Interactive Neural Network Robot User Investigation for Medical Image Segmentation

Mario Amrehn[1], Maddalena Strumia[2], Markus Kowarschik[2], Andreas Maier[1,3]

[1]Pattern Recognition Lab, Friedrich-Alexander University
Erlangen-Nürnberg (FAU), Germany
[2]Siemens Healthcare GmbH, Forchheim, Germany
[3]Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany
mario.amrehn@fau.de

Abstract. Interactive image segmentation bears the advantage of correctional updates to the current segmentation mask when compared to fully automated systems. Especially in the field of inter-operative medical image processing of a single patient, where high accuracies are an uncompromisable necessity, a human operator guiding a system towards an optimal segmentation result is a time-efficient constellation benefiting the patient. There are recent categories of neural networks which can incorporate human-computer interaction (HCI) data as additional input for segmentation. In this work, we simulate this HCI data during training with state-of-the-art user models, also called robot users, which aim to act similar to real users given interactive image segmentation tasks. We analyze the influence of chosen robot users, which mimic different types of users and scribble patterns, on the segmentation quality. We conclude that networks trained with robot users with the most spread out seeding patterns generalize well during inference with other robot users.

1 Introduction

The trans-catheter arterial chemoembolization (TACE) [1] is a minimally invasive procedure to treat hepatocellular carcinoma (HCC). During the treatment, volumetric cone-beam C-arm computed tomography (CBCT) [2] images of the patient's abdomen are generated. The physician maximizes the efficacy of the operation selecting all cancerous cells while reducing the toxicity of the treatment by omitting surrounding healthy tissue during lesion segmentation. Therefore, a crucial step during the intervention is the accurate segmentation of liver lesions in order to precisely isolate the conspicuous tissue's cells from the oxygen supply of the liver.

In recent years, fully-automatic segmentation systems based on convolutional neural networks (CNN) like the U-net [3] outperformed more traditional learning based approaches to medical image segmentation. In 2017, interactive CNNs were published [4,5] which, to some degree, include guidance from a human user for their final segmentation result. The guidance is provided by post-processing the current segmentation result. In that year, Amrehn et al. [6] and Wang et al. [7]

demonstrated the potential of rule-based seed drawing robot users and feasibility of a combination of interaction input data with traditionally fully-automatic CNN segmentation systems.

All of these systems model the user in a specific way via a set of fixed rules. Kohli et al. [8] described a way to realistically simulate some groups of users. However, most often, the similarity analysis of a simulated user to actual humans interacting with the system is omitted when a new interactive method with its custom interacting robot user are presented. In this work, we quantify the similarity between proposed robot users and illustrate their differences.

2 Materials and Methods

The network topology used is a fully convolutional neural network based on U-net [3] with $3.12 \cdot 10^7$ trainable parameters as depicted in **Fig.** 1. The proposed network utilizes three input channels, with size of 256^2 pixels each, to encode gray-valued image data as well as user provided seed information. Convolution operations are performed utilizing $3 \times 3 \times n$ kernels, where $n \in \{2^6, 2^7, 2^8, 2^9, 2^{10}\}$ depending on the depth of the network. A 2×2 neighborhood is used for pooling. Three input channels encode the gray-valued C-arm CT image data as well as user provided seed information. The seeding channels consist of background respective foreground seeds transformed by the Euclidean distance function. A distance transform as a pre-processing step on the sparse seed images decreases the necessary size of the network's minimum receptive field, which is especially important for its initial layers to capture the seed information as context to the gray-valued input image [9]. Utilizing a distance transform, the seed formation is spread over the whole input channel and seed information is preserved even with small kernel sizes.

The robot user mimics the interaction of a real user. It is assumed, that a human user sets additional seed points during segmentation based on the structures seen on the gray-valued input image, the previously set foreground and background seeds, the current segmentation mask image, as well as a notion of the segmentation ground truth which the physician has from their domain

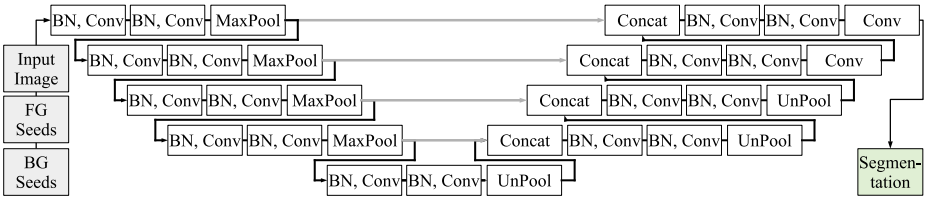

Fig. 1. Schematic representation of a U-net convolutional neural network topology. The input channels include foreground (FG) and background (BG) seed information. Skip-connections are depicted as links in gray. Before each convolution, batch normalization (BN) is applied. The outcome is a dense segmentation mask of size 256^2 pixels (green).

knowledge. These five inputs are also commonly used for a rule-based robot user, as depicted in **Fig.** 2. In the following analysis, five different robot users are evaluated.

Random Sampling Over Whole Image (rand) Seeds are placed at random on the seed input channels. Here, a fraction of $r_{rand} = 0.1$ of seeds are drawn with the label inverted i. e. these seeds are misplaced.

Random Sampling From GT (rand_gt) This robot user samples seed point positions at random and copies labels from the ground truth. Note that *rand_gt* equals *rand* with $r_{rand} = 0.0$. Here, the number of seeds per interaction is $n_{rand\_gt} \in \{1, 5, 10\}$.

Robot User by Kohli et al. (kohli12) Proposed in [8] and selected for user simulation in [5], this robot user utilizes the current segmentation image and the ground truth in order to place one seed point in the center of the largest, wrongly labelled image area.

Robot User by Xu et al. (xu16) The robot user proposed in [4] samples $f_{xu16} \in \{1, 5, 10\}$ foreground and $b_{xu16} \in \{1, 5, 10\}$ background seed points at random constrained by a minimum distance to established seeds (such that $2 \leq n_{xu16} \leq 20$). Possible background seed locations are either sampled inside a 20 pixel wide margin around the GT object's contour line (called *strategy 1* in the original paper), or in the entire background region (*strategy 2*), depending on parameter $s_{xu16} \in \{1, 2\}$.

Robot User by Wang et al. (wang17) In [7,6] the robot user utilized places seed points at random on wrongly labeled image areas. This behaviour is similar to *kohli12*, but not limited to the center of the image areas. Whether a region is ignored during placement of additional seed points is determined by an area size threshold $t_{wang17} \in \{10, 20, 30, 40\}$ in pixels.

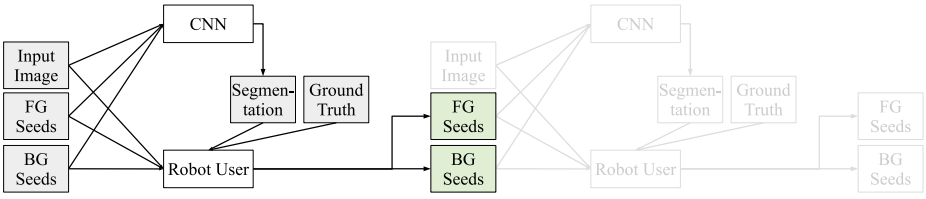

Fig. 2. A robot user bases its seed placement decision process on up to five different inputs (gray): the gray-valued input image, the previous foreground and background seeds, the current segmentation mask, and the ground truth segmentation mask. The outcome of a robot user system is a new set of proposed seed points (green).

When training a new network with robot user interaction input, a classical chicken or the egg causality paradox emerges. A fully trained network would be needed in order to segment the input image. Thereafter, additional correcting seed points may be selected by a robot user, which leads to an updated segmentation result. This interaction data may be used for training the new network. However, a fully trained network would need exactly these steps to be trained first. Therefore, in this work, we initialize the new network with interaction training data acquired by a non-learning-based method. Here, robot user interactions are recorded via iterative segmentation utilizing GrowCut [10]. In preliminary experiments, we determined that segmentation methods like GrabCut, which are more robust and therefore more independent of user input patterns do not qualify for the proposed initialization of a new network. The GrowCut method is chosen due to its well known tendency to benefit from careful placement of large quantities of seed points. The figure of merit for segmentation quality is a Dice score, also known as intersection over union (IoU), generated after each GrowCut iteration step, as depicted in **Fig.** 3.

3 Experiments

The data utilized in the experiments are 2-D slices of volumetric CBCT images of liver lesions depicting HCC. The lesions in the volumetric images are fully annotated by medical experts. Subsequently, the image data are cropped to a volume of interest (VOI), with voxel resolutions from $0.46^3\,mm^3$ to $0.68^3\,mm^3$. All annotated lesions are smaller than $117^3\,mm^3$ which allows for a (VOI) of 256^3 voxels depicting the largest lesion outlines. For training and testing, 90 slice images are drawn from the 38 3-D VOI images. 90 % of images are used for training, 10 % for testing.

One network $\mathbb{M}_i$ is trained for every robot user and every parameter configuration tested for a robot user as described in **Sec.** 2, where $i \in [0, 27)$. The quality of their segmentation outcome is analyzed via the Dice score for the current segmentation mask with the ground truth. It is analyzed, which robot user input patterns during training will generate networks able to generalize to other input patterns during inference.

4 Results

For the evaluation, 27 CNN models were trained with seeding data from one of the 27 robot user configurations each. The Dice scores for the test set are depicted in **Fig.** 4. Each of the 27 models $m(x)$ are trained only on robot user x's seeding training data. A mean Dice score is computed for each of the 27 trained segmentation models $m(.)$ after segmenting the 9 test images with seed input data from one of the 27 robot users.

5 Discussion and Outlook

It becomes apparent from **Fig.** 4b), that (1) CNNs trained with robot users based on rules to place seeds almost at random (*rand*, *rand_gt*, *xu16*) yield similar segmentation results when other user input patterns are utilized during inference. (2) Robot user input with more distinct seeding patterns like *wang17* generates trained networks which are better adjusted to their seeding (see **Fig.** 4a) *wang17*), but not generalizing well to other input patterns.

An interpretation of this result is, that when improving on randomized seeds for training, it is not feasible to train on generalized user input patterns for all use cases, due to (1). Therefore, it is a necessity to train on personalized seeding patterns formalized as individual robot users (2), where a high similarity to the input patterns of the real user operating the system is imperative.

References

1. Lewandowski RJ, Geschwind JF, Liapi E, Salem R. Transcatheter intraarterial therapies: rationale and overview. Radiology. 2011;259(3):641–657.
2. Strobel N, Meissner O, Boese J, Brunner T, Heigl B, Hoheisel M, et al. 3D imaging with flat-detector C-arm systems. Multislice CT. 2009; p. 33–51.
3. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015; p. 234–241.

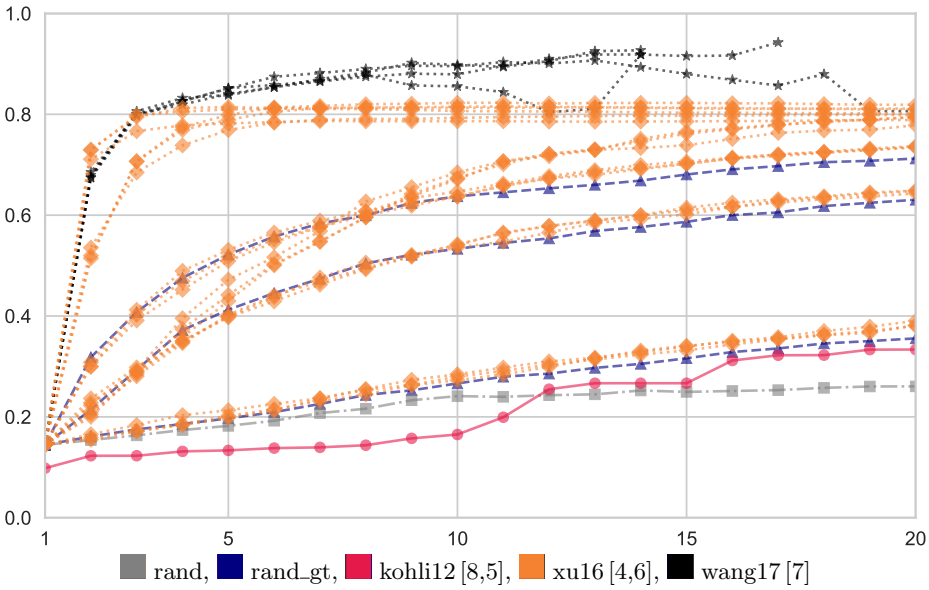

Fig. 3. The mean Dice scores per robot user over all input images and per interaction is depicted. Each robot user provides seeds during interactive segmentation. A segmentation's quality is measured as Dice score after each GrowCut [10] iteration step.

4. Xu N, Price B, Cohen S, Yang J, Huang TS. Deep interactive object selection. Computer Vision and Pattern Recognition (CVPR). 2016; p. 373–381.
5. Liew JH, Wei Y, Xiong W, Ong SH, Feng J. Regional interactive image segmentation networks. Computer Vision (ICCV). 2017; p. 2746–2754.
6. Amrehn MP, Gaube S, Unberath M, Schebesch F, Horz T, Strumia M, et al. UI-Net: Interactive Artificial Neural Networks for Iterative Image Segmentation Based on a User Model. Visual Computing for Biology and Medicine (VCBM). 2017; p. 143–147.
7. Wang G, Zuluaga MA, Li W, Pratt R, Patel PA, Aertsen M, et al. DeepIGeoS: a deep interactive geodesic framework for medical image segmentation. Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 2018;.
8. Kohli P, Nickisch H, Rother C, Rhemann C. User-centric learning and evaluation of interactive segmentation systems. Computer Vision (IJCV). 2012;100(3):261–274.
9. Meyer MI, Galdran A, Mendonça AM, Campilho A; Springer. A Pixel-Wise Distance Regression Approach for Joint Retinal Optical Disc and Fovea Detection. Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2018; p. 39–47.
10. Vezhnevets V, Konouchine V. GrowCut: Interactive multi-label ND image segmentation by cellular automata. Computer Graphics and Applications (Graphicon). 2005; p. 150–156.

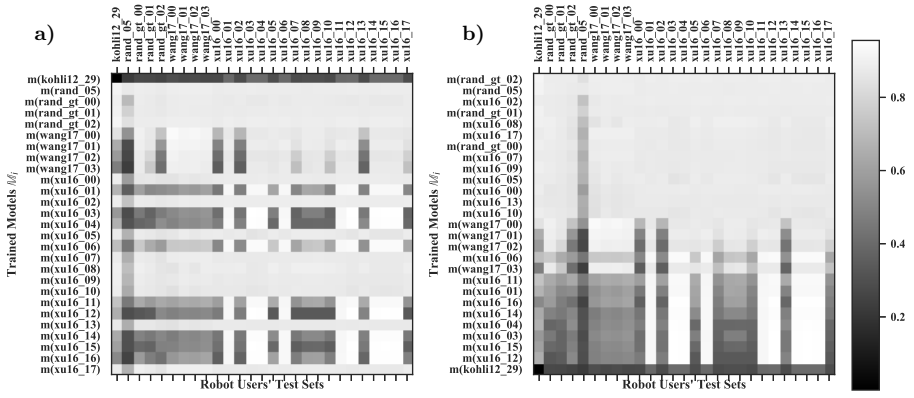

Fig. 4. (a) Each of the 27×27 cells represents the segmentation quality in Dice score given a trained segmentation model $m(.)$ (row) and a robot user's (column) seed input data for the test set. Each model $m(x)$ was trained beforehand only on robot user x's seeding training data. In (b) the rows are sorted by sum of Dice scores per row descending.