Computers in Biology and Medicine 104 (2019) 111–116

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation

Philipp Tschandl*, Christoph Sinz, Harald Kittler

ViDIR Group, Department of Dermatology, Medical University of Vienna, Vienna, Austria

ARTICLE INFO

keywords:
Segmentation
Classification
Fully convolutional networks
Dermatoscopy

ABSTRACT

Background and objective: Fully convolutional neural networks have been shown to perform well for automated skin lesion segmentation on digital dermatoscopic images. Our concept is that transferring encoder weights from a network trained on a classification task on images of the same domain may contain useful information for segmentation.

Methods: We trained a fully convolutional network where ResNet34 layers are reused as encoding layers of a U-Net style architecture. We entered the encoding layers i) with He uniform ("random") initialization, ii) pre-trained ImageNet weights, or iii) after fine-tuning ResNet34 for skin lesion classification. After transferring the layers to the fully convolutional network architecture we trained for a binary segmentation task using official ISIC 2017 challenge data.

Results: Pretraining of ResNet34-layers with either ImageNet or fine-tuning for skin lesion classification achieved a higher Jaccard than random initialization (0.763 and 0.768 vs 0.740) on the ISIC 2017 test-set. This improved performance warrants further exploration on how to implement cross-task learning for skin lesion segmentation. In additional experiments we found that post-processing with fully connected conditional random fields consistently decreased Jaccard on ISIC 2017 test-set images despite reasonable visual results. Further exploration of the test-set revealed that conditional random field - post-processing decreased segmentation performance only if ground truth annotations consisted of simple shapes but increased it if shapes were complex.

Conclusions: Our findings suggest that domain specific pretraining of encoders can be helpful when there are only few ground truth masks available for segmentation training, but may not be of additional benefit to ImageNet pretraining given enough segmentation training data. Complexity of ground truth annotations have a large impact on segmentation metrics and should be taken into account in skin lesion segmentation research.

1. Introduction

Segmentation of skin lesions in dermatoscopic images has been used as a preprocessing step for feature extraction and automated diagnosis [1,2]. While image processing techniques such as edge detection or Otsu thresholding were used in the past (see comprehensive review of techniques in Ref. [3]), current implementations take advantage of neural network based approaches (see below) or more advanced deformable models [4]. Since the first convolutional layer of convolutional neural networks commonly detects basic shapes and colors [5], deep learning based segmentation may use similar pathways to detect a lesion, without the overhead of explicitly coding filters and selecting them for a task at hand. The implementation of convolutional neural networks as end-to-end solutions made the use of segmentation less important, if not even unnecessary [6–10]. There is, however, an increasing interest in the scientific community for automated segmentation of skin lesions, either as a generic segmentation task or as way to improve classification [11].While there has been some focus on neural network based skin lesion segmentation specifically, most current methods based on deep learning are equally applicable for other segmentation tasks and originate from other fields of research.

Implementations of fully convolutional networks using decoders with transposed convolutions made segmentation problems applicable for end-to-end learning in deep neural networks [12]. The *U-Net* architecture, which was implemented for cell nucleus segmentation and re-uses encoder feature maps on different scales in the decoder section, is commonly used for this purpose [13] because it provides fine-grained segmentation results. More recent advances include the approach by Codella et al. [14], who used an implementation with Dense Blocks for segmenting skin area on clinical images. Iglovikov et al. described a U-

* Corresponding author.
E-mail addresses: philipp.tschandl@meduniwien.at.at (P. Tschandl), christoph.sinz@meduniwien.at.at (C. Sinz), harald.kittler@meduniwien.at.at (H. Kittler).

https://doi.org/10.1016/j.compbiomed.2018.11.010
Received 3 September 2018; Received in revised form 13 November 2018; Accepted 13 November 2018

Net implementation re-using layers of a ResNet34-network pretrained on ImageNet data (*LinkNet34*), an approach winning several segmentation challenges [15,16].Other researchers proposed conditional random fields (CRF) to enhance segmentation results. Gaussian edge potentials were used efficiently for segmentation post-processing [17], and recently CRFs were used successfully in the realm of medicine as part of a neural network architecture for the segmentation of histopathologic images [18].Even with current architectures, however, segmentation masks often still may not look satisfactory with disconnected areas and holes. Mirikharaji and Hamarneh [19] implemented a star shape prior into the loss function specifically for skin lesion segmentation to enhance segmentation results on those problems, and showed promising results on ISIC 2016 challenge images.

In the present study we will combine and expand current segmentation techniques in the following ways:

- Use a classification task for pretraining weights of FCN encoders, with the intuition that the weights will be useful for segmentation of skin lesions
- Expansion of the *LinkNet34* architecture to re-use larger ResNet152 layers, which we correspondingly term *LinkNet152*
- Post-process skin lesion segmentation masks with CRF

2. Materials and methods

2.1. Datasets

2.1.1. Classification

To initialize encoder weights of the segmentation network by pretraining on a classification task, we used the images and labels from the 10015 images, which are part of the HAM10000 dataset [20] that integrates the seven most common and important pigmented skin lesion classes. A randomly shuffled split (stratified for diagnoses) of 20% was used for validation during training, without any further hold-out test-set. Classification training was performed using a crossentropy loss. Since the HAM10000 data is highly imbalanced, with nevi representing more than 50% of the cases, the training process would be prone to over-fit to predicting the nevus class. We therefore weighted the loss function with the inverse class-frequencies in the dataset scaled by the number of available label classes.

2.1.2. Segmentation

To allow comparison of the proposed method to previous work (Table 1), we chose the official ISIC 2017 challenge data for segmentation-training [21]. For enhancing performance in the newer ISIC 2018 challenge benchmark we collected all lesion segmentation data available to us at the time of conducting experiments:

- All publicly available segmentations of the ISIC-archive available either through their Web-API[1] or previous and current challenges [21] (Fig. 1 left and middle). For simplicity, the first listed segmentation was taken for every image regardless of the experience of the annotator. We removed any images also present in the test- and validation sets of the ISIC segmentation challenges.
- Manual segmentations of 1200 images of author PT for the HAM10000 dataset made by manual line-tracings (Fig. 1 right).
- Publicly available segmentations of the PH[2]-dataset [22].

For experiments using all datasets combined, we used a randomly shuffled split of 20% of the ISIC 2018 training data for validation during training, leaving a total of 13355 segmentation masks for training purposes. For experiments using ISIC 2017 data (Table 1) we trained on the official training set (n = 2000), used the official validation set (n = 150) for validation, and performed final evaluation on the official test-set (n = 600).

Table 1
Segmentation performance on the **ISIC 2017 test-set**, with training on **ISIC 2017 training** data only. For comparison, we report baseline comparison of a basic UNet architecture and previous literature on the same dataset.

Network	Encoder Weights	Dice	Jaccard	Jaccard-TS
UNet16	Random	0.798	0.698	0.562
	ImageNet	0.807	0.716	**0.603**
	HAM10000	**0.808**	**0.718**	0.599
LinkNet34	Random	0.829	0.740	0.643
	ImageNet	0.847	0.763	0.677
	HAM10000	**0.851**	**0.768**	**0.686**
LinkNet152	Random	0.825	0.735	0.630
	ImageNet	0.850	0.767	**0.689**
	HAM10000	**0.853**	**0.770**	0.685
UNet [13]	Random	0.778	0.683	0.561
II-FCN [28]		0.794	0.699	
Auto-ED [29]		0.824	0.738	
LIN [30]		0.839	0.753	
Navarro et al. [31]		0.854	0.769	

Bold values denote the highest achieved metric for a single network architecture.

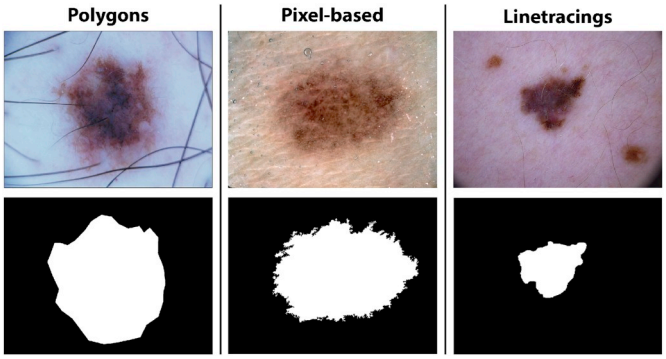

Fig. 1. Examples of training images and their corresponding ground-truth segmentation masks. Cases from the ISIC training sets are either made with polygons (left) or pixel based features (middle), whereas our own HAM10000 segmentation masks (right) are based on line-tracings.

2.2. Network architectures

LinkNet [23] reuses encoder blocks of known networks in a U-Net architecture and enables implementation of different CNN architectures like ResNet34 [15] or WideResNet [16]. Our concept was to pre-train these CNNs (see Fig. 2 left) on a classification task and then transfer corresponding layers as encoders into the LinkNet model (red Blocks in Fig. 2 right). We trained on the *UNet16* and *LinkNet34* architectures [15], which re-use layers from VGG16 and ResNet34, respectively. To expand the idea of the modular LinkNet architecture, we further re-used encoder blocks originating from a ResNet152 architecture, and correspondingly referred to the resulting network as *LinkNet152*. Training was implemented within the PyTorch framework, with CNN network architectures and ImageNet weights taken from torchvision,[2] where "random" weights refer to He uniform initialization [24].

2.3. Training implementation

All training and inference runs were performed on a custom Ubuntu

[1] https://isic-archive.com/api/v1.
[2] https://pytorch.org/docs/stable/torchvision/models.html.

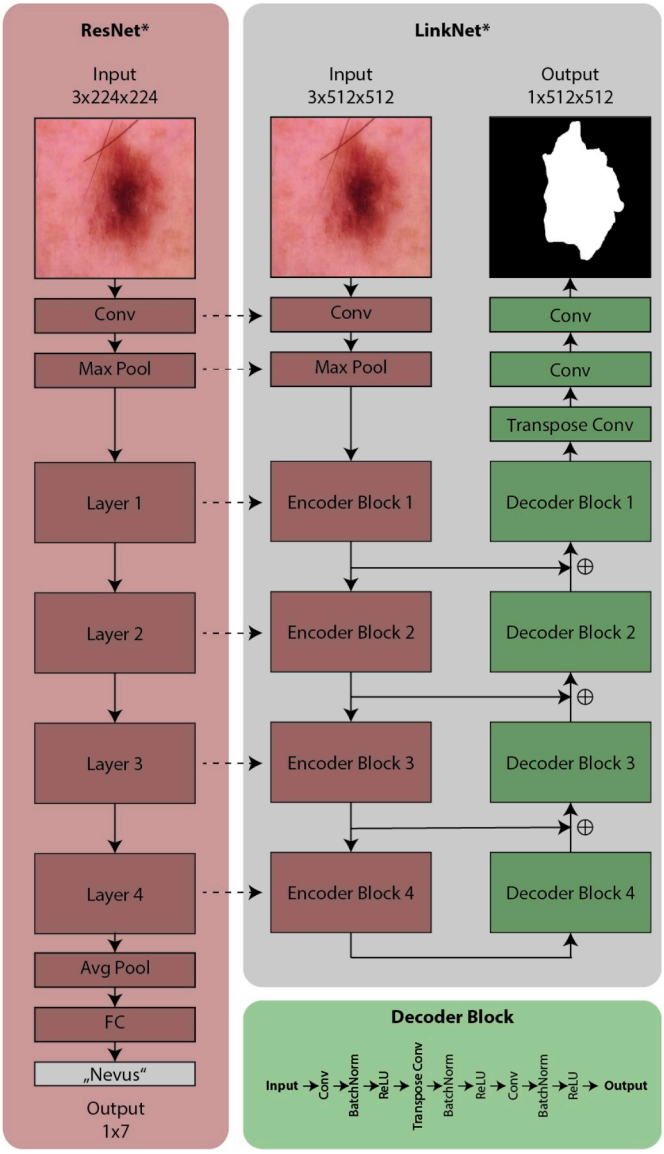

Fig. 2. Schematic architecture of a ResNet (left/red) and the LinkNet (right/gray) architecture adapted from Ref. [23]. The ResNet architecture can be pretrained in its original form for classification, and layers afterwards directly reused within the LinkNet network as Encoders (Blocks in red). This layer transfer is generic and can be applied to any ResNet size. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

workstation with 3 NVIDIA GeForce 1080Ti GPUs and 2 Intel Xeon E5-2637 CPUs. Average training time of a single model with this setup was 27 min for ISIC 2017 data, and 4 h when combining all datasets. We resized all available masks and images to 512×512 pixels before training with nearest neighbour (binary masks) or bilinear interpolation (RGB images) resampling. Common augmentation techniques such as horizontal and vertical mirroring, as well as random rotation of 90° were applied during training, but no normalization through mean pixel subtraction or color constancy [25].

A Jaccard-loss has been shown to be effective in binary skin lesion segmentation tasks [26] previously. As described by Ref. [15], we combined binary cross-entropy and the negative logarithm of the Jaccard index (Equation (1)), which served as loss function during training, where y is the two-dimensional binary ground truth segmentation mask and $\hat{y}$ the sigmoid activations of the pixelwise lesion-probability network output. i denotes a single of n pixels of those masks, and ε is fixed to $1e-15$.

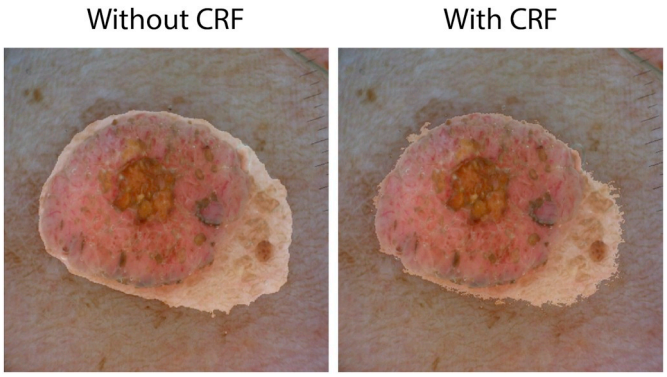

Fig. 3. Alpha overlay of segmentation masks from network output (left) and after post-processing with CRF (right). The positive segmentation pixels more closely adapted to the lesion, but the segmentation mask outline became more jagged.

$$\ell(\hat{y}, y) = -\frac{1}{n} * \sum_{i=1}^{n} [y_i * log(\hat{y}_i) + (1 - y_i) * log(1 - \hat{y}_i)] - log\left(\frac{\varepsilon + \sum_{i=1}^{n} \hat{y}_i y_i}{\varepsilon + \sum_{i=1}^{n} \hat{y}_i + \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \hat{y}_i y_i}\right) \quad (1)$$

We used *AdamW* [27] as an optimizer with an initial learning rate set to $1e-4$, and a learning rate step-down if validation loss did not improve for 2 epochs (gamma 0.1, minimum learning rate $1e-08$). Training was limited to 50 epochs but stopped early if validation loss did not improve over more than 4 epochs. In case of encoders with ImageNet or classification-pretrained weights, the first 6 epochs were run with the encoder weights frozen and the learning rate fixed to $1e-4$.

2.4. Postprocessing

Sigmoid-activated probability maps of lesion-segmentations were thresholded at a cutoff of 0.5. Resulting binary masks were post-processed optionally with fully connected conditional random fields (CRF), which have been reported beneficial for segmentation tasks in general [17] and recently for medical image segmentation in particular [18] (Fig. 3). We reused a previous implementation[3] of CRF as post-processing step which uses pydensecrf[4] as a python wrapper for fully connected CRFs with Gaussian edge potentials by Krähenbühl et al. [17]. Applying CRF post-processing increased the average inference time of a single image from 0.6 to 1 s.

As a last step of post-processing we restricted segmentations to the largest continuous area if more than one was present, also in experiments in which CRF was not applied.

We report *Dice* and *Jaccard* index, as well as a thresholded Jaccard counting 0 if the resulting intersection over union is < 0.65) as *Jaccard-TS* as outcome metrics.

3. Results

3.1. Different weight initialization

Restricting training to ISIC 2017 images and initializing encoders with ImageNet-weights increased segmentation performance in comparison to random initialization of the LinkNet34 architecture (0.763 vs. 0.740), which was further improved through pretraining of encoder layers on the HAM10000 dataset (0.768, Table 1). Repeating training runs for other architectures we saw a similar increase of the Jaccard

[3] https://github.com/milesial/Pytorch-UNet.
[4] https://github.com/lucasb-eyer/pydensecrf.

Table 2
Segmentation performance on the **ISIC 2017 test-set**, with **training on combined datasets**.

Network	Encoder Weights	CRF	Dice	Jaccard	Jaccard-TS
UNet16	Random	✓	0.837	0.752	0.646
		–	0.840	0.752	0.655
	ImageNet	✓	0.855	0.775	0.682
		–	0.872	0.796	0.724
	HAM10000	✓	0.856	0.776	0.682
		–	**0.876**	**0.801**	**0.729**
LinkNet34	Random	✓	0.842	0.758	0.665
		–	0.851	0.768	0.679
	ImageNet	✓	0.875	0.799	0.741
		–	**0.890**	**0.817**	**0.767**
	HAM10000	✓	0.871	0.794	0.738
		–	0.886	0.813	0.766
LinkNet152	Random	✓	0.844	0.762	0.664
		–	0.850	0.765	0.674
	ImageNet	✓	0.872	0.796	0.729
		–	**0.886**	**0.813**	**0.759**
	HAM10000	✓	0.868	0.791	0.726
		–	0.882	0.808	**0.759**

Bold values denote the highest achieved metric for a single network architecture.

metric with HAM10000 pretraining, but generally lower values for *UNet16*, and not consistently higher values for *LinkNet152*.

3.2. Combined datasets

Because CNN architectures usually perform better with more training data, we repeated training of the same architectures with a more diverse dataset as described in *Materials and Methods*. The absolute evaluation metrics on the ISIC 2017 test-set were higher than with ISIC 2017-training data (Table 2). Interestingly, training on a larger segmentation dataset removed the positive effect of HAM10000 encoder pretraining of the LinkNet architectures, in which ImageNet weights performed better (Jaccard 0.817 vs. 0.813 for LinkNet34 and 0.813 vs. 0.808 for LinkNet152).

3.3. Postprocessing with CRF

Post-processing with CRF consistently decreased performance on the ISIC 2017 test-set (Table 2). Comparing raw and CRF-processed segmentation outputs (Fig. 3), CRF provided a more close, but also more jagged outline.

Because of the visual difference and more desirable close lesion boundary in combination with the decreased Jaccard in the ISIC 2017 test-set we tried to get more exploratory insight. Previous reports on ISIC segmentation tasks described differences in annotation techniques and quality, which, according to our hypothesis, may have had an impact on CRF improvement. Through manual inspection of annotation masks we identified two main groups, polygonal shapes (*simple*) and those with jagged outlines supposedly marked up via superpixels or pixel intensities (*complex*) [21]. In order to distinguish between those two groups automatically we measured perimeter length of the ground-truth shape in relation to the perimeter of the shapes' convex hull, where polygons have a value of ~ 1 (e.g. Fig. 1 left: 1.04) and complex shapes >1 (e.g. Fig. 1 middle: 6.54). Resulting values were split at an arbitrary cutoff of 1.5 for further analyses (black line in Fig. 4). We found that CRF improvement increased Jaccard only if ground-truth shapes were complex but decreased if they were simple (Fig. 5).

3.4. ISIC 2018 benchmark

The authors of this manuscript did not have access to the ISIC 2018 segmentation challenge validation- and test-set ground truths during experiments or at the time of manuscript submission, but authors PT and HK were part of the organisational team, which precluded official participation. In order to compare our results with other state-of-the-art approaches, predictions of this model on the unknown test images for task 1 of the ISIC2018 challenge were nevertheless submitted (limited to three entries for the test-set) to obtain performance metrics. Our best model tied position 1, and our second best position 2 on the final test-set (Table 3; numbers given as returned by the challenge site[5]). Though,

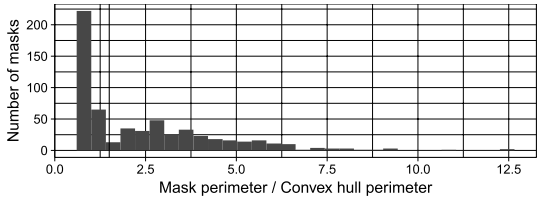

Fig. 4. Histogram of segmentation mask complexities within the ISIC 2017 test-set. Complexity is measured as the relation of mask perimeter length to the perimeter length of its convex hull. For analyses, a value of 1.5 (black vertical line) was used to split *simple* from *complex* masks.

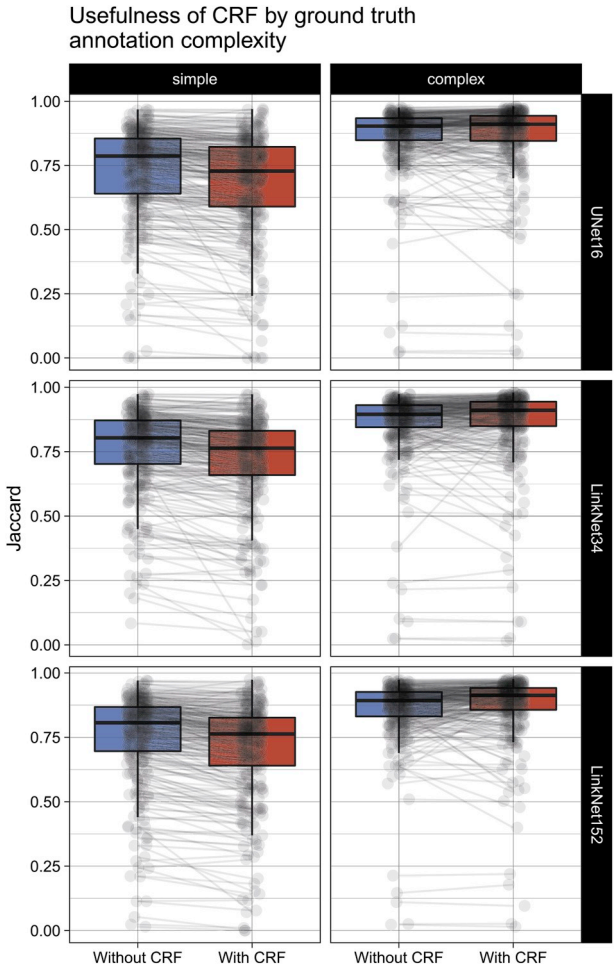

Fig. 5. Jaccard values on HAM10000 pretrained segmentation networks shown in Table 2. Connected points denote segmentation mask prediction of the same image with and without CRF, boxplots their aggregate values. CRF post-processing decreases performance on simple shapes ($p < 0.001$ for all), and increases it on complex ones ($p < 0.05$ for all). p-values were calculated via paired Wilcoxon signed rank test.

[5] https://challenge.kitware.com/#phase/5b1c193356357d41064da2ec.

Table 3
Jaccard-TS values on the ISIC 2018 segmentation sets with models trained on all available datasets combined, including the ISIC 2017 test-set. *Trained without partial unfreezing of encoder layers. **Pretrained encoders frozen for the first 5 epochs.

Network	Encoder Weights	CRF	*Validation*	*Test*
LinkNet152	ImageNet**	–	0.787	–
		✓	**0.823**	**0.799**
	HAM10000*	–	**0.795**	**0.769**
		✓	**0.820**	**0.802**

Rows in bold were models submitted to the ISIC 2018 challenge.

since our submissions used significantly more training data, direct conclusions about comparability cannot be drawn. Because the test-set ground truth is not available at the time of submission we could not perform additional experiments with restricted data.

3.5. Comparison to human annotations

With respect to the high performance of our model we wanted to know if automated segmentation masks of the best performing network (Table 3) are distinguishable from human annotations. To answer this, we compared segmentations of our model with segmentations of dermatologists with multiple years of experience in dermatoscopy (all three authors of this manuscript). Therefore, we performed manual segmentations of 100 lesions via line-tracings (drawing a perimeter line free-hand without any prior) within the ImageJ software. During annotation the authors were blinded with respect to masks of the other annotators. The two co-authors were blinded to the purpose of the segmentations at the time of annotation and asked to *include the full lesion with as little normal skin as possible*. The lesion images were randomly selected out of a separate dataset of dermatoscopic images originating from a MoleMax™ system at the Department of Dermatology at the Medical University of Vienna. We calculated Jaccard values for humans using each other human rater as the ground truth and were able to show that human raters did not achieve significantly higher values than our model (Table 4).

4. Discussion

A diversity of image processing techniques were used for automated skin lesion segmentation in earlier years [3], and used as preprocessing for automated feature extraction [32]. More recent though have shown superiority of neural networks in that task, especially fully convolutional neural networks [12]. Multiple techniques have been described to enhance performance, such as introducing priors to the loss function [19], augmenting training data through other sources [14], or using dense blocks [33]. Iglovikov et al. [16] successfully improved their models by reusing ImageNet pretrained weights within the LinkNet architecture, whereas Li and Ping [18] successfully implemented conditional random fields for segmenting cancer metastases in digital pathologic images. In this study, we show that combining multiple enhancements, namely gathering a large dataset, pretraining of encoders by classification tasks, expansion to the *LinkNet152* architecture, and use of CRFs our model achieves a competitive segmentation performance on the ISIC 2018 test set (Jaccard-TS 0.802 and 0.799). Without post-processing (see Table 1), our approach performed on par with a full resolutional convolutional networks (Jaccard of 0.77) [34], and outperformed a current dense deconvolutional network (Jaccard of 0.765) [33] on ISIC 2017 training- and test-data.

Table 4
Agreement of segmentation, measured as mean Jaccard over 100 lesion segmentations, between humans and a HAM10000-pretrained LinkNet152 network with CRF post-processing. *denotes a statistical significant difference between a human reader and LinkNet152, tested via paired Wilcoxon signed rank test of single lesion segmentations.

	Annotator 1	Annotator 2	Annotator 3	LinkNet152
Annotator 1	1	0.932	0.905*	0.921
Annotator 2	0.932*	1	0.917*	0.930
Annotator 3	0.905	0.917	1	0.909

The increased performance in a segmentation task with domain-near classification pretraining, but lack of a clear effect with more training data in a larger network, warrants further research. That cross-task learning could be beneficial in general is in line with more recent reports on building damage assessment [35]. Although the integration of classification and segmentation is reminiscent of Mask-RCNN [36], segmentation and classification labels may not always be available for the same images. A previous report tackled this problem with the implementation of Mask-X-RCNN [37], which allowed for different types of annotations in training data, but corresponding class labels and segmentations should be present for at least a subset of images. Our proposed method is feasible for different datasets, and may be easier to implement as a type of pretraining for any kind of fully convolutional networks.

Given the results of highly differing segmentation accuracy in complex vs. simple shapes with CRF, we suspect that published performance metrics of networks depend on ground-truth annotation methods. This is especially interesting since many high-performing entries of the ISIC 2018 lesion segmentation challenge used CRFs [38]. Future work in the realm of skin lesion segmentation, especially using the ISIC challenge as benchmarks, will need to take this into account. The current results are only a snapshot of current Fully Connected CRFs. In a follow-up work we want to evaluate how CRFs integrated into a network architecture [18,39] behave for the task of skin lesion segmentation. It will also be interesting to explore if addition of adversarial learning improves medical image segmentation [40] and has an additional impact on performance metrics.

4.1. Limitations

This work, like most neural network - based dermatoscopic segmentation studies, suffers from the restriction of available ground truth data to a limited number of pigmented diagnoses: nevi, melanomas and seborrheic keratoses. Therefore, results of this study can only show performance on lesions that are generally of brown to black color, and have a limited set of patterns. Also, currently available data largely consists of single lesions centered in an image where boundaries are more or less clearly visible. More challenging classes such a non-pigmented tumors, solar lentigines or actinic keratoses surrounded by sundamaged skin are not part of the current sets and should be researched on in future work to estimate generalisability of FCNs for skin lesion segmentation. Notably, previous work has shown good results of image processing in BCCs in propietary datasets [41]. Since there is a critical number of BCCs available within the public HAM10000 dataset, future studies may gain insight how neural network based segmentation compares to these methods in less pigmented classes.

5. Conclusion

We demonstrated that a combination of multiple enhancements to bare fully convolutional networks can produce a state-of-the-art segmentation network that is equivalent to human annotations by line-tracings. Therefore we consider this problem close to being solved - at least for pigmented skin lesions. We are convinced that future work should focus on implementation of automated segmentation in the clinical workflow, rather than incremental improvements on the metric which most probably will have no value in clinical practice.

References

[1] M. Binder, H. Kittler, A. Seeber, A. Steiner, H. Pehamberger, K. Wolff, Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network, Melanoma Res. 8 (3) (Jun 1998) 261–266.

[2] I. Maglogiannis, C.N. Doukas, Overview of advanced computer vision systems for skin lesions characterization, IEEE Trans. Inf. Technol. Biomed. 13 (5) (2009) 721–733.

[3] R.B. Oliveira, M.E. Filho, Z. Ma, J.P. Papa, A.S. Pereira, J.M.R. Tavares, Computational methods for the image segmentation of pigmented skin lesions: a review, Comput. Methods Progr. Biomed. 131 (2016) 127–141 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260716303418.

[4] Z. Ma, J.M.R.S. Tavares, A novel approach to segment skin lesions in dermoscopic images based on a deformable model, IEEE J. Biomed. Health Informat. 20 (2) (2016) 615–623.

[5] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Proceedings of the 25th International Conference on Neural Information Processing Systems, 1 2012, pp. 1097–1105 ser. NIPS'12.

[6] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (02 2017) 115–118.

[7] P. Tschandl, H. Kittler, G. Argenziano, A pretrained neural network shows similar diagnostic accuracy to medical students in categorizing dermatoscopic images after comparable training conditions, Br. J. Dermatol. 177 (3) (2017) 867–869.

[8] A. Menegola, J. Tavares, M. Fornaciali, L. Tzy Li, S. Avila, E. Valle, RECOD Titans at ISIC Challenge 2017, (Mar 2017) ArXiv e-prints.

[9] M.A. Marchetti, N.C.F. Codella, S.W. Dusza, D.A. Gutman, B. Helba, A. Kalloo, N. Mishra, C. Carrera, M.E. Celebi, J.L. DeFazio, N. Jaimes, A.A. Marghoob, E. Quigley, A. Scope, O. Yelamos, A.C. Halpern, Results of the 2016 international skin imaging collaboration international symposium on biomedical imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images, J. Am. Acad. Dermatol. 78 (2) (02 2018) 270–277.

[10] S.S. Han, M.S. Kim, W. Lim, G.H. Park, I. Park, S.E. Chang, Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm, J. Invest. Dermatol. (2018).

[11] J. Kawahara, G. Hamarneh, Fully convolutional networks to detect clinical dermoscopic features, CoRR abs/1703.04559 (2017) [Online]. Available: http://arxiv.org/abs/1703.04559.

[12] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2017) 640–651.

[13] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, CoRR abs/1505.04597 (2015) [Online]. Available: http://arxiv.org/abs/1505.04597.

[14] N.C.F. Codella, D. Anderson, T. Philips, A. Porto, K. Massey, J. Snowdon, R.S. Feris, J.R. Smith, Segmentation of both diseased and healthy skin from clinical photographs in a primary care setting, CoRR abs/1804.05944 (2018) [Online]. Available: http://arxiv.org/abs/1804.05944.

[15] A. Shvets, A. Rakhlin, A.A. Kalinin, V. Iglovikov, Automatic Instrument Segmentation in Robot-assisted Surgery Using Deep Learning, (2018) 1803.01207.

[16] V. Iglovikov, S. Seferbekov, A. Buslaev, A. Shvets, Ternausnetv2: fully convolutional network for instance segmentation, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2018.

[17] P. Krähenbühl, V. Koltun, Efficient inference in fully connected crfs with Gaussian edge potentials, in: J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, vol. 24, Curran Associates, Inc., 2011, pp. 109–117.

[18] Y. Li, W. Ping, Cancer metastasis detection with neural conditional random field, Medical Imaging with Deep Learning, 2018.

[19] Z. Mirikharaji, G. Hamarneh, Star Shape Prior in Fully Convolutional Networks for Skin Lesion Segmentation, (Jun. 2018) ArXiv e-prints.

[20] P. Tschandl, C. Rosendahl, H. Kittler, The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, Sci. Data 5 (2018) 180161.

[21] N.C.F. Codella, D. Gutman, M.E. Celebi, B. Helba, M.A. Marchetti, S.W. Dusza, A. Kalloo, K. Liopyris, N.K. Mishra, H. Kittler, 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), CoRR abs/1710.05006 (2017).

[22] T. Mendona, P.M. Ferreira, J.S. Marques, A.R.S. Marcal, J. Rozeira, Ph2 - a dermoscopic image database for research and benchmarking, 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC), 2013, pp. 5437–5440.

[23] A. Chaurasia, E. Culurciello, Linknet: exploiting encoder representations for efficient semantic segmentation, CoRR abs/1707.03718 (2017) [Online]. Available: http://arxiv.org/abs/1707.03718.

[24] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on imagenet classification, 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1026–1034.

[25] C. Barata, M.E. Celebi, J.S. Marques, Improving dermoscopy image classification using color constancy, IEEE Journal of Biomedical and Health Informatics 19 (3) (2015) 1146–1152.

[26] Y. Yuan, M. Chao, Y. Lo, Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance, IEEE Trans. Med. Imag. 36 (9) (2017) 1876–1886.

[27] I. Loshchilov, F. Hutter, Fixing weight decay regularization in adam, CoRR abs/1711.05101 (2017) [Online]. Available: http://arxiv.org/abs/1711.05101.

[28] H. Wen, II-FCN for skin lesion analysis towards melanoma detection, CoRR abs/1702.08699 (2017) [Online]. Available: http://arxiv.org/abs/1702.08699.

[29] M. Attia, M. Hossny, S. Nahavandi, A. Yazdabadi, Spatially aware melanoma segmentation using hybrid deep learning techniques, CoRR abs/1702.07963 (2017) [Online]. Available: http://arxiv.org/abs/1702.07963.

[30] Y. Li, L. Shen, Skin lesion analysis towards melanoma detection using deep learning network, Sensors 18 (2) (Feb 2018).

[31] F. Navarro, M. Escudero-Vinolo, J. Bescos, Accurate segmentation and registration of skin lesion images to evaluate lesion change, IEEE J Biomed Health Inform (2018).

[32] M. Binder, A. Steiner, M. Schwarz, S. Knollmayer, K. Wolff, H. Pehamberger, Application of an artificial neural network in epiluminescence microscopy pattern analysis of pigmented skin lesions: a pilot study, Br. J. Dermatol. 130 (4) (Apr 1994) 460–465.

[33] H. Li, X. He, F. Zhou, Z. Yu, D. Ni, S. Chen, T. Wang, B. Lei, Dense deconvolutional network for skin lesion segmentation, IEEE Journal of Biomedical and Health Informatics (2018) 1–1.

[34] M.A. Al-masni, M.A. Al-antari, M.-T. Choi, S.-M. Han, T.-S. Kim, Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks, Comput. Methods Progr. Biomed. 162 (2018) 221–231.

[35] K.R. Nia, G. Mori, Building damage assessment using deep learning and ground-level image data, Canadian Conference on Computer and Robot Vision, CRV), 2017.

[36] K. He, G. Gkioxari, P. Dollar, R. Girshick, Mask r-cnn, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018 1–1.

[37] R. Hu, P. Dollr, K. He, T. Darrell, R. Girshick, Learning to segment every thing, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[38] I. S. I. Collaboration, ISIC2018 task 1 leaderboard, [Online]. Available, 2018. https://challenge.kitware.com/#phase/5b1c193356357d41064da2ec.

[39] M.T.T. Teichmann, R. Cipolla, Convolutional crfs for semantic segmentation, CoRR abs/1805.04777 (2018) [Online]. Available: http://arxiv.org/abs/1805.04777.

[40] L. Bi, D. Feng, J. Kim, Dual-path adversarial learning for fully convolutional network (fcn)-based medical image segmentation, Vis. Comput. 34 (6) (2018) 1043–1052.

[41] R. Kaur, R. LeAnder, N.K. Mishra, J.R. Hagerty, R. Kasmi, R.J. Stanley, M.E. Celebi, W.V. Stoecker, Thresholding methods for lesion segmentation of basal cell carcinoma in dermoscopy images, Skin Res. Technol. 23 (3) (2017) 416–428.