Arroyo, "Erfnet: Efficient residual factorized convnet for real-time semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 263–272, 2018.