ESANN 2022 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 5-7 October 2022, i6doc.com publ., ISBN 978287587084-1. Available from http://www.i6doc.com/en/.

A weakly supervised approach to skin lesion segmentation

Simone Bonechi[1,2]

1 - Department of Social, Political and Cognitive Sciences,
University of Siena, Siena, Italy
2 - Department of Information Engineering and Mathematics,
University of Siena, Siena, Italy

Abstract. Early detection of skin cancers greatly increases patients' chances of recovery. To support dermatologists in this diagnosis, many decision support systems based on Convolutional Neural Networks have recently been proposed to segment the lesion and classify it. The use of the information coming from the segmentation, as an additional input to the classifier, proved to be fundamental to increase its performance and, in fact, the shape of the lesion is of diagnostic importance unanimously recognized by clinicians. However, in the ISIC database, the public reference dataset that collects a huge number of skin lesion images, all samples are labeled for classification but only a very small fraction of them are also labeled for segmentation. To overcome this limitation, the present paper proposes a weakly supervised approach to extract the segmentation label maps of approximately 43,000 ISIC images, used to train a segmentation network, with very promising performance.

1 Introduction

Recently, deep learning methods, and in particular Convolutional Neural Networks (CNNs), have had a huge impact in numerous application fields, from image classification [1, 2] and semantic segmentation [3, 4] to object detection [5, 6]. CNNs are also becoming very popular in medical image analysis [7] and many decision support systems have been developed, for example, for automatic reporting of medical exams [8] and analysis of retinal fundus images [9]. Moreover, CNNs were also successfully employed in skin lesion analysis to classify and segment nevi and melanomas [10, 11, 12, 13, 14]. Melanoma is an aggressive form of cancer, triggered by an uncontrolled proliferation of melanocytes, pigment–producing cells of neuroectodermal origin. Cutaneous melanoma is the 20th most common cancer worldwide and occurs most frequently in adults, aged between 40 and 60, while it is rarely observed before puberty [15]. Although cutaneous melanoma comprises less than 5% of all skin tumor cases, it causes the majority (75%) of deaths. Employing CNNs can allow to speed–up the diagnosis of an early melanoma and drastically increase the positive outcome of the disease. To train state–of–the–art CNN models for this task, a large set of supervised data is needed, available thanks to the International Skin Imaging Collaboration (ISIC) Archive [16]. The ISIC database, which contains more than 60,000 images, labelled for classification, is widely used for lesion analysis. Unfortunately, only a small fraction of the ISIC images was also labelled at the pixel–level, since the high cost of image tagging by medical experts made the collection of large sets of segmentation label maps very difficult. However, lesion segmentation also proved very useful in improving the classification accuracy [13, 14], as the lesion boundaries are also used by dermatologists for their diagnosis. To this end, a weakly–supervised approach, inspired by [17], was employed to automatically extract segmentation supervisions for approximately 43,000 ISIC images. However, unlike the original method, which applied to images equipped with a bounding–box supervision — not available in ISIC —, a detection network (YOLO [6]) was previously used to extract the bounding–box corresponding to each lesion. Then, a segmentation network was trained to recognize the background and foreground (lesion) within the bounding–box. The rationale behind this approach is that a segmentation network focused on a bounding–box performs a simpler task than using the entire image. Finally, the segmentation supervision is generated using the network output on the bounding–box. To consider the fact that the lesion boundaries are blurred and not easy to delimit precisely, an uncertain region has been defined, which often falls on the boundaries, according to the network output probability. The generated supervisions were first compared with the label maps available for the small image subset of the ISIC dataset to assess their quality. Additionally, the supervisions have been also employed to train a segmentation network to demonstrate that, with their use, the lesion segmentation accuracy can be increased. The document is organized as follows. In Section 2, the weakly supervised method is described together with the ISIC dataset; then, in Section 3, the experimental setup and the results are presented. Finally, Section 4 collects conclusions and future perspectives.

2 Materials and Methods

The ISIC dataset, which is used in our experiments, will be described in Section 2.1. Instead, the weakly–supervised method employed to automatically extract the segmentation supervision for about 43,000 images of the ISIC database is presented in Section 2.2.

2.1 ISIC Archive

The International Skin Imaging Collaboration (ISIC) [16] is a joint academia and industry project designed to facilitate the application of digital skin imaging to help reduce melanoma mortality. In this context, an open source archive of skin images has been collected to promote and ease the development of automated diagnostic systems. The database is updated regularly to include new images from various sources and, at the moment, it contains over 60,000 images of skin lesions along with different metadata (diagnosis, clinical attributes, image type, etc.). In addition, in 2018, a segmentation challenge on ISIC was lunched [18, 19], for which the segmentation supervisions of 2694 images (2594 for training and 100 for validation) were released. The ground truth segmentation masks for these images were generated using several techniques (manual, fully or semi–automatic), but all data were reviewed and curated by practicing dermatologists with expertise in dermoscopy. The segmentation challenge was not reopened in 2020 and, therefore, no further segmentation masks were released.

2.2 Weakly–Supervised Method

Inspired by the approach proposed in [17], which was also applied in scene text segmentation [20, 21], we developed a weakly–supervised pipeline to automatically extract the segmentation supervisions for the ISIC dataset. Since the ISIC archive does not provide the bounding–box supervision for each lesion, an additional step was added to automatically extract bounding–boxes using a detection network (YOLOv3 [6]). The overall pipeline of our approach is shown in Figure 1. Specifically, the YOLO network was trained based on the bounding–boxes extracted from the ground truth images of the ISIC 2018 challenge. The trained YOLO network was then used to extract the bounding–boxes from the entire ISIC database. These bounding–boxes were then selected using the following criteria. Only one bounding–box per image is considered; in case of multiple bounding–boxes, the one with the highest network output probability was selected. Additionally, only bounding–boxes with a network output probability grater that 0.5 were selected, discarding only those that actually have a very low probability to contain a lesion. In this way, bounding–boxes for a subset of images of the ISIC dataset were obtained. In the next phase of our pipeline, a background/foreground network was trained to extract the most relevant object (the lesion) from the bounding–box. The Segmentation Multiscale Attention Network (SMANet) [21], with a ResNet50 backbone pretrained on ADE20k dataset, was trained to this aim using the 2018 ISIC segmentation challenge images cropped around the lesions. The SMANet learned to extract the pixels belonging to the lesion from image crops containing it. Subsequently, the bounding–boxes predicted with YOLO were exploited to extract the image crops containing the lesions. These crops are then used as input for the SMANet to obtain the segmentation of each lesion within the bounding–box. The SMANet outputs on image crops were finally post–processed in order to create the segmentation of the entire image. In particular, all the image pixels external to the bounding–boxes were labelled as background; instead, the pixels inside the bounding–box were labelled according to the SMANet output probability map in the following way[1]:

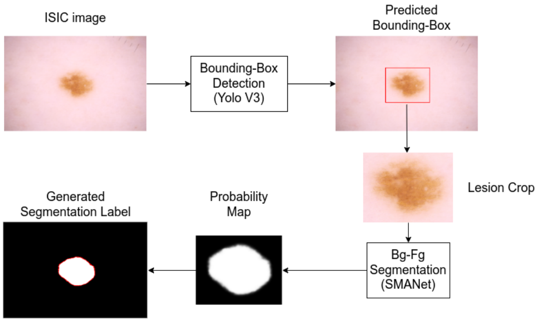
ISIC image
Bounding-Box Detection (Yolo V3)
Predicted Bounding-Box
Lesion Crop
Bg-Fg Segmentation (SMANet)
Probability Map
Generated Segmentation Label

Fig. 1: Overall pipeline of the proposed approach. In the generated segmentation, pixels belonging to the uncertain region are shown in red.

[1] Thresholds are chosen based on [17]

- Foreground (lesion) — if the probability is higher than 0.7;
- Background — if the probability is lower than 0.3;
- Uncertain region — if the probability is between 0.3 and 0.7.

The definition of an uncertain region allows to consider the fuzzyness of the lesion boundaries, which are difficult to be precisely delimited. Indeed, the use of the uncertain region has proved to be effective, avoiding the gradient propagation in regions where the lesion could be misclassified with the background.

3 Experimental Results

In this section, the quality of the generated supervisions is evaluated. In Section 3.1, the generated label maps were compared with the ones provided for the ISIC 2018 segmentation challenge. Furthermore, in Section 3.2, the generated supervisions were employed to train a segmentation network to demonstrate that, with their inclusion, it is possible to improve its performance.

3.1 Comparison of label maps

To evaluate the quality of the generated label maps, the procedure described in Section 2.2 was employed to extract the segmentation supervisions for the 100 images of the validation set of the 2018 ISIC segmentation challenge. In particular, the YOLO detector has provided valid bounding–boxes, according to the criteria described above, for 86 out of 100 images. These bounding–boxes are then used to generate the related label maps, which have been compared with the segmentation ground truth released for the 2018 challenge. The Intersection over Union (IoU) is employed to compare the two segmentation supervision obtaining a Mean IoU (MIoU) of 85.57% on the 86 images. Examples of the generated label maps along with the original ground truth are reported in Figure 2.

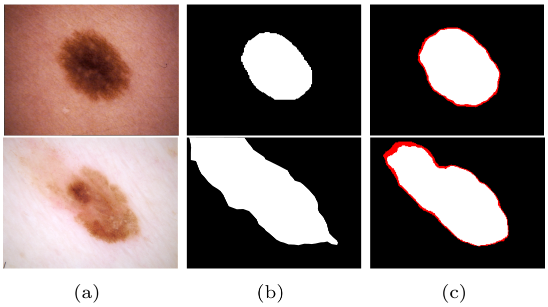

Fig. 2: In (a), original ISIC images, in (b) the related ground truth segmentation and, in (c), the generated supervisions (with the uncertain region in red).

3.2 Segmentation of Skin Lesions

The generated supervisions of the ISIC images were employed to train a deep segmentation network (the SMANet [21]). In particular, to prove the quality of the extracted label maps, the following experimental setup was devised:

- ISIC Ground Truth (ISIC_GT) — The SMANet was trained only with the 2594 images of the ISIC 2018 segmentation challenge;
- Weakly Supervised label maps (WS) — Only the generated label maps for 43,885 images of the ISIC database were employed to train the SMANet;
- WS + ISIC_GT — The segmentation network was first trained on the weakly supervised label maps and then fine tuned on the original ground truth supervisions provided for the 2018 segmentation challenge.

In all the experiments, the network was trained on random crops of 377×377 pixels using flipping and rotation to augment the training dataset. To stop the network training, a subset of about 120 images, extracted from the training set, were employed as validation set in all the setups. Finally, the network has been evaluated, using a sliding window approach, on the validation set of the ISIC 2018 segmentation challenge (see Table 1).

Setups	**Mean Accuracy**	**Mean IoU**
ISIC_GT	86.69% ± 3.19%	74.09% ± 4.18 %
WS	87.30% ± 3.24%	74.89% ± 3.83%
WS+ISIC_GT	**91.23% ± 2.39%**	**81.12% ± 3.43%**

Table 1: Mean Accuracy and IoU (along with 95% confidence interval) on the 100 images of the validation set of the ISIC 2018 segmentation challenge in the three experimental setups.

We can observe that, using solely the generated labels (WS setup), the results are comparable (slightly better) to that obtained with only the original ground truth supervisions (ISIC_GT). Instead, when the network is first pre–trained with weakly supervised labels and then fine tuned on the original labels (WS+ISIC_GT) the MIoU increases by approximately 7%. These results are quite surprising and demonstrate the quality of the generated supervisions, proving their importance to improve the precision of a lesion segmentation network.

4 Conclusions

In this paper, a weakly supervised approach has been proposed to generate the segmentation supervision of about 43,000 images of the ISIC database. A YOLO detector was employed to predict the bounding–box containing the lesion and, then, a segmentation network was trained to extract the foreground (lesion) pixels within the bounding–box. Finally, to create the label maps, the pixels outside the bounding box were labeled as background while, instead, the pixels inside were annotated based on the network output. The generated label maps proved to be quite useful if included during the training of a segmentation network. It will be a matter of further research the creation of an iterative procedure that iteratively retrain the YOLO detector to extract valid bounding–box for all the lesions in the ISIC archive.

References

[1] A. Krizhevsky et al. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25:1097–1105, 2012.

[2] K. He et al. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on CVPR*, pages 770–778, 2016.

[3] J. Long et al. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on CVPR*, pages 3431–3440, 2015.

[4] L. Chen et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2017.

[5] K. He et al. Mask R–CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969, 2017.

[6] J. Redmon and A. Farhadi. YOLOv3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[7] G. Litjens et al. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[8] S. Bonechi et al. Segmentation of Petri plate images for automatic reporting of urine culture tests. In *Handbook of Artificial Intelligence in Healthcare*, pages 127–151. Springer, 2022.

[9] P. Andreini et al. A two–stage GAN for high–resolution retinal image generation and segmentation. *Electronics*, 11(1):60, 2021.

[10] S. Bonechi et al. Fusion of visual and anamnestic data for the classification of skin lesions with deep learning. In *International Conference on Image Analysis and Processing*, pages 211–219. Springer, 2019.

[11] L. Tognetti et al. A new deep learning approach integrated with clinical data for the dermoscopic differentiation of early melanomas from atypical nevi. *Journal of Dermatological Science*, 101(2):115–122, 2021.

[12] A. Esteva et al. Dermatologist–level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[13] K. Hasan et al. DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation. *Informatics in Medicine Unlocked*, page 100819, 2022.

[14] P. Thapar et al. A novel hybrid deep learning approach for skin lesion segmentation and classification. *Journal of Healthcare Engineering*, 2022, 2022.

[15] M. Rastrelli et al. Melanoma: epidemiology, risk factors, pathogenesis, diagnosis and classification. *In vivo*, 28(6):1005–1011, 2014.

[16] International Skin Imaging Collaboration. Siim-isic 2020 challenge dataset. *International Skin Imaging Collaboration https://doi.org/10.34970/2020-ds01*, 2020.

[17] S. Bonechi et al. Generating bounding box supervision for semantic segmentation with deep learning. In *IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, pages 190–200. Springer, 2018.

[18] N. Codella et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC). *arXiv preprint:1902.03368*, 2019.

[19] P. Tschandl et al. The HAM10000 dataset, a large collection of multi–source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018.

[20] S. Bonechi et al. COCO_TS dataset: pixel–level annotations based on weak supervision for scene text segmentation. In *International Conference on Artificial Neural Networks*, pages 238–250. Springer, 2019.

[21] S. Bonechi et al. Weak supervision for generating pixel–level annotations in scene text segmentation. *Pattern Recognition Letters*, 138:1–7, 2020.