Slim-YOLO: A Simplified Object Detection Model for the Detection of Pigmented Iris Freckles as a Potential Biomarker for Cutaneous Melanoma

D. Nathasha U. Naranpanawa*, Yanyang Gu†, Shekhar S. Chandra†, Brigid Betz-Stablein‡, Richard A. Sturm‡, H. Peter Soyer‡ and Anders P. Eriksson†
*†*School of Information Technology and Electrical Engineering, The University of Queensland, Australia*
‡*Dermatology Research Centre, The University of Queensland Diamantina Institute, The University of Queensland, Australia*
*nathasha.naranpanawa@uq.edu.au

2021 Digital Image Computing: Techniques and Applications (DICTA) | 978-1-6654-1709-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/DICTA52665.2021.9647150

***Abstract*—Melanomas are the most dangerous form of skin cancer, accounting for a majority of mortality among all skin cancers. As melanomas tend to go unnoticed without constant supervision, it is important that steps are taken to identify and prevent their spread before they reach more severe stages. Many recent clinical trials and research have identified various indicators of melanoma that may be used for early detection. In this work, we explore the use of Convolutional Neural Networks (CNN) to localize and detect one such indicator - a strong correlation between the number of pigmented freckles in a person's iris and their risk of developing melanoma on the skin. We model this task of detecting pigmented iris freckles as a single-class, one-sized object detection problem. For this, we propose Slim-YOLO, a lighter and simpler object detection model based on YOLOv3. The simplifications of Slim-YOLO are introduced through reducing the model computations by removing the need for multiple detection scales and classification. We also remove the constraints applied by anchor boxes. The experimental results show that Slim-YOLO is capable of achieving comparable performance (90.7% in mAP) with YOLOv3 (93.7%) while yielding a smaller model size of two-third the size of YOLOv3. These may prove highly beneficial to better facilitate deployment of the model on mobile devices in the future. Thus, we automate the iris freckle detection process successfully to help provide insights to practitioners, and contribute to the use of deep learning methods in detection of anomalies in medical imaging.**

***Index Terms*—melanoma, deep learning, object detection, pigmented iris freckles**

I. Introduction

The primary cause of skin cancers is considered to be the prolonged exposure to harmful ultraviolet radiation from sunlight which results in uncontrolled growth of abnormal cells in the epidermis, the outermost layer of the skin [1]. Among the main types of skin cancer, Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC) are considered to be more common and less threatening. However, cutaneous melanoma is the least common, but the most dangerous type due to its ability to spread into other organs faster.

The global incidences of skin cancers as a collective have been increasing over the past few decades, with nearly 300,000 melanoma skin cancers reported in 2018 alone [2]. The highest rates of melanoma cases are seen in Australia and New Zealand with more than 33 cases per 100,000 of the population [3]. New Zealand has the highest mortality rate with 4.8 deaths per 100,000 of the population [4]. Overall, malignant melanomas account for 1% of skin cancers but 60% mortality due to skin cancers.

The survival rate for melanoma is strongly associated with the depth of growth of the cancer. If the melanoma has grown deeper and thicker, it might be more difficult to

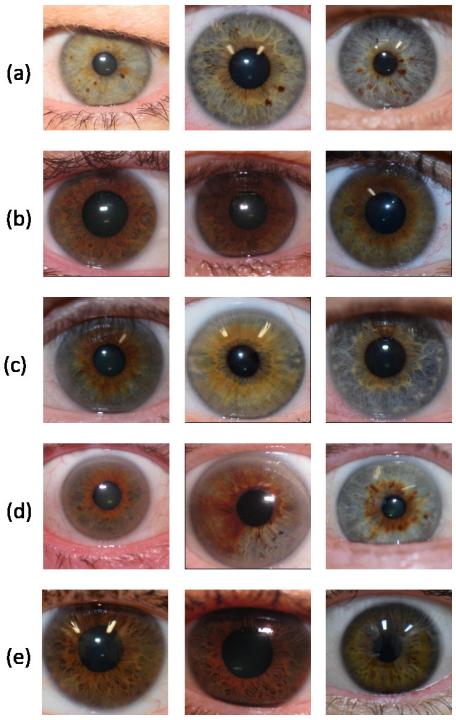

Fig. 1. (a) Eyes with clearly visible freckles. However, there can be complications in identifying iris freckles including (b) darker eyes where freckles might not be clearly visible, (c) eyes with pupillary rings but no freckles, (d) eyes with both pupillary ring and freckles, and (e) eyes with muscle fibres and crypts clearly showing but with no visible freckles.

978-1-6654-1709-9/21/$31.00 ©2021 IEEE

treat. Therefore, detection of melanoma at early stages of its development is critical. According to the Skin Cancer Foundation, the 5-year survival rate of melanoma is 99% when detected early [5]. However, most countries currently do not have screening programs that are accessible to citizens as a method of identifying who might be at risk. Instead, people are expected to seek expert medical attention if they notice that a mole or a lesion has been showing unnatural growth over time. This leads to most cases being overlooked as some people might not be trained well enough to make an educated guess by themselves. Therefore, to mitigate the growth of melanoma before reaching critical stages, and to reduce the disease burden of a patient at a later stage in life, automatic and early detection of melanoma is critical in the present.

As a potential method of early detection of skin cancer, a previous dermatological study [6] identified a strong correlation between the number of pigmented iris lesions (iris naevi and iris freckles) in a person's eyes and their risk of developing melanoma on the skin. The original study considered 1254 European-background Australians consisting of case participants who had a personal history of melanoma, and control participants who had no such history. The study concluded that a person's risk of developing melanoma was found to be increased by 1.45-fold if they had 3 or more iris lesions in their eyes (Fig. 1(a)). Moreover, in younger case participants (below 40 years of age), the number of iris lesions was greater than that of control participants of the same age group. This indicated a window of opportunity to assess and identify people at risk of developing melanoma potentially before they reach 40 years of age, and to use iris lesions as a potential bio marker for early detection.

While the above study has great importance in employing non-traditional methods of early skin cancer detection, it poses following unique challenges:

- First, trained personnel were required to manually count these iris lesions by observing digital images of patient eyes. While this might lead to more accurate detections, it is an inefficient method of doing so. The requirement of trained skill and high levels of supervision would mean that the chance of this method directly being used by a clinician or the general public would be low.
- Second, while we are going to be detecting iris freckles automatically, the freckle detection problem is difficult to define because there is currently no agreement as to what an iris pigmented freckle is. Irises have many textural and coloring complications, and other features that may look visually similar to pigmented freckles (Fig. 1(c)-(e)). In addition, pigmented freckles are clearly visible only in light-colored eyes, including blue and grey eyes. However, these pigmented freckles may be present in brown and green/hazel eyes as well (Fig. 1(b)).

To address the above challenges, we propose a simple, one-stage object detection model - Slim-YOLO - by improving upon the popular YOLOv3 model [7]. For this, we focus on some specific characteristics of the freckle detection problem. Firstly, we only need to detect a single class of objects in a homogeneous environment. Secondly, the objects are similar in shape and size, where the scales of the objects lie in a small range. Therefore, the proposed method would focus on object detection without classification, and without any bells and whistles that usually weigh down object detection models trying to detect differently-scaled objects. With these improvements, Slim-YOLO is capable of automatically detecting pigmented iris freckles with minimal supervision. Our main objective is to provide a system that can be used by clinicians to make educated and accurate predictions of freckle count in a patient without manually counting them. We also provide our own definition of a pigmented iris freckle, which might serve as a bio-marker for early-stage melanoma.

The rest of the paper is organized as follows. Section II summarises related work that has been done in improving YOLO-based models for object detection. Section III describes the proposed methodology, followed by Section IV where the experimental settings are presented and results are discussed. Finally, Section V concludes the paper.

II. Related Work

In recent years, object detection methods based on deep convolutional neural networks (CNNs) have progressed rapidly due to their capabilities in learning robust and higher-level representations of images automatically. Modern object detectors are categorized into two types based on their pipelines: two-stage detectors and one-stage detectors.

Two-stage detectors have two distinct and individual modules; the first is for region proposal, and the second is using a CNN for feature extraction and classification. The R-CNN family of object detector, including R-CNN [8], Fast R-CNN [9], and Faster R-CNN [10], are popular two-stage detectors. Some of the issue of R-CNN detectors were further resolved by SPPNet [11] and RFCN [12].

One-stage detectors abandon the previous pipeline of proposal detection and verification, and instead apply a single neural network to the complete image. This leads to a simpler and faster model architecture for object detection. State-of-the-art one-stage detectors, including SSD [13], YOLO [14], YOLOv2 [15] and YOLOv3 [7], apply regression to perform both detection and classification simultaneously. While these one-stage detectors might not have a great localization accuracy compared to two-stage detectors, they are faster, more efficient, and simpler in their architecture.

For application-specific object detections, improvements based upon the original YOLO models have been used widely due to their high efficiency. PYolo [16] was improved upon YOLOv3 by generating custom anchor boxes and adding dilated convolutions to increase the perception field in order to detect pneumonia lesions on lung X-ray images. Modified YOLO models have also been used for pedestrian detections [17], shuttlecock detection in badminton [18], and apple detection at different growth stages [19]. However, more computations were added to these models in order to make

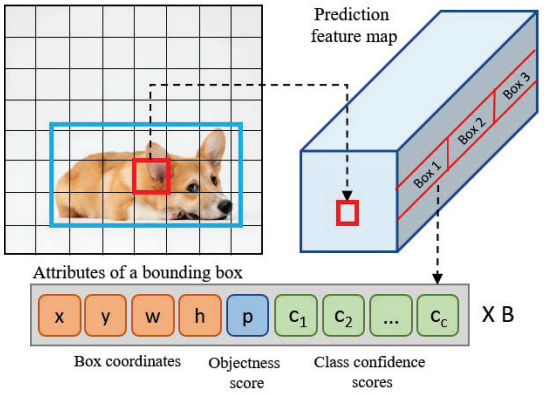

Fig. 2. The output tensor for one bounding box for a feature map in YOLOv3. The red cell is responsible for detecting the object.

them robust and suitable for the specific applications, making them denser and computationally heavier.

Small object detection with YOLO models have also seen great improvements. Pham *et al*. [20] focused on replacing the coarser detection levels of YOLOv3 with finer detection levels with lower subsampling factors. Their model YOLO-fine was successfully used for small object detection in remote sensing images. Similarly, UAV-YOLO [21] optimized the network structure of YOLOv3 to expand the receptive fields to detect small objects on images as viewed from Unmanned Aerial Vehicles (UAVs).

Fang *et al*. [22] reduced the model parameters of Tiny-YOLO-V3 to reduce model size, and merged feature maps to maintain the detection accuracy. This allowed their model Tinier-YOLO to be deployed in constrained environments such as embedded devices. YOLO Nano [23] was also designed for embedded object detection by employing a a human-machine collaborative design strategy and improving upon the Tiny-YOLO model.

Most of these improvements and modifications to the original YOLO models increased detection accuracy, especially for small object detection. However, all of them maintain the three-level detection introduced in YOLOv3, as well as classification for single class problems. The different levels of detection in YOLOv3 are for aiding the prediction of differently-scaled objects in the same image. For a single-class detection of objects that are similar in shape and size, three levels of detections can be computationally heavy and unnecessary. Thus, by stripping off these frills, we argue that we should be able to improve YOLOv3 to be lighter and simpler with comparable accuracy to the original model for a single-class, one-sized object detection problem.

III. Methodology

A. Definition of Pigmented Iris Freckles

The difficulty in defining pigmented iris freckles arises because they might look similar to iris naevi. According to the original study [6] iris naevi are 'benign melanocytic proliferations of the iris; they are raised, well-circumscribed brown lesions which form a nodule that replaces the iris stroma'. However, iris freckles are only present on the surface of the iris, and will not disrupt the normal structure of it [24]. Distinguishing between iris freckles and iris naevi would require a slit-lamp examination [25].

According to our observations, another architecture of the iris that hinders the identification of an iris freckle is the peripupillary ring of the eye, which is a brown hyperpigmented zone immediately around the pupil of the eye [26]. The boundary of the peripupillary ring is not even, and would sometimes transfer to the iris in a freckling manner at some places while still being connected to the ring (Fig. 2(c)). However, it is important to distinguish the peripupillary ring from actual iris freckles as the peripupillary ring does not associate with melanoma [6].

Considering all these medical definitions and our manual inspection of digital images, we define a pigmented iris freckle as following for our problem of automatically detecting them.

A pigmented iris freckle is:

- Different in colour to that of the peripupillary ring
- Separated from the peripupillary ring
- Darker in colour than the iris, usually brown
- Not visibly different in texture to the rest of the iris surface
- Not visibly raised, but flat
- Similar in size to other freckles in the iris
- Not spread into the pupil, or the sclera

B. Weakly-supervised Annotation

Following the above definitions, we annotate iris freckles on digital image with point annotations using LabelBox [27]. As most detectors require bounding boxes as ground truth annotation, the point annotations for the iris freckles are converted into pseudo bounding boxes with a fixed-size in the training dataset. These pseudo bounding boxes are not guaranteed to enclose the entire boundary of a freckle or an estimation of its actual size. It is only used to create a representative set of bounding boxes without having to perform expensive manual annotations. Furthermore, as most freckles are smaller and closely positioned to other freckles, using points to annotate is more representative than drawing a bounding box around each one of them.

C. Original YOLOv3 Architecture

The YOLOv3 model is an incremental improvement to its predecessors: YOLO and YOLOv2. The feature extractor used by YOLOv3 is Darknet-53. This has 53 convolutional layers, and includes residual blocks, skip connections and upsampling to improve accuracy.

The model first divides the entire image into a grid of SxS cells, and applies a unified convolutional network on the full image. If the centre of an object in the image falls within a cell, that particular cell is responsible for detecting that particular object. It would also predict bounding boxes and confidence scores for that object. YOLOv3 also enables detection at three different scales to improve performance. For

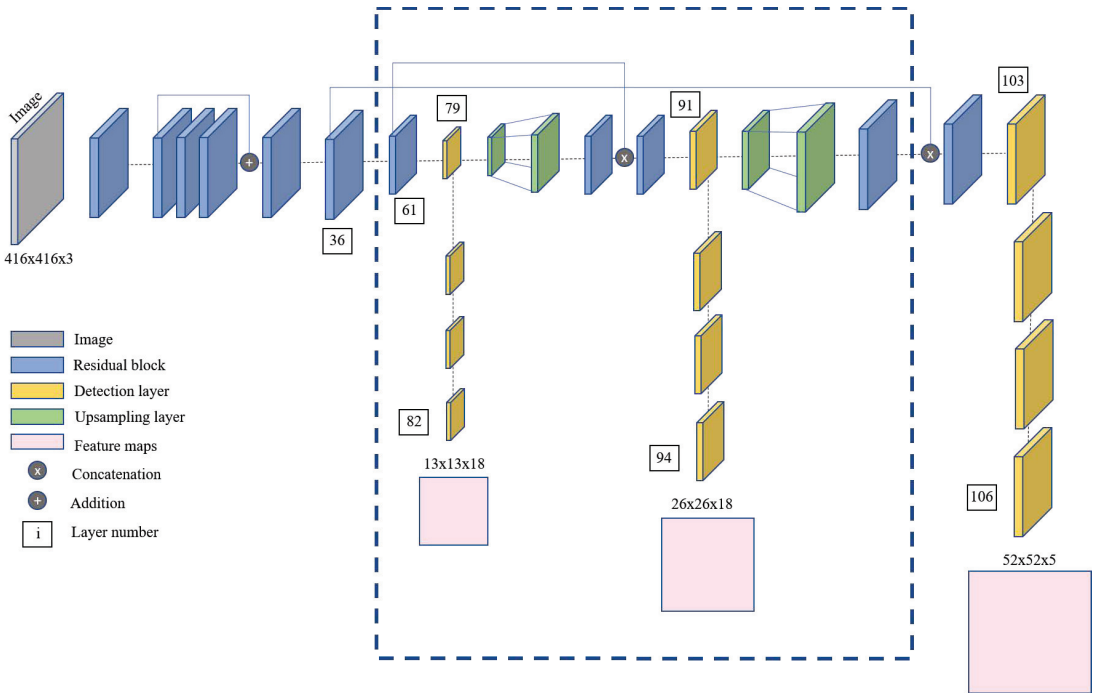

Fig. 3. The Slim-YOLO architecture. The blue dotted-line box shows the removed layers of the original YOLOv3.

example, if an image is 416x416, the three scales are 52x52, 26x26, and 13x13. These scales detect small, medium, and large objects respectively. Each grid cell is assigned with a number of predetermined anchor boxes to allow for the case of multiple objects being present in a given cell. Anchor boxes are generated by running k-means clustering on all ground truth bounding box labels across the entire dataset before training. Generally, the 9 most common sizes of objects are assigned as anchor boxes. These anchor boxes may vary in their aspect ratios, and three boxes each are assigned for different YOLO detection layers. At training, the detection layer determines which anchor box has the highest IoU with the ground truth box, and then uses that anchor box to predict offsets for the bounding box prediction.

The final prediction for a given object is a tensor of size $S \times S \times B \times (5+C)$, which is obtained by applying a 1x1 kernel on the final feature maps at the three scales. Here, B is the number of bounding boxes that can be predicted at each scale, and C is the number of conditional class probabilities. Each bounding box is attributed with 5 predictions: x, y, w, h, and objectness score (Fig. 2). The (x,y) coordinate is the centre of the bounding box with respect to the grid cell, and width (w) and height (h) are predicted with respect to the entire image. The objectness score is represented by the IoU between the predicted box and the ground truth box.

Finally, YOLOv3 applies non-max suppression (NMS) to clean up multiple detections of the same object. For each detected object, the bounding boxes with the class confidences and objectness scores below a certain threshold are discarded. From the remaining bounding boxes, the box with the highest confidence is chosen, and every box that has an IoU of more than 0.5 with that box are discarded as well.

D. Overview of the Proposed Architecture

Based on YOLOv3, we simplify the network to fulfil our single-scale and single-class iris freckle detection by reducing the scales of detection, eliminating classification, and removing anchors. Hence, non-maximum suppression and loss function are modified accordingly.

1) Reducing the Scales of Detection: Considering that there is no great difference in the scales of iris freckles, we determine that one scale of detection is enough for this problem. The original scales 13x13, 26x26, and 52x52 separate the image grid into cells with stride 32x32, 13x13 and 8x8 respectively for an image of 416x416 in size. The three default anchor boxes associated with the smallest scale are 10x13, 16x30, and 33x23. This indicates that the smallest detection scale is sufficient to detect the iris freckles given that the pseudo bounding box annotations provided are 20x20 in size.

Thus, we remove the two coarser detection scales for large and medium sized objects and any intermediate layers that became redundant (Fig. 3). All upsampling layers and routing are adjusted accordingly. Originally, the number of detections for 3 scales with 3 anchor boxes for each scale (B=3) and for 1 class of objects (C=1) were:

$$\begin{aligned} outputs &= (S1 \times S1 \times B \times (5+C)) \\ &\quad + (S2 \times S2 \times B \times (5+C)) \\ &\quad + (S3 \times S3 \times B \times (5+C)) \\ &= (13 \times 13 \times 3 \times 6) + (26 \times 26 \times 3 \times 6) \\ &\quad + (52 \times 52 \times 3 \times 6) \\ &= 63,882. \end{aligned} \tag{1}$$

TABLE I
MODEL CONFIGURATIONS

Model	Number of detection layers	Detection layers scales (S)	Number of Convolutional layers	Number of anchor boxes	Anchor box sizes	Output tensor size
YOLOv3	3	13x13, 26x26, 52x52	75	9	10x13, 16x30, 33x23, 30x61, 62x45, 59x119, 116x90, 156x198, 373x326	SxSx3x(5+1)
YOLOv3-nc	3	13x13, 26x26, 52x52	64	9	10x13, 16x30, 33x23, 30x61, 62x45, 59x119, 116x90, 156x198, 373x326	SxSx3x(5+0)
YOLOv3-tiny	3	13x13, 26x26, 52x52	16	9	10x13, 16x30, 33x23, 30x61, 62x45, 59x119, 116x90, 156x198, 373x326	SxSx3x(5+1)
Slim-YOLOv3-tiny	1	52x52	16	0	N/A	SxSx1x(5+0)
Slim-YOLO-wc	1	52x52	33	3	10x13, 16x30, 33x23	SxSx3x(5+1)
Slim-YOLO-nc	1	52x52	33	3	10x13, 16x30, 33x23	SxSx3x(5+0)
Slim-YOLO	1	52x52	33	0	N/A	SxSx1x(5+0)

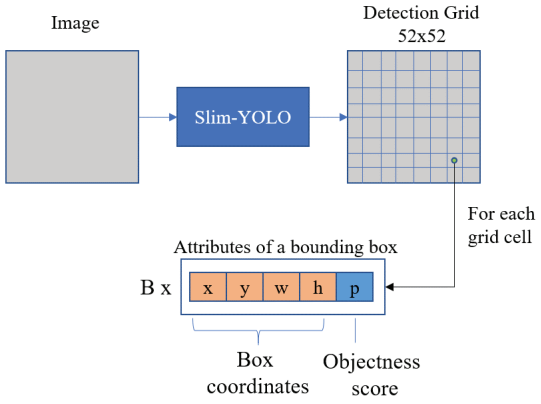

Fig. 4. The output tensor size of Slim-YOLO.

Now that the scales have been reduced to one, the number of detections is:

$$\begin{aligned} new\ outputs &= (S3 \times S3 \times B \times (5+C)) \\ &= (52 \times 52 \times 3 \times 6) \\ &= 48,672. \end{aligned} \tag{2}$$

2) Eliminating Classification: Since we are only interested in detecting a single-class object, classification is unnecessary for our detection model. Therefore, Slim-YOLO eliminates the class predictions within the network. In the original YOLOv3, the YOLO layers have a configuration to set the number of classes to be detected. We set this to zero, and remove all class prediction related calculations from the network. The number of filters applied on the layer prior to the YOLO layer is $3 \times (1+5)$ in the case of a single-class problem with the original YOLOv3. Since we are ignoring the class predictions, this filter is now set to be $3 \times (0+5)$. These updates reshape the final output tensor of a given detection layer of size S to be $S \times S \times B \times 5$. Given that we only have one detection scale in Slim-YOLO, the output tensor is of size $52 \times 52 \times 3 \times 5$ (Fig. 4 with B=3). This further reduces the number of detections to 40,560.

3) Eliminating Anchors: When there are a number of different objects of various sizes to be detected in an image, predicting bounding boxes directly without any template would make the model susceptible to errors. Predictions might favour bounding boxes with larger dimensions, and the training process will be unstable because the range of values to be predicted can vary significantly. This is where the concept of anchor boxes come into play in YOLOv3. These are predetermined boxes of different aspect ratios, primarily used for controlling the prediction of bounding boxes. Anchor boxes that have the highest IoU with the ground truth annotations are chosen to calculate the offsets of network predictions.

However, if the objects are of only one size, the need for anchors become redundant, similar to not needing different scales. All ground truth objects are of the same aspect ratio in our problem of detecting iris freckles. Hence, it is not necessary to constraint the calculations with anchors.

Based on this, we remove the anchor boxes from the YOLOv3 configurations by removing the 'anchors' parameter in the YOLO detection layer. Instead, the ground truth bounding box dimensions are directly used in place of anchor box dimensions where necessary, completely eliminating the need for anchors. This is equivalent to having only one anchor box per YOLO detection layer. Thus the final output tensor becomes $S \times S \times 1 \times 5$ where S is 52, leading to 13,520 detections only (Fig. 4 with B=1).

4) Non-maximum Suppression: Non-max suppression helps remove multiple detections of the same object. Originally, only the bounding boxes with the highest class confidences are kept. Then, all bounding boxes that have objectness scores lower than a specified threshold are discarded. If there are still multiple detections remaining, the bounding box with the largest objectness score is picked and the predictions that has an IoU larger than a specified threshold with this bounding box are discarded.

By removing classification, the calculation of non-max suppression also changes. Since we do not have class confidence scores anymore, we only consider the objectness scores to filter out multiple detections. The objectness score threshold used in YOLOv3 was reused in Slim-YOLO without any changes.

5) Loss Function: The loss function of original YOLOv3 is composed of three parts. The first part is the regression loss where it penalises the bounding box attributes $(\hat{x}, \hat{y}, \hat{w}, \hat{h})$. The second part is the object confidence loss where it penalises the objectness score $(\hat{C})$ for bounding boxes that have objects as well as boxes that have no objects. The third part is the classification score that penalises the class prediction score

TABLE II
RESULTS FROM TRAINING ALL MODELS

Model	Mean average precision (%)	Training time (hours)	Inference time (milliseconds)	Memory (Gb)
SSD	77.5	1.828	30.65	9.9
Faster RCNN	84.7	11.30	64.65	4.0
YOLOv3	93.7	3.679	23.3	8.12
YOLOv3-nc	93.2	3.689	19.5	8.21
YOLOv3-tiny	84.0	1.231	13.0	1.89
Slim-YOLOv3-tiny	81.9	1.363	13.4	1.87
Slim-YOLO-wc	90.5	2.292	18.3	4.89
Slim-YOLO-nc	90.9	2.296	17.6	5.27
Slim-YOLO	**90.7**	**2.261**	**15.3**	**5.37**

for the object.

In the proposed Slim-YOLO architecture, we exclude the penalisation of class prediction from the loss function as we no longer predicts the class confidence scores. The new loss function is presented in Equation 3:

$$\begin{aligned} loss &= loss_{regression} + loss_{objectness} \\ &= \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ &+ \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\ &+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} (C_i - \hat{C}_i)^2 \\ &+ \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{noobj} (C_i - \hat{C}_i)^2, \end{aligned} \tag{3}$$

where λ is the hyper-parameter that balances the influence of bounding boxes with and without objects, and 1_{ij}^{obj} denote if bounding box prediction j in cell i is responsible for the predicted object.

IV. EXPERIMENTS AND RESULTS

A. Dataset

We were given access to the digital photographs of the eyes from 1,117 participants from the original study [6] and an additional set of eye images that were not included in the original study. For a subset of the images, metadata containing the eye colour and the manual count of iris freckles for each participant were also provided. All images were pre-processed to crop out each eye area separately as most images contained the full face and both eyes. This reduced the resolution of the eye images significantly. All images are of size 416x416 pixels. The training set comprises 2502 eye images, while the validation set contains 626 images. All of these images contain at least one freckle. The test set contains 1,126 images out of which only 643 images have freckles according to metadata from the original study. For training, the iris freckles were annotated using Labelbox [27] with bounding boxes of size 20x20 pixels.

B. Experimental Setting

We designed the following experiments to compare the performance of:

- 3 scale YOLO - no classification (YOLOv3-nc)
- 1 scale YOLO - with classification (Slim-YOLO-wc)
- 1 scale YOLO - no classification (Slim-YOLO-nc)
- 1 scale YOLO - no classification and no anchors (Slim-YOLO)

Each of these implementation were compared with the following object detection models:

- SSD
- Faster R-CNN
- YOLOv3
- YOLOv3-tiny

The configurations of all YOLOv3 variants are presented in Table I.

YOLOv3-tiny is a lighter, more efficient version of YOLOv3, containing only 13 convolutional layers [28]. We apply the modifications of removing scales, classification, and anchors on YOLOv3-tiny as well (Slim-YOLOv3-tiny) for further comparison with Slim-YOLO.

All experiments were carried out on a NVIDIA Tesla V100 32GB GPU. All YOLO variant models were trained for 300 epochs, while SSD and Faster R-CNN were trained for 100 epochs. The IoU threshold and object confidence threshold for evaluation were 0.6 and 0.3 respectively. For each of the experiments, we report the training time, inference time, detection accuracy with mean average precision (mAP) as the metric, and the size of the network.

C. Results

There were 9 different training trials that were carried out as experiments. Table II presents the mAP and durations for training YOLOv3 variations, SSD, and Faster RCNN.

From the results, it is evident that Slim-YOLO is able to maintain comparable detection accuracy to YOLOv3 despite being 60% smaller in the memory footprint and 40% faster with training/inference times. While YOLOv3 showed a mAP of 93.7%, Slim-YOLOv3 also achieved a mAP of 90.7%. Fig. 5 shows the difference between mAP achieved for YOLOv3 and Slim-YOLO. Both trends are similarly increasing and stable. Fig. 6 illustrates the training loss for YOLOv3 and Slim-YOLO, indicating that the Slim-YOLO was able to converge

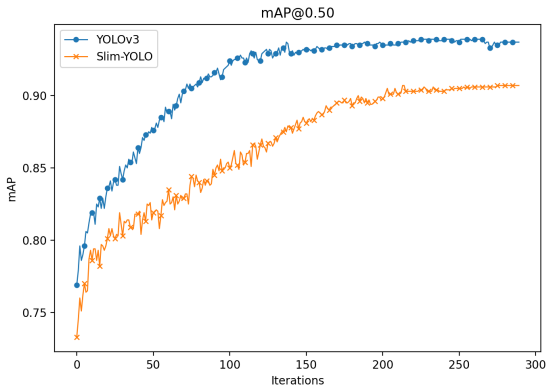

Fig. 5. Comparison of Mean Average Precision between YOLOv3 and Slim-YOLO

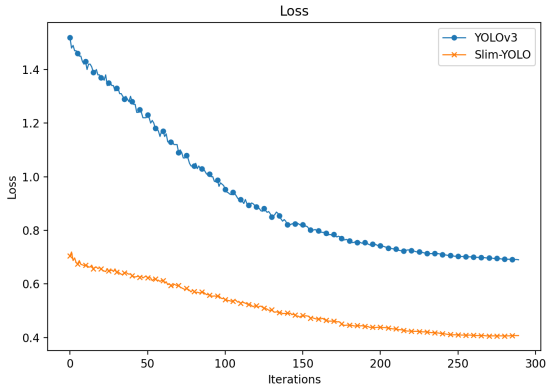

Fig. 6. Comparison of training loss between YOLOv3 and Slim-YOLO

more quickly and smoothly with less overfitting. This reveals that the removal of classification and detection scales have not affected the performance of the model negatively, considering that the freckles we wanted to detect are fairly small objects. It is also clear that anchor boxes become redundant for detection of single-class objects regardless of object size.

Both training time and inference time for Slim-YOLO variations are lower than that of default YOLOv3, indicating the model has become more lightweight and efficient. Removing just classification from YOLOv3 did not have much of an impact on YOLOv3-nc, but reducing detection scales with Slim-YOLO variations have significantly reduced the memory consumption of the model, along with training and inference durations.

Even though YOLOv3-tiny is much more compact and efficient than both YOLOv3 and Slim-YOLO, it suffers in its detection accuracy. However, Slim-YOLOv3-tiny is able to maintain a comparable mAP with default YOLOv3-tiny. This indicates that the modifications we introduce in this work are applicable across many YOLOv3 models for detecting just one class of objects in images.

On the other hand, the state-of-the-art object detectors SSD and Faster R-CNN did not achieve a good detection accuracy

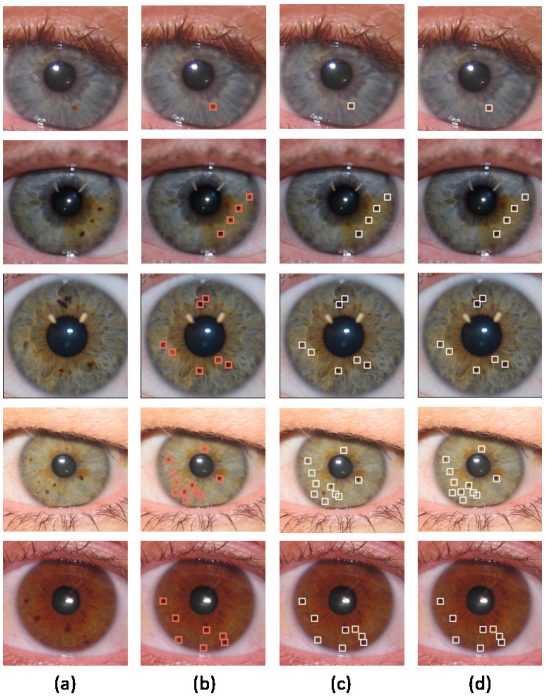

Fig. 7. (a) Original images, (b) ground truth, and predictions of pigmented iris freckles. (c) with YOLOv3, (d) with Slim-YOLO

compared to the the YOLOv3 variants, and are much slower with inference. This could be due to the fact that these models are constrained in their performance of detecting small objects. Therefore, SSD and Faster R-CNN can be considered unsuitable for the application at hand.

According to the visual results illustrated in Fig. 7 Slim-YOLO is capable of detecting pigmented iris freckle detection as accurately as YOLOv3.

V. Conclusion

In this paper, we present a simplified YOLO network, Slim-YOLO, for pigmented iris freckle detection. Our main contributions are concluded as below.

First, Slim-YOLO was able to achieve its objective of improving YOLOv3 to be simpler and lighter while maintaining comparable detection accuracy for single-class small object detection. We do this strategically, by first reducing the number of detection scale, then by removing the need for computing classification loss, and finally by avoiding the restrictions applied by anchor boxes in YOLOv3.

Secondly, we show that Slim-YOLO is highly capable of automatically detecting pigmented iris freckles accurately and with minimal supervision. Not only can it detect freckles in darker coloured eyes, Slim-YOLO also avoids misinterpreting textural complications of the iris as freckles. While we do not perform extensive evaluations of the model's performance on mobile devices, if this light-weight model is to be deployed on mobile devices and the general public would be able to perform self-checks of their eyes, the predictions can be forwarded to clinicians. Thus, Slim-YOLO can potentially be of

very good use for clinicians in making educated and automatic predictions of iris freckles without having to manually count them. Our contributions can therefore serve to improve the use of pigmented iris freckles as a bio-marker for early cutaneous melanoma detection. The question of how accurate this work is from a clinical point of view with regard to melanoma detection remains to be answered, but it would require a larger scale clinical study that is out of scope for our current work.

As future work, Slim-YOLO can be simplified further by removing the need for the calculation of prediction bounding box width and height, given that the predictions are approximately the same size. This will fully utilize Slim-YOLO as an object detection model based on weakly-supervised point annotations only. Furthermore, the simplified, lighter model can be deployed on mobile devices with simple cameras. This will allow access to the general public in rural or remote areas to self-check iris freckles and identify signs of melanoma earlier.

VI. Acknowledgements

This work has been funded by the Merchant Charitable Foundation and the Australian Research Council through grant FT170100072.

HPS holds an NHMRC MRFF Next Generation Clinical Researchers Program Practitioner Fellowship (APP1137127).

References

[1] D. L. Narayanan, R. N. Saladi, and J. L. Fox, "Ultraviolet radiation and skin cancer," *International Journal of Dermatology*, vol. 49, no. 9, pp. 978–986, sep 2010. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-4632.2010.04474.x

[2] J. Ferlay, M. Colombet, I. Soerjomataram, C. Mathers, D. M. Parkin, M. Piñeros, A. Znaor, and F. Bray, "Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods," *International Journal of Cancer*, vol. 144, no. 8, pp. 1941–1953, apr 2019.

[3] International Agency for Research in Cancer; World Health organization, "Global Cancer Observatory," 2018. [Online]. Available: https://gco.iarc.fr/

[4] EHINZ - Environmental Health Indicators New Zealand, "Melanoma." [Online]. Available: https://www.ehinz.ac.nz/indicators/uv-exposure/melanoma/

[5] The Skin Cancer Foundation, "Skin Cancer Facts & Statistics," 2021. [Online]. Available: https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/

[6] A. M. Laino, E. G. Berry, K. Jagirdar, K. J. Lee, D. L. Duffy, H. P. Soyer, and R. A. Sturm, "Iris pigmented lesions as a marker of cutaneous melanoma risk: an Australian case–control study," *British Journal of Dermatology*, vol. 178, no. 5, pp. 1119–1127, may 2018.

[7] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018. [Online]. Available: https://pjreddie.com/yolo/.http://arxiv.org/abs/1804.02767

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[9] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Computer Vision - ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 346–361.

[12] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 379–387. [Online]. Available: https://github.com/daijifeng001/r-fcn.

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37.

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[15] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517–6525.

[16] S. Yao, Y. Chen, X. Tian, R. Jiang, and S. Ma, "An improved algorithm for detecting pneumonia based on yolov3," *Applied Sciences*, vol. 10, no. 5, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/5/1818

[17] V. V. Molchanov, B. V. Vishnyakov, Y. V. Vizilter, O. V. Vishnyakova, and V. A. Knyaz, "Pedestrian detection in video surveillance using fully convolutional YOLO neural network," in *Automated Visual Inspection and Machine Vision II*, vol. 10334, 2017, p. 103340Q. [Online]. Available: https://www.researchgate.net/publication/317967088

[18] Z. Cao, T. Liao, W. Song, Z. Chen, and C. Li, "Detecting the shuttlecock for a badminton robot: A YOLO based approach," *Expert Systems with Applications*, vol. 164, p. 113833, feb 2021.

[19] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Computers and Electronics in Agriculture*, vol. 157, pp. 417–426, feb 2019.

[20] M. T. Pham, L. Courtrai, C. Friguet, S. Lefèvre, and A. Baussard, "YOLO-fine: One-stage detector of small objects under various backgrounds in remote sensing images," *Remote Sensing*, vol. 12, no. 15, pp. 1–26, 2020.

[21] M. Liu, X. Wang, A. Zhou, X. Fu, Y. Ma, and C. Piao, "Uav-yolo: Small object detection on unmanned aerial vehicle perspective," *Sensors (Switzerland)*, vol. 20, no. 8, pp. 1–12, 2020.

[22] W. Fang, L. Wang, and P. Ren, "Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments," *IEEE Access*, vol. 8, pp. 1935–1944, 2020.

[23] A. Wong, M. Famuori, M. J. Shafiee, F. Li, B. Chwyl, and J. Chung, "Yolo nano: a highly compact you only look once convolutional neural network for object detection," in *2019 Fifth Workshop on Energy Efficient Machine Learning and Cognitive Computing - NeurIPS Edition (EMC2-NIPS)*, 2019, pp. 22–25.

[24] C. Schwab, C. Mayer, I. Zalaudek, R. Riedl, M. Richtig, W. Wackernagel, R. Hofmann-Wellenhof, G. Richtig, G. Langmann, L. Tarmann, A. Wedrich, and E. Richtig, "Iris Freckles a Potential Biomarker for Chronic Sun Damage," *Investigative Ophthalmology & Visual Science*, vol. 58, no. 6, pp. BIO174–BIO179, 07 2017. [Online]. Available: https://doi.org/10.1167/iovs.17-21751

[25] R. Z. Csoma, E. Tóth-Molnár, A. Varga, H. Szabó, H. Orvos, L. Kemény, and J. Oláh, "Risk factors and relationship of cutaneous and uveal melanocytic lesions in monozygotic and dizygotic twin pairs," *PLOS ONE*, vol. 11, no. 8, pp. 1–14, 08 2016. [Online]. Available: https://doi.org/10.1371/journal.pone.0160146

[26] R. A. Sturm and M. Larsson, "Genetics of human iris colour and patterns," *Pigment Cell and Melanoma Research*, vol. 22, no. 5, pp. 544–562, 2009.

[27] LabelBox Inc., "LabelBox." [Online]. Available: https://labelbox.com/

[28] P. Adarsh, P. Rathi, and M. Kumar, "YOLO v3-Tiny: Object Detection and Recognition using one stage improved model," in *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*. Institute of Electrical and Electronics Engineers Inc., mar 2020, pp. 687–694.