PROJECT REPORT

Segmentation of Chronic Wounds

Cay Rahn
6255648

Course: KEN4244 Deep Learning for Image & Video Processing
Academic Year: 2023/24

December 11, 2023

1 Introduction

1.1 Motivation

- many people affected by chronic wounds that need to be monitored
- Why is automatic Wound Segmentation so important? And why is it a complex problem?
- Manual segmentation by experts expensive and very time consuming
- experts differ in their segmentation
- different types of wounds have different characteristics
- changing lighting conditions, distance to camera, camera angle, different cameras have impact on result
- controlled environment not feasible in clinical setting
- ideally, we want to be able to take pictures with a smartphone without overly complicated instructions for the person taking the picture
- experience as photographer should not be required, clinical proffessionals should be able to take pictures that are then segmented correctly

1.2 Research Questions

- can the results be reproduced?
- what influence does the input image size have? Can we rescale the images and are able to transfer what is learned
- how robust is the model/architectures to transformations/distortions on the input
- XAI

2 Dataset

2.1 Available Datasets

- not many medical datasets on chronic wounds publicly available [13]
- often focus on specific type of chronic wounds - often diabetic or pressure ulcer
- Example for such a dataset: data from Diabetic Foot Ulcer Challenge 2022 [8] → unfortunately only available after application, therefore not appropriate for this project with limited timescope
- other dataset of Foot ulcer wounds is available as part of the Foot Ulcer Segmentation Challenge 2021 [16]

2.2 WSNET Dataset

- used dataset consists of 2686 wound images with their corresponding masks introduced by Oota et al. [13].
- 8 different wound types represented in dataset: venous ulcer, trauma wound, diabetic ulcer, surgical wound, arterial ulcer, cellulitis, pressure ulcer and a not further specified group of other wounds
- unfortunately, the wound classification is not available

3 State of the Art

3.1 Semantic Segmentation

- semantic segmentation (pixel-wise classification)
- deep learning methods dominant in the last years (since we have CNNs)
- most techniques for segantic segmantion use encoder-decoder as base for network architecture, inspired by auto-encoders [3].
- encoder: information in feature space /context information [3, 10]
- decoder: information mapped into spatial categorization [3, 10]
- encoder subsampling, decoder upsampling [11]
- "encoder network weights typically pre-trained on the large ImageNet object classification dataset" [2]
- in case of wound segmentation we have two classes: foreground, the wound and background

4 different segmentation models used in this project: Unet, Linknet, FPN and PSPNet, all fully convolutional models.

Fully convolutional Neural Networks

- deconvolution to upsample high-dimensional features maps [11]
- retain spatial information
- in base form: loss of information

U-Net U-Net is a convolutional network developed for Biomedical Image Segmentation [14].

- encoder/decoder (in paper called contracting and expansive path)
- encoder (contracting path) is CNN: repeated application of two 3x3 convolutions (unpadded), followed by ReLU and 2x2 max pooling operation (stride 2) for downsampling $\rightarrow$ number of features is doubled

- expansive path: repeated step consisting of an upsampling of the feature map followed by 2x2 convolution (halving number of feature channels), concatenation with the correspondingly cropped feature map from contracting path and two 3 x 3 convolutions
- each step followed by ReLU
- cropping neccessary due to loss of border pixels in every convolution
- final layer: 1x1 convolution to map 64 component feature vector to disired number of classes
- in total 23 convolutional layers
- input-tile size must be chosen s.t. all 2x2 max-pooling operations are applied to a layer with an even x and y size
- large niput tiles favored
- energy function: pixel-wised soft-max over final feature map combined with cross entropy loss function
- data augmentation for robustness witch few training samples
- "U-net contains a context path to learn context information and a spatial path to preserve spatial information" [10]
- encoder is usually backbone [10]
- skip connections
- its ability to generalize to multi-scale information is limited [11]

Linknet

- encoder: initial block (convolutiuon with 7x7 kernel and stride 2, spatial max-pooling 3x3 with stride 2), later blocks (convolution 3x3 with stride 2, conv 3x3, TODO) [3]
- decoder: earlier blocks (TODO), last block (full conv with 3x3 kernel with upsampling of factor 2, convolution 3x3, full-conv (2x2 with upsampling of factor 2)
- implementation used has 4 skip connections instead of the original 4 [7]
- "LinkNet sends spatial information directly from the encoder to the matching decoder, conserving as much of the image's spatial information as feasible."
- "directly connects shallow feature map in encoder module to the decoder module of the corresponding size" → accurate position information on shallow layer, avoids redundant parameters and computations [11]

FPN

- long for Feature Pyramid Network
- creating feature maps of various layers and sizes [11]
- bottom-up pathway: "feature hierarchy consisting of feature maps at several scales with scaling step of 2"[9]

- top-down-pthway and lateral connections": "higher resolution features by upsampling spatially coarser, but semantically stronger, fea- ture maps from higher pyramid levels", enhance with features from bottom-up pathway via lateral connections $\rightarrow$ merge feature maps of same spatial size from both pathways[9]
- bottom-up feature map: lower-level sementatics but activations more accurately localized (fewer subsampling)[9]
- creation of top-down feature maps: upsample and then merge[9]
- final feature contains local and global context information[9]

PSPNet

- long for Pyramid Scene Parsing Network
- feature map extracted with pretrained backbone
- pyramid pooling to get context information
- pyramid pooling: fusion of features under four different pyramid scales (global pooling and sub-regions for different locations), 1x1 convolution to maintain weight of global feature after each pyramid, upsampling of output to get same size as original feature map
- "different levels of features concatenated as final pyramid pooling feature"
- final prediction by convolution layer which input isoriginal feature map concatenated with pyramid pooling output
- motivation: pyramid pooling provides levels of information, more helpful than global pooling

3.1.1 Evaluation

There exist several methods to evaluate how good a predicted segmentation is. Since semantic segmentation performs a pixel-wise classification, resulting in a segmentation mask, classical metrics such as accuracy and precision are available. Two performance metrics that are commonly used in semantic segmentation in medical imaging are the Dice Coefficient and the Intersetcion over Union (IoU) score. They indicate the segmentation quality better than pixel-wise accuracy [5].

- loss function often uses pixel-wise (weighted) cross-entropy loss even though differentiable approximations of the two metrics exist [5]

IOUScore The IoU-Score (Intersection over Union), also known as the Jaccard index J describes the ratio between the intersection of the ground truth mask y and the predicted mask $\tilde{y}$ and the union of the predicted and the ground truth mask. By this it compares the similarity of the two masks [4].

$$\mathrm{IoU}(y, \tilde{y} := \frac{\text{Area of overlap}}{\text{Area of union}} \tag{1}$$

$$= \frac{|y \cap \tilde{y}|}{|y \cup \tilde{y}|} \tag{2}$$

Dice Coefficient The Dice coefficient is the F1 score calclulated for the image masks. In terms of intersection and union, this means it calculates the ratio between two times the overlap between ground truth y and predicted mask $\tilde{y}$ and the total area.

$$\mathrm{Dice}(y, \tilde{y}) := 2 \cdot \frac{\text{Area of overlap}}{\text{Total area}} \tag{3}$$

$$= 2 \cdot \frac{|y \cap \tilde{y}|}{|y| + |\tilde{y}|} \tag{4}$$

To gain more insight into the type of the errors the model makes, the rate of false positives and false negatives can be used do differentiate Type I and Type II errors [8].

3.2 Wound Segmentation

- one type: diabetic foot ulcers → are monitored to ensure healing process is optimal and there is no infection, normally long time span [8]

- wounds have complex structure containing different types of tissue with different colour and texture → different regions with borders in between [1]
- heterogeneous wound images

- before deep networks: features describing color and texture, algorithms such as region growing and optimal thresholding or classical machine learning models, e.g. Support Vector Machines [15]
- Convolutional Neural Networks then used, manually extracted features replaced by the ones the CNN learns autonomously [15]
- some methods include pre-processing steps to remove background (User interaction, manual feature engineering to detect background pixels, standardizing background in advance before taking feature) → not automatic
- Diabetic Foot Ulcer Challenge 2022 used FCN, U-Net and SegNet with different backbones as baseline for their challenge (categorical cross-entropy loss) [8]
- generally often classical models used with minor adaptions
- following standing out

- Scebba et al. proposed two step method: object detector that produces bounding boxes containing the wounds and then segmentation on those areas (u-net, convNet, DeepLapV3 with ResNet-101 backbone and FCN with VGG16 backbone, pixel-wise weighted binary cross entropy loss, weighting term was computed as the ratio between the total number of wound bed and background pixels of each training set fold)
- Oota et al. claim they set a new state of the art, their method will described in more detail in the following

3.2.1 WSNET

- based on the four before described segmentation architectures: U-Net, LinkNet, PSPNet and FPN
- experimented with different backbones, in the scope of this project MobileNet [6] is used since it is the smallest one and allows faster training
- all backbones with ImageNet pre-trained weights
- they performed Wound-Domain Adaptive Pretraining by classifying the wound images in 5 ulcer types
- data augmentation on the training data and corresponding masks, not on test data
- augmentation consists of horizontal flip, random rotation, optical distortion, grid distortion, blur, random brightness contrast, and transpose

Global-Local Architecture

- motivation: obtain global signals from entire image and local signals from smaller patches for details
- only local might cause incomplete segmentation for large wounds
- local architecture: split image in 16 non-overlapping patches (48x48x3), stacking results in 48x48x(3x16) volume
- parallel 16 local models with shared weights
- combined to full-size mask at end
- stack output of global and local model to output of size (192x192x2)
- 1x1 convolution to get final mask

- interesting that they use segmentation models that already use methods to localize method, e.g. FPN already considers different context sizes

Reported Results

- pretraining on wound images improves results
- data augmentation leads to improvements
- local only models significantly worse than global model

4 WSNET

4.1 Code availability and reproduction of the results

Although the code for WSNET [13] is stated to be publicly available, a closer inspection of the linked GitHub repository shows, that this is only partially the case. A lack of documentations makes it hard to make use of the code, especially since the code seems to contain some errors.

In the scope of this project, the code was used to create runable models again. Unfortunately, the classes of the wounds are not available, which makes it impossible to perform pre-training as it was described in the original paper [13]. In total there are eight models available: A local model and a combined global-local model for each of the segmentation models Unet, PSPNet, FPN, and Linknet. The Python library used for the segmentation models is `segmentation_models` [7]. The implementation processed showed some differences to the described model architecture. In particular, it was claimed that the wound images were split up in parts of 48 px times 48 px. However, some of the used models only allow input sizes that are divisable by 32 and the GitHub showed a size of 64 px was used.

Information about the size of training, validation and test set is not given in the paper or code. In the scope of this project, a split of 70 % training, 15 % validation and 15 % test data is used.

	Unet		Linknet		PSPNet		FPN	
	IoUe	Dice	IoU	Dice	IoU	Dice	IoU	Dice
Global-Local model	0.739	0.856	0.752	0.854	0.729	0.847	0.761	0.861
Local model	0.546	0.831	0.540	0.831	0.536	0.805	0.557	0.817

Table 1: IoUe-Scores and Dice Coefficients for the four different models with each Global-Local and Local architecture. The backbone used is mobilenet.

5 Technical Information

5.1 Prior Experience

I have a strong programming background, consisting of a B.Sc. in Computer Science and three years of work experience in Web Developement with Python. Beside the content of the course Advanced Concepts of Machine Learning, I have no prior experience with Deep Learning.

5.2 Code and Data Availability

The code produced in the scope of the project is available on GitHub: `https://github.com/Zianor/DLIV-chronic-wound-segmentation`. Package versions are included to ensure reproducability.

The used data is available on GitHub as well: `https://github.com/subbareddy248/WSNET/` [12, 13]. Availability on a later point of time cannot be guaranteed.

5.3 Used Hardware

All computations are performed on a MacBook Air (24 GB RAM, Apple M2 Chip with an 8-core GPU). The package versions for GPU-utilization on MacOS are included in the package versions on GitHub.

References

[1] Mohammad Faizal Ahmad Fauzi et al. "Computerized segmentation and measurement of chronic wound images". In: *Computers in Biology and Medicine* 60 (2015), pp. 74–85. ISSN: 0010-4825. DOI: `https://doi.org/10.1016/j.compbiomed.2015.02.015`.

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.12 (2017), pp. 2481–2495. DOI: `10.1109/TPAMI.2016.2644615`.

[3] Abhishek Chaurasia and Eugenio Culurciello. "LinkNet: Exploiting encoder representations for efficient semantic segmentation". In: *2017 IEEE Visual Communications and Image Processing (VCIP)*. 2017, pp. 1–4. DOI: `10.1109/VCIP.2017.8305148`.

[4] Yeong-Jun Cho. "Weighted Intersection over Union (wIoU): A New Evaluation Metric for Image Segmentation". In: *ArXiv* abs/2107.09858 (2021).

[5] Tom Eelbode et al. "Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index". In: *IEEE Transactions on Medical Imaging* 39.11 (2020), pp. 3679–3690. DOI: `10.1109/TMI.2020.3002417`.

[6] Andrew G. Howard et al. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. 2017. arXiv: `1704.04861 [cs.CV]`.

[7] Pavel Iakubovskii. *Segmentation Models*. `https://github.com/qubvel/segmentation_models`. 2019.

[8] Connah Kendrick et al. *Translating Clinical Delineation of Diabetic Foot Ulcers into Machine Interpretable Segmentation*. 2022. arXiv: `2204.11618 [eess.IV]`.

[9] Tsung-Yi Lin et al. "Feature Pyramid Networks for Object Detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017.

[10] Yujian Mo et al. "Review the state-of-the-art technologies of semantic segmentation based on deep learning". In: *Neurocomputing* 493 (2022), pp. 626–646. ISSN: 0925-2312. DOI: `https://doi.org/10.1016/j.neucom.2022.01.005`.

[11] Abderrahim Norelyaqine, Rida Azmi, and Abderrahim Saadane. "Architecture of deep convolutional encoder-decoder networks for building footprint semantic segmentation". In: *Scientific Programming* 2023, 8552624 (2023). DOI: https://doi.org/10.1155/2023/8552624.

[12] Subba Reddy Oota et al. "HealTech - A System for Predicting Patient Hospitalization Risk and Wound Progression in Old Patients". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Jan. 2021, pp. 2463–2472.

[13] Subba Reddy Oota et al. "WSNet: Towards an Effective Method for Wound Image Segmentation". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Jan. 2023, pp. 3234–3243.

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: (May 2015).

[15] Gaetano Scebba et al. "Detect-and-segment: A deep learning approach to automate wound image segmentation". In: *Informatics in Medicine Unlocked* 29 (2022), p. 100884. ISSN: 2352-9148. DOI: https://doi.org/10.1016/j.imu.2022.100884.

[16] Chuanbo Wang et al. *Fully automatic wound segmentation with deep convolutional neural networks*. en. Dec. 2020. DOI: 10.1038/s41598-020-78799-w.