A Weakly Supervised Approach for the Segmentation of Skin Lesions in Chronic Graft Versus Host Disease

VANDERBILT INSTITUTE FOR SURGERY AND ENGINEERING

Tahsin Reasat[1], Kelsey Parks[2,3], Xiaoqi Liu[1], Jianing Wang[1], Eric R. Tkaczyk[2,3], Benoit M. Dawant[1]
(1) Dept. of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, US
(2) Dermatology Service and Research Service, Tennessee Valley Healthcare System, Department of Veterans Affairs, Nashville, TN, US
(3) Vanderbilt Dermatology Translational Research Clinic, Department of Dermatology, Vanderbilt University Medical Center, Nashville, TN, US

ABSTRACT

Chronic graft-versus-host-disease (cGVHD) is a common occurrence after hematopoietic stem cell transplantation (HCT) and the primary affected organ is skin. The extent of the disease is measured in terms of affected body surface area (BSA). An automatic image segmentation algorithm would allow a convenient assessment of the BSA in clinics where there is a scarcity of expert dermatologists. However, the current deep learning algorithms used for segmentation require a large amount of annotated data. It is difficult to collect pixel-level annotations from experts because it is time-consuming, and there is inter-rater variability in the case of pixel-level annotations. However, a patch level annotation scheme would significantly reduce annotation time. Hence, we propose a weakly supervised segmentation (WSS) approach which can perform image segmentation using the patch level information. The solution involves training a convolutional neural network classifier and extracting the class activation maps (CAM) of the trained network using Gradient-weighted Class Activation Mapping (Grad-CAM)[1]. We test our algorithm on images of skin cGVHD and show that the class activation maps can focus on the affected body surface area.

DATA

- 91 cross-polarized 3D images were acquired from the back and abdomen of 19 cGVHD patients. Each image was annotated by two annotators (only one annotation was used in this experiment).
- The 3D images (Fig. 1a) were captured and annotated using a Vectra H1 camera (inset) and associated software. 2D images (Fig. 1b) and the corresponding ground truth (Fig. 1c) were extracted. The images were color normalized under a grey world[2] assumption (Fig. 1d). Patches were taken from both the image and ground truth (Fig. 1e) and label creation rules were applied. Here, **pp** = positive pixel percentage, and **th** = threshold.
- The variations in the visual appearance (hyperpigmentation in Fig. 2a and erythema in Fig. 2b). The variations create confusion among annotators. In Fig. 3, two different annotators (green and blue) are demarcating different regions for the same image.

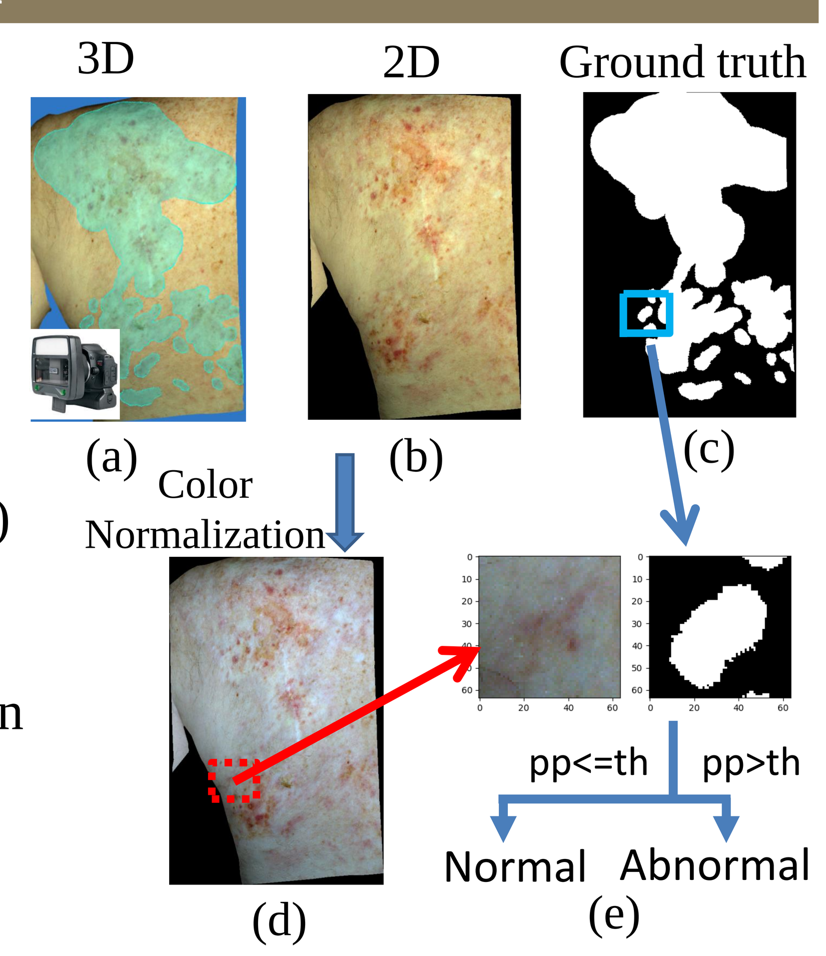

Fig. 1. Data processing steps.

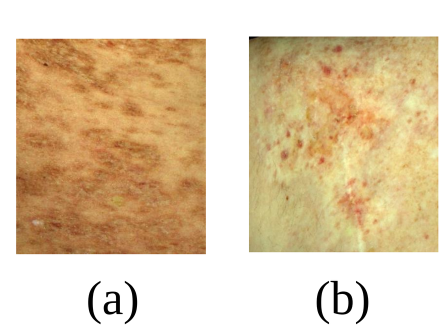

Fig. 2. Variations of abnormalities

Fig. 3. Disagreement between annotators

METHOD

The detailed structure of the VGG11[3] classifier network is shown in Fig. 4a. During training, random image patches are extracted and the patch level label is used to train a classification network (Fig. 4a). During testing, patches with a 50% overlap were used. For each patch, the activation map (CAM) is computed by backpropagating the signal from the abnormal class to the last convolution block (Fig. 4b).

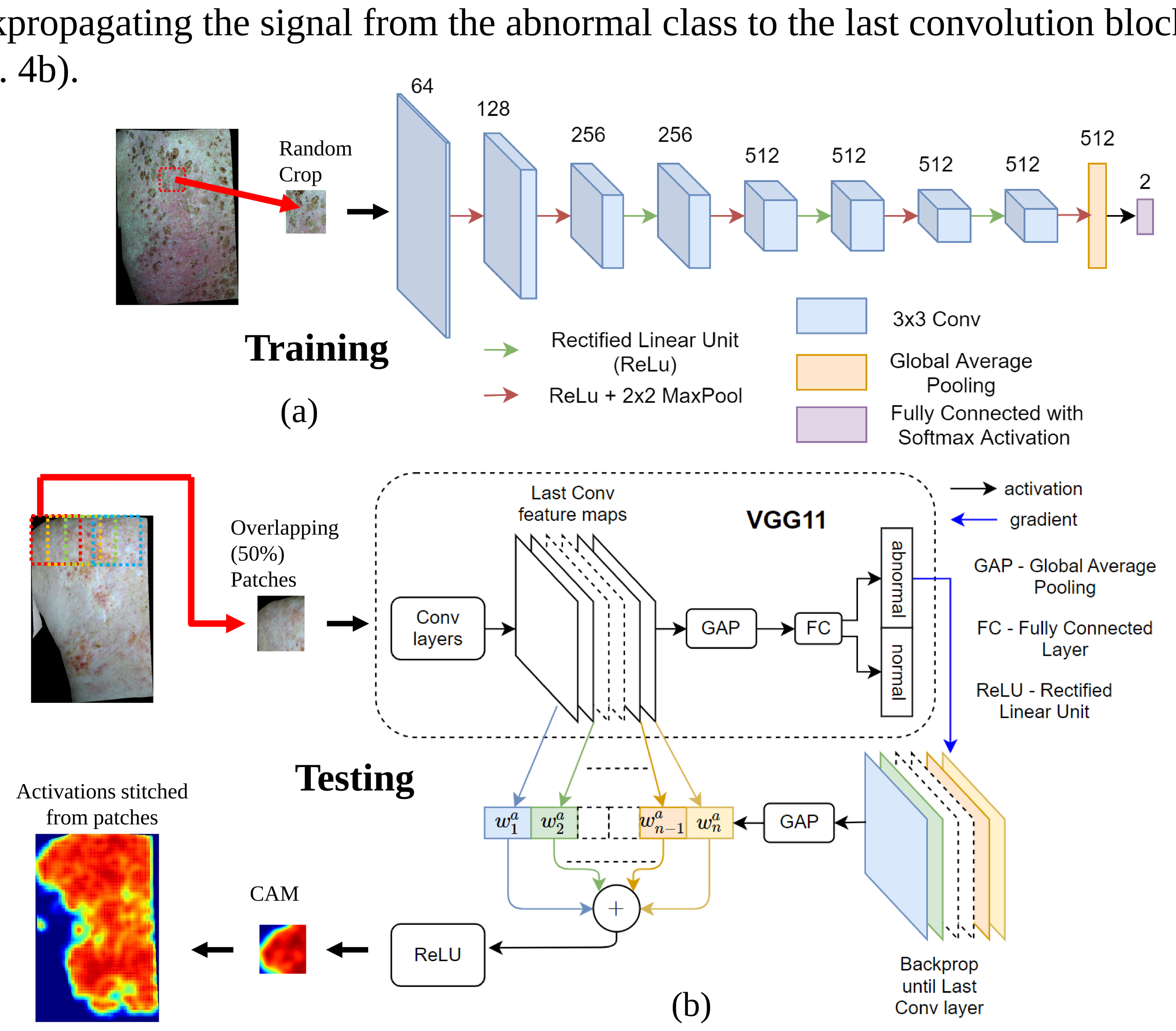

Fig. 4. The proposed training and testing method.

EXPERIMENTAL SETUP

Training Patients: 16, 63 images => hyperpigmentation only=13, erythema only=38, both hyperpigmentation and erythema=12,
Testing Patients: 3, 28 images => hyperpigmentation only=12, erythema only=16
Augmentation: Rotation [0, 360°] and Contrast Shift [0.6, 1.2]
Classifier Loss: Binary Crossentropy = $-\{y\log(p) + (1-y)\log(1-p)\}$, $y \in \{0, 1\}$ is the patch level ground truth, p is the estimated probability)
Segmentation Metric: Dice = $\frac{X \cap Y}{|X|+|Y|}$,
X is the estimated binary segmentation mask, Y is the ground truth mask.

- For each activation map, 0.5 of the maximum value was used as the threshold to compute the binary segmentation mask.

RESULTS

- The model gives the best dice (0.62) for a patch size of 64 and a threshold of 0.2. The network performs relatively well (dice 0.73) in case of erythema (Fig. 5a) when compared to hyperpigmentation (dice 0.47).
- Some images have dark areas in the peripheral region which look like hyperpigmentation, e.g., the upper shoulder area in Fig. 5b. In Fig. 5c, the neck area has sun-damaged skin that resembles hyperpigmentation which is picked up by the network.
- The human inter-rater agreement between the ground truth and annotations from a second annotator is 0.76.

Effect of input patch size

Patch Size	Mean Dice
64	0.62
128	0.61
256	0.47
512	0.34

Effect of threshold (th)

Threshold (th)	Mean Dice
0.1	0.61
0.2	0.62
0.3	0.59
0.4	0.56
0.5	0.56

Examples of Segmentation

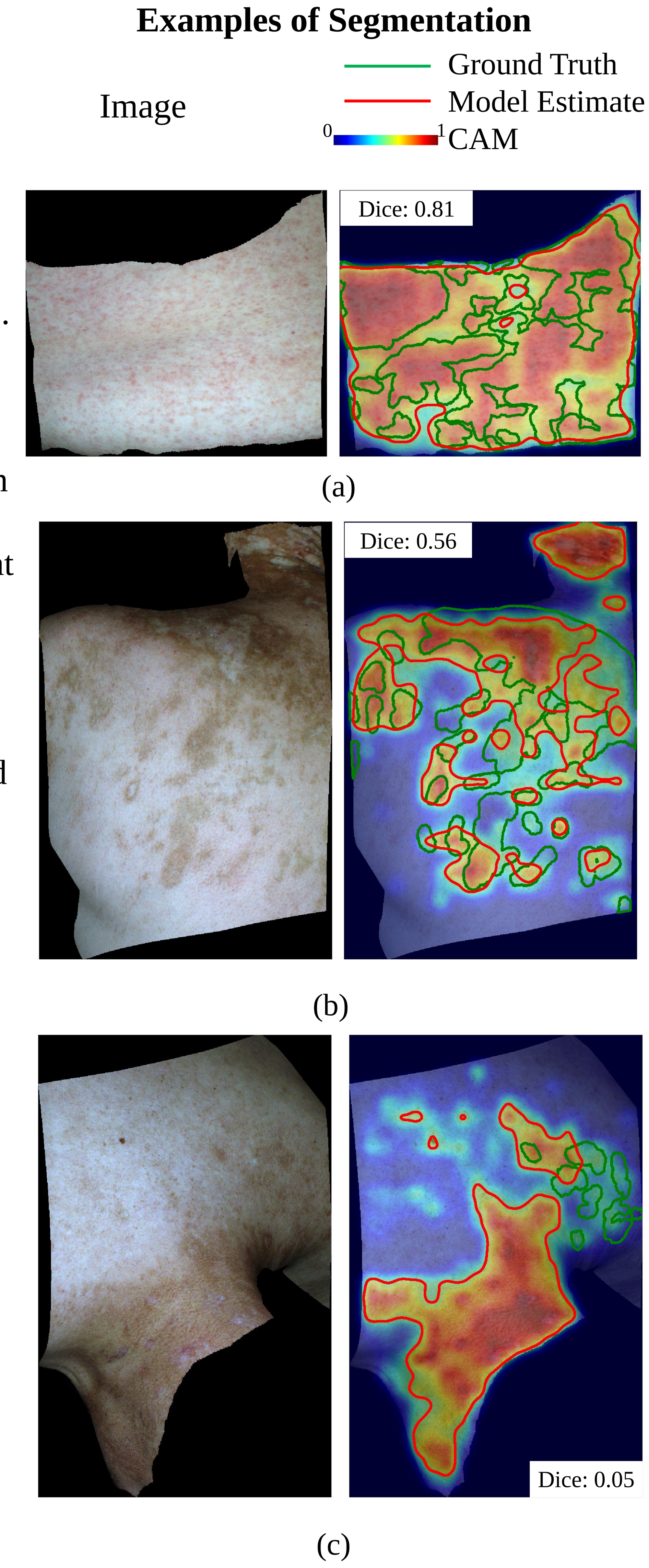

Fig 5. Examples of good, average, and bad segmentation are shown.

CONCLUSION

Although the human-machine agreement is less than the human inter-rater agreement, the results suggest that the extracted class activations maps can focus on areas of abnormalities on the skin. Hence, the patch level activation masks could be used as a recommendation system for the clinicians. Also for future annotations, the activation outputs could be used as a suggestion to the annotator which could potentially decrease inter-rater variability seen in the pixel level annotations.

Reference

1. Selvaraju, Ramprasaath et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618-626. 2017.
2. Buchsbaum, Gershon. "A spatial processor model for object colour perception." *Journal of the Franklin institute* 310, no. 1 (1980): 1-26.
3. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).