Analysis of image classification, object detection and instance segmentation in terms of robustness to artefacts in pigmented skin lesion classification

Florian Katsch*, Christoph Rinner† PhD, Philipp Tschandl‡ MD PhD

* Student at the Medical Informatics Master's Curriculum, Medical University of Vienna, Vienna, Austria
† Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria
‡ Department of Dermatology, Medical University of Vienna, Vienna, Austria

Objective

High incidence rates of malignant melanoma and the linkage between early diagnosis and survival rates foster developments in the field of automated skin lesion diagnosis. Automated diagnosis is affected by artefacts in dermatoscopic images. Skin lesions in dermatoscopic images can be identified, amongst other methods, by *image classification*, *object detection* and *instance segmentation*. With increasing complexity, models of these categories either classify the image as a whole, find several types of objects and their approximate position or individually segment and classify several instances of objects within one image. We picked three popular CNN architectures of these categories to analyse their robustness to artefacts.

Data and Methods

We first identified three common artefacts in dermatoscopic images (gel bubbles, ink markings, rulers) and established a method to superimpose artefacts in dermatoscopic images. The HAM10000 dataset (10.015 train and 1.511 dermatoscopic test images) [1] was augmented with artificially inserted artefacts. We ran several trainings on the ResNet-34, Faster R-CNN and Mask R-CNN network architectures, followed by a comprehensive testing to analyse their characteristics and robustness against present artefacts in images. To visualize the activations of network models, gradient-weighted class activation mappings were used.

Results

Our results suggest that ResNet and Faster R-CNN models perform worse than the Mask R-CNN when tested on images with superimposed artefacts. Artefacts in all tested images led to a decrease in area under the precision-recall curve values of 0.030 for ResNet and 0.045 for Faster R-CNN in comparison to only 0.011 for Mask R-CNN (see Figure 1). However, changes in models performance only became significant once 40% or more of the images had superimposed artefacts in it. Compared to real world scenarios this number is disproportionately high, and is not expected in clinical routine data . We also observed severe losses in performance when introducing artefacts selectively, for images that are assigned to a specific diagnosis, during training. Nevertheless, by including artefacts in all training data, the attention could be directed towards lesion areas. We examined this with gradient-weighted class activation mappings, but here we owe more detailed studies.

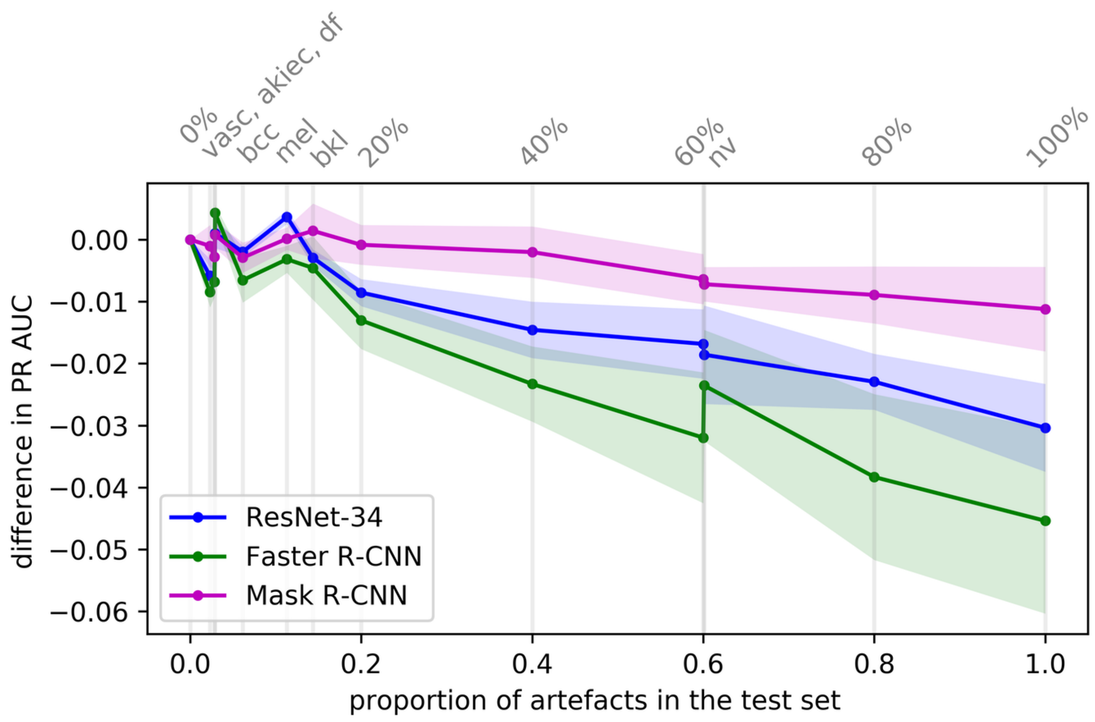

Figure 1: Degradation of PR-AUC values for the three different architectures when tested on datasets with different proportions of superimposed artefacts. Used datasets are labelled with the name of the class in which artefacts were inserted or with the percentage of images with superimposed artefacts if artefacts were inserted independently from the class label. A more robust behaviour is shown by Mask R-CNN; the values are averaged over five folds and the shaded areas show deviations within one standard deviation.

Conclusion

Results indicate that the use of instance segmentation architectures might help to counter the effects of artefacts on diagnostic accuracy. Latest research revealed a linkage between ink skin markings and an increase in false-positive predictions in melanoma recognition. This increase in false-positive predictions is especially troublesome regarding routine clinical data, where classes are highly unbalanced towards benign lesions and falsely classified images might lead to unnecessarily excisions of lesions [2]. Using instance segmentation methods might help to cut down the impact of artefacts on classification results with off-the-shelf methods and further research on other architectures of this family should be promoted.

References

[1] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," Scientific data, vol. 5, p. 180161, 2018.
[2] J. K. Winkler, et al., "Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition," JAMA dermatology, vol. 155, no. 10, pp. 1135–1141, 2019.