Deeply Supervised Rotation Equivariant Network for Lesion Segmentation in Dermoscopy Images

Xiaomeng Li(✉), Lequan Yu, Chi-Wing Fu, and Pheng-Ann Heng

Department of Computer Science and Engineering,
The Chinese University of Hong Kong, Shatin, Hong Kong
xmli@cse.cuhk.edu.hk

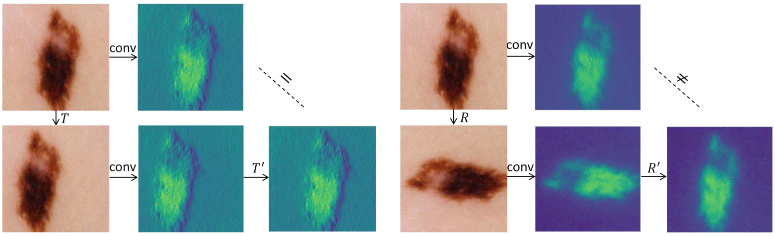
Abstract. Automatic lesion segmentation in dermoscopy images is an essential step for computer-aided diagnosis of melanoma. The dermoscopy images exhibits rotational and reflectional symmetry, however, this geometric property has not been encoded in the state-of-the-art convolutional neural networks based skin lesion segmentation methods. In this paper, we present a deeply supervised rotation equivariant network for skin lesion segmentation by extending the recent group rotation equivariant network. Specifically, we propose the G-upsampling and G-projection operations to adapt the rotation equivariant classification network for our skin lesion segmentation problem. To further increase the performance, we integrate the deep supervision scheme into our proposed rotation equivariant segmentation architecture. The whole framework is equivariant to input transformations, including rotation and reflection, which improves the network efficiency and thus contributes to the segmentation performance. We extensively evaluate our method on the ISIC 2017 skin lesion challenge dataset. The experimental results show that our rotation equivariant networks consistently excel the regular counterparts with the same model complexity under different experimental settings. Our best model also outperforms the state-of-the-art challenging methods, which further demonstrate the effectiveness of our proposed deeply supervised rotation equivariant segmentation network.

1 Introduction

Skin cancer has become the most prevalent cancer in the United States [12], and melanoma is the most deadly form of skin cancer, leading to over 9,000 deaths in the Unite States in 2017 [13]. A common technique used by dermatologists for diagnosing skin diseases is the dermoscopy, which enables observation by enhancing the visual effect of pigmented skin lesions. Lesion segmentation in dermoscopy images is an essential component in the diagnosis of skin diseases. However, segmenting skin lesions by dermatologists is time-consuming and error-prone to inter- and intra-observer variabilities. Moreover, due to the growing shortage of dermatologists per capita, the automatic lesion segmentation in dermoscopy images would be beneficial to more people [8]. Convolutional

D. Stoyanov et al. (Eds.): OR 2.0/CARE/CLIP/ISIC 2018, LNCS 11041, pp. 235–243, 2018.
https://doi.org/10.1007/978-3-030-01201-4_25

neural networks (CNNs) have proven to be very powerful models for a board array of image recognition tasks. In the domain of skin lesion segmentation, all leading methods adopted CNN-based methods [2,16,17]. For example, Yuan et al. [17] proposed a deep convolutional neural network (DCNN), trained it with multiple color spaces, and achieved the best performance in the ISIC 2017 skin lesion segmentation challenge. Yu et al. [16] explored the network depth property and proposed a deep residual network with more than 50 layers for automatic skin lesion segmentation.



(a) Conv is translation equivariant

(b) Conv is not rotation equivariant

Fig. 1. Convolution layer is translation equivariant (a); but convolution is not rotation equivariant (Zoom in to see the detailed comparison), as shown in (b).

The success of these CNN-based models can be partially attributed to the effectiveness of weights sharing in the convolution layer, where the translation equivariance is preserved. To be specific, translating a layer's input produces the corresponding translation in the layer's output. As shown in Fig. 1(a), shifting the input of the convolution leads to the predictable shifting in the output. This translation equivariance property of convolution is effective in most perception tasks, where the same weights can be used to encode the local spatial pattern and reduce the model parameter to avoid overfitting. Unlike natural images, dermoscopy images exhibit not only translation symmetry but also rotation and flipping symmetry as well. However, if one rotates the convolution input, the generated output does not necessarily rotate in a predictable manner, as shown in Fig. 1(b). Previous works utilized data augmentation technique like rotation and flipping, to encourage the network to learn rotation and flipping covariance. Even though this strategy could regularize the network to learn the equivariance on the training set, there is no guarantee that the equivariance property will generalize to other images. Moreover, forcing the network to learn the redundant knowledge introduced by different data transformations would reduce the model efficiency. Specifically, with the same level of model complexity, the regular CNN needs to learns not only the discriminative features but only the input rotations and reflections. Furthermore, comparing with natural images, the biomedical images are scarce and more difficult to obtain, and it is highly demanded to design an efficient network to improve the model efficiency.

We consider to improve the network efficiency by encoding the rotation and flipping equivariance into the network, in which the network preserves the equivariance inherent without relying on data augmentation. Recently, there are some works have made significant progress for rotation equivariant networks [6,10]. Cohen et al. [6] explored rotation and reflection equivariant inherent network for classification problems, where the feature learned in the G space exhibits rotation equivariance. In this paper, we propose a deeply supervised rotation equivariant network by extending G-CNN [6] for skin lesion segmentation. Our network encodes the translation, rotation and flipping symmetry of dermoscopy images, and thus improves the skin lesion segmentation performance. Specifically, we design the G-upsampling layer and the G-projection layer for the segmentation task with the G-convolution layer. The G-upsampling layer upsamples the features in the G space and the G-projection layer performs average pooling over the rotation dimension and then projects features from the G space to $\mathbb{Z}$ space, making the whole network rotation equivariant. To better stabilize the learning processing of the proposed network, we also integrated the deep supervision [4,9] in our network to further improve the performance. Compared with the plain convolution neural networks, our network enjoys a substantially higher degree of weight sharing, and increases the expressive capacity of the network without increasing the number of parameters. We extensively evaluate our method on the ISIC 2017 skin lesion segmentation challenge. The results demonstrate the efficiency of our proposed rotation equivariant segmentation network, and our method outperforms other state-of-the-art methods on the challenging dataset. Several works [1,14,15] also explore the rotation equivariant network in the biomedical image domain. However, our work further explores the equivariant segmentation networks with deep supervision scheme [4,9] for automatic lesion segmentation in dermoscopy images.

2 Method

In this section, we first introduce the concept of group equivariant convolution (G-convolution), and then describe the proposed G-upsampling and G-projection layers for the segmentation task. Finally we present our proposed deeply supervised rotation equivariant framework.

2.1 G-convolution

The regular first convolution layer is a function that maps the input to feature maps with K channels $f : \mathbb{Z}^2 \rightarrow \mathbb{R}^K$. The function can be described as Eq. 1.

$$[f * \varphi](x) = \sum_{y \in \mathbb{Z}^2} \sum_{k} f_k(y) \varphi_k(x - y), \tag{1}$$

where φ_k denotes the convolution kernel.

To encode rotation equivariance in the network, Cohen et al. [6] proposed to conduct convolution on groups, where the group $p4$ consists of all compositions of translations and rotations by 90° about any center of rotation in the grid, and

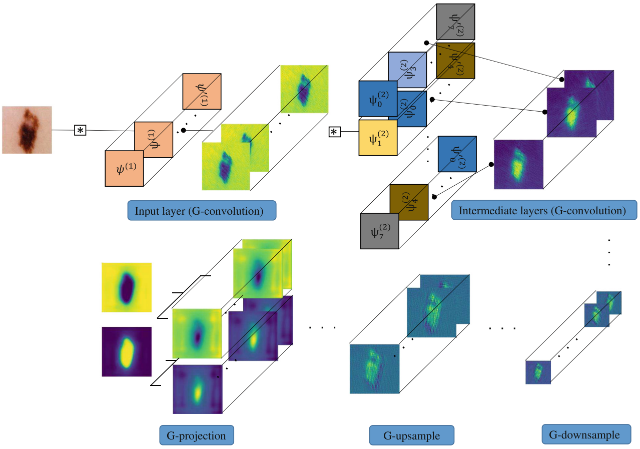


Fig. 2. The illustration of the G-convolution, G-upsampling and G-projection operation. Except the G-projection layer, we only show 1 channel in all other layers to simplify the illustration.

the group $p4m$ additionally includes reflections. Specifically, for the input layer, the $(\mathbb{Z}^2 \rightarrow G)$ convolution is defined as

$$[f * \varphi](g) = \sum_{y \in \mathbb{Z}^2} \sum_{k} f_k(y) \varphi_k(g^{-1}y), \tag{2}$$

where g is a transformation in the predefined group $p4$ or $p4m$. Then, in the following layers, feature maps and filters are both functions on G and the $(G \rightarrow G)$ convolution can be described as

$$[f * \varphi](g) = \sum_{h \in G} \sum_{k} f_k(h) \varphi_k(g^{-1}h) \tag{3}$$

2.2 G-upsampling and G-projection for Segmentation Problem

In the segmentation problem, the down-sampled feature maps need to be upsampled in the G space for pixel-level prediction, and thus we design the G-upsampling layer. The convention upsampling layer performs upsample operation for feature maps at the spatial dimension. In the G space, the G-upsampling layer performs upsample operation over all eight rotations (for group $p4m$) at each spatial position, as shown in Fig. 2.

To enable the equivariant network to produce final score maps for skin lesion segmentation, we also define the $(G - \mathbb{Z}^2)$ projection layer.

$$f_k(y) = \frac{1}{|G|} \sum_{G} (f_k(h)), \tag{4}$$

where $|G|$ denotes the number of element in group G. For example, it equals to 4 for group $p4$ and 8 for group $p4m$. With the G-upsampling layer and the G-projection layer, we can design a segmentation network, which is equivariant to the input symmetric transformations.

2.3 Deeply Supervised *G*-FCNs

The deeply supervised rotation equivariant network is based on the ResNet34 [7] architecture, where we replace the convolution layer, upsampling layer to the G-convolution, G-upsampling and G-projection layers. As shown in Fig. 3, we use three 2×2 G-upsampling layers and one G-projection layer following the feature maps generated by ResNet34. We also adopt the U-net like long-skip connections to preserve the low-level features. The deep supervision mechanism is performed by upsampling at three different spatial resolution of features, and the final result is the weighted combination of three segmentation predictions. Since all the elements in the network are equivariant to 90° rotation and reflection of the input, the whole framework also preserves the rotation equivariant property. In other words, if one clock-wise rotates the input image 90°, the network output will rotate in the same manner. Readers can find more details about the network architecture from our code[1].

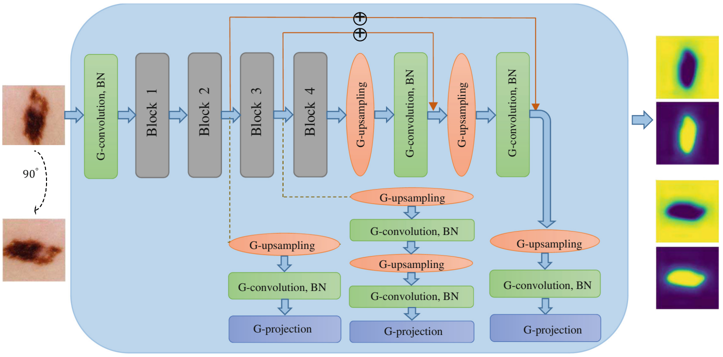


Fig. 3. The framework of our proposed rotation equivariant network for skin lesion segmentation. The network is based on ResNet34 backbone, and is integrated with deep supervision and U-Net connections. All the regular operations are replaced to G-convolution, G-upsampling, and G-projection operations. The whole architecture is equivariant to input symmetric transformation. In other words, if one rotate the input for 90°, then the prediction score would rotate in the same way. Note that we omit the pooling operation, ReLU activations to simplify the illustration.

[1] https://github.com/xmengli999/Deeply-Supervised-Rotation-Equivariant-Network-for-Lesion-Segmentation.

3 Experiments and Results

3.1 Dataset and Evaluation Metrics

We evaluate our method on the dataset of ISIC 2017 skin lesion segmentation challenge [5], which consists of a training set with 2000 annotated dermoscopic images, a validation set with 150 images, and a testing set with 600 images. The image size ranges from 540 × 722 to 4499 × 6748. To keep balance between segmentation performance and computational cost, we first resize all the images to 224 × 224 using bicubic interpolation. For evaluation metric, we follow the challenge instructions to employ five evaluation metrics, including jaccard index (JA), dice coefficient (DI), pixel-wise accuracy (AC), sensitivity (SE) and specificity (SP). Note that the final rank is determined according to JA in the ISIC 2017 skin lesion segmentation challenge.

Table 1. Ablation study of the deeply supervised rotation equivariant network.

Model	No. of para	Evaluation metrics				
		JA	DI	AC	SE	SP
ResnetFCN34*	22.8M	71.27	80.21	91.39	78.31	96.78
(RE)-ResnetFCN34*	22.8M	74.54	83.27	92.58	81.05	**97.59**
DS-U-ResnetFCN34*	23.2M	74.38	83.06	92.51	82.52	97.14
(RE)-DS-U-ResnetFCN34* (ours)	23.2M	76.65	85.00	93.27	84.61	96.80
(RE)-DS-U-ResnetFCN34 (ours)	23.5M	**77.23**	**85.60**	**93.55**	**85.40**	97.15

3.2 Implementation Details

All the experiments were implemented using PyTorch [11], and were trained with stochastic gradient descent (SGD) algorithm (momentum is 0.9) from scratch. The learning rate is set to 0.01 and decays at epoch 60. All the models are trained for 70 epochs. As for experiments with the plain convolution, we employed data augmentation like 90° rotation and flipping. The main loss function and the deep supervision branches are trained with cross entropy loss. The weights for main loss and deep supervision are 0.7, 0.2 and 0.1 respectively.

3.3 Ablation Study

Table 1 shows the segmentation performance on the test dataset with different configurations. ResnetFCN34* refers to the FCN-based Resnet34 network, while (RE)-ResnetFCN34* and DS-U-ResnetFCN34* are the rotation equivariant and deeply supervised with long range U-Net connections counterparts, respectively. The * denotes that we remove the first pooling layer from the original Resnet34 network, following the setting in [6]. Note that all the rotation equivariant networks are performed with group $p4m$ [6]. To analyze the effectiveness of rotation equivariant network fairly, all the comparison are performed with the same model complexity. Specifically, compared with the original filter numbers in Resnet34, the number of filters is divided by roughly $\sqrt{8}$ in each G-convolution layer.

From the comparison in Table 1, we can see that the rotation equivariant network largely excels the plain counterpart, with 3.27% improvement on JA. The deeply supervised version also improve the JA performance significantly. When integrate the deep supervision with U-Net connections into the rotation equivariant network ((RE)-DS-U-ResnetFCN34*), we can further improve the segmentation performance (2.27% on JA). To better adapt the network for our skin lesion segmentation task, we replace the first pooling layer of ResnetFCN34 with a G-convolution with stride of 2 and denoted the deeply supervised rotation equivariant version as (RE)-DS-U-ResnetFCN34. It is observed that (RE)-DS-U-ResnetFCN34 achieves the best performance on the all evaluation metrics excepting for SP, demonstrating the superiority and effectiveness of rotation equivariant networks under same level of model complexity.

Table 2. Comparison with state-of-the-art methods on the ISIC 2017 test dataset.

Team	JA	DI	AC	SE	SP
Our Method	**0.772**	**0.856**	**0.936**	**0.854**	0.972
Yuan and Lo [17]	0.765	0.849	0.934	0.825	0.975
Berseth [2]	0.762	0.847	0.932	0.820	0.978
Bi et al. [3]	0.760	0.844	0.934	0.802	**0.985**
RECOD	0.754	0.839	0.931	0.817	0.970
Jer	0.752	0.837	0.930	0.813	0.976
NedMos	0.749	0.839	0.930	0.810	0.981
INESC	0.735	0.824	0.922	0.813	0.968
Shenzhen U (Lee)	0.718	0.810	0.922	0.789	0.975

3.4 Comparison with Other Methods

We compare our result with state-of-the-art results on the ISIC 2017 testing dataset. There are totally 21 submissions and the top results are listed in Table 2. Yuan et al. [17] trained a CNN network with multiple color spaces and achieves the best performance on the skin lesion segmentation challenge. Our best model, trained from scratch on the single RGB color space, outperforms other state-of-the-arts in the test dataset of the ISIC challenge. This comparison validates the effectiveness of our proposed deeply supervised rotation equivariant network in the skin lesion segmentation task.

4 Conclusion

In this paper, we present a deeply supervised rotation equivariant segmentation network for skin lesion segmentation by utilizing the recent findings on rotation equivariant CNNs. We design the G-upsampling and G-projection layers to enable our network for the segmentation task, and introduce the deep supervision mechanism to improve performance. Our network encodes the rotation and reflection symmetry of dermoscopy images, and significantly improves the skin lesion segmentation performance. Our method has achieved the best performance on the ISIC 2017 skin lesion segmentation challenge dataset. Future works include the extension of equivariance to arbitrary rotation and scaling.

Acknowledgments. The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (Project no. GRF 14225616) and a grant from Hong Kong Innovation and Technology Commission (Project no. ITS/426/17FP). We special thank Dr. Taco Cohen for fruitful discussions, kindly help and encouragement in our exploration.

References

1. Bekkers, E.J., Lafarge, M.W., Veta, M., Eppenhof, K.A., Pluim, J.P.: Roto-translation covariant convolutional networks for medical image analysis. arXiv preprint arXiv:1804.03393 (2018)
2. Berseth, M.: ISIC 2017-skin lesion analysis towards melanoma detection. arXiv preprint arXiv:1703.00523 (2017)
3. Bi, L., Kim, J., Ahn, E., Feng, D.: Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv preprint arXiv:1703.04197 (2017)
4. Chen, H., Qi, X., Yu, L., Heng, P.A.: DCAN: deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 2487–2496 (2016)
5. Codella, N.C., Gutman, D., Celebi, M.E., et al.: 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). arXiv preprint arXiv:1710.05006 (2017)
6. Cohen, T., Welling, M.: Group equivariant convolutional networks. In: International Conference on Machine Learning, pp. 2990–2999 (2016)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
8. Kimball, A.B., Resneck, J.S.: The us dermatology workforce: a specialty remains in shortage. J. Am. Acad. Dermatol. **59**(5), 741–745 (2008)
9. Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z.: Deeply-supervised nets. In: Artificial Intelligence and Statistics, pp. 562–570 (2015)
10. Marcos, D., Volpi, M., Komodakis, N., Tuia, D.: Rotation equivariant vector field networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 5058–5067. IEEE (2017)
11. Paszke, A., et al.: Automatic differentiation in PyTorch (2017)
12. Rogers, H.W., Weinstock, M.A., Feldman, S.R., Coldiron, B.M.: Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the us population, 2012. JAMA Dermatol. **151**(10), 1081–1086 (2015)
13. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2017. CA Cancer J. Clin. **67**(1), 7–30 (2017). https://doi.org/10.3322/caac.21387

14. Veeling, B.S., Linmans, J., Winkens, J., Cohen, T., Welling, M.: Rotation equivariant CNNs for digital pathology. arXiv preprint arXiv:1806.03962 (2018)
15. Winkens, J., Linmans, J., Veeling, B.S., Cohen, T.S., Welling, M.: Improved semantic segmentation for histopathology using rotation equivariant convolutional networks. Med. Imaging Deep Learn. 330–341 (2018)
16. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans. Med. Imaging **36**(4), 994–1004 (2017)
17. Yuan, Y., Lo, Y.C.: Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks. IEEE J. Biomed. Health Inform. (2017). https://doi.org/10.1109/JBHI.2017.2787487