



**CZECH TECHNICAL
UNIVERSITY
IN PRAGUE**

F3

**Faculty of Electrical Engineering
Department of Cybernetics**

Bachelor's Thesis

Detection, classification and matching of skin lesions

Samuel Šúr

Cybernetics and Robotics

May 2023

Supervisor: prof. Dr. Ing. Jan Kybic

I. Personal and study details

Student's name: **Šúr Samuel**

Personal ID number: **492386**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Cybernetics**

Study program: **Cybernetics and Robotics**

II. Bachelor's thesis details

Bachelor's thesis title in English:

Detection, Classification and Matching of Skin Lesions

Bachelor's thesis title in Czech:

Detekce, klasifikace a hledání korespondencí kožních lézí

Guidelines:

The objective of this thesis is to develop a method for the detection, matching and classification of skin marks and skin lesions (objects of interest) from medium-scale skin images.

1. Get familiar with the state of the art in this domain and with available datasets, especially Skin3D and the ISIC Challenge (<https://challenge.isic-archive.com/>).
2. Using existing state of the art methods for object detection (eg. Faster R-CNN, YOLO), train and evaluate their performance on the 2D images from the Skin3D dataset.
3. Using existing state of the art methods for object classification (eg. ResNet, DenseNet), train and evaluate their performance on the available data. Study the performance as a function of image quality - resolution, blur, color changes.
4. Optional: Find correspondences between objects of interest in different images using existing methods such as (SIFT, SURF, RANSAC, Coherent Point Drift) for object matching.
5. Optional: Acquire a dataset using a mobile phone camera and evaluate the developed methods on this data.

Bibliography / sources:

- [1] Zhao, M. et al. 'Skin3D: Detection and Longitudinal Tracking of Pigmented Skin Lesions in 3D Total-Body Textured Meshes'. Medical Image Analysis 77, 2022
- [2] Korotkov, K.. 'Automatic Change Detection in Multiple Pigmented Skin Lesions'. Ph.D. Thesis, Universitat de Girona, 2014.
- [3] Ahmedt-Aristizabal, D. et al. Monitoring of Pigmented Skin Lesions Using 3D Whole Body Imaging, arXiv:2205.07085, 2022.
- [4] Chum, Matas: Matching with PROSAC – Progressive Sample Consensus. CVPR 2005
- [5] Myronenko et al. Point Set Registration: Coherent Point Drift. IEEE Trans. PAMI, 2010
- [6] Sime-Serra et al: Discriminative Learning of Deep Convolutional Feature Point Descriptors, ICCV 2015.

Name and workplace of bachelor's thesis supervisor:

prof. Dr. Ing. Jan Kybic Biomedical imaging algorithms FEE

Name and workplace of second bachelor's thesis supervisor or consultant:

Date of bachelor's thesis assignment: **16.01.2023** Deadline for bachelor thesis submission: **26.05.2023**

Assignment valid until: **22.09.2024**

prof. Dr. Ing. Jan Kybic
Supervisor's signature

prof. Ing. Tomáš Svoboda, Ph.D.
Head of department's signature

prof. Mgr. Petr Páta, Ph.D.
Dean's signature

III. Assignment receipt

The student acknowledges that the bachelor's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the bachelor's thesis, the author must state the names of consultants and include a list of references.

Date of assignment receipt

Student's signature

Acknowledgement / Declaration

I would like to thank my supervisor prof. Dr. Ing. Jan Kybic for his support and help during the preparation of this thesis. I would also like to thank my parents, Janka and Martin, for their continuous support, during my studies. I would like to thank my uncle Vaclav Pejchal who always kept me motivated. And finally, I would like to thank my friends for their support and help during my studies.

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

In Prague, 20 May 2023

.....

Abstrakt / Abstract

Cielom tejto práce je vytvoriť pipeline na detekciu a klasifikáciu kožných lézií v rozsiahlych snímkach pomocou modelov klasifikácie obrazu a detekcie objektov založených na konvolučných neurónových sieťach a hlbokom učení. Prvá časť práce je zameraná na teoretické východiská počítačového videnia a metód hlbokého učenia pri klasifikácii obrazu a detekcii objektov. Potom opisujeme state-of-the-art v diagnostike kože pomocou metód počítačového videnia. Trénujeme a vyhodnocujeme výkonnosť state-of-the-art metód klasifikácie obrazu a modelov detekcie objektov na snímkach kožných lézií. Vytvárame vlastný dataset snímok kožných lézií a testujeme výkonnosť modelu detekcie objektov na tomto súbore údajov. Nakoniec spojíme modely klasifikácie obrazu a detekcie objektov do jedného pipeline-u, ktorý sa dá použiť na detekciu klasifikáciu kožných lézií na snímkach kože bez akejkoľvek manuálnej intervencie používateľa.

Kľúčové slová: CNN, kožné lézie, klasifikácia obrázkov, detekcia objektov, hlboké učenie, počítačové videnie

This thesis aims to create a pipeline for the detection and classification of skin lesions in wide-area images using CNN-based image classification and object detection models. The first part of the thesis is focused on the theoretical background of computer vision and deep learning methods in image classification and object detection. Then we describe the state-of-the-art in skin diagnosis using computer vision methods. We train and evaluate the performance of state-of-the-art image classification and object detection models on skin lesion images. We also create our own dataset of wide-area skin lesion images and test the performance of the object detection model on this dataset. Lastly, we combine the image classification and object detection models into a fully-automated pipeline, which can be used to classify skin lesions in wide-area images without any manual intervention.

Keywords: CNN, skin lesions, image classification, object detection, deep learning, computer vision

Contents /

1 Introduction	1
1.1 Problem definition	1
1.2 Structure of the thesis	2
2 Object Detection	3
2.1 Traditional detectors	3
2.2 Neural network detectors	3
2.2.1 Two stage detectors	4
2.2.2 One stage detectors	4
3 Image classification	8
3.1 Ensemble models	9
4 State of the art	10
4.1 Detection of PSLs	10
4.2 Classification of PSLs	11
4.2.1 Traditional methods	12
4.2.2 Deep learning methods ..	12
4.3 Self-examination systems for PSLs	13
4.4 Summary	15
5 Data	16
5.1 Datasets	16
5.2 DermaNude dataset	17
6 PSL detection implementation ..	21
6.1 Our approach	21
6.1.1 Centroid metric	23
6.2 Results	24
6.3 PSL Detection on the Der- maNude dataset	26
6.4 Conclusion	28
7 PSL classification implementa- tion	32
7.1 Our approach	32
7.1.1 Ensemble models	33
7.2 Results	33
7.2.1 Individual models	33
7.2.2 Ensemble models	35
7.3 Conclusion on the perfor- mance of the models	37
7.4 Experiments to study ro- bustness of the models	38
7.4.1 Resolution	38
7.4.2 Brightness	38
7.4.3 Contrast	38
7.4.4 Blur	40
7.4.5 Color	40
7.4.6 Experiments with un- known class	41
7.5 Conclusion of experiments	43
8 System for self-examination of skin lesions	46
9 Conclusion	51
9.1 Comparison of the original assignment and the final re- sult	51
9.2 Future work	51
References	53
A Consent form - DermaNude dataset	61
B DermaNude Dataset License Form	62
C Technical details	66
C.1 Classification	67
C.2 PSL detection	67
C.3 Self-examination of skin le- sions	67
C.4 Datasets	67

Tables / Figures

<p>5.1. Specifications of the DermaNude dataset. 18</p> <p>6.1. Results of the YOLOv7 model on the DermaNude dataset. . 27</p> <p>6.2. Results of the model on test dataset using COCO metrics (AP and AR (Average Precision and Average Recall)). 29</p> <p>6.3. Average Precision (mAP) (Pascal VOC 2012). 30</p> <p>6.4. Average Precision (mAP) using centroid metric. 30</p> <p>6.5. Average Precision (mAP) using centroid metric on the DermaNude dataset. 30</p> <p>7.1. Ensemble models. 33</p> <p>7.2. Performance of ResNet50 on test set. 35</p> <p>7.3. Performance of EfficientNet-B0 on test set. 35</p> <p>7.4. Performance of EfficientNet-B7 on test set. 36</p> <p>7.5. Performance of Res50-EffB0 on test set. 36</p> <p>7.6. Performance of Res50-EffB7 on test set. 36</p> <p>7.7. Performance of Effb0-Effb7 on test set. 37</p> <p>7.8. Performance of ensemble of all models on test set. 37</p> <p>7.9. Performance of the models 38</p> <p>7.10. Percentage of PSLs classified as “Unknown”. 41</p> <p>7.11. Unknown classification on images not containing any of the 8 classes—3dBodyTex.v1 dataset. 44</p> <p>7.12. Unknown classification on images not containing any of the 8 classes—DermaNude dataset. 44</p>	<p>5.2. Image taken by NIKON D3400 camera (24MP) and its close-up (1MP) 18</p> <p>5.3. Image taken by Realme 8 smartphone (16MP) and its close-up (1MP). 20</p> <p>6.2. Texture images from 3dBodyTex.v1 dataset with bounding boxes from Skin3D dataset. 21</p> <p>6.1. Original image and its first 15 overlapping patches 22</p> <p>6.3. Preset learning rate for 30 epochs. 23</p> <p>6.4. Centroid metric. 24</p> <p>6.5. Precision x Recall curve for annotations A1. 24</p> <p>6.6. Precision x Recall curve for annotations A2. 25</p> <p>6.7. Precision x Recall curve for annotations A3. 25</p> <p>6.8. Bounding box sizes in test dataset. 26</p> <p>6.9. Precision x Recall curve for DermaNude—camera images. . 27</p> <p>6.10. Precision x Recall curve for DermaNude—smartphone images. 28</p> <p>6.11. Comparison of ground truth and predicted bounding boxes on the DermaNude dataset. . 28</p> <p>6.12. False positives on the DermaNude dataset. 29</p> <p>6.13. Prediction (red) and ground truth (blue) bounding boxes for annotations A2, with confidence scores (red text) 30</p> <p>7.1. Training progress of the EfficientNet-B0 model, model evaluated on train set and validation set. 33</p> <p>7.2. Training progress of the ResNet-50 model, model evaluated on train set and validation set. 34</p>
--	--

7.3.	Training progress of the EfficientNet-B7 model, model evaluated on train set and validation set.	34
7.4.	Impact of the resolution change on the balanced accuracy of the models.	39
7.5.	Preset learning rate for 30 epochs.	39
7.6.	Impact of the contrast change on the balanced accuracy of the models.	40
7.7.	Impact of the blur change on the balanced accuracy of the models.	41
7.8.	Impact of the color change on the balanced accuracy of the models.	42
7.9.	Example of the cropped images from 3dBodyTex.v1 and DermaNude dataset.	43
7.10.	Examples of changes applied to the test dataset.	45
8.1.	Output of the CLI App	47
8.2.	Cropped bounding boxes by CLI app.	47
8.4.	Output of the CLI App—changed conditions	48
8.3.	Output of the CLI App—changed conditions	49
8.5.	Output of the CLI App	50

Chapter 1

Introduction

Skin diseases are among the most common and costly health conditions worldwide [1] and according to WHO, every third cancer diagnosed is skin cancer¹. There were approximately 57000 deaths caused by melanoma in 2020², which is one of the most common malignant skin lesions.

This thesis focuses on the diagnosis of skin lesions using deep learning methods for image classification and object detection. The goal of this thesis is to develop a method for automatic detection and classification of skin lesions. The method is able to detect skin lesions in wide-area images and classify them into one of the common types of skin lesions. A new dataset of wide-area skin lesion images is created for the purpose of this thesis.

Dermatologists examine skin lesions by visual inspection and by using dermoscopy, which is a non-invasive technique that allows for a magnified view of the skin lesion. The other method is a biopsy, which is a surgical procedure that involves the removal of a small sample of the skin lesion for further analysis. Another method, called the ABCD rule can be used by non-specialists to distinguish between benign and malignant skin lesions. The rule is based on the following characteristics of the skin lesion: asymmetry, border irregularity, color variation, and diameter [2].

Detecting melanoma, at an early stage can significantly improve the patient's prognosis [3]. Therefore, it is important to develop a method that can detect skin lesions automatically, which can help not only dermatologists but also ordinary people to detect melanoma in the early stages. Such methods include Computer-Aided Diagnosis (CAD) systems. Developing CAD systems is a challenging task; while dermoscopy photos are often of high quality, with the same lighting conditions, and taken from the same distance or angle of view, the photos of skin lesions taken by patients can be of different distances, angles, lighting conditions, and quality. Therefore, it is important to develop a method that can diagnose skin lesions in different conditions.

1.1 Problem definition

The objective of this thesis is to develop a method for automatic detection and classification of skin marks and skin lesions (objects of interest) from medium-scale skin images. The thesis is focused on the following tasks:

- Using existing state-of-the-art methods for object detection, train and evaluate their performance on the 2D images from the Skin3D dataset (see “Sec. 5.1”).
- Using existing state-of-the-art methods for image classification, train and evaluate their performance on the available data. Study the performance as a function of image quality - resolution, blur, color changes, etc.

¹ [https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-\(uv\)-radiation-and-skin-cancer](https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv)-radiation-and-skin-cancer)

² <https://gco.iarc.fr/today/data/factsheets/cancers/16-Melanoma-of-skin-fact-sheet.pdf>

- Acquire a dataset using different types of cameras (smartphone, DSLR) and evaluate the developed methods on this data.
- Build a pipeline that combines both image classification and object detection into a fully automated system

1.2 Structure of the thesis

The thesis is divided into 9 chapters. In the second chapter, we discuss the related work in the field of object detection. The third chapter is dedicated to the related work in the field of image classification. In the fourth chapter, we discuss the related work in the field of skin lesion diagnosis, which includes both image classification and object detection methods, as well as matching methods. Lastly this chapter also discusses applications available for users to self-diagnose skin lesions. The fifth chapter is dedicated to the available datasets of skin lesions and we introduce our own dataset of wide-area skin lesion images, called DermaNude. In next 3 chapters, we discuss our approach to the problem of skin lesion diagnosis—detection, classification and creation of a pipeline that combines both into a fully automated system. In the ninth chapter, we discuss the achieved goals of the thesis and discuss the possible future work.

Chapter 2

Object Detection

One of the goals of this thesis is to localize PSLs (Pigmented skin lesions) in images of skin. To understand the state-of-the-art of localization of PSLs in images, it is necessary to understand the current state-of-the-art in object detection generally. Object detection is a computer vision task of localizing objects in images by predicting bounding boxes around them. The problem has been studied for many years and many different algorithms have been developed. Object detection can be split into two main categories: object detection based on methods not using neural networks and object detection based on methods using neural networks.

2.1 Traditional detectors

One of the most notable algorithms for object detection of the first category is the Viola-Jones detector [4]. The Viola-Jones detector consists of three main stages: the first stage is the Haar feature selection, the second stage is the AdaBoost training and the third stage is the cascade classifier. The Haar-like feature selection is a method for extracting features from images. The Haar-like features are rectangular patterns that capture the contrast between the pixels within the feature area. The algorithm applies these features to every possible location in the image and calculates the difference between the sum of the pixels in the white area and the sum of the pixels in the black area. The next step is to train the classifier using the AdaBoost algorithm. The AdaBoost algorithm is used to select the most important features and to distinguish between object and non-object regions. The last step is to combine the classifiers into a cascade of classifiers. The cascade of classifiers is used to reduce the number of false positives.

The next algorithm for object detection from the first category is the HOG (Histogram of Oriented Gradients) detector [5]. The HOG detector is based on the HOG descriptor. The HOG descriptor is a feature vector that describes the shape of an object. The HOG descriptor is calculated by dividing the image into cells and calculating the gradient histogram for each cell. The local gradient histograms are then normalized and concatenated into a single feature vector. The HOG detector is trained using the SVM (Support Vector Machine) algorithm. The SVM algorithm is used to find the optimal hyperplane that separates the object and non-object regions.

The next algorithm for object detection from the first category is the DPM (Deformable Part Models) detector. The DPM detector was firstly introduced by Felzenszwalb et al. in 2009 [6]. It was developed as a successor to the HOG detector. It is based on the method of breaking down the objects into smaller parts and analyzing them separately.

2.2 Neural network detectors

Neural network detectors can be divided into two groups: the first group is the one-stage detectors and the second group is the two-stage detectors. The one-stage detectors use

a single CNN stage to directly predict the location of the objects in an image. The two-stage detectors have an initial stage to propose regions of interest and a second stage to classify the regions of interest.

2.2.1 Two stage detectors

In 2014, R. Girshick et al. [7] introduced RCNN (Regions with CNNs). They achieved state-of-the-art results on the PASCAL VOC 2007 dataset—58.5% mAP compared to 33.7% mAP by previous state-of-the-art method [6]. The RCNN firstly uses a selective search algorithm to generate region proposals. For each region proposal, a feature vector is extracted using the CNN. The feature vectors are then passed through a linear SVM classifier to determine the presence of an object as well as the class of the object it belongs to. The RCNN’s main drawback is the speed. The RCNN is very slow because it uses 2000 region proposals per image. RCNN is not suitable for real-time applications.

In the next years, new models were introduced improving the speed and accuracy of object detection. In 2014, K. He et al. [8] introduced the SPPNet (Single Shot Multibox Detector). SPPNet is 24–102 times faster than RCNN. The mAP of SPPNet is 59.2% on the PASCAL VOC 2007 dataset. In 2015, R. Girshick [9] introduced the Fast R-CNN followed by the Faster R-CNN [10]. The Faster R-CNN increased both, the speed and the accuracy—73.2% mAP on the Pascal VOC 2007 dataset. The architecture of the Faster R-CNN may be seen in “Fig. 2.1”.

In 2017, T. Lin et al. [12] introduced the FPN (Feature Pyramid Networks). The FPN is a top-down architecture with lateral connections. It generates a pyramid of feature maps with strong semantics at all levels from a single input image. The FPN serves as a backbone network for region proposal networks (RPNs). FPN provides multi-scale feature pyramids where each level corresponds to a different scale of the input image, allowing for better detection of objects at different sizes. The FPN is used as a backbone network for many object detection algorithms.

2.2.2 One stage detectors

In 2016, J. Redmon et al. [13] introduced the YOLO (You Only Look Once) algorithm as a first one-stage detector. The YOLO model processes images at a speed of 45 frames per second (fps). The mAP of YOLO is 63.4% on Pascal VOC 2007 dataset. During the following years, many improvements to the YOLO algorithm were introduced, improving speed, accuracy, changing the architecture of the network. The YOLOv1 [13] process of detection starts with dividing the input image into $S \times S$ grid cells. Each grid cell predicts B bounding boxes and confidence scores for each of the bounding boxes. It also predicts C class probabilities for each bounding box. These predictions form a tensor described as $S \times S \times (B \times 5 + C)$. Each bounding box is described by 5 variables: x, y, w, h, c . The variables x, y are the coordinates of the center of the bounding box. The variables w, h are the width and height of the bounding box. The variable c is the confidence score of the bounding box. Low confidence score bounding boxes are filtered out using the predefined confidence threshold. The bounding boxes are then sorted by their confidence score and the non-maximum suppression algorithm is applied to remove overlapping bounding boxes. The basic workflow of the YOLOv1 model is shown in “Fig. 2.2”. The YOLOv1 network consists of 2 parts—the backbone network and the head. The backbone network is a convolutional neural network that extracts features from the input image. The head of the network is a fully connected layer that turns the features extracted by the backbone network into the prediction tensor. The backbone network is a 24-layer convolutional neural network.

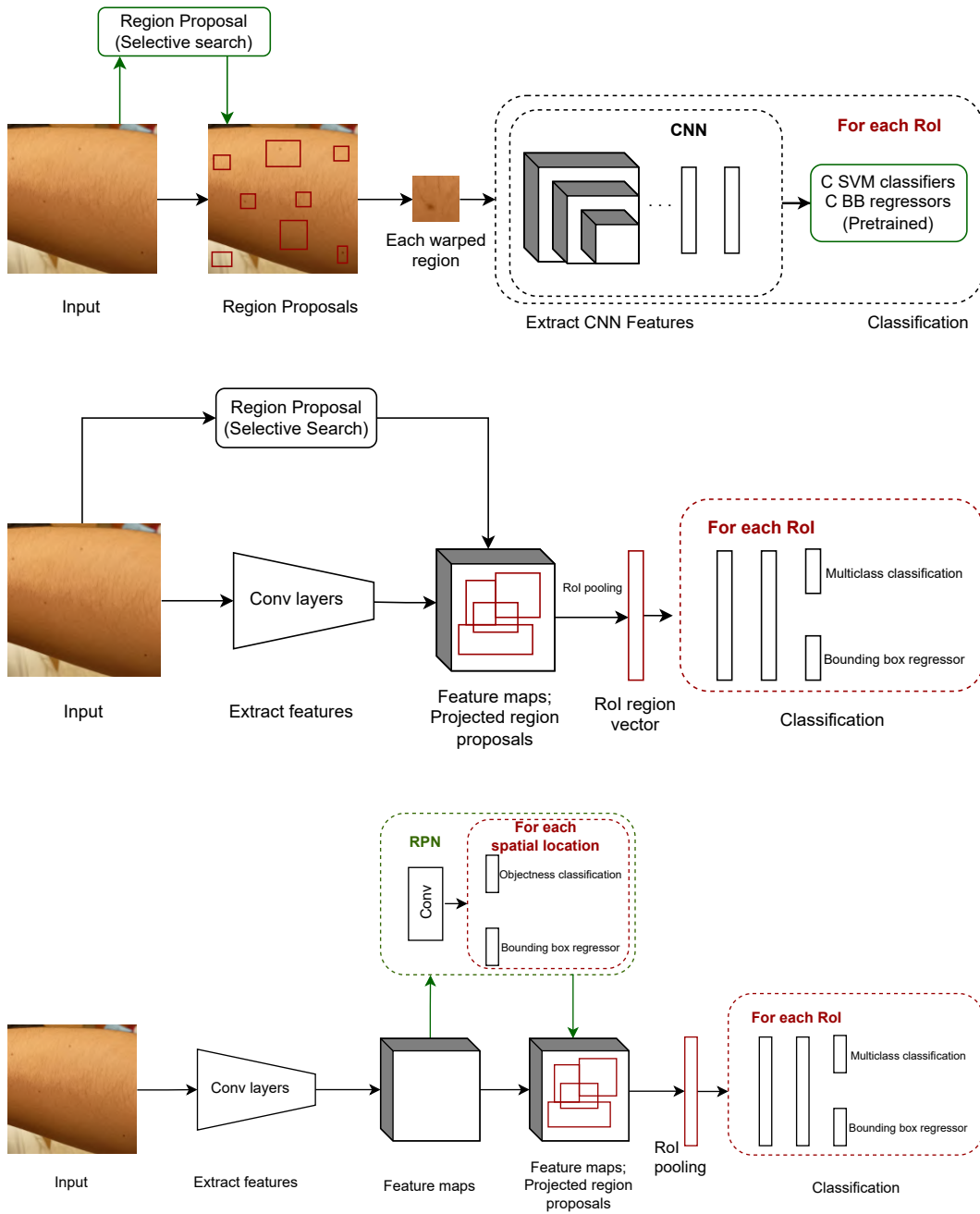


Figure 2.1. General architecture of the R-CNN models. From top to bottom: RCNN, Faster RCNN and Fast RCNN. Modified image from [11].

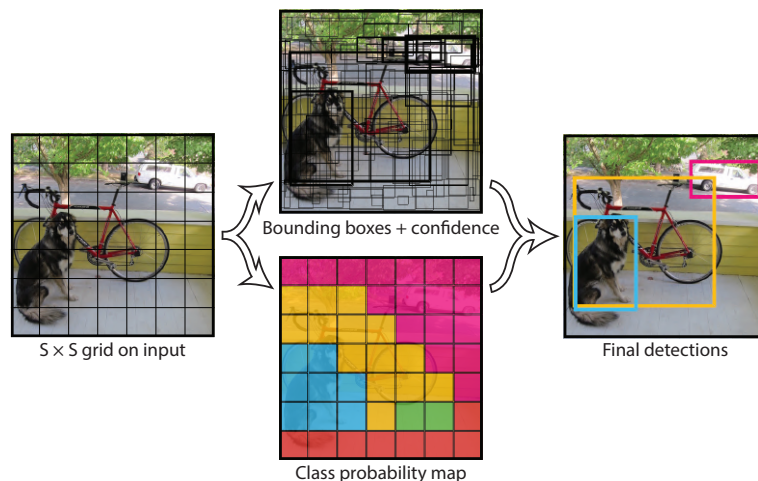


Figure 2.2. Basic workflow of YOLOv1 model. [13]

The authors of the YOLOv1 introduced also the YOLOv2 (YOLOv9000) [14] and YOLOv3 networks [15].

For a backbone network of the YOLOv3 algorithm, the authors decided to use the Darknet-53, while YOLOv2 uses the Darknet-19. The latter uses 19 convolutional layers, while the former uses 53 convolutional layers. YOLOv3 further distinguishes itself from YOLOv2 and YOLOv1 by using a similar architecture to the FPN. The format of the predictions is also different. YOLOv3 predicts three bounding boxes per grid cell at different scales. The prediction tensors are described as $S \times S \times (3 \times 5 + C)$, then $(2 \cdot S) \times (2 \cdot S) \times (3 \times 5 + C)$ and finally $(4 \cdot S) \times (4 \cdot S) \times (3 \times 5 + C)$. This can be combined and written as $N \times N \times (3 \times 5 + C)$. The YOLOv3 increased detection accuracy for small objects. [15]

In 2020, Alexey Bochkovskiy et al. introduced the YOLOv4 algorithm. It is the first YOLO version that is not created by the authors of the original YOLO algorithm. The YOLOv4 algorithm uses the CSPDarknet53 backbone network. The CSPDarknet53 is a modified version of the Darknet53 network. The second part of the YOLOv4 network is the Neck. The Neck consists of the SPP (Spatial Pyramid Pooling) block and the PAN (Path Aggregation Network) block. The SPP block is used to extract features from the input image at different scales. The PAN block is used for parameter aggregation from different backbone levels. The last part—the YOLOv3 head. [16]

In 2022, Alexey Bochkovskiy et al. upgraded the YOLOv4 which yielded the YOLOv7 algorithm. The YOLOv7 introduces Extended Efficient Layer Aggregation Network (EELAN). This paper also introduces the concept of Trainable Bag of Freebies (TBoF) which is a reparameterization algorithm. [17]

In 2015, W. Liu et al. [18] introduced the SSD (Single Shot MultiBox Detector) network. The algorithm generates a set of default boxes over different aspect ratios and scales per feature map location. The SSD network then predicts the class probabilities and bounding box offsets for each of the default boxes.

In 2017, F. Lin et al. [19] introduced the RetinaNet network. The paper introduces the new loss function for object detection, the Focal Loss. The Focal Loss is a generalization of the cross-entropy loss. The Focal Loss is used to address the problem of the class imbalance between the foreground and background classes.

In 2019, H. Law et al. [20] introduced the CornerNet network. The CornerNet network brings a new approach to object detection. The CornerNet detects objects as a set of keypoints—top left and bottom right corners of the bounding box. By this

method, there is no need to use anchor boxes. CornerNet achieves state-of-the-art results on the COCO dataset.

In 2019, X. Zhou et al. [21] introduced the CenterNet network. The CenterNet network considers objects to be the center points of the bounding boxes. The network then regresses other parameters such as size, orientation, pose, 3D location.

In 2020, M. Carion et al. [22] introduced the DETR network (DEtection TRansformer). The DETR network uses transformer architecture which is a new state-of-the-art type of neural network architecture. Their approach does not use Non-Maximum Suppression (NMS) and anchor boxes.

Chapter 3

Image classification

One of the goals of this thesis is to classify PSLs into classes. To understand the state-of-the-art of classification of PSLs in images, it is necessary to understand the current state-of-the-art in image classification generally. The image classifiers for classification can be divided into two main categories: the approaches using the traditional computer vision techniques and the approaches using deep learning. The traditional approaches include k-Nearest Neighbors, Support Vector Machines, Naive Bayesian algorithm, etc.

The deep learning approaches are based on convolutional neural networks (CNN).

In 2012, A. Krizhevsky et al. [23] introduced the AlexNet architecture, which achieved state-of-the-art results on the ImageNet dataset.

K. Simonyan, from Visual Graphics Group (VGG) from Oxford University introduced the VGGNet architecture in 2014. One of the key contributions of VGGNet was the introduction of the concept of **deepness** in convolutional neural networks. The model consisted of up to 19 layers, which was much deeper than previous state-of-the-art models such as AlexNet. [24].

Ch. Szegedy et al. [25] introduced GoogleNet also known as InceptionNet in 2014. This architecture was designed to address the limitations of previous state-of-the-art models, such as VGGNet, which had a large number of parameters and required a lot of computational power. Another key contribution of InceptionNet was the introduction of the concept of **inception modules** which allowed the model to learn both local and global features.

ResNet or Residual Network was introduced by K. He et al. in 2015. The main contribution of the ResNet architecture was the introduction of the residual connections, which allows the model to learn from the residual error between the input and the output of the layer. This architecture was designed to address the problem of vanishing gradients, which was a major problem in the previous state-of-the-art models [26].

In 2016, S. Xie et al. [27] introduced ResNeXt architecture, which is based on the ResNet architecture. ResNeXt builds upon the ResNet architecture by using a split-transform-merge strategy to learn more diverse and powerful feature representations. This strategy involves creating a large number of parallel pathways (called **cardinality**) within the network and then aggregating the information from all these pathways using a merge operation. By doing so, ResNeXt can capture a broader range of features and achieve state-of-the-art performance.

In 2017 S. Howard et al. introduced MobileNet family of models [28], which are based on the concept of depthwise separable convolutions. These convolutions are designed to reduce the number of parameters and computational power required by the model. Therefore the MobileNet models are suitable for mobile devices. Instead of standard convolutional layers, the MobileNet models perform two types of convolutions: depthwise convolution and pointwise convolution. The depthwise convolution applies a single filter to each input channel, pointwise convolution combines the output of the depthwise convolution using 1x1 convolutional filters.

In 2019 M. Tan and Q. V. Le introduced a family of EfficientNet models [29], which are based on the concept of compound scaling. The idea is to scale the width and depth of the network in a compound way, which allows to achieve better results with less computational power. The family of EfficientNet models consists of 8 different models, which differ in the scaling factor of the width and depth of the network. The EfficientNet networks achieve state-of-the-art results on the ImageNet dataset.

3.1 Ensemble models

The common approach to improve the performance of the image classification models is to use ensemble models. The ensemble models combine the predictions of multiple models into a single prediction. The general equation for the ensemble model is as follows:

$$Y^*(x) = \sum_{i=1}^N w_i y_i^*(x)$$

where $Y^*(x)$ is the ensemble model's prediction, $y_i^*(x)$ is the i -th model's prediction and w_i is the weight of the i -th model. The weights determine the importance or contribution of each model to the final prediction. The ensemble models generally improve the performance of the models and are more robust. However they are more computationally expensive.

Chapter 4

State of the art

In this chapter we shall describe the state of the art in the field of computerized diagnosis of PSLs. This includes the available datasets, types of the datasets and the methods used for PSL (Pigmented skin lesion) diagnosis. Firstly we describe the state-of-the-art in the field of object detection of PSLs, then we describe the state-of-the-art in the field of image classification of PSLs. Lastly we describe the state-of-the-art in the field of self-examination systems for PSLs.

4.1 Detection of PSLs

In 1989, D.A. Perednia et al. [30] used Laplacian-of-a-Gaussian (often referred to as the Marr-Hildreth filter), to detect borders of the skin lesions. Later, in 1991, they introduced the method, that considers the images to be 3D images where brightness is considered to be height, the skin features appear as 'pits'. These brightness pits are then used for detection by kNN or Fisher's MDA algorithm [31]. Later, they combined multiresolution hierarchical segmentation with the region-growing algorithm and neural network for the detection of PSLs, which resulted in human-like performance in [32].

B. McGregor [33] explores the use of a computerized image registration technique to align and compare non-dermoscopy images of pigmented skin lesions. Their method has been restricted to the images of torso.

Tim K. Lee [34] proposed a method for counting the skin lesions of the wide area photo of the back. The method involves pre-processing the image, applying mean shift filtering and region-growing techniques to identify candidate moles.

In 2007, T.S. Cho et al. [35] presented a framework to detect PSLs. Firstly, they detect pixels containing skin, then the skin regions are processed with a Gaussian filter to propose the candidate regions. The candidate regions are then classified using the SVM classifier. If the hair is present in the image, it is removed to increase accuracy by using various filter techniques.

J.S. Pierrard et al., [36] in 2007 presented a framework for skin lesion detection on images of a face. The model detection is based on normalized cross-correlation matching where the Laplacian-of-Gaussian filter is applied beforehand.

K. Korotkov in his research [37–39] explored the way to match and track skin lesions in time from whole body 2D scans. 2D scans of the whole body are taken from 23 turntable positions. The research matches the skin lesion by executing two methods. Firstly by finding intra-exploration and later inter-exploration correspondences. Intra-exploration is a method to match skin lesions on images acquired during one exploration. Inter-exploration is a method to match skin lesions on images from successive scans. The intra-exploration in Korotkov's thesis is divided into several steps. Firstly the skin lesions are detected. For detection, the research works with MSER (maximally stable extremal regions). MSER [40] is a feature extraction technique which works by identifying regions of an image which are stable under small intensity variations. The method requires a lot of manual tuning of parameters and preprocessing of the images. Then,

the homography transformation of MSER patches around detected moles is computed by using SIFT (Scale-invariant feature transform). The SIFT correspondences are then loaded to RANSAC (Random sample consensus) algorithm. The SIFT followed by the RANSAC algorithm for skin lesion registration is described in detail in a study by Anagnostopoulos et al. [41]. The stereo pairs are created from the images of the same skin lesions from different cameras, then the stereo pairs are triangulated. The stereo pairs are created and triangulated for every turntable position. Then, the mole sets are acquired from skin lesions in each turntable position. Then, the triangulated moles can be grouped into mole sets. Finally, the mole sets from adjacent positions are merged to create a mole map of the whole body. Korotkov’s work uses a naive approach for inter-exploration skin lesion matching. It assumes, that the skin lesion which has been photographed by camera C at position P in one screening session is at the same/similar C-P relation at second screening. The SIFT and RANSAC algorithm is used for lesion matching across regions of interest. The author also states the matching problem of skin lesions in case of patients with a lot of body hair.

Skin3D of Mengliu Zhao et al. [42] uses 3D images of the whole body and uses new algorithms for detection and tracking of skin lesions. The contribution of this research is also in the form of manual annotation of 3DBodyTex dataset discussed in “Sec. 5.1”. The 3D meshes are expanded into 2D texture images. The skin lesions are detected using R-CNN on these 2D texture images using the bounding boxes. The unwrapping of images to 2D has been chosen due to the time consumption of the 3D annotation process by human annotators. For tracking lesions in time it is necessary to map detected skin lesions back into 3D coordinates. For anatomically corresponding vertices across scans the work relies on 3D-Coded [43]. The research works with the optimization of 3D geodesic distances between vertices of registered meshes. However, research by Bogo et al. [44] suggested to register the meshes to a mesh template with unified UV mapping, then the pixel locations are compared in the texture images which leads to identifying the skin lesions. The problem with this method is, that sometimes anatomically close distances between lesions may translate into large distances in 2D texture images. Zhao et al. further expands on the problem of appearing and disappearing skin lesions, either due to lack of visibility or natural disappearance/appearance, by using a “dummy node”.

Ahmedt-Aristizabal et al. [45] also researched the way of detection and longitudinal tracking of skin lesions. Their approach uses a compilation of whole-body databases—FAUST, 3DBodyTex, 3D human models from Renderpeople (see “Sec. 5.1”), and synthetic images generated from these models. This work uses Scaled-YOLOv4 for skin lesion detection. For anatomically corresponding vertices across scans the research uses LoopReg [46] instead of 3D-Coded, which is used in Skin3D work by Mengliu Zhao et al. The work then matches the skin lesions using the geodesic distance. The work does not explore the problem regarding the appearing and disappearing skin lesions.

Works by Korotkov, Zhao et al., and Ahmedt-Aristizabal et al. are focused on the detection and tracking of skin lesions in time. However these, works do not explore the problem of skin lesion classification.

4.2 Classification of PSLs

The Section about state-of-the-art in the classification of PSLs is divided into two parts. The first part describes the traditional methods used for the classification of PSLs. The second part describes the deep learning methods used for the classification of PSLs.

4.2.1 Traditional methods

F. Xie et al. developed a novel method for the classification of PSLs [47] into benign and malignant. Their method consists of three steps. Firstly, lesions are extracted using a self-generative neural network. The second step was focusing on feature extraction such as color, texture, lesion border. In total, 57 features are extracted. They use Principal components analysis (PCA) to reduce the dimensionality of the features. In the last, the third step, skin lesions are classified using an ensemble of neural networks. Their work also compares the proposed method with other common methods used for classification such as Random forest, kNN, SVM, etc. In conclusion, their method achieved a better score (by 7.5%) than common classifiers. The work uses PH2 (see “Sec. 5.1”), EDRA datasets, as well as dataset provided by Chinese People’s Liberation Army General Hospital.

A. Masood et al. [48] compare three different ANN (Artificial neural networks) methods for image classification applied on classification of PSLs. The methods are Levenberg-Marquardt, Resilient Backpropagation and Scaled Conjugate Gradient. Most of the images in their dataset come from the Melanoma Diagnostic Centre, Royal Prince Alfred Hospital and Hôpital Charles Nicolle CHU Rouen in France. They rescale the images into 150x150 with a bit depth of 24. Levenberg-Marquardt method performed best with a specificity score of 95.1%.

W.F. Cueva et al. [49] proposed a system for the classification of PSLs based on extraction of features according to the ABCD rule. PSLs are segmented using the method called Mumford-Shah. It is used to detect the area, center, and diameter of binary images. These parameters are then used to determine if the PSLs are symmetric or asymmetric. Border irregularity is extracted by the Harris-Stephens algorithm. Color extraction is done by using HSV color space. The system uses backpropagation feed-forward ANN and classifies PSLs into 3 classes—common, non-common, and melanoma.

J.A. Jaleel et al. [50] proposed ANN based system for the classification of PSLs. The system classifies dermoscopy images in a few steps. The first step is image processing, followed by image segmentation to remove healthy skin from the mole. The image segmentation is done by thresholding. The system uses a 2D-wavelet transform to extract features from the segmented images (mean, standard deviation, mean absolute deviation, L1 norm, L2 norm). The last step is to classify PSLs, this is done by using backpropagation feed-forward ANN. The system uses 2 classes—cancerous and non-cancerous.

R.B. Aswin et al. [51] introduced a new method using for dermoscopy PSLs classification using the genetic algorithm and ANN. Their method consists of 4 steps. Firstly, the images are pre-processed by using hair removal software called Dullrazor [52]. The second step-segmentation uses Otsu color thresholding. The third step is a feature extraction. Two feature extraction methods were employed—GLCM (Gray level co-occurrence matrix) and normalized red, blue, green color features. Seven features are extracted and passed to ANN. The ANN used in this work is optimized by the Genetic algorithm. The output of the classifier is cancerous or non-cancerous class. The accuracy achieved by this system is 88%.

4.2.2 Deep learning methods

Previously, the research works used mostly hand-crafted features for the classification of PSLs or used shallow neural networks. L.Yu et al. [53] proposed a novel method for the classification of melanoma using deep learning. The deeper neural network is used to extract richer features from the image of PSLs. The proposed work consists of

two main steps—segmentation and classification. The segmentation is done by Fully Convolutional Residual Network (FCRN) which has 16 residual blocks. The classification is done by using a similar network as FCRN for segmentation. The classification is given by average of Softmax and SVM. The accuracy of this automated method is 85.5 % with segmentation and 82.8 % without segmentation. The dataset used in this work is ISIC 2016 [54]. This work won the classification part of the ISIC2016 challenge with 64% balanced accuracy.¹

K. Matsunaga et al. [55] used the approach of ensembling models. They formed a classification model which classified PSLs into 3 classes—melanoma, nevocellular nevus, and seborrheic keratosis. Their approach formed modified ResNet-50 for binary classification of Seborrheic keratosis and the rest of the classes. Then they used modified ResNet-50 for binary classification of melanoma and rest. Metadata such as age and sex are used as input after the images are passed through the ResNet-50. The last step is to classify the images into 3 classes. The proposed method achieved first place with a balanced accuracy of 91.1% in the ISIC2017 challenge.²

D.B. Mendes et al. [56] proposed a deep CNN based on ResNet152 architecture to classify 12 different types of PSLs. The work uses a combination of the MedNode dataset and the Dermofit Image Library, both discussed in see “Sec. 5.1”. The accuracy of this method is 78%.

U.O. Dorj et al. [57] propose a method for the classification of PSLs into 4 classes. AlexNet followed by SVM is used for classification. The work formed a custom dataset by searching for images on the internet (the work does not specify the source of the images). The research reports average accuracy between 91.8% and 95.1% for the 4 classes.

The best solution in the ISIC2018³ for the classification by A. Nozdryn-Plotnicki et al. [58] trained more than 15 different models, including different versions of ResNet, SeResNeXt, InceptionNet etc. Then the models were ensembled. Then the yielded probabilities are considered as features, passed to the XGBoost classifier for training. Their best model (ensemble of 10 models) achieved an accuracy of 88.5% in the ISIC2018 challenge.

N. Gessert et al. [59], the winning solution of the ISIC2019⁴ challenge, trained various models (mostly models from the EfficientNet family) and formed an ensemble from them. They report a sensitivity of 72.5% and an AUC of 95.4%. Among 9 classified classes was also an unknown class of lesions. Their second approach using images with metadata achieved sensitivity of 74.2%.

Q. Ha et al. [60] trained 18 models: various instances of EfficientNet-B3, B4, B5, B6, SE-ResNext101, ResNet101. The models’ output size was 9, however, the objection was to classify benign and malignant lesions, so the output was later reduced to 2. The method also used 14 metadata features. The metrics reported include only AUC which is 0.96. The solution ranked 1st in the ISIC2020 challenge⁵.

4.3 Self-examination systems for PSLs

In this section, we shall describe the state of the art in self-examination systems for PSLs, which can be used by anybody to diagnose their PSLs.

¹ <https://challenge.isic-archive.com/leaderboards/2016/>

² <https://challenge.isic-archive.com/leaderboards/2017/>

³ <https://challenge.isic-archive.com/leaderboards/2018/>

⁴ <https://challenge.isic-archive.com/leaderboards/2019/>

⁵ <https://www.kaggle.com/competitions/siim-isic-melanoma-classification/leaderboard>

Dermofit library and Atlas dataset (19398 images in total). The research developed a smartphone app (not specified closer) and a web application¹. The reported average accuracy of the model is 86.4% on the Dermofit library. The work does not explore the detection of PSLs.

AI Dermatologist² is a commercial system for the classification of PSLs. It is available for iOS and Android devices, as well as web application. It classifies a single lesion without the detection of PSLs. The authors do not report any metrics, nor the dataset used for training, or the model used.

4.4 Summary

In this chapter, we have described the state-of-the-art in the field of computerized diagnosis of PSLs.

The field of the classification of the PSLs is advancing rapidly, thanks to the ISIC challenges, which provided large datasets of PSLs in years the 2016, 2017, 2018, 2019, and 2020. The best results are achieved using deep learning methods, which are able to achieve better than human-level performance.

The situation is different in the field of detection of PSLs, where the research is not advancing as fast. The main reason is the lack of datasets of whole-body images of PSLs, every work uses its own dataset, which are mostly not publicly available.

The works that focus on detection, do not continue the classification of the detected skin lesions. The works that created a system for self-examination of PSLs are focusing only on the classification of the PSLs. The user needs to take a picture of a single lesion or crop the image to select the lesion, which he wants to classify.

Another problem is that freely available self-examination systems are stale, they are not updated for several years, or they are taken down from the app stores. Although this is not the case of the commercial systems, their performance is mostly not reported.

Therefore, a fully automated system is needed for the detection and classification of PSLs. With the least amount of user interaction. The system needs to be able to detect whether the image is of sufficient quality for detection and classification. Then be able to detect and classify PSLs in the images provided by the user.

¹ <https://b2017.modelderm.com/>

² <https://ai-derm.com/>

Chapter 5

Data

This chapter discusses available datasets for the classification and detection of PSLs. We also, introduce our own dataset of wide-area skin lesion images called DermaNude dataset.

5.1 Datasets

The datasets for the classification can be split into two subcategories: the datasets using clinical images and dermoscopic images. The clinical images are taken by a normal camera such as one in the smartphone. These types of images are usually of low quality, different lighting, angle, etc. The dermoscopic images are more useful for CAD systems, because of the better quality, more information can be extracted. These images are taken with dermoscope, which is an instrument used by dermatologists to examine the skin. The dermoscope creates uniform lighting and magnification of the skin lesion, which allows to see the details of the skin lesion.

Although there are many datasets available for the classification of PSLs, the situation is different for the detection of PSLs. This is caused mainly by ISIC challenges¹, which were focused on the classification of PSLs. Each year, the ISIC challenge provided a dataset of PSLs for the classification. However, the dataset for detection of PSLs was not provided. While the images of datasets for the classification of PSLs contain only one skin lesion, the images of datasets for detection of PSLs contain multiple skin lesions.

The dataset called Dermofit Image Library² is a collection of 1300 dermoscopic images of pigmented skin lesions. It contains 10 different types of skin lesions including Actinic Keratosis, Basal Cell Carcinoma, Melanocytic Nevus (mole), Seborrheic Keratosis, Squamous Cell Carcinoma, Intraepithelial Carcinoma, Pyogenic Granuloma, Haemangioma, Dermatofibroma, Malignant Melanoma.

Dataset PH2³ consists of 200 dermoscopic images of pigmented skin lesions. It contains 3 different types of skin lesions: common nevi, atypical nevi, and melanoma.

Dataset HAM10000⁴ contains 10015 dermoscopic images of pigmented skin lesions. It contains 7 different types of skin lesions: Actinic Keratosis, Basal Cell Carcinoma, Benign Keratosis, Dermatofibroma, Melanocytic Nevus, Melanoma, Vascular Lesion.

Challenge ISIC2019⁵ was focused on skin lesion classification. They provide a dataset made out of 3 different datasets: HAM10000 [71], MSK [72], and BCN20000[73]. The dataset contains 25 331 dermoscopic JPEG images from 8 different classes, same as in the HAM10000 but one more class—Squamous cell carcinoma.

Challenge ISIC2020 features a dataset of 33 126 images of pigmented skin lesions[74]. It features 5 classes—nevus, melanoma, seborrheic keratosis, lentigo, and also unknown

¹ <https://challenge.isic-archive.com/>

² <https://licensing.edinburgh-innovations.ed.ac.uk/product/dermofit-image-library>

³ <https://www.fc.up.pt/addi/ph2%20database.html>

⁴ <https://challenge2018.isic-archive.com/>

⁵ <https://challenge2019.isic-archive.com/>

class. This dataset contains fewer classes than the previous ISIC challenges as the ISIC2020 challenge focused on the melanoma classification: the goal was to classify the images into benign or malignant type of lesions.

DermNet image library¹ is a collection of clinical images of various lesions. Although the dataset consists of text description of the lesions, it is not suitable for computerized diagnosis of PSLs as is, because the images are not labeled. Further preprocessing is required to label the images. However, this dataset consists a wide range of skin diseases and is a good source of clinical images for PSL classification, as well as for PSL detection.

Interactive atlas of dermoscopy² [75] contains more than 2000 dermoscopic images of various skin lesions. This dataset is not free of charge.

Another notable dataset that contains clinical images of PSLs is the MED-NODE dataset³. This dataset has 100 images of naevus and 70 images of melanoma.

3dBodyTex.v1 is a dataset [76] which consists of 400 scans of 200 people (100 males, 100 females) in close-fitting clothing in various positions. Moreover, each scan is accompanied by texture. The dataset does not contain any annotations of the skin lesions. The annotations, called Skin3D, were later created by the authors Mengliu Zhao, Jeremy Kawahara et al. [42] and are available on GitHub⁴. Their work provides over 25 000 manually annotated skin lesions which are visible on textures of 3dBodyTex.v1 dataset. Their work divided the 3dBodyTex.v1 dataset into 3 groups - train set, validation set, and test set. The test set has been simultaneously annotated by 3 different annotators and is divided into 3 subsets accordingly. The newest addition to the 3DBodyTex.v1 dataset is the 3DBodyTex.v2 dataset⁵. It contains 3000 static 3D human scans of 500 subjects. The dataset lacks annotations of the skin lesions. The 3DBodyTex datasets are publicly available after signing the license agreement.

Renderpeople⁶ is a company that provides 3D models of people. They also provide 3D scans of people in swimsuits. These scans may be used for the detection of PSLs. However, these scans are not free of charge.

5.2 DermaNude dataset

In this section, we will discuss the creation of the DermaNude dataset. This dataset has been created to reflect the real-world conditions of skin lesion detection. Another reason, for the creation of this dataset, is that the skin is more complex (the skin pores, hair, etc.), and our dataset captures this complexity better than the 3DBodyTex.v1 dataset where the quality of the skin images is lower. Better quality of images means that more information can be extracted from the images, which can lead to better performance of the models, trained on this dataset. Lastly, the only publicly available annotated dataset for the detection of PSLs is 3DBodyTex.v1, therefore a new publicly available dataset is needed. We have collected 102 images of skin from 10 persons or subjects as referred later. We have annotated the images in the form of bounding boxes. There are a total of 2120 bounding boxes. The annotation tool we have used is called Makesense⁷, which is a free online tool licensed under GPL-3.0 License. We also collect the age of

¹ <https://dermnetnz.org/image-library>

² <https://dermoscopy.org/>

³ https://www.cs.rug.nl/~imaging/databases/melanoma_naevi/

⁴ <https://github.com/jeremykawahara/skin3d>

⁵ <https://cvi2.uni.lu/3dbodytexv2/>

⁶ <https://www.renderpeople.com/>

⁷ <https://www.makesense.ai/>

the subjects and their gender. The dataset is divided into 2 parts: images taken by a DSLR camera and images taken by a smartphone. The specifications of the dataset can be seen in “Tab. 5.1”. It is important to note, that all subjects have given their consent for the use of their images in this dataset, and the dataset follows the EU General Data Protection Regulation 2016/679 of 27 April 2016. As the dataset includes the images of people, but the faces are not visible, the images are not considered personal data according to the GDPR. However, age and gender of the subjects are considered personal data according the Article 4 of the GDPR. The consent form can be seen in “App. A”. It is also important to note, that not all subjects have given their consent to share their information in the dataset to the third parties, therefore only a subset of the dataset is publicly available. To get access to the dataset, one needs to fill the license form in “App. B” and send it to the author of this thesis. The license form, as well as contact information is available on the official website we have created for the dataset¹. This website also features the consent form in case one wants to contribute to the dataset. The desktop and smartphone version of the website is shown in “Fig. 5.1”.

	camera	smartphone
number of subjects	8	10
number of images	47	55
males	7	8
females	1	2
camera type	NIKON D3400	Realme 8, Iphone 12
number of skin lesions	898	1222
image type	.jpg	.jpg

Table 5.1. Specifications of the DermaNude dataset.

This dataset has been created for the purpose of skin lesion detection and classification. The example of the dataset is shown in “Fig. 5.2” and in “Fig. 5.3”. It can also be seen, that images contain backgrounds. We have decided to keep the backgrounds in the dataset.

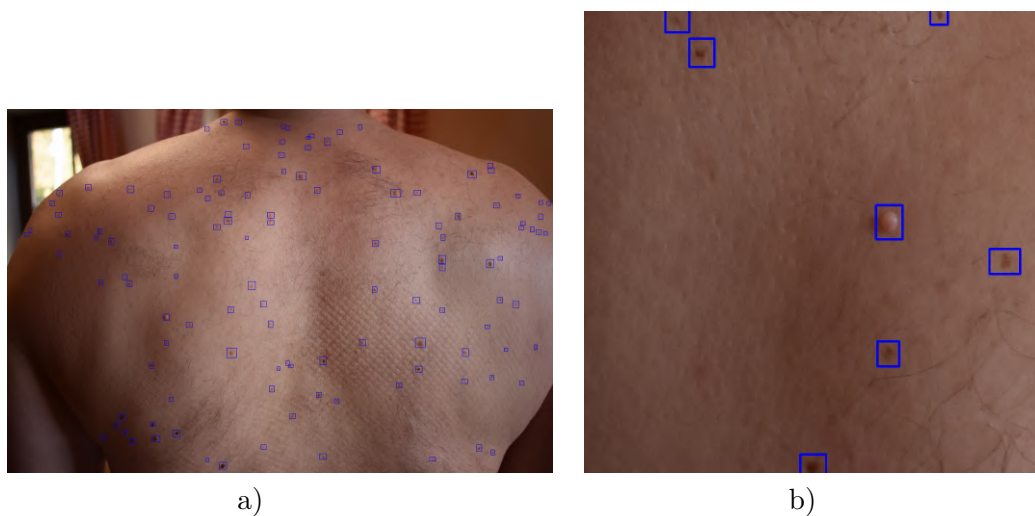
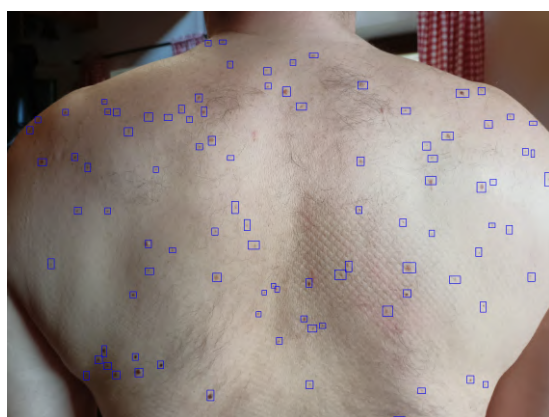


Figure 5.2. Image taken by NIKON D3400 camera (24MP) and its close-up (1MP) + annotations.

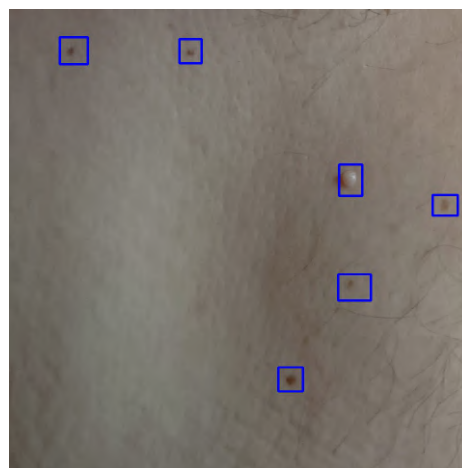
¹ <https://samuelsur.xyz/dermanude>



Figure 5.1. The official website of the DermaNude dataset. TOP: desktop layout. BOTTOM: smartphone layout.



a)



b)

Figure 5.3. Image taken by Realme 8 smartphone (16MP) and its close-up (1MP) + annotations.

Chapter 6

PSL detection implementation

In this section, we discuss our approach to the PSL detection. As the goal of this thesis is to build an automated system for the detection and classification of the PSLs, the first step is to build the detection part. The goal of this chapter is to train and evaluate the performance of the deep learning model, which will locate the PSLs in the image. Firstly, we discuss the datasets and their preprocessing.

6.1 Our approach

We acquired the 3dBodyTex.v1 dataset discussed in “Sec. 5.1”. We used the Skin3D annotations also discussed in “Sec. 5.1”. The Skin3D annotations divide the 3dBodyTex.v1 into 3 groups: train set, validation set and, test set. Each image in the test set has been simultaneously annotated by 3 different annotators. Therefore, the test set is divided into 3 subsets accordingly. Each subset can be used to evaluate the performance of the model independently.

The dataset preprocessing is the first step in our approach. Each texture image has 4096x4096 pixels. As you can see in “Fig. 6.2”, the quality of individual skin lesions is already low, therefore rather than resizing the image, we have chosen an approach to divide each texture image into 49 overlapping patches of size 800x800 pixels. An example of the overlapping patches can be seen in “Fig. 6.1”.

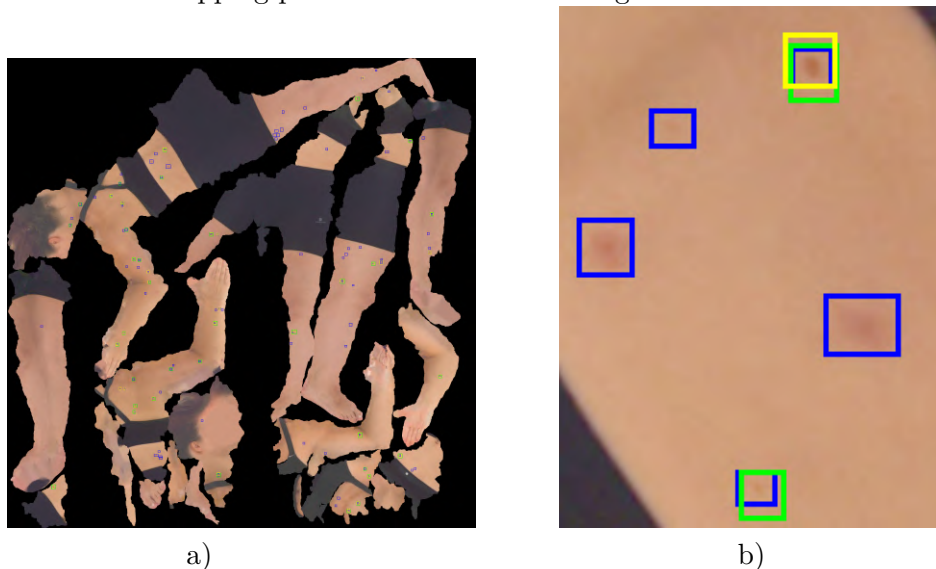


Figure 6.2. Texture images from 3dBodyTex.v1 dataset with bounding boxes from Skin3D dataset. (left: original image, right: close-up image)

The method of overlapping patches is a commonly used technique in neural networks for processing large images or datasets, and it offers several advantages over alternative methods. One major advantage is that it allows for better utilization of available computational resources by breaking down the input data into smaller, overlapping patches,



Figure 6.1. Original image (bottom) and its first 15 overlapping patches (top).

which reduces the memory and computation requirements of the network. Using overlapping patches rather than just slicing the image into non-overlapping patches can also help reduce the impact of boundaries that can arise when processing large images. This can lead to more accurate and reliable results.

For training of neural network on the dataset, we use the YOLOv7 with official pretrained weights. In this work, we rely on the implementation of the YOLOv7 by Chris Hughes¹. His implementation is far more adjustable for custom datasets than the original version². It is easier to use with custom datasets. To optimize the training process, we utilized Adam optimizer with a base learning rate of 0.001 and a weight decay of 0.0005. We tried different learning rates, both above and below 0.001, but we found that a learning rate ~ 0.001 works best for our dataset. We implemented a method of Non-maximum suppression with a Intersection over union (IoU) threshold of 0.01 to filter out overlapping bounding boxes. We evaluated the model’s performance using mean Average Precision (mAP) or Average Precision (AP), as we are detecting only one class, with an IoU threshold of 0.5, as described in the COCO competition³. We applied early stopping with a patience of 5 epochs, and the training would stop if the mAP on the validation dataset did not improve at least by 0.0001 for five consecutive epochs. Lastly, we utilized a cosine annealing learning rate scheduler, as described in this paper [77], but without restarts and with a warm-up period of 2 epochs. We tried to use also stable learning rate, as well as a cosine learning rate without warm-up period, but the results were better with a warm-up period in a cosine learning rate. The preset learning rate as defined by our learning rate scheduler for 30 epochs is shown in “Fig. 6.3”.

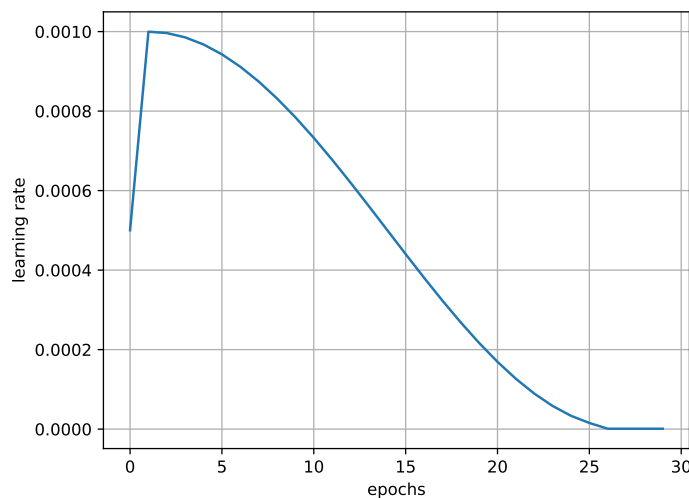


Figure 6.3. Preset learning rate for 30 epochs.

6.1.1 Centroid metric

For the purpose of evaluation of the model’s performance, we have implemented a centroid metric. The centroid metric is a metric, which checks whether the predicted bounding box contains the centroid of the ground truth bounding box and whether the ground truth bounding box contains the centroid of the predicted bounding box. If both conditions are met, the prediction is considered as true positive. If more bounding boxes meet the conditions, the bounding box with the highest confidence score is considered as true positive. After all ground truth bounding boxes were checked for match with

¹ <https://github.com/Chris-hughes10/Yolov7-training>

² <https://github.com/WongKinYiu/yolov7>

³ <https://cocodataset.org/#detection-eval>

prediction boxes, the ones without a match are considered as false negatives. The predictions without a match are considered as false positives. The average precision is then calculated, similarly to the Pascal VOC 2012 metric, as the area under the precision x recall curve. The “Fig. 6.4” demonstrates the centroid metric.

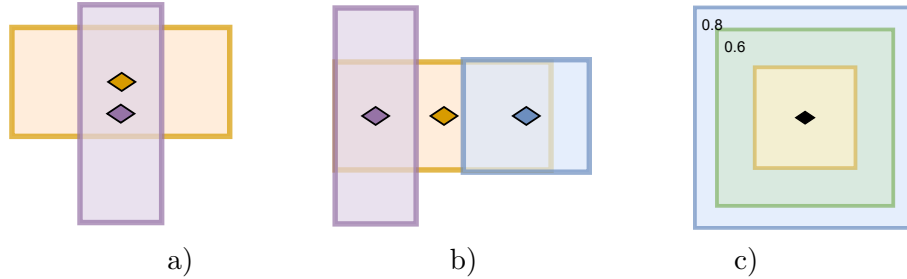


Figure 6.4. Centroid metric. For the purpose of this figure, we assume that the yellow box is the ground truth bounding box. The rhomboids represent the centroids of the bounding boxes. a) Both boxes contain the centroid of the other box. The pink box is therefore considered as TP. b) The pink box has the same IoU with ground truth bounding box as in a), but does not contain the centroid of the ground truth bounding box. c) Both green and blue bounding boxes contain the centroid of the ground truth bounding box and ground truth bounding box contains the centroid of both green and blue bounding boxes. However the blue box, although the IoU is worse than in the green box, has higher confidence score. Therefore the blue box is considered as TP and the green box as FP. boxes contain the centroid of the yellow box and yellow.

6.2 Results

After the model was trained using the parameters described in “Sec. 6.1”, we evaluated its performance on the test dataset. The model stopped training in epoch 10 after 5 epochs with no improvement in mAP. The evaluation metrics used on the test dataset are different from the ones used during training. We used Non-maximum suppression with IoU threshold of 0.01 to filter out overlapping predictions. On “Fig. 6.5”, “Fig.6.6”, “Fig.6.7” we can see the Precision x Recall curves of skin lesions predicted by our model for test datasets by each annotator. The Precision x Recall curves are calculated in Pascal VOC 2012 manner, which means that the Precision x Recall curve is calculated for each confidence threshold with IoU threshold of 0.5.

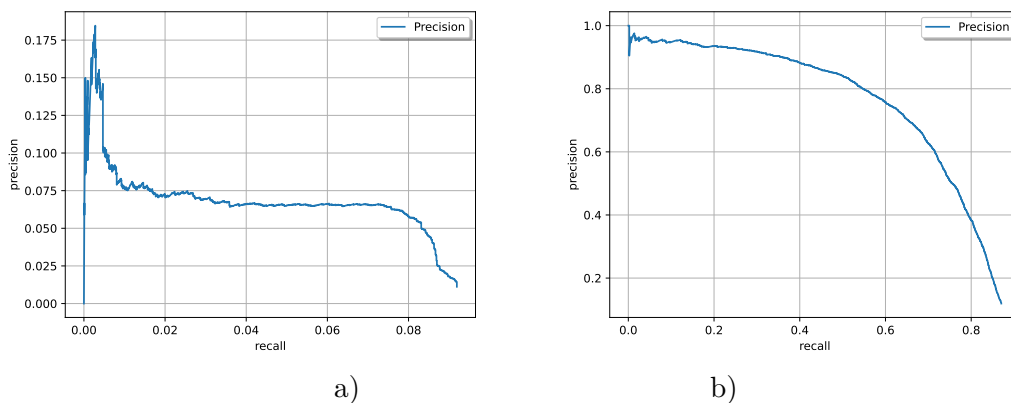


Figure 6.5. Precision x Recall curve for annotations A1. VOC 2012 metric (a) and centroid metric (b).

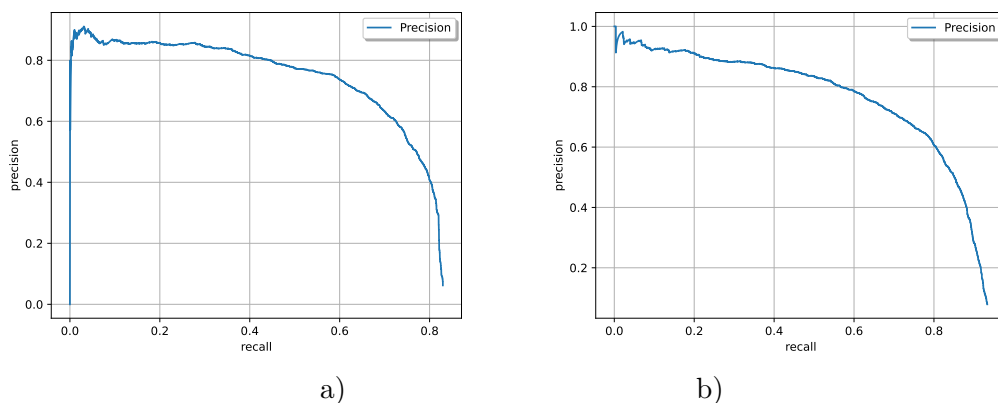


Figure 6.6. Precision x Recall curve for annotations A2. VOC 2012 metric (a) and centroid metric (b).

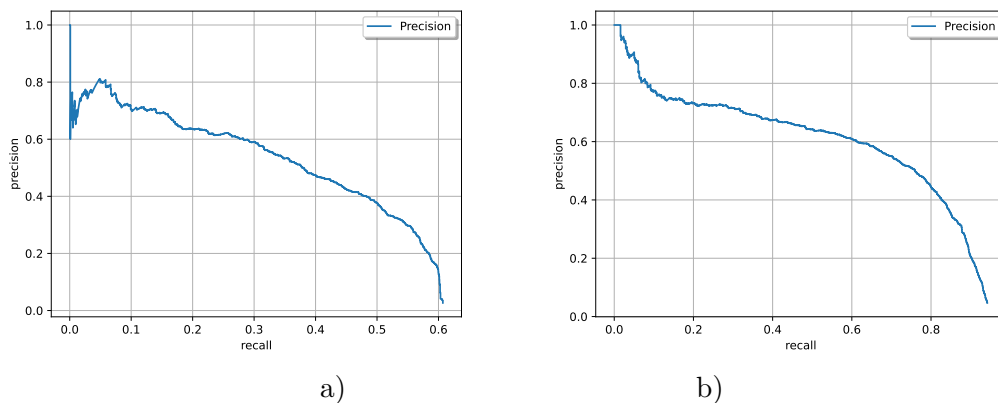


Figure 6.7. Precision x Recall curve for annotations A3. VOC 2012 metric (a) and centroid metric (b).

We provide the results computed using Pascal VOC 2012 metric—AP (IoU = 0.5) through all confidence thresholds “Tab. 6.3”. The Average Precision (AP) in Pascal VOC 2012 manner is computed as the area under the Precision x Recall curve [78][79]. We must note, that AP and mAP are in our case the same metric, as we are detecting only one class (PSL or not). It may be seen that the model does not perform well on the test dataset annotated by A1, with AP only 0.65%. The same is shown on Precision x Recall curve “Fig. 6.5”. The performance on the test dataset annotated by A2 “Fig. 6.6” and A3 “Fig. 6.7” is significantly better. Therefore we evaluate model performance also using AP and AR (average recall) using COCO metric. We calculate it using 2 different confidence thresholds c , which removes low confidence predictions. The results may be seen in “Tab. 6.2”. These results prove that the model does not perform well on the test dataset annotated by A1, similarly to the results in “Tab. 6.3”. Therefore, we study the annotations of A1 and compare them with the annotations of A2 and A3. The average size of the test set bounding boxes (through all image patches): 215 pixels for A1, 667 pixels for A2 and 372 pixels for A3. From this we can see that the bounding boxes in A1 are significantly smaller than in A2 and A3. These differences in bounding box sizes in test dataset may be seen in “Fig. 6.8”. Therefore, we compute results using the centroid metric, which we have introduced in “Sec 6.1.1”.

The results show, that the model performs significantly better on the annotations of A1—mAP 67.94%. The results for annotations A1 are even better than A3, while for VOC 2012 metric it was the opposite. The Precision x Recall curves for centroid metric may be seen in “Fig. 6.5”, “Fig. 6.6”, “Fig. 6.7”.

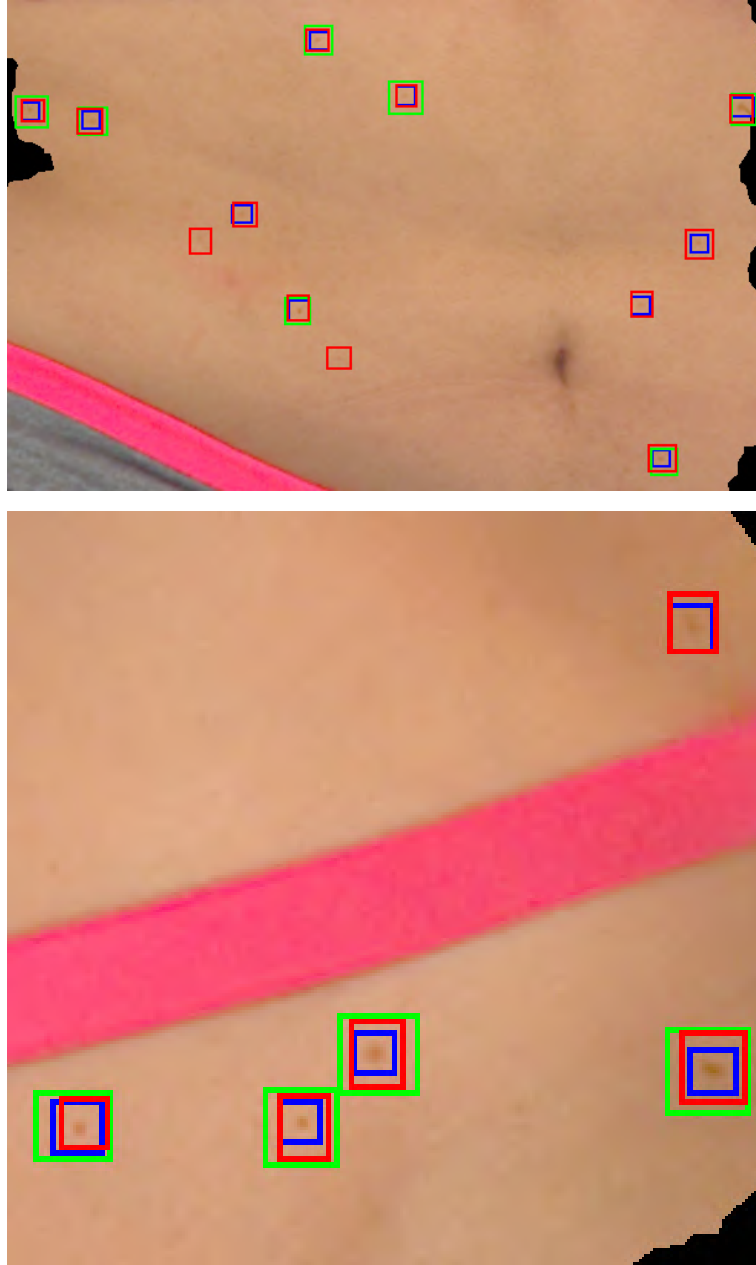


Figure 6.8. Bounding box sizes in test dataset (two different patches). Blue: A1, green: A2, red: A3.

6.3 PSL Detection on the DermaNude dataset

The next step is to evaluate the model’s performance on our DermaNude dataset. We used the same model as in the previous section, but we changed the input image size to 1600 pixels for the longer side of the image.

The problem with the evaluation of the model on the DermaNude dataset is that, the annotations for the 3dBodyTexv1 dataset, on which the model has been trained, are not tight around the skin lesion. The annotations for the DermaNude dataset are much closer to the skin lesion. As a consequence, the IoU threshold of 0.5, which is generally used for computing the mAP, is too high for the DermaNude dataset. The predictions and ground truth bounding boxes can be seen in an example from the DermaNude dataset in “Fig. 6.11”. For this reason, we evaluated our model on the DermaNude dataset with various IoU thresholds. Then we applied non-maximum suppression to filter out overlapping predictions. The results for camera and phone images are shown in “Tab. 6.1”. From this table, we can see that the model performs better on images made with smartphone than on images made with a DSLR camera. However, a similar problem arises as in the previous section. The predicted boxes may not be tight around the skin lesion. Therefore we compute the results using the centroid metric, which we have introduced in “Sec 6.1.1”. The results for camera and phone images are shown in “Tab. 6.5”. The results show, that the model performs similarly on the images made with smartphone and DSLR camera. This is expected as the smartphone cameras are of a good quality nowadays, however, the DSLR cameras are still better (less noise, better lighting, etc.).

The Precision x Recall curves for centroid metric, as well as Pascal VOC 2012 metric, may be seen in “Fig. 6.10” for smartphone images and in “Fig. 6.9” for camera images.

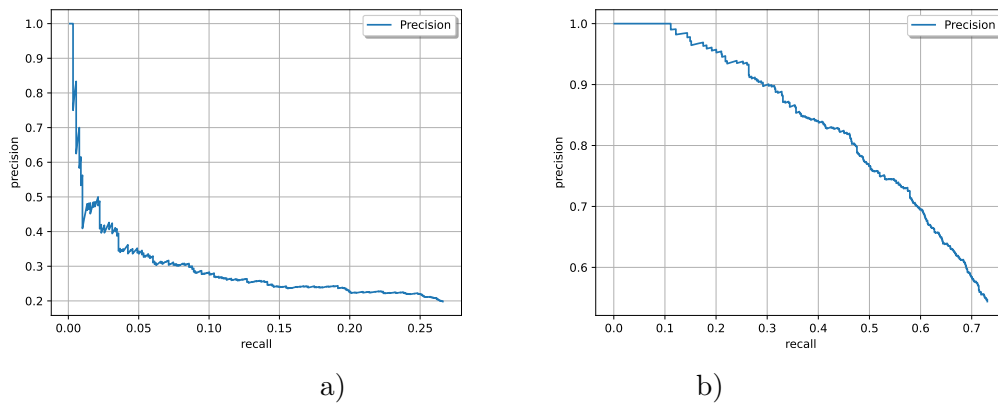


Figure 6.9. Precision x Recall curve—DermaNude camera. VOC 2012 metric (a) and centroid metric (b).

If the background is well lit and the color of the background is similar to skin color and contains dark spots similar to PSLs, the model tends to predict these black spots in the background (although with low confidence). This is shown in “Fig. 6.12”.

IoU threshold	mAP_{phone}	mAP_{camera}
0.5	21.62%	8.03%
0.4	35.36%	17.92%
0.3	47.39%	34.42%
0.2	56.51%	50.15%
0.15	58.15%	59.31%

Table 6.1. Results of the YOLOv7 model on the DermaNude dataset.

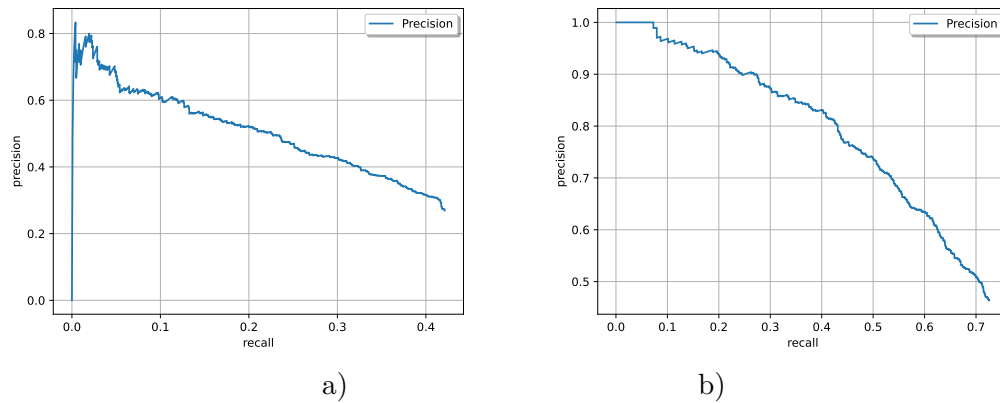


Figure 6.10. Precision x Recall curve—DermaNude smartphone. VOC 2012 metric (a) and centroid metric (b).

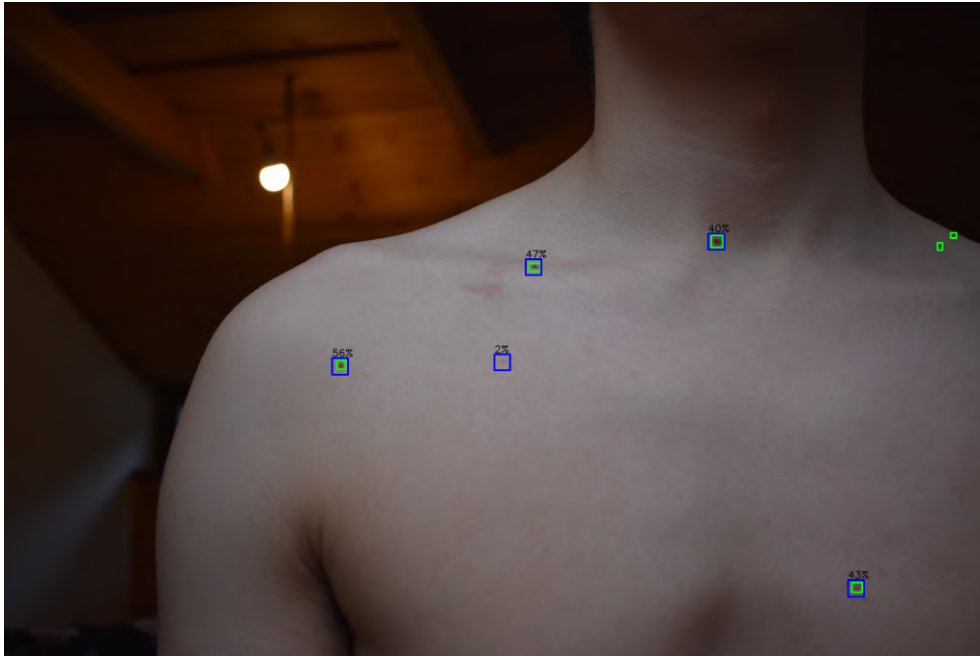


Figure 6.11. Comparison of ground truth (green) and predicted (blue) bounding boxes on the DermaNude dataset. The numbers above the bounding boxes are the confidence scores.

6.4 Conclusion

The most similar work to ours is the work of J. Kawahara et al [42]. Their work uses the 3DBodyTex.v1. For detecting skin lesions, they use Faster RCNN. They do not use overlapping patches, but they use a sliding window approach. We must note that we use overlapping patches but then we do not consider the overlapping regions as one, but as two different regions. Average Precisions, as stated in the paper by J. Kawahara et al. are compared with our results in “Tab. 6.3”. The table also includes our results on the DermaNude dataset, but as stated before, the IoU threshold of 0.5 is too high for both, 3DBodyTex.v1 and DermaNude dataset. The centroid metric is much more suitable for the evaluation of the model’s performance on PSL detection as the PSL boundaries are not always clear.



Figure 6.12. False positives on the DermaNude dataset. The model predicts the black spots in the background as PSLs, however, their confidence is low. Blue predictions, green ground truth.

Their work uses the same metric for evaluation on test dataset as ours. We can see that our work reaches a better score in the case of annotator A2, but worse in the case of annotator A3 and A1. Their work also proposes a metric of overlapping centroid.

On the “Fig. 6.13” we can see the predictions of the model and ground truth bounding boxes of annotator A2.

Annotator	AP_{50}		AR	
	$c = 0.25$	$c = 0.5$	$c = 0.25$	$c = 0.5$
A1	0.26%	0.14%	12.65%	4.34%
A2	47.07%	17.94%	45.78%	16.19%
A3	26.68%	16.02%	43.27%	21.10%

Table 6.2. Results of the model on test dataset using COCO metrics (AP and AR (Average Precision and Average Recall)).

Another work similar to ours is the paper by Ahmed-Aristizabal et al. [45]. As mentioned earlier, they use similar model to ours—Scaled YOLOv4. They also use 3dBodyTex.v1 dataset among others with custom made ground-truth data. They split

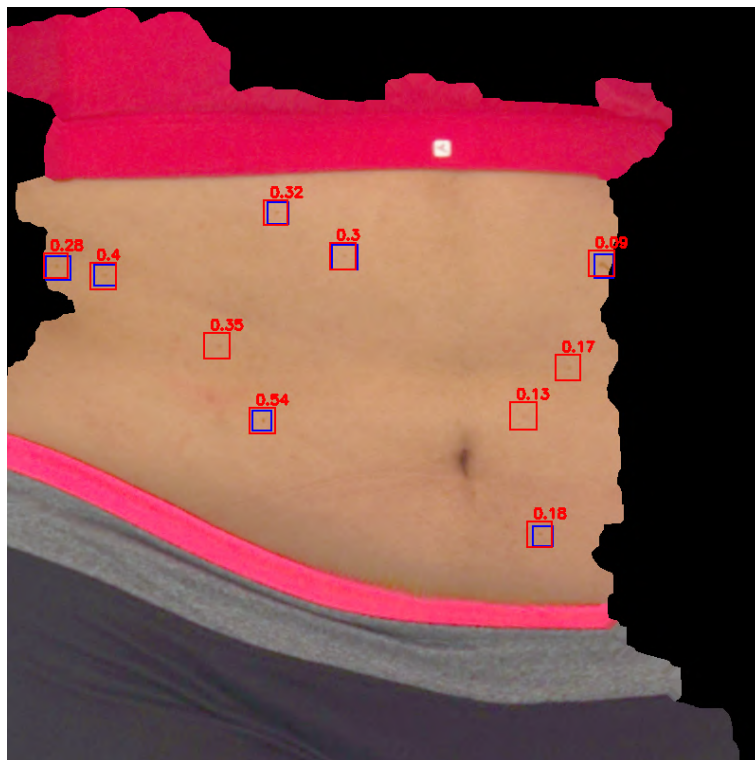


Figure 6.13. Prediction (red) and ground truth (blue) bounding boxes for annotations A2, with confidence scores (red text).

Annotator	$mAP_{Our\ model}$	$mAP_{J.Kawahara}$	$mAP_{our\ model, DermaNude\ dataset}$
A1	0.65%	8%	camera: 8.03% phone: 21.62%
A2	63.10%	59%	
A3	33.32%	56%	

Table 6.3. Comparison of Average Precision (mAP) (Pascal VOC 2012) of our model and model by J. Kawahara et al.

Annotator	$mAP_{Our\ model}$	$mAP_{J.Kawahara}$
A1	67.94%	75%
A2	72.73%	84%
A3	58.37%	77%

Table 6.4. Comparison of Average Precision (mAP), using the centroid metric, of our model and model by J. Kawahara et al.

mAP	camera	phone
	61.35%	58.49%

Table 6.5. Comparison of Average Precision (mAP) using centroid metric on the DermaNude dataset.

the dataset images into overlapping patches of a size 680x680. They use Soft-NMS (a novel type of NMS filter) to suppress bounding boxes [80]. The use of confidence score to filter out bounding boxes is mentioned, but not specified in the paper. Their results are $AP_{50} \approx 92\%$ (the type of mAP used is not reported in the work, therefore we do

not include it in the tables). Although this method differs from our approach in terms of ground-truth data, it introduces high precision results.

Chapter 7

PSL classification implementation

In this section, we will discuss our approach to the PSL classification. This is the second part of the fully automated system for PSL diagnosis. The classification program needs a single lesion image as an input and outputs the probability of the classes: Actinic Keratosis, Basal Cell Carcinoma, Benign Keratosis, Dermatofibroma, Melanocytic Nevus, Melanoma, Vascular Lesion, Squamous Cell Carcinoma. These classes are common in the available datasets. However, in practice, PSLs can be from different classes or not present at all. Therefore the “Unknown” class is also included in the classification program.

7.1 Our approach

We train three state-of-the-art CNN models for image classification. Two models are trained from scratch—Efficientnet-B0 and ResNet50 and one model using pretrained weights—EfficientNet-B7. We do not use layer freezing, we train all layers of the models. We use train, validation, test split in the ratio 60:20:20 using the preprocessed ISIC 2019 dataset¹. It features the same images as in the ISIC 2019 dataset, but they are preprocessed and resized to 512x512 pixels (the center crop of the original image). The dataset includes 8 classes, we present their names and abbreviations, which we use in the rest of the report: Actinic Keratosis (AK), Basal Cell Carcinoma (BCC), Benign Keratosis (BK), Dermatofibroma (DF), Melanocytic Nevus, Melanoma (MEL), Vascular Lesion (VASC), Squamous Cell Carcinoma (SCC). All of the models are trained on the train set. We use the validation set for checking the performance and overfitting of the models. Then we form different ensembles of the models and evaluated them with the individual models on the test set.

We use a number of augmentations during training to simulate real world data and therefore improve the generalization of the model. These augmentations include image transposition, flip, shift, scale, rotation, brightness change, contrast change, hue change, saturation change, cutout, etc. Although we use 512x512 images, we eventually resize them to 256x256. The training dataloader uses the weighted random sampler to balance the classes. We use different batch sizes for each model, to use the GPU efficiently. We use the Adam optimizer with learning rate of ≈ 0.001 . We discovered that the learning rate of ≈ 0.001 works well, therefore we did not use other learning rates. We follow the balanced accuracy metric on a validation set during training. If the model does not improve its balanced accuracy for 10 epochs, we stop the training. We save the model based on the best balanced accuracy on the validation dataset.

In practice, the input PSL can be of a different class than the classes in the dataset we use for training. It is also possible to get an image which does not contain a PSL. Therefore we add the “Unknown” class. We add it in postprocessing. If the maximum of the predicted probabilities is less than 0.5, we classify the PSL as “Unknown”.

¹ <https://www.kaggle.com/datasets/cdeotte/jpeg-isic2019-512x512>

7.1.1 Ensemble models

Our ensemble strategy is to average the predicted probabilities of the individual models.

$$Y^*(x) = \frac{1}{N} \sum_{i=1}^N y_i^*(x)$$

Their combinations and names are shown in “Tab. 7.1”.

Model combination	Model name
ResNet50 + EfficientNet-B0	Res50-Effb0
ResNet50 + EfficientNet-B7	Res50-Effb7
EfficientNet-B0 + EfficientNet-B7	Effb0-Effb7
ResNet50 + EfficientNet-B0 + EfficientNet-B7	Res50-Effb0-Effb7

Table 7.1. Ensemble models.

7.2 Results

In this section, we are presenting the performance of the models on the test set. The results in this section are computed for 8 classes, without the “Unknown” class.

When referring to sensitivity, we use a macro average of recall, which is the average of recall obtained per each class.

7.2.1 Individual models

The training progress of the EfficientNet-B0 model is shown in “Fig. 7.1”. The training progress of the ResNet-50 model is shown in “Fig. 7.2”. The training of EfficientNet-B7 can be seen in “Fig. 7.3”.

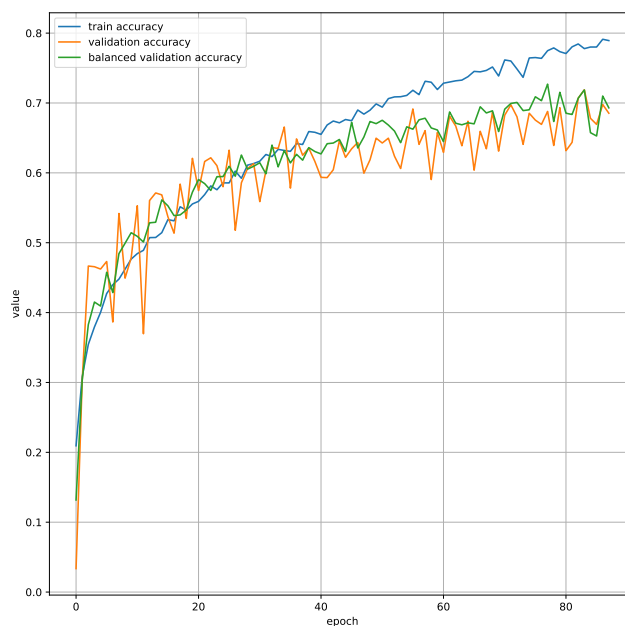


Figure 7.1. Training progress of the EfficientNet-B0 model, model evaluated on train set and validation set.

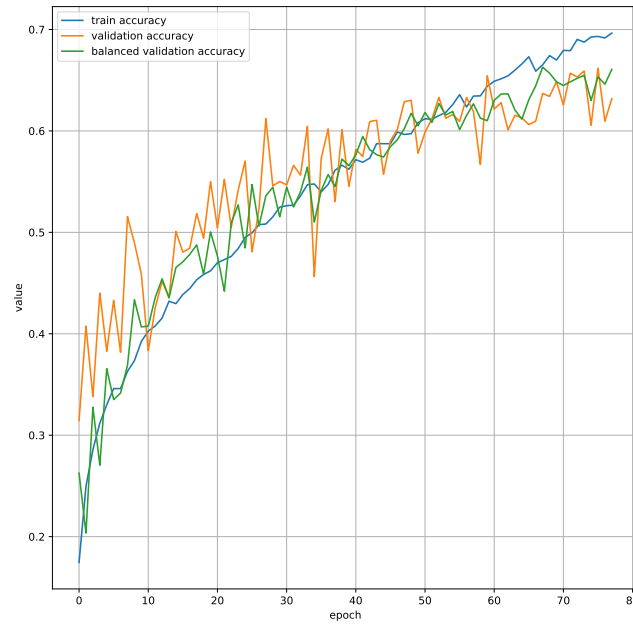


Figure 7.2. Training progress of the ResNet-50 model, model evaluated on train set and validation set.

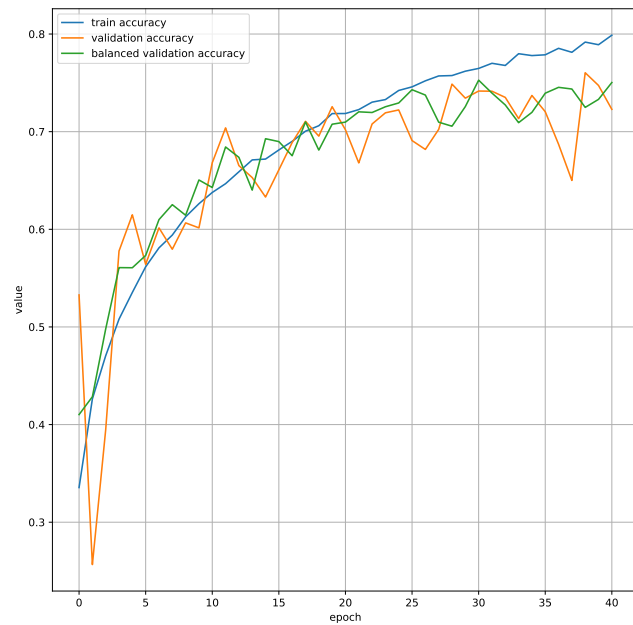


Figure 7.3. Training progress of the EfficientNet-B7 model, model evaluated on train set and validation set.

We can see, that the training of all models is relatively stable and the models do not overfit. Next, we shall present the performance on the test set.

In “Tab. 7.2” we can see the performance of the ResNet50 model on the test set without augmentations. The model achieved an accuracy of 0.63 and balanced accuracy of 0.64.

Class	Precision	Recall	F1-score	Support
AK	0.27	0.45	0.33	167
BCC	0.69	0.56	0.62	673
BKL	0.35	0.63	0.45	516
DF	0.23	0.89	0.37	45
MEL	0.66	0.40	0.50	910
NV	0.87	0.74	0.80	2568
SCC	0.23	0.57	0.33	128
VASC	0.54	0.90	0.68	60
accuracy			0.63	5067
macro avg	0.48	0.64	0.51	5067
weighted avg	0.71	0.63	0.65	5067

Table 7.2. Performance of ResNet50 on test set.

In “Tab. 7.3” we can see the performance of the EfficientNet-B0 model on the test set without augmentations. The model achieved an accuracy of 0.68 and balanced accuracy of 0.71.

Class	Precision	Recall	F1-score	Support
AK	0.42	0.59	0.49	167
BCC	0.71	0.72	0.71	673
BKL	0.51	0.60	0.55	516
DF	0.56	0.80	0.66	45
MEL	0.49	0.73	0.59	910
NV	0.92	0.66	0.77	2568
SCC	0.39	0.68	0.49	128
VASC	0.90	0.88	0.89	60
accuracy			0.68	5067
macro avg	0.61	0.71	0.64	5067
weighted avg	0.74	0.68	0.69	5067

Table 7.3. Performance of EfficientNet-B0 on test set.

In “Tab. 7.4” we can see the performance of the EfficientNet-B7 model on the test set without augmentations. The model achieved an accuracy of 0.73 and balanced accuracy of 0.74.

7.2.2 Ensemble models

In “Tab. 7.5” we can see the performance of the Res50-Effb0 model on the test set without augmentations. The model achieved an accuracy of 0.70 and balanced accuracy of 0.72.

Class	Precision	Recall	F1-score	Support
AK	0.36	0.71	0.48	167
BCC	0.79	0.70	0.74	673
BKL	0.64	0.57	0.60	516
DF	0.33	0.84	0.47	45
MEL	0.61	0.67	0.64	910
NV	0.88	0.79	0.84	2568
SCC	0.55	0.65	0.59	128
VASC	0.49	0.98	0.65	60
accuracy			0.73	5067
macro avg	0.58	0.74	0.63	5067
weighted avg	0.76	0.73	0.74	5067

Table 7.4. Performance of EfficientNet-B7 on test set.

Class	Precision	Recall	F1-score	Support
AK	0.40	0.57	0.47	167
BCC	0.73	0.71	0.72	673
BKL	0.47	0.64	0.54	516
DF	0.50	0.87	0.63	45
MEL	0.60	0.61	0.61	910
NV	0.89	0.75	0.82	2568
SCC	0.37	0.72	0.49	128
VASC	0.79	0.90	0.84	60
accuracy			0.70	5067
macro avg	0.60	0.72	0.64	5067
weighted avg	0.74	0.70	0.72	5067

Table 7.5. Performance of Res50-EffB0 on test set.

Class	Precision	Recall	F1-score	Support
AK	0.40	0.72	0.51	167
BCC	0.79	0.74	0.77	673
BKL	0.57	0.64	0.60	516
DF	0.39	0.93	0.55	45
MEL	0.66	0.61	0.63	910
NV	0.89	0.81	0.85	2568
SCC	0.53	0.70	0.60	128
VASC	0.64	0.97	0.77	60
accuracy			0.74	5067
macro avg	0.61	0.77	0.66	5067
weighted avg	0.77	0.74	0.75	5067

Table 7.6. Performance of Res50-Effb7 on test set.

In “Tab. 7.6” we can see the performance of the Res50-Effb7 model on the test set without augmentations. The model achieved an accuracy of 0.74 and balanced accuracy of 0.77.

Class	Precision	Recall	F1-score	Support
AK	0.49	0.69	0.57	167
BCC	0.78	0.79	0.79	673
BKL	0.64	0.63	0.63	516
DF	0.55	0.91	0.68	45
MEL	0.59	0.73	0.65	910
NV	0.90	0.77	0.83	2568
SCC	0.59	0.75	0.66	128
VASC	0.75	0.97	0.85	60
accuracy			0.75	5067
macro avg	0.66	0.78	0.71	5067
weighted avg	0.77	0.75	0.76	5067

Table 7.7. Performance of Effb0-Effb7 on test set.

In “Tab. 7.7” we can see the performance of the Effb0-Effb7 model on the test set. The model achieved an accuracy of 0.75 and balanced accuracy of 0.78.

Lastly, we made an ensemble from all individual models. The performance can be seen in “Tab. 7.8”. This model achieved an accuracy of 0.75 and balanced accuracy of 0.77.

Class	Precision	Recall	F1-score	Support
AK	0.46	0.68	0.55	167
BCC	0.77	0.78	0.78	673
BKL	0.57	0.64	0.61	516
DF	0.53	0.91	0.67	45
MEL	0.62	0.66	0.64	910
NV	0.89	0.79	0.84	2568
SCC	0.54	0.76	0.63	128
VASC	0.78	0.95	0.86	60
accuracy			0.75	5067
macro avg	0.65	0.77	0.70	5067
weighted avg	0.77	0.75	0.75	5067

Table 7.8. Performance of ensemble of all models on test set.

7.3 Conclusion on the performance of the models

From the previous, we can form “Tab. 7.9” which ranks the models based on their performance and compares them with state-of-the-art models.

The best solution in the ISIC 2019 challenge by N. Gessert et al. [59] trained various models and formed an ensemble from them. They report a sensitivity of 72.5% and AUC 95.4%. However, they trained their models also on “Unknown” class which is not present in our dataset. Only model that is comparable to our models from this work is EfficientNetB0, which achieves a sensitivity of 66.7% and AUC 94.0%. This is the only model trained on 8 classes from their work. Our EfficientNetB0 achieves a sensitivity of 70.7% and AUC 93.7.

Second place in the ISIC 2019 by S. Zhou et al. [81] trained multiple models and formed an ensemble from them. This work presents results for 8 classes. Their best-

Model	Balanced accuracy
Effb0-Effb7	0.78
Res50-Effb0-Effb7	0.77
Res50-Effb7	0.77
EfficientNetB7	0.74
Res50-Effb0	0.72
EfficientNetB0	0.71
ResNet	0.64
Gessert et al. [59] (8 classes)	0.68
Zhou et al. [81]	0.74

Table 7.9. Performance of the models.

performing individual model—SeResNeXt101 achieves a sensitivity of 73.9 %. Our best-performing individual model, EfficientNetB7 achieves the same sensitivity of 73.9%. Their best-performing ensemble formed from models EfficientNet-B2, EfficientNet-B4 and DenseNet121 achieves a sensitivity of 75.3%. Our best-performing ensemble Effb0-Effb7 is better with a sensitivity of 77.9%.

7.4 Experiments to study robustness of the models

While performance on the test set is important, it is also important to see how the model behaves to changes of the image parameters, which simulates the real world data. We study the impact of resolution, brightness, contrast, blur and color change on the performance of the models. Therefore we run the models on whole test dataset with particular change applied to it, then we evaluate the performance and increase/decrease the particular change and evaluate the performance again. We describe the results of the experiments and discuss them.

7.4.1 Resolution

In this section we study the impact of resolution change. This is important, as the image from the camera may have a different resolution than the images in the dataset. We accomplished this by applying downscaling followed by upscaling of the particular images in the test dataset.

The results of the resolution change may be seen in figure “Fig. 7.4”. The results show that the models are robust to the resolution change, and the ensemble models tend to perform better than the individual models.

7.4.2 Brightness

In this section, we study the impact of the brightness of the images on the model performance. We simulated the change of brightness of the images by applying the brightness change to the test dataset through the range of $[-1, 1]$, where -1 is the smallest brightness, and 1 is the maximum brightness, with step 0.05. The results in the form of balanced accuracy on the test dataset are shown in figure “Fig. 7.5”.

7.4.3 Contrast

In this section, we study the impact of the change in contrast in the images by applying the contrast change to the test dataset. We specified different values at which the

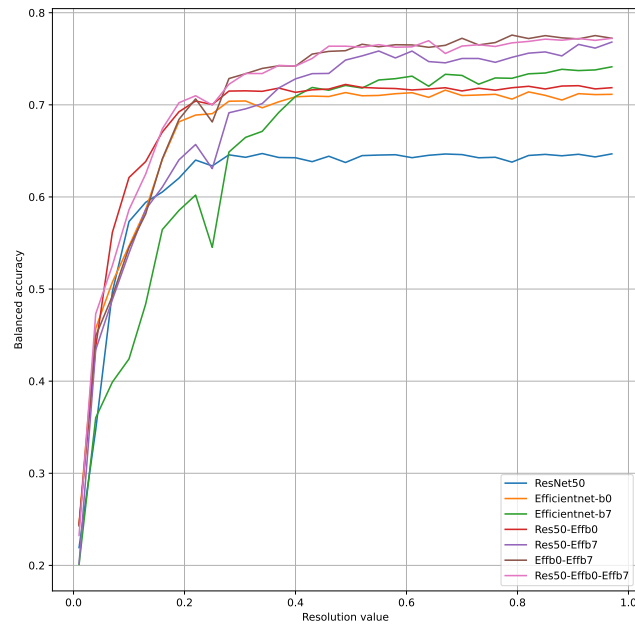


Figure 7.4. Impact of the resolution change on the balanced accuracy of the models.

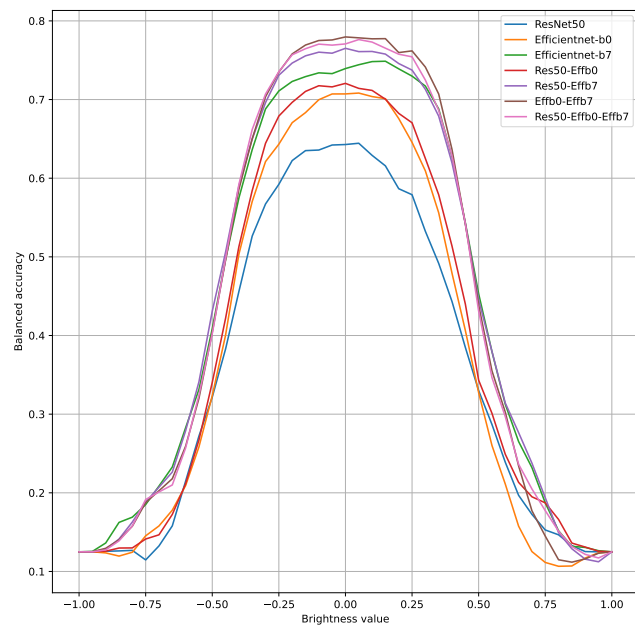


Figure 7.5. Impact of the brightness change on the balanced accuracy of the models.

contrast change is applied, in a range $[-1, 1]$ (-1 minimum, 1 maximum) with a step of

0.05. The results in the form of balanced accuracy on the test dataset are shown in “Fig. 7.6”.

The results show that, while the images with low contrast are harder to classify, the slight increase in contrast may improve the balanced accuracy.

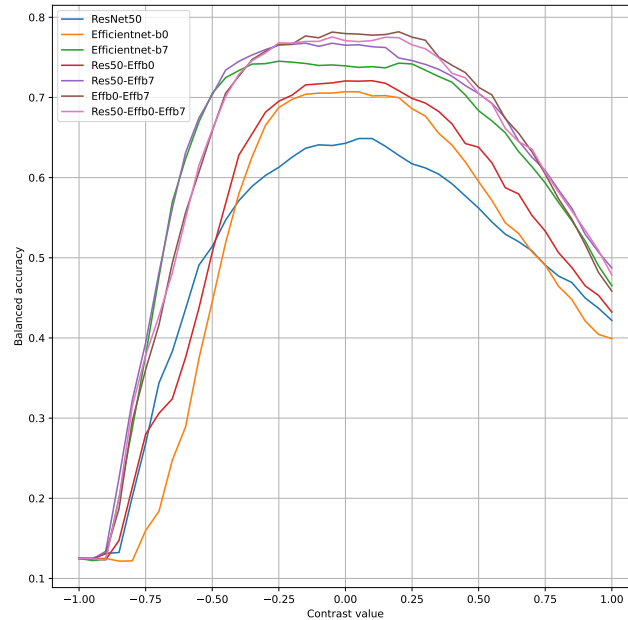


Figure 7.6. Impact of the contrast change on the balanced accuracy of the models.

7.4.4 Blur

We simulated the impact of blur, by changing the blur parameter of images in the test dataset. We specified different values at which the blur is applied, in a range $[1, 25]$ with a step of 1.

The results in the form of balanced accuracy on the test dataset are shown in figure “Fig. 7.7”. It may be seen, that the blur has a significant impact on the performance of the models. The worst performance is achieved by the EfficientNetB7 model, which has the best performance on test set, but the worst performance on the blurred test set after the value of 5.

7.4.5 Color

We simulated the change of color of the images by applying the RGB shift to the test dataset. We tested with shift of 0 or ± 50 applied to various channels. We tested all combinations of the shifts. The results in the form of balanced accuracy on the test dataset are shown in figure “Fig. 7.8”.

We can see, that the EfficientNetB7 model’s response to the color change is significantly better than the response of the other models. In some cases, it is better than the response of the ensemble models. The highest balanced accuracy is achieved when we shift all channels the same amount in the same direction. It is expected, as changing all channels the same amount in the same direction is equivalent to changing the brightness of the image. The worst performance is achieved when 2 channels are shifted in the same direction, and the third channel is shifted in the opposite direction.

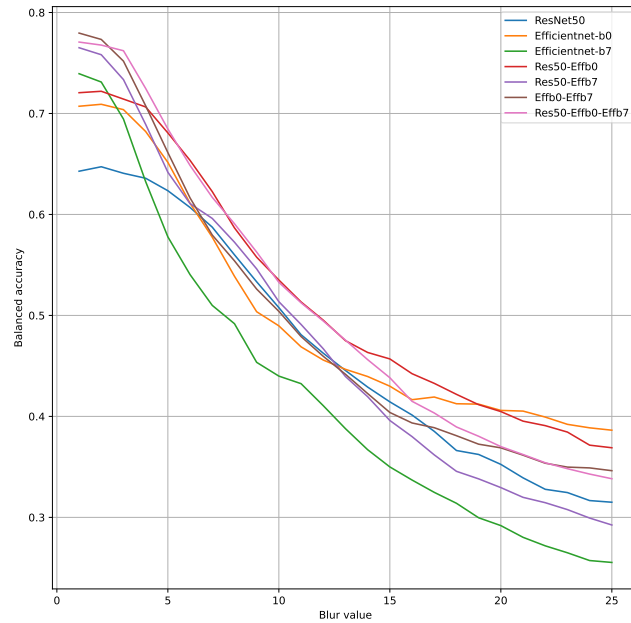


Figure 7.7. Impact of the blur change on the balanced accuracy of the models.

7.4.6 Experiments with unknown class

While 8 classes are supported by our classification models, we add one more class, during postprocessing, which is called “Unknown”. Here we study how the models behave to the “Unknown” class. The percentage of PSLs the models classified as “Unknown” is shown in table “Tab. 7.10”. We can see, that EfficientNetB7 model has the lowest percentage of PSLs classified as “Unknown”. While the ResNet50 model has the highest percentage of PSLs classified as “Unknown”.

Model	Percentage of Unknowns
EfficientNetB7	10.18%
Effb0-Effb7	11.88%
EfficientNetB0	14.01%
Res50-Effb0-Effb7	15.59%
Res50-Effb7	16.12%
Res50-Effb0	19.32%
ResNet50	24.22%

Table 7.10. Percentage of PSLs classified as “Unknown” from test set.

The next step is to study how the models behave when the image does not contain any of the 8 classes, on which the models were trained. Therefore we use images of plain skin, hair, background and other cropped images from the 3DBodyTex.v1 dataset and DermaNude dataset. Then we run the models on the images and count the number of images classified as “Unknown”.

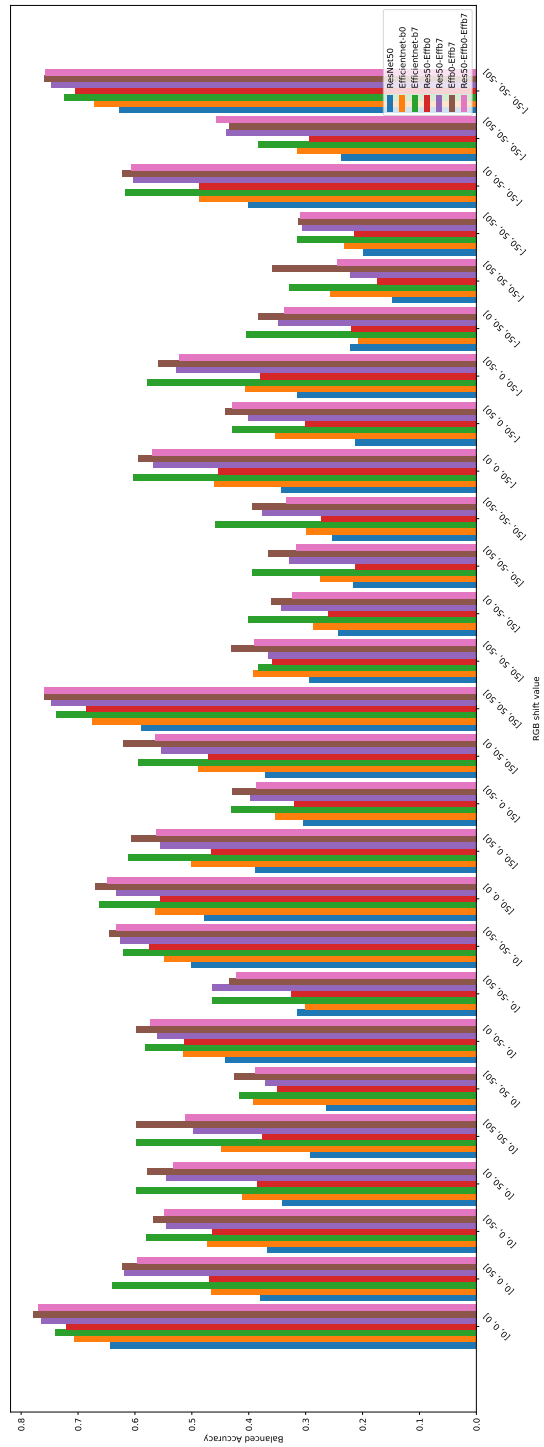


Figure 7.8. Impact of the color change on the balanced accuracy of the models.

We crop 37589 of bounding boxes from 3dBodyTex.v1, which do not overlap with the ground truth bounding boxes. Then we run the classification models on the cropped bounding boxes. The results are shown in table “Tab. 7.11”.

We do the same for the DermaNude dataset (phone and camera). The results are shown in table “Tab. 7.12”. The example of the cropped images is shown in figure “Fig. 7.9”.

The best ensemble model is Res50-Effb7, which achieves 91.48% of PSLs classified as “Unknown” on the 3dBodyTex.v1 dataset. The best individual model is EfficientNetB7, which achieves 86.37% of PSLs classified as “Unknown” on the 3dBodyTex.v1 dataset. In case of images by DermaNude camera, the best ensemble model is Res50-Effb0 with 85.08% of PSLs classified as “Unknown” and best individual model is EfficientNetB7 with 63.48% of PSLs classified as “Unknown”. In case of images by DermaNude phone, the best individual model is Resnet50 with 64.19% of PSLs classified as “Unknown” and best ensemble model is Res50-Effb7 with 56.49% of PSLs classified as “Unknown”.



Figure 7.9. Example of the cropped images from DermaNude camera (top), DermaNude phone (middle) and 3dBodyTex.v1 (bottom).

7.5 Conclusion of experiments

We trained various models on the ISIC 2019 dataset and formed an ensemble from them. Then we evaluated the performance of the models on the test set containing 8 classes (without the “Unknown” class). The best performing individual model, in terms of sensitivity is EfficientNetB7, which achieves a sensitivity of 73.9%. The best performing ensemble is EffB0-Effb7 and achieves a sensitivity of 77.9%. These results are better than the results of the best two solutions in the ISIC 2019 challenge. Although, ISIC2019 challenge includes classification of Unknown class, which is not

dataset	3dBodyTex.v1	
model	unknowns	percentage
Resnet50	28442	75.67%
EfficientNetB0	8569	22.80%
EfficientNetB7	32467	86.37%
Res50-Effb0	18502	49.22%
Res50-Effb7	34385	91.48%
EffB0-Effb7	8851	23.55%
Res50-Effb0-Effb7	15311	40.73%
total	37589	

Table 7.11. Unknown classification on images not containing any of the 8 classes—3dBodyTex.v1 dataset.

dataset	DermaNude—phone		DermaNude—camera	
model	unknowns	percentage	unknowns	percentage
Resnet50	2787	64.19%	3246	58.72%
EfficientNetB0	2125	48.94%	3735	67.57%
EfficientNetB7	1665	38.35%	3509	63.48%
Res50-Effb0	1862	42.88%	4703	85.08%
Res50-Effb7	2453	56.49%	4355	78.78%
EffB0-Effb7	1310	30.17%	2956	53.47%
Res50-Effb0-Effb7	1862	42.88%	4102	74.20%
total	4342		5528	

Table 7.12. Unknown classification on images not containing any of the 8 classes—DermaNude dataset.

present in their dataset, the best performing models report scores also without the Unknown class. In case of macro-averaged precision, the best performing individual model is EfficientNetB0 with 61%. The best performing ensemble is EffB0-Effb7 and achieves a macro-averaged precision of 66%. We have also studied the robustness of the models to various parameter changes of the images in the dataset. Examples of these changes are shown in figure “Fig. 7.10”. We found that the models are robust to the resolution change. A slight increase in contrast may improve the balanced accuracy of some models. The brightness change has significant impact on the performance of the models, when the brightness is $\pm 50\%$ the balanced accuracy drops below 50% for all models. The models’ response to blur change results in drop of the balanced accuracy for all models, starting from low values of blur. During the experiments with color, we found, that models are connecting the colors with the classes, it is expected as the color of the lesions is one of the most important features also used in the medical diagnosis.

We found that ensemble models are generally more robust to the changes in image parameters than the individual models. We also studied the the impact of the unknown class. The results show, that ResNet50 (individually or in ensemble) tends to classify more PSLs as “Unknown” than the other models, when the unknown class is not present in the dataset. In case of images not containing PSLs, it is similar—ResNet50 individually or in ensemble tends to classify more images as “Unknown” than the other models.

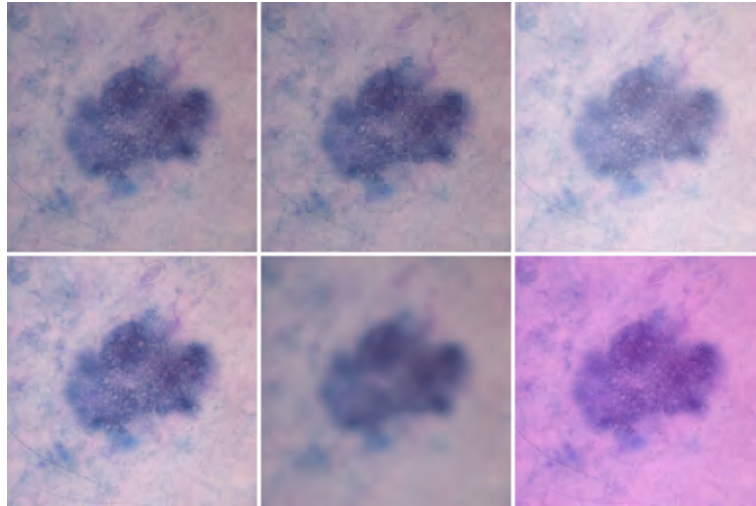


Figure 7.10. Examples of augmentations applied to the test dataset. From left to right; Upper row: original image, resolution change(0.4, equivalent to $\sim 100 \times 100$), brightness (+20%), Lower row: contrast (+20%), blur (10), color shift (+50, 0, +50)

Chapter 8

System for self-examination of skin lesions

We developed a system for self-examination of skin lesions in the form of CLI (Command Line Interface) application. The user inputs an image where he wants to check for the type of the lesion. The system controls the quality of the image using the BRISQUE algorithm. We rely on implementation by S. Kastruylin et al. [82], which is freely available¹. The BRISQUE algorithm is a no-reference image quality assessment algorithm, which outputs a score in a range [0, 100] (0 best, 100 worst). The mean average BRISQUE score on the DermaNude dataset shot by DSLR camera is ~ 10 with the worst value of ~ 24 . The mean average BRISQUE score on the DermaNude dataset shot on a smartphone camera is 42. The highest value is 54. Therefore we require the BRISQUE score to be at most 60. Another quality control is the size of the image. We require the image to be at least 24MP, which is easily achievable with modern smartphones. If the image passes the quality control, it is resized, so that the longer side has 1600 pixels, while preserving the aspect ratio. The image is then passed to the object detection model, which outputs the bounding boxes of the lesions. The output boxes are rescaled to the original input image size accordingly, the proposed bounding boxes are cropped and are passed to the classification model, which outputs the type of the lesions. The user can also specify the confidence threshold, which is used to filter out the detections with low confidence.

```
user@server:~/cli_app$ python3 main.py -i test.jpg -c 0.2
The results folder is not empty. Do you want to delete the files in the results folder? (y/n) y
checking image quality...
image size: 24 MP...ok
brisque score 17...ok

Image quality is good.
100%|#####| 4/4 [00:01<00:00, 2.97it/s]
image_name  detection_scores  NV      MEL      BCC      BKL      AK      SCC      DF      VASC      probs  unknown  preds
0  box_0      0.545997  0.973160  0.009893  0.001308  0.011684  0.000068  0.000265  0.001613  0.002008  0.844758  False  NV
1  box_1      0.511678  0.844758  0.057237  0.010067  0.015781  0.000700  0.001359  0.009953  0.060146  0.844758  False  NV
2  box_2      0.481734  0.884654  0.024471  0.005835  0.007909  0.000153  0.000620  0.063905  0.012453  0.884654  False  NV
3  box_3      0.400107  0.878126  0.015961  0.015796  0.011822  0.000379  0.000397  0.056616  0.020904  0.878126  False  NV
```

Listing 8.1. Example of the user-CLI interaction.

The output of the CLI application is shown in “Lis. 8.1”. The program outputs the results to the terminal and saves the results to the results folder, including image with bounding boxes, index, detection score, classification score and classified type of lesion. The results folder contains also the cropped bounding boxes, which are used for classification. The output dataframe consists of the image name, which is in form of “box_{*i*}”, where *i* is the index of the bounding box. The detection score is the confidence of the detection model, which is used to filter out the detections with low confidence. The next 8 columns are the classification scores of the 8 classes. The last column is the predicted class. The “unknown” column is used to indicate whether the model classified lesion as “Unknown”. The “probs” column is the maximum of the classification scores, which is used to determine the predicted class.

¹ <https://github.com/photosynthesis-team/piq>

The CLI application features loading animation, which is shown while the image is being processed or brisque score is being calculated. This feature is for better user experience, as these processes can take longer time.

We run couple of experiments, with different images. The first experiment is with the image taken by camera “Fig. 8.1”, where we can see the output of the CLI App. The model has found 4 PSLs.

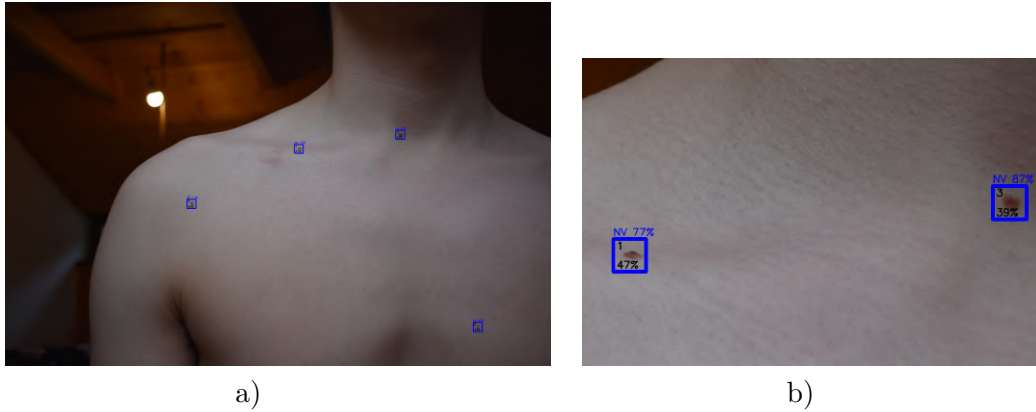


Figure 8.1. Output of the CLI App. a) Output of the CLI App. b) Close-up of the output image. The black text is the index and detection score of the bounding box. The blue text is the classification score and the predicted class.

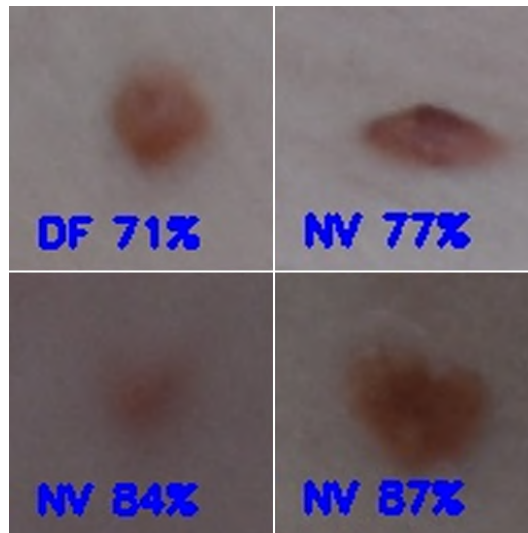


Figure 8.2. Collage from cropped bounding boxes by CLI app.

The application predicted, that there is one Dermatofibroma and 3 occurrences of Melanocytic Nevus. We run another experiment, with similar image.

This time with a smartphone, from different angle, different lighting conditions and background. The output of the CLI application is shown in “Fig. 8.3”. We can see, that although we have used low confidence for detection, the application successfully filtered/set to “Unknown” the detection of the background.

Same PSLs, which were predicted on both photos, are shown in “Fig. 8.4”. Although, the angle is different, the classifier classified the top-row PSLs as the same class (with different probabilities).

The bottom-row PSLs were predicted as completely different. The difference may be due to the better quality of the photo from the phone (in this case). Also the

lighting conditions seem to be better in the photo from the phone. The angle is also different, but is similar for the lesions in bottom row. We can conclude, that if we had correspondences between PSLs by using the earlier discussed methods, we could use the correspondences to improve the classification. The correspondences could be used to improve the classification by using multiple PSL images as input to the classification model. Then the output would be averaged or weighted according to image quality score by BRISQUE algorithm.

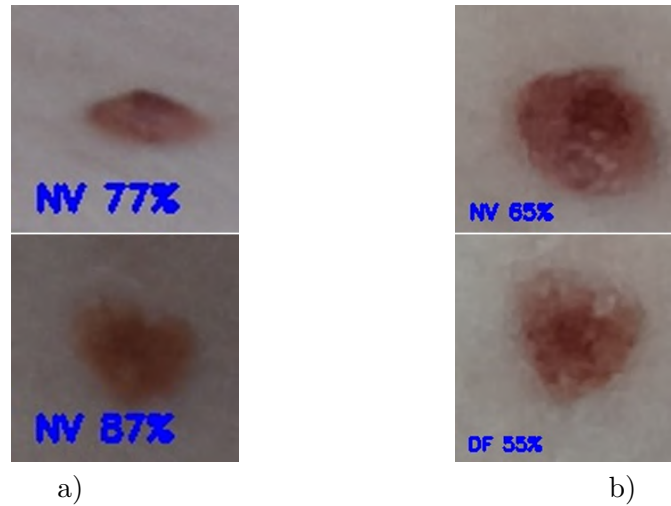


Figure 8.4. Same PSLs predicted by both photos. a) Photo taken by DSLR camera. b) Photo taken by smartphone. Although, the classifier detected the same lesions, the predicted classes by the classifier are different in case of bottom images.

Another example of the output of the CLI application is shown in “Fig. 8.5”. The number of lesions is much greater than in “Fig. 8.1” and “Fig. 8.3”. The application found mostly Melanocytic Nevus, but also a Vascular Lesion.

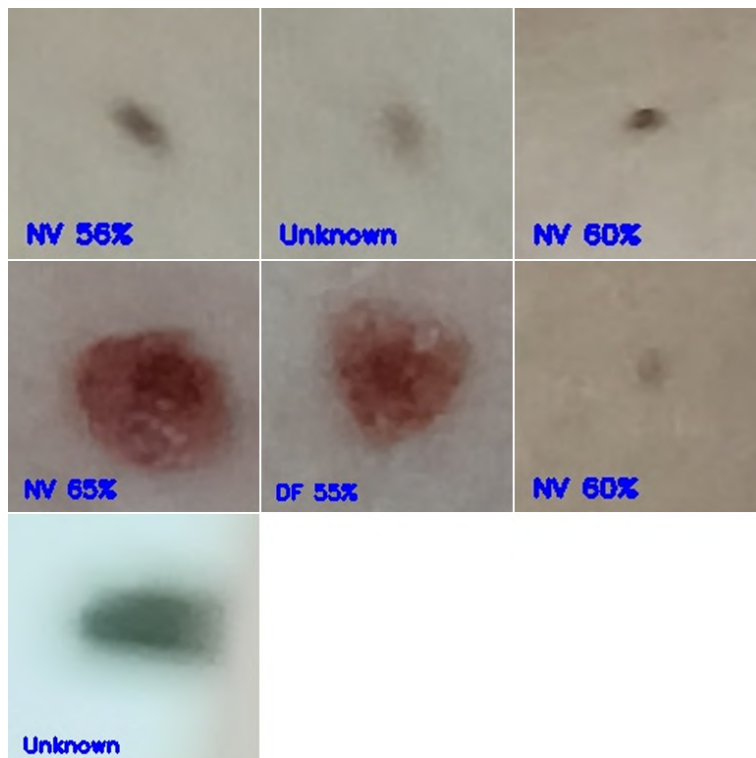


Figure 8.3. Output of the CLI App with changed conditions. Top: Output of the CLI App. Bottom: Collage from the cropped bounding boxes with the predicted classes and scores.

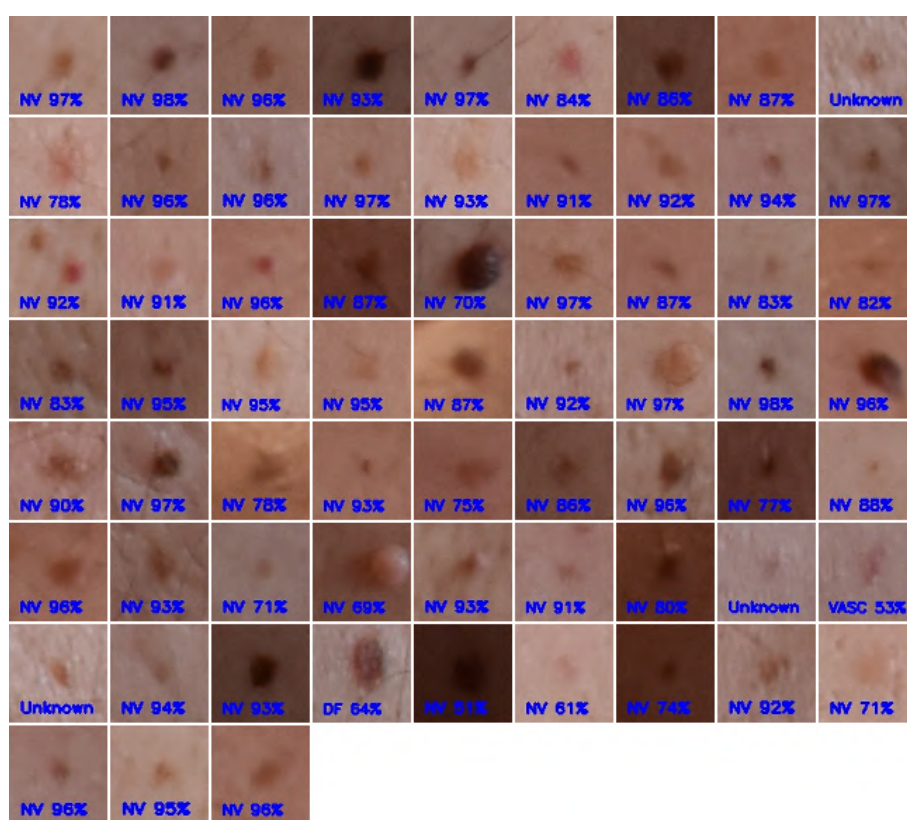
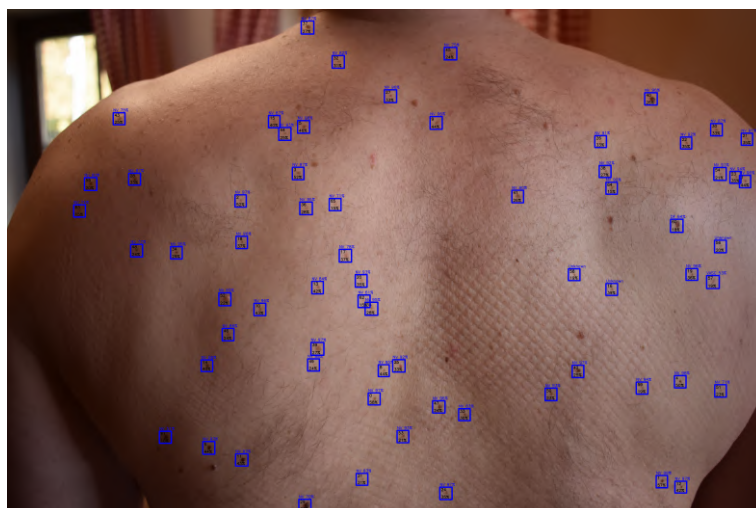


Figure 8.5. Output of the CLI App. Top: Output of the CLI App. Bottom: Collage from the cropped bounding boxes with the predicted classes and scores.

Chapter 9

Conclusion

We trained and evaluated the performance of the image classification and object detection state-of-the-art models on skin lesion images. We also created our own dataset of wide-area skin lesion images and tested the performance of the object detection YOLOv7 model on this dataset. Lastly, we combined the image classification and object detection models into a pipeline, which can be used to classify skin lesions in wide-area images. The advantage of this approach is that the user does not need to perform any manual cropping of the images, which is required by the image classification models.

9.1 Comparison of the original assignment and the final result

We have achieved all the main goals of the thesis. We studied available datasets and state-of-the-art in the diagnosis of PSLs. Using the existing state-of-the-art methods for object detection we trained and evaluated the performance of the YOLOv7 model on the 2D images from the Skin3D dataset. Using the existing state-of-the-art for object detection, we trained 3 individual models and created 4 ensemble models. Evaluated their performance as a function of image quality - resolution, blur, color changes. We also studied the contrast and brightness changes. The original assignment also mentions 2 optional goals. We have achieved one of them. We acquired a dataset using a mobile phone and DSLR camera and evaluated the developed methods on this data. The second optional goal is to find correspondences between objects of interest in different images using existing methods such as (SIFT, SURF, RANSAC, Coherent Point Drift) for object matching. We did not achieve this goal.

We have achieved one more goal, which was not mentioned in the original assignment: the creation of the fully automated system, which combines PSL detection and PSL classification.

9.2 Future work

In the future, the performance of the object detection model could be improved by introducing our own dataset into the training process and training more state-of-the-art models for object detection. The PSL classification can be done by introducing the “Unknown” class during training. We have the “Unknown” class, but we introduce it in the postprocessing. Another improvement could be to use SIFT/SURF and RANSAC to find correspondences between skin lesions on various images. These correspondences can help to classify the images in fully automated application more accurately. These correspondences could also be used later to track the skin lesions over time. Lastly, a new updated dataset of wide-area skin lesion images could be created using the photos of the same people as in our dataset, to create a dataset that could be used for tracking

lesions across time. Finally, a smartphone app could be created, which would make the whole process of taking the photos and diagnosing the skin lesions much easier for the user.

References

- [1] Divya Seth, Khatiya Cheldize, Danielle Brown, and Esther E. Freeman. Global Burden of Skin Disease: Inequities and Innovations. *Current Dermatology Reports*. 2017, 6 (3), 204–210. DOI 10.1007/s13671-017-0192-7.
- [2] Franz Nachbar, Wilhelm Stolz, Tanja Merkle, Armand B. Cognetta, Thomas Vogt, Michael Landthaler, Peter Bilek, Otto Braun-Falco, and Gerd Plewig. The ABCD rule of dermoscopy: High prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology*. 1994, 30 (4), 551–559. DOI 10.1016/S0190-9622(94)70061-3.
- [3] Kalyan Saginala, Adam Barsouk, John Sukumar Aluru, Prashanth Rawla, and Alexander Barsouk. Epidemiology of Melanoma. *Medical Sciences*. 2021, 9 (4), 63. DOI 10.3390/medsci9040063.
- [4] P. Viola, and M. Jones. *Rapid object detection using a boosted cascade of simple features*. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. 2001. I–I. ISSN: 1063-6919.
- [5] N. Dalal, and B. Triggs. *Histograms of oriented gradients for human detection*. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. 2005. 886–893 vol. 1.
- [6] Pedro Felzenszwalb, David McAllester, and Deva Ramanan. *A discriminatively trained, multiscale, deformable part model*. In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. 2008. 1–8.
- [7] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. *Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation*. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014. 580–587. ISSN: 1063-6919.
- [8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015, 37 (9), 1904–1916. DOI 10.1109/TPAMI.2015.2389824. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [9] Ross Girshick. *Fast R-CNN*. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015. 1440–1448. ISSN: 2380-7504.
- [10] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017, 39 (6), 1137–1149. DOI 10.1109/TPAMI.2016.2577031. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [11] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision*. 2020, 128 (2), 261–318. DOI 10.1007/s11263-019-01247-4.

- [12] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. *Feature Pyramid Networks for Object Detection*. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. 936–944. ISSN: 1063-6919.
- [13] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. *You Only Look Once: Unified, Real-Time Object Detection*. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. 779–788. ISSN: 1063-6919.
- [14] Joseph Redmon, and Ali Farhadi. *YOLO9000: Better, Faster, Stronger*. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. 6517–6525. ISSN: 1063-6919.
- [15] Joseph Redmon, and Ali Farhadi. *YOLOv3: An Incremental Improvement*. . 2018. arXiv.
<http://arxiv.org/abs/1804.02767>.
- [16] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. . 2020. arXiv.
<http://arxiv.org/abs/2004.10934>.
- [17] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*. 2022.
<https://arxiv.org/abs/2207.02696v1>. Publication Title: arXiv.org.
- [18] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. *SSD: Single Shot MultiBox Detector*. 2016.
<http://arxiv.org/abs/1512.02325>.
- [19] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. *Focal Loss for Dense Object Detection*. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017. 2999–3007. ISSN: 2380-7504.
- [20] Hei Law, and Jia Deng. CornerNet: Detecting Objects as Paired Keypoints. *International Journal of Computer Vision*. 2020, 128 (3), 642–656. DOI 10.1007/s11263-019-01204-1.
- [21] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. *Objects as Points*. . 2019. arXiv.
<http://arxiv.org/abs/1904.07850>.
- [22] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. *End-to-End Object Detection with Transformers*. . 2020. arXiv.
<http://arxiv.org/abs/2005.12872>.
- [23] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. 2017, 60 (6), 84–90. DOI 10.1145/3065386.
- [24] Karen Simonyan, and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. . 2015. arXiv.
<http://arxiv.org/abs/1409.1556>.
- [25] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. *Going deeper with convolutions*. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015. 1–9. ISSN: 1063-6919.

- [26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Deep Residual Learning for Image Recognition*. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. 770–778. ISSN: 1063-6919.
- [27] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. *Aggregated Residual Transformations for Deep Neural Networks*. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. 5987–5995. ISSN: 1063-6919.
- [28] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*. . 2017. arXiv. <http://arxiv.org/abs/1704.04861>.
- [29] Mingxing Tan, and Quoc V. Le. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. . 2020. arXiv. <http://arxiv.org/abs/1905.11946>.
- [30] D. A. Perednia, R. G. White, and R. A. Schowengerdt. Localization of cutaneous lesions in digital images. *Computers and Biomedical Research, an International Journal*. 1989, 22 (4), 374–392. DOI 10.1016/0010-4809(89)90032-3.
- [31] Raymond G. White, Douglas A. Perednia, and Robert A. Schowengerdt. Automated feature detection in digital images of skin. *Computer Methods and Programs in Biomedicine*. 1991, 34 (1), 41–60. DOI 10.1016/0169-2607(91)90081-4.
- [32] D. P. Filiberti, J. A. Gaines, P. Bellutta, P. Ngan, and D. A. Perednia. Efficient segmentation of large-area skin images: a statistical evaluation. *Skin Research and Technology*. 1997, 3 (1), 28–35. DOI 10.1111/j.1600-0846.1997.tb00156.x.
- [33] Bruce McGregor. Automatic registration of images of pigmented skin lesions. *Pattern Recognition*. 1998, 31 (6), 805–817. DOI 10.1016/S0031-3203(97)00107-6.
- [34] T.K. Lee, M.S. Atkins, M.A. King, S. Lau, and D.I. McLean. Counting moles automatically from back images. *IEEE Transactions on Biomedical Engineering*. 2005, 52 (11), 1966–1969. DOI 10.1109/TBME.2005.856301.
- [35] Taeg Sang Cho, William T Freeman, and Hensin Tsao. *A reliable skin mole localization scheme*. In: *2007 IEEE 11th International Conference on Computer Vision*. 2007. 1–8.
- [36] Jean-Sebastien Pierrard, and Thomas Vetter. *Skin Detail Analysis for Face Recognition*. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 2007. 1–8.
- [37] Konstantin Korotkov. *Automatic change detection in multiple pigmented skin lesions*. Ph.D. Thesis, Universitat de Girona. 2014. <https://www.tdx.cat/handle/10803/260162>.
- [38] Konstantin Korotkov, Josep Quintana, Susana Puig, Josep Malvehy, and Rafael Garcia. A new total body scanning system for automatic change detection in multiple pigmented skin lesions. *IEEE transactions on medical imaging*. 2015, 34 (1), 317–338. DOI 10.1109/TMI.2014.2357715.
- [39] Konstantin Korotkov, Josep Quintana, Ricard Campos, America Jesus-Silva, Pablo Iglesias, Susana Puig, Josep Malvehy, and Rafael Garcia. An Improved Skin Lesion Matching Scheme in Total Body Photography. *IEEE journal of biomedical and health informatics*. 2019, 23 (2), 586–598. DOI 10.1109/JBHI.2018.2855409.

- [40] J Matas, O Chum, M Urban, and T Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*. 2004, 22 (10), 761–767. DOI 10.1016/j.imavis.2004.02.006.
- [41] Christos Nikolaos E. Anagnostopoulos, Dimitrios D. Vergados, and Panagiotis Mintzias. *Image registration of follow-up examinations in digital dermoscopy*. In: *13th IEEE International Conference on BioInformatics and BioEngineering*. 2013. 1–4.
- [42] Mengliu Zhao, Jeremy Kawahara, Kumar Abhishek, Sajjad Shamanian, and Ghasan Hamarneh. Skin3D: Detection and Longitudinal Tracking of Pigmented Skin Lesions in 3D Total-Body Textured Meshes. *Medical Image Analysis*. 2022, 77 102329. DOI 10.1016/j.media.2021.102329.
- [43] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C. Russell, and Mathieu Aubry. *3D-CODED: 3D Correspondences by Deep Deformation*. In: Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, eds. Cham: Springer International Publishing, 2018. 235–251. ISBN 978-3-030-01215-1 978-3-030-01216-8.
- [44] Federica Bogo, Javier Romero, Enoch Peserico, and Michael J. Black. Automated Detection of New or Evolving Melanocytic Lesions Using a 3D Body Model. 2014, 8673 593–600. DOI 10.1007/978-3-319-10404-1_74.
- [45] David Ahmedt-Aristizabal, Chuong H. Nguyen, Lachlan Tychsen-Smith, Ashley Stacey, Shenghong Li, Joseph Pathikulangara, Lars Petersson, and Dadong Wang. Monitoring of Pigmented Skin Lesions Using 3D Whole Body Imaging. *Computer methods and programs in biomedicine*. 2022, 232 107451.
- [46] Bharat Lal Bhatnagar, C. Sminchisescu, C. Theobalt, and Gerard Pons-Moll. LoopReg: Self-supervised Learning of Implicit Surface Correspondences, Pose and Shape for 3D Human Mesh Registration. *ArXiv*. 2020,
- [47] Fengying Xie, Haidi Fan, Yang Li, Zhiguo Jiang, Rusong Meng, and Alan Bovik. Melanoma Classification on Dermoscopy Images Using a Neural Network Ensemble Model. *IEEE Transactions on Medical Imaging*. 2017, 36 (3), 849–858. DOI 10.1109/TMI.2016.2633551.
- [48] Ammara Masood, Adel Ali Al-Jumaily, and Tariq Adnan. *Development of Automated Diagnostic System for Skin Cancer: Performance Analysis of Neural Network Learning Algorithms for Classification*. In: Stefan Wermter, Cornelius Weber, Włodzisław Duch, Timo Honkela, Petia Koprinkova-Hristova, Sven Magg, Günther Palm, and Alessandro E. P. Villa, eds. *Artificial Neural Networks and Machine Learning – ICANN 2014*. Cham: Springer International Publishing, 2014. 837–844. ISBN 978-3-319-11179-7.
- [49] Wilson F. Cueva, F. Muñoz, G. Vásquez, and G. Delgado. *Detection of skin cancer "Melanoma" through computer vision*. In: *2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*. 2017. 1–4.
- [50] J. Abdul Jaleel, and Sibi Salim. Artificial Neural Network Based Detection of Skin Cancer. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Energy*. 2012, 1 200–205.
- [51] R. B Aswin, J. Abdul Jaleel, and Sibi Salim. *Hybrid genetic algorithm — Artificial neural network classifier for skin cancer detection*. In: *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*. 2014. 1304–1309.

- [52] Tim Lee, Vincent Ng, Richard Gallagher, Andrew Coldman, and David McLean. Dullrazor®: A software approach to hair removal from images. *Computers in Biology and Medicine*. 1997, 27 (6), 533–543. DOI 10.1016/S0010-4825(97)00020-6.
- [53] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Pheng-Ann Heng. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. *IEEE Transactions on Medical Imaging*. 2017, 36 (4), 994–1004. DOI 10.1109/TMI.2016.2642839.
- [54] David Gutman, Noel C. F. Codella, Emre Celebi, Brian Helba, Michael Marchetti, Nabin Mishra, and Allan Halpern. *Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)*. . 2016. arXiv. <http://arxiv.org/abs/1605.01397>.
- [55] Kazuhisa Matsunaga, Akira Hamada, Akane Minagawa, and Hiroshi Koga. Image Classification of Melanoma, Nevus and Seborrheic Keratosis by Deep Neural Network Ensemble. *ArXiv*. 2017, abs/1703.03108
- [56] Danilo Barros Mendes, and Nilton Correia da Silva. *Skin Lesions Classification Using Convolutional Neural Networks in Clinical Images*. . 2018. arXiv. <http://arxiv.org/abs/1812.02316>.
- [57] Ulzii-Orshikh Dorj, Keun-Kwang Lee, Jae-Young Choi, and Malrey Lee. The skin cancer classification using deep convolutional neural network. *Multimedia Tools and Applications*. 2018, 77 (8), 9909–9924. DOI 10.1007/s11042-018-5714-1.
- [58] Aleksey Nozdryn-Plotnicki, Jordan Yap, and William Yolland. Ensembling Convolutional Neural Networks for Skin Cancer Classification.
- [59] Nils Gessert, Maximilian Nielsen, Mohsin Shaikh, René Werner, and Alexander Schlaefer. Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. *MethodsX*. 2020, 7 100864. DOI 10.1016/j.mex.2020.100864.
- [60] Qishen Ha, Bo Liu, and Fuxu Liu. *Identifying Melanoma Images using EfficientNet Ensemble: Winning Solution to the SIIM-ISIC Melanoma Classification Challenge*. . 2020. arXiv. <http://arxiv.org/abs/2010.05351>.
- [61] Dan E. Webster, Christine Suver, Megan Doerr, Erin Mounts, Lisa Domenico, Tracy Petrie, Sancy A. Leachman, Andrew D. Trister, and Brian M. Bot. The Mole Mapper Study, mobile phone skin imaging and melanoma risk data collected using ResearchKit. *Scientific Data*. 2017, 4 (1), 170005. DOI 10.1038/sdata.2017.5.
- [62] Muhammad Aleem Taufiq, Nazia Hameed, Adeel Anjum, and Fozia Hameed. *m-Skin Doctor: A Mobile Enabled System for Early Melanoma Skin Cancer Detection Using Support Vector Machine*. 2017.
- [63] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. GrabCut: Interactive Foreground Extraction Using Iterated Graph Cuts. *ACM Trans. Graph.*. 2004, 23 309–314. DOI 10.1145/1186562.1015720.
- [64] Tiago M de Carvalho, Eline Noels, Marlies Wakkee, Andreea Udrea, and Tamar Nijsten. Development of Smartphone Apps for Skin Cancer Risk Assessment: Progress and Promise. *JMIR Dermatology*. 2019, 2 (1), e13376. DOI 10.2196/13376.
- [65] A. Udrea, G.d. Mitra, D. Costea, E.c. Noels, M. Wakkee, D.m. Siegel, T.m. de Carvalho, and T.e.c. Nijsten. Accuracy of a smartphone application for triage of skin

- lesions based on machine learning algorithms. *Journal of the European Academy of Dermatology and Venereology*. 2020, 34 (3), 648–655. DOI 10.1111/jdv.15935.
- [66] T. Maier, D. Kulichova, K. Schotten, R. Astrid, T. Ruzicka, C. Berking, and A. Udrea. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. *Journal of the European Academy of Dermatology and Venereology*. 2015, 29 (4), 663–667. DOI 10.1111/jdv.12648.
- [67] Ioannis Kousis, Isidoros Perikos, Ioannis Hatzilygeroudis, and Maria Virvou. Deep Learning Methods for Accurate Skin Cancer Recognition and Mobile Application. *Electronics*. 2022, 11 (9), 1294. DOI 10.3390/electronics11091294.
- [68] Cahyo Adhi Hartanto, and Adi Wibowo. *Development of Mobile Skin Cancer Detection using Faster R-CNN and MobileNet v2 Model*. In: *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (IC-ITACEE)*. 2020. 58–63.
- [69] Parvathaneni Naga Srinivasu, Jalluri Gnana SivaSai, Muhammad Fazal Ijaz, Akash Kumar Bhoi, Wonjoon Kim, and James Jin Kang. Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM. *Sensors*. 2021, 21 (8), 2852. DOI 10.3390/s21082852.
- [70] Seung Seog Han, Myoung Shin Kim, Woohyung Lim, Gyeong Hun Park, Ilwoo Park, and Sung Eun Chang. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm. *Journal of Investigative Dermatology*. 2018, 138 (7), 1529–1538. DOI 10.1016/j.jid.2018.01.028.
- [71] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*. 2018, 5 (1), 180161. DOI 10.1038/sdata.2018.161.
- [72] Noel C. F. Codella, David Gutman, M. Emre Celebi, Brian Helba, Michael A. Marchetti, Stephen W. Dusza, Aadi Kalloo, Konstantinos Liopyris, Nabin Mishra, Harald Kittler, and Allan Halpern. *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)*. . 2018. arXiv. <http://arxiv.org/abs/1710.05006>.
- [73] Marc Combalia, Noel C. F. Codella, Veronica Rotemberg, Brian Helba, Veronica Vilaplana, Ofer Reiter, Cristina Carrera, Alicia Barreiro, Allan C. Halpern, Susana Puig, and Josep Malvehy. *BCN20000: Dermoscopic Lesions in the Wild*. . 2019. arXiv. <http://arxiv.org/abs/1908.02288>.
- [74] Veronica Rotemberg, Nicholas Kurtansky, Brigid Betz-Stablein, Liam Caffery, Emmanouil Chousakos, Noel Codella, Marc Combalia, Stephen Dusza, Pascale Guitera, David Gutman, Allan Halpern, Brian Helba, Harald Kittler, Kivanc Kose, Steve Langer, Konstantinos Liopyris, Josep Malvehy, Shenara Musthaq, Jabpani Nanda, Ofer Reiter, George Shih, Alexander Stratigos, Philipp Tschandl, Jochen Weber, and H. Peter Soyer. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific Data*. 2021, 8 (1), 34. DOI 10.1038/s41597-021-00815-z.
- [75] Peter A. Lio, and Paul Nghiem. Interactive Atlas of Dermoscopy. *Journal of the American Academy of Dermatology*. 2004, 50 (5), 807–808. DOI 10.1016/j.jaad.2003.07.029.

-
- [76] Alexandre Fabian A. Saint, Eman Ahmed, Abd El Rahman Shabayek, Kseniya Cherenkova, Gleb Gusev, Djamila Aouada, and Björn Ottersten. 3DBodyTex: Textured 3D Body Dataset. 2018,
- [77] Ilya Loshchilov, and Frank Hutter. *SGDR: Stochastic Gradient Descent with Warm Restarts*. . 2017. arXiv.
<http://arxiv.org/abs/1608.03983>.
- [78] Rafael Padilla, Sergio L. Netto, and Eduardo A. B. da Silva. *A Survey on Performance Metrics for Object-Detection Algorithms*. In: *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*. 2020. 237–242.
- [79] Rafael Padilla, Wesley L. Passos, Thadeu L. B. Dias, Sergio L. Netto, and Eduardo A. B. da Silva. A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. *Electronics*. 2021, 10 (3), 279. DOI 10.3390/electronics10030279.
- [80] Navaneeth Bodla, Bharat Singh, Rama Chellappa, and Larry S. Davis. *Soft-NMS – Improving Object Detection With One Line of Code*. . 2017. arXiv.
<http://arxiv.org/abs/1704.04503>.
- [81] Steven Zhou, Zhuang Yixin, and Meng Rusong. *Multi-Category Skin Lesion Diagnosis Using Dermoscopy Images and Deep CNN Ensembles*.
- [82] Sergey Kastruyulin, Jamil Zakirov, Denis Prokopenko, and Dmitry V. Dylov. *PyTorch Image Quality: Metrics for Image Quality Assessment*. 2022.
<https://arxiv.org/abs/2208.14818>.

Appendix A

Consent form - DermaNude dataset

CONSENT FORM FOR INCLUDING PHOTOS IN DermaNude DATASET

I, [insert name], hereby give my consent for including photos of me in the DermaNude dataset. I understand that my photos will be used to create a proprietary dataset (the "Dataset") for the purposes of performing non-commercial scientific and/or academic research or education (teaching).

The terms "Personal Data" and "Processing of Personal Data" in this section refer to the definitions provided in Article 4 of the EU General Data Protection Regulation 2016/679 of 27 April 2016. Please note that age and gender information is considered personally identifiable information (PII) under the GDPR, and the processing of such data is subject to its requirements. However, if the faces of the participants have been blurred or are not present, the processing of PII will be minimized. Furthermore, only the necessary minimum of personal data, such as age and gender, will be included in the dataset.

I understand that the Dataset will be owned, produced, and maintained by the Licensor, and that I am not entitled to any ownership or compensation for my photos.

I agree that the Dataset may be used for non-profit research only, and that any use of the Dataset in the development of a commercial product, or the distribution or sharing of the data with a third party without prior written authorization of the Licensor, is strictly prohibited. However, I may optionally allow the sharing of my information (age, sex, photos without faces) with third parties for non-commercial research purposes by checking the box below:

- I agree to allow my data, such as age, sex and photos, to be shared with third parties for non-commercial research purposes.

I understand that my photos will be accessed by full-time or part-time employees (including staff, researchers, or members of a research project(s) maintained or made at the Licensee), independent contractors retained for the research project(s) or Licensee's students (collectively, "Authorized Persons"), and I agree to bear responsibility for such use of the Dataset by its authorized users.

I have read and understand the terms and conditions of this consent form and hereby agree to its contents.

Date:

Signature:

Appendix B

DermaNude Dataset License Form

Data License Agreement for DermaNude Dataset

Between: Samuel Šúr, hereinafter referred to as "Licensor".

And:

(ORGANIZATION NAME, ADDRESS), SIGNATORY NAME AND TITLE
hereinafter referred to as "Licensee".

This Dataset License Agreement (« Agreement ») becomes effective upon the date of the last signature below ("Effective Date").

1. DEFINITIONS

1.1 "Dataset" means the proprietary Dataset specified above which is owned, has been produced and maintained by the Licensor.

1.2 "Permitted Uses" means a non-exclusive, non-transferable, non-assignable, free of charge right to use the Dataset solely for the purposes of performing non-commercial scientific and/or academic research by the Licensee or education (teaching).

1.3 "Updates" means any revisions of, additions to, or deletions from the Dataset.

2. LICENSE

2.1 Subject to the terms and conditions of this Agreement, Licensor hereby grant to Licensee a non-exclusive, royalty-free, non-transferable license right (the "License") for the term of this Agreement to use the Dataset for the Permitted Uses, but only as long as the Licensee complies with the terms and conditions of this Agreement. The Dataset is licensed to the Licensee, not sold.

2.2 The Dataset may be used for non-for-profit research only. Any use of the Dataset in the development of a commercial product, or the distribution or sharing of the data with a third party without is strictly prohibited.

2.3 Licensee shall agree to inform the Licensor in written form about academic research projects for which Dataset is being used.

2.4 Licensee shall ensure that no persons other than full-time or part-time employees (including staff, researchers or members of a research project(s) maintained or made at the Licensee, independent contractors retained for the research project(s)) of Licensee, regardless of their physical location; or Licensee's students ("Authorized Persons"), will get access to the Dataset. Use of the Dataset by the Authorized Persons of the Licensee mentioned in this Section 2.4. shall be deemed use of the Dataset by the Licensee, and Licensee agrees to bear responsibility for such use of the Dataset by its authorized users. Licensee shall inform its authorized users about the terms and conditions of this license and on Permitted Uses of the Dataset. In the event of any unauthorized use of the Dataset happen, the Licensor may (i) terminate such authorized user access to the Dataset , or (ii) terminate this Agreement for cause as specified in clause 6.2 below.

2.5 Any publication or work using this dataset, in whole or in part, should cite the following Bachelor's thesis:

["Šúr, S. \(2023\). Detection, classification and matching of skin lesions. Czech Technical University in Prague."](#)

2.6 For the purpose of this section, the term "Personal Data" and "Processing of Personal Data" shall have the meaning as described in article 4 of the EU General Data Protection Regulation 2016/679 of 27 April 2016. Licensee acknowledges that the Dataset provided under this Agreement contains Personal Data, including age and gender information. Licensee agrees to comply with all applicable laws and regulations related to the Processing of Personal Data, including the GDPR. Licensee shall take all necessary technical and organizational measures to ensure the security and confidentiality of Personal Data contained in the Dataset, and to prevent unauthorized access, disclosure, alteration of such data.

3. DELIVERY

As soon as practicable after the Effective Date, Licensors will provide to Licensee one (1) copy of the Dataset by electronic means to an email of Licensee.

4. UPDATING

During the term of this Agreement, Licensor may provide Licensee with Updates to the Dataset as they become available and as reasonably practicable. Unless otherwise expressly indicated herein, all references in this Agreement to Dataset shall include all Updates thereto.

5. ARCHIVE

During the term of the Agreement specified in the following clause 6.1, Licensor grant the Licensee the right to archive one (1) complete copy of the Dataset . Licensee shall use such copy of the Dataset strictly in accordance with the provisions of this Agreement.

6. TERM AND TERMINATION

6.1 This Agreement shall have an initial term of three (3) years commencing on the Effective Date. During this period, this Agreement may be terminated by Licensor, by registered letter with acknowledgment of receipt, at any time and automatically, in consideration of thirty (30) days' notice. Licensee shall have the option to prolong the effective date of this Agreement by submitting a written request to Licensor prior to the expiration of the initial term. If the Licensor approves such request, this Agreement shall be extended for an additional term to be agreed upon by the parties in writing. If the Licensor does not approve such a request, this Agreement shall expire at the end of the initial term as specified above.

6.2 This Agreement may be terminated by Licensor if the following event shall occur: Licensee has materially breached any term or covenants of this Agreement, or if Licensor believe that Licensee or its authorized users have used the Dataset beyond the scope of this license and such material breach shall continue uncured for 30 days after written notice thereof from Licensor detailing such material breach.

7. POST TERMINATION PROVISIONS

Except as provided below, as soon as practicable after termination of this Agreement for any reason whatsoever, without regard to the fault, but no later than 30 days after such termination, Licensee shall:

- (a) Destroy or deliver to Licensor, at Licensor's option, any and all copies of the Dataset and all Updates hereto, and other machine-readable or human-readable material delivered to Licensee pursuant to this Agreement and then in Licensee's possession or under Licensee's control, containing all or any part of the Dataset provided to Licensee pursuant to this Agreement.
- (b) Expunge from any data storage facility owned by or under control of the Licensee any data contained in or derived from the Dataset delivered to the Licensee pursuant to this Agreement
- (c) Certify to Licensor in writing that Licensee has complied with the requirement of this section.

8. LIMITED WARRANTY

8.1 Licensor promises that it has the right to give Licensee permission to use the Dataset described in this Agreement, and that the Dataset will not violate any third party's copyright, trademark, trade secret, or other intellectual property rights. If there are any infringement claims, Licensor will either remove the infringing information from the Dataset, modify the Dataset so it's not infringing, or terminate this Agreement.

8.2 Except for what's stated in Section 8.1, Licensor does not give any other warranties, including warranties of merchantability or fitness for a particular purpose. Licensor doesn't guarantee that the information in the Dataset is accurate, adequate, or complete, or that the Dataset will meet Licensee's requirements. No one is allowed to make any statements that change or add to the warranties or limitations in this Agreement, and Licensee can't rely on any such statements.

9. LIMITATION OF LIABILITY

Licensor is not responsible for any special, indirect, incidental, or consequential damages that happen because of this Agreement or because of anything related to it. This includes damages from contract, warranty, negligence, tort, strict liability, or any other reason.

10. PROPRIETARY RIGHTS

10.1 Licensor owns all the intellectual property rights to the Dataset and all its parts, including the copyright. Licensee acknowledges this fact and agrees that Licensor will remain the owner of the Dataset and all its property rights. This Agreement does not give Licensee any ownership rights to the Dataset.

10.2 Licensees shall not alter, remove, obscure or obstruct the display of any copyright, trademark or other proprietary notice placed upon Dataset by Licensor.

10.3 Licensee shall execute and deliver to Licensor from time to time, at Licensor's request any further instruments or documents that may reasonably be required by Licensor to vest and confirm such proprietary rights to Licensor.

10.4 Licensee is only permitted to use the Dataset in accordance with the terms outlined in this Agreement. Licensee is strictly prohibited from (a) copying, downloading or otherwise reproducing the Dataset in any medium, in whole or in part, except for limited portions resulting from discrete searches that may be printed or downloaded; (b) modifying, adapting, or altering the Dataset, including decompiling, disassembling, or creating derivative works, without prior written approval from Licensor, or (c) selling, licensing, or distributing the Dataset or any parts thereof to third parties, including Licensee's subsidiaries and affiliates.

10.5 Licensee acknowledge that the Dataset is the product of Licensor's extensive gathering and coordination of information, selection therefrom of information considered by Licensor to be relevant and useful, and original arrangement of selected information.

10.6 Licensee acknowledges that any use of the Dataset by Licensee which is not in accordance with the terms of this Agreement or any violation of the restrictions imposed on Licensee's use of the Dataset would cause irreparable harm to Licensor.

Date: ___ / ___ / ___

As Licensor:

.....
Samuel Šúr

For Licensee:

Signed by:

.....
Name of the signatory: _____

Title: _____

Appendix C

Technical details

All of the code for this thesis can be found in the GitLab repository¹ The structure of the repository is shown in “Lis. C.1”.

```
|--- classification/
|   |--- classification_code
|       |--- main.py
|       |--- stats/
|       |--- requirements.txt
|       |--- model_efficientnet-b0.pth
|       |--- model_efficientnet-b7.pth
|       |--- model_resnet50.pth
|   |--- trained_models
|       |--- model_efficientnet-b0.pth
|       |--- model_efficientnet-b7.pth
|       |--- model_resnet50.pth
|       |--- ...
|--- detection/
|   |--- eval_dermanude/
|       |--- utils/
|       |--- ...
|       |--- main.py
|       |--- models/
|           |--- best_model.pt
|   |--- yolov7/
|       |--- main.py
|       |--- dataset_wrapper.py
|       |--- generate_txt_for_eval.py
|       |--- preprocess.py
|       |--- json_to_csv.py
|       |--- my_nms.py
|       |--- requirements.txt
|--- cli_app/
|   |--- main.py
|   |--- models/
|       |--- ...
|   |--- results/
|   |--- utils/
|       |--- custom_loader.py
|   |--- Readme.md
|--- Readme.md
```

Listing C.1. Structure of the repository.

¹ https://gitlab.fel.cvut.cz/sursamue/bakalarska_praca.

The 3 main parts of the code are classification, detection, and CLI application. They are described in the following sections. We also summarize the location of the datasets used in this thesis.

C.1 Classification

The code for classification is located in the *classification/classification/* folder. The code is in the *main.py* file.

The required packages are listed in the *requirements.txt* file. The *Read.me* file contains all the information about the code and how to run it.

C.2 PSL detection

The code for PSLs detection in the 3DBodyTex.v1 is located in the *detection/yolov7* folder. The code is in the *main.py* file. The *Read.me* file contains all the information about the code and how to run it. The code for PSLs detection in the DermaNude dataset is located in the *detection/eval_dermanude* folder. The code is in the *main.py* file.

C.3 Self-examination of skin lesions

The code for self-examination of skin lesions is located in the *cli_app* folder. The code is in the *main.py* file located in the *cli_app* folder. The *read.me* file contains all the information about the code and how to run it.

C.4 Datasets

3DBodyTex.v1 dataset can be obtained after signing the license form¹. The PSL annotations for 3DBodyTex.v1 are available on GitHub². The ISIC 2019 dataset can be obtained from Kaggle³. The DermaNude dataset can be obtained after signing the license form and sending it to the author of this thesis. The license form, as well as contact information, is available on the official website⁴.

¹ <https://cvi2.uni.lu/3dbodytexv1/>

² <https://github.com/jeremykawahara/skin3d>

³ <https://www.kaggle.com/datasets/cdeotte/jpeg-isic2019-512x512>

⁴ <https://samuelsur.xyz/dermanude>