algorithms

MDPI

Article

Detectron2 for Lesion Detection in Diabetic Retinopathy

Farheen Chincholi * and Harald Koestler

Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 91054 Erlangen, Germany

* Correspondence: farheen.s.chincholi@fau.de

Abstract: Hemorrhages in the retinal fundus are a common symptom of both diabetic retinopathy and diabetic macular edema, making their detection crucial for early diagnosis and treatment. For this task, the aim is to evaluate the performance of two pre-trained and additionally fine-tuned models from the Detectron2 model zoo, Faster R-CNN (R50-FPN) and Mask R-CNN (R50-FPN). Experiments show that the Mask R-CNN (R50-FPN) model provides highly accurate segmentation masks for each detected hemorrhage, with an accuracy of 99.34%. The Faster R-CNN (R50-FPN) model detects hemorrhages with an accuracy of 99.22%. The results of both models are compared using a publicly available image database with ground truth marked by experts. Overall, this study demonstrates that current models are valuable tools for early diagnosis and treatment of diabetic retinopathy and diabetic macular edema.

Keywords: diabetic retinopathy; diabetic macular edema; deep learning; machine learning; deep neural networks

Citation: Chincholi, F.; Koestler, H. Detectron2 for Lesion Detection in Diabetic Retinopathy. *Algorithms* **2023**, *16*, 147. https://doi.org/10.3390/a16030147

Academic Editors: Bharatendra Rai and S.A. Senthil Kumar

Received: 19 January 2023
Revised: 19 February 2023
Accepted: 2 March 2023
Published: 7 March 2023

1. Introduction

Diabetes is a significant contributor to blindness among people aged 20 to 74 in the United States, according to a study conducted by the National Health and Nutrition Examination Survey (NHANES) at the Centers for Disease Control and Prevention (CDC) [1]. The study, which was published in the Journal of the American Medical Association [2], found a link between diabetes and failing eyesight in people with the disease.

Diabetes can lead to two serious eye conditions: diabetic retinopathy (DR) and diabetic macular edema (DME). DR is assessed by grading the presence and severity of retinopathy in the macula and peripheral retina of each eye [3]. DR is divided into non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR) and is graded based on the presence of microaneurysms, hemorrhages, cotton wool spots, and hard exudates, as illustrated in Figure 1 [4]. Meanwhile, DME is identified by the presence of blot hemorrhages and hard exudates within a 2-disc diameter from the center of the macula.

The treatment for DR and DME is determined by the severity of the condition [5]. With mild or moderate DR, progression can often be prevented through good blood sugar control. Early detection of DR and DME is crucial as evidence suggests that appropriate management at an early stage, including regulation of blood pressure, glucose levels, and lipid profiles, can greatly slow the progression of DR and even reverse moderate NPDR to DR-free stage. DME is treated with anti-VEGF medications and Focal Laser Treatment. For PDR, the main therapy is panretinal photocoagulation (PRP) [6].

In recent years, the utilization of deep learning models has gained popularity in the analysis of retinal images and detection of DR and DME. Despite its effectiveness, previous research in the automatic detection of DR and DME from digital fundus images has faced criticisms due to its black-box approach that relies on various representations for predictions without explicitly displaying the diabetic retinopathy lesions like microaneurysms

and retinal hemorrhages [5]. This has raised concerns among physicians regarding the acceptability of the method for clinical use. However, the new deep learning-based screening software presents an opportunity to address these concerns by filling the gap in detecting both DR and DME simultaneously, while providing a more transparent method for predictions.

This paper develops a prototype of a deep learning-based screening software for the early detection and location of diabetic retinopathy (DR) and diabetic macular edema (DME) from digital fundus images. The prior research in this field, which encompasses classical machine learning, deep learning, and deep neural networks, is critically reviewed and discussed in Section 2. The research methodology, which includes the dataset used, the baseline models (Faster R-CNN (R50-FPN) and Mask R-CNN (R50-FPN)), process flow, and training and loss functions, is detailed in Section 3. The results of the research, including evaluation metrics, analysis of accuracy, false Negative Rate, and convergence speed, and analysis of the baseline models in object detection and instance segmentation, are showcased in Section 4. Finally, Section 5 delves into the discussion and future work related to this research.

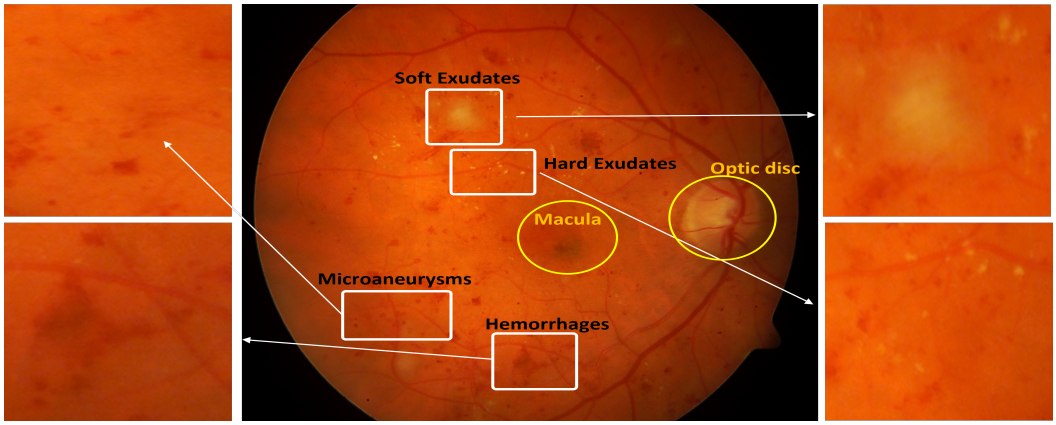


Figure 1. The retinal fundus can exhibit various types of lesions, each with its distinct appearance. Soft exudates appear as white, feathery or fluffy spots, while hard exudates are harder, uneven and have a white or yellowish appearance. Microaneurysms are small, round and have defined borders, measuring less than 3 mm in diameter, and are red in color. Hemorrhages are indications of bleeding in the retina and can manifest as dots, blots or flames. Additionally, the macula and optic disc are also indicated.

2. Related Work

This section provides an overview of the various approaches used for DR detection and grading. The section is divided into three subsections. The first Section 2.1 discusses classical machine learning approaches used for DR detection and grading, including their limitations. The second Section 2.2 focuses on deep learning methods used for grading DR, which have the advantage of handling large and complex data and learning directly from the data. Section 2.3 discusses the use of deep learning methods specifically for detecting DME, which is a common complication of DR.

2.1. Classical ML

In classical machine learning approaches for DR and DME detection, various algorithms have been used, such as linear regression, logistic regression, support vector machine, principal component analysis, decision tree, random forest, and naive Bayes. In some cases, these algorithms have been combined with optimization techniques like moth-flame optimization to enhance their performance.

In a study by Pragathi Nagaraja Rao et al. [7], an integrated machine learning approach was proposed for early detection of diabetic retinopathy. The approach combined the use of support vector machine, principal component analysis, and moth-flame optimization

techniques. The individual performance of decision tree, support vector machine, random forest, and naive Bayes algorithms was initially evaluated on a diabetic retinopathy dataset. The integration of PCA improved the performance of the decision tree algorithm, but reduced the performance of the other algorithms. The integration of moth-flame optimization with SVM and PCA resulted in improved performance with an average of 85.61%.

A study [8] focuses on early detection of diabetic retinopathy using an ensemble machine learning model composed of Random Forest, Decision Tree, Adaboost, K-Nearest Neighbor, and Logistic Regression algorithms. The diabetic retinopathy dataset was normalized using min-max normalization before training the ensemble model, and the results showed that the ensemble model outperformed the individual machine learning algorithms.

According to the study by Alsaih et al. [9], a machine learning framework was developed for DME using optical coherence tomography (OCT) volumes. The dataset used in the study consisted of 32 OCT volumes obtained from the Singapore Eye Research Institute using the CIRRUS SD-OCT device. The study employed pre-processing, feature detection, feature representation, and classification steps to classify the images as normal or DME. The best results were achieved using LBP16-ri vectors and linear-support vector machine in combination with PCA and bag of words representations, with a sensitivity and specificity of 87.5% each.

The studies have shown that classical ML algorithms can be useful in DR and DME detection, however, they have limitations such as difficulty in handling large and complex data, and dependence on the quality of hand-engineered features. These limitations have led to a shift towards deep learning methods in recent years.

2.2. Deep Learning (DL) Methods for DR Grading

In this category of studies, deep neural networks are used to grade DR into different stages, including non-DR, mild NPDR, moderate NPDR, severe NPDR, and PDR.

A study by Dai et al. [6] described the development of a deep learning system called DeepDR for the early detection and grading of diabetic retinopathy. The system was trained on a large dataset from a single ethnic cohort of patients with diabetes. The system was trained using a large dataset from a single ethnic cohort of patients with diabetes. The performance of DeepDR was high in detecting microaneurysms, cotton-wool spots, hard exudates, and hemorrhages, with area under the curve (AUC) scores ranging from 0.901 to 0.967. The grading of diabetic retinopathy was also successful with AUC scores ranging from 0.943 to 0.972. However, further validation in multiethnic and multicenter cohorts is needed to confirm the robustness of lesion detection and grading.

Another study by Ting, D.S.W [5] aimed to evaluate the diagnostic performance of a deep learning system (DLS) for detecting referable diabetic retinopathy using retinal images. The DLS achieved a high accuracy with an area under the curve of 0.936, a sensitivity of 90.5%, and a specificity of 91.6%. However, it should be noted that the focus of this study was only on detecting referable diabetic retinopathy, which may not encompass all cases of diabetic retinopathy. Additionally, the study has limitations, including varying reference standards and a "opaque" effect, which may impact its acceptance by physicians.

In the study [10], the authors present a machine learning-based approach to automate the classification of diabetic retinopathy (DR). Convolutional neural networks (CNNs), such as VGG-16 and VGG-19 [11], were employed to analyze fundus images and categorize the DR severity into five levels ranging from 0 (no DR) to 4 (proliferative DR). The system was evaluated using various performance metrics including accuracy, sensitivity, specificity, and AUC, and the results showed that the system achieved a high accuracy of 82%, sensitivity of 80%, specificity of 82%, and AUC of 0.904.

These studies demonstrate the effectiveness of deep learning algorithms in grading DR, which can be useful in the diagnosis and prompt treatment of the disease.

2.3. DL Methods for DME Detection

In the literature, various DL-based approaches have been proposed for detecting and grading DME in digital fundus images. In [12], a deep learning-based approach was proposed for grading DME. A deep convolutional neural network (DCNN) was trained on a large dataset of digital fundus images. The DCNN was trained to distinguish between normal, mild, moderate, and severe DME using supervised learning approach, which means that the network was trained to make predictions based on the labeled examples in the training dataset. The training process involved iteratively adjusting the weights of the network to minimize the prediction error on the training set, until the error converged to a minimum. The results showed that the DCNN was able to achieve an accuracy of 90.6% in grading DME.

The study [13] presents an end-to-end deep learning approach for the detection and grading of DME in digital fundus images. The authors used a convolutional neural network (CNN) to analyze the fundus images and classify them into different stages of DME, ranging from normal to severe. The performance of the system was evaluated on a large dataset of digital fundus images and results showed that the proposed approach achieved an accuracy of 91.5% in detecting and grading DME.

The study by Chen et al. [14] presents a deep transfer learning-based approach for the detection of DME in digital fundus images. The authors used a pre-trained CNN, which was fine-tuned on a dataset of fundus images with DME. The results of the study showed that the proposed approach was accurate in detecting DME, with an accuracy of 95.7%. This study highlights the potential of transfer learning in medical image analysis, especially in the detection of DME.

These studies demonstrate the effectiveness of deep learning-based approaches for detecting and grading DME in digital fundus images. However, further research is needed to improve the accuracy and robustness of these methods.

2.4. Previous Studies: A Summary of Techniques, Focus, and Challenges

The reviewed literature explored a range of ML and DL techniques for DR detection and summarized the different methods of the models in Table 1. Approximately 70% of the previous studies were focused on detecting the stages of DR or classifying images as DR or non-DR [15], as depicted in Figure 2. In contrast, only 27% of the studies aimed at classifying NPDR and PDR, and a mere 6% aimed at detecting and classifying lesions in fundus images into NPDR or PDR. Despite the various techniques and focus areas, the current challenge in the field remains to create a reliable system that can detect both DR and DME. This challenge has been the emphasis of recent studies and is a crucial area of ongoing research in the field of DR detection.

Table 1. Table summarizing classical machine learning and deep learning methods for diabetic retinopathy and diabetic macular edema detection and grading

Field	Ref	Method	Database
Classical Machine Learning	[7]	Integrated SVM, PCA, and Moth-Flame Optimization	Own dataset
	[8]	Ensemble of Random Forest, Decision Tree, Adaboost, K-NN, Logistic Regression	DR dataset
	[9]	LBP16-ri vectors and SVM with PCA and BoW representations	-
DL methods for DR grading	[6]	DeepDR	Single ethnic dataset
	[5]	DLS	Single ethnic dataset
	[10]	CNNs (VGG-16 and VGG-19)	Fundus images
DL methods for DME detection	[12]	Deep Convolutional Neural Network	Own dataset
	[13]	Convolutional Neural Network	Own dataset
	[14]	Transfer Learning-based Convolutional Neural Network	Own dataset

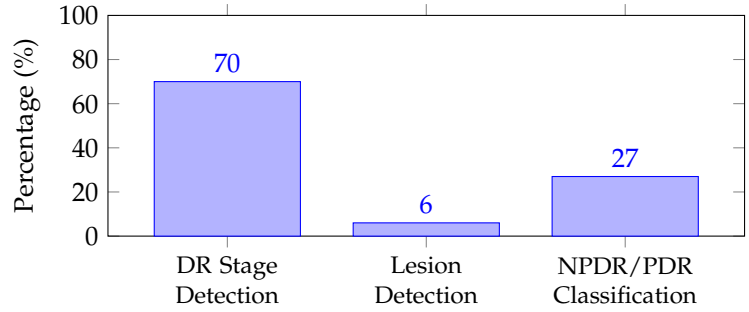


Figure 2. Percentage of studies with a focus on different aspects of DR, highlighting the lower percentage dedicated to lesion detection.

3. Methodology

3.1. Process Flow Explanation

Figure 3 provides a visual representation of the process of using the models to detect hemorrhages within an input image and determine the presence of either DME or DR. The input image is fed into the base deep learning models trained to detect hemorrhages and locate them within the image. The models output a prediction on the presence of hemorrhages within the image. If a hemorrhage is detected, the location of the hemorrhage with respect to the macula is calculated. The location of the hemorrhage is then used to classify the images. If the hemorrhage is found within the macula, it is determined that DME is present. Otherwise, DR is present.

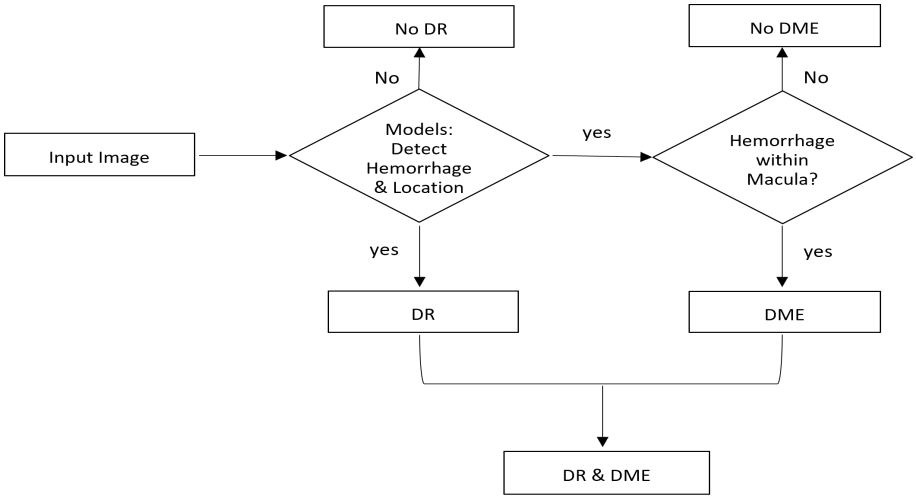


Figure 3. The step-by-step process of detecting hemorrhages and locating DME/DR with Faster R-CNN (R50-FPN) and Mask R-CNN (R50-FPN) models.

3.2. Dataset

The study used 89 color fundus images from a publicly available database DIARETDB1 (89) [16], with 84 images having at least mild non-proliferative diabetic retinopathy (NPDR) and 5 images of normal eyes. Medical experts marked the images as ground truth. Annotations were created using Labelme, which were then converted to the Common Objects in Context (COCO) [17] format using a local script or the tool Roboflow [18]. The original data set was divided into a training set of 28 images and a test set of 61 images, with different confidence levels marking the affected areas of the images. The training set had 18 images with hard exudates, 6 images with soft exudates, 19 images with microaneurysms, and 21 images with hemorrhages, while the test set had 20 images with hard exudates, 9 images with soft exudates, 20 images with microaneurysms, and 18 images with hemorrhages. The ground truth confidence level in the DIARETDB1 data set was set to 0.75.

Publicly available datasets such as MESSIDOR [19], APTOS 2019 Blindness Detection [20], and the kaggle [21] diabetic-retinopathy-detection database are often utilized for detecting diabetic retinopathy (DR) and diabetic macular edema (DME) in public domain. However, one drawback of these datasets is that they solely offer a severity rating ranging from 0 to 4 for each image, and do not include precise annotations of the lesions present in the images. This can pose difficulties when utilizing these datasets for tasks that require pinpointing the location of the lesions.

3.3. Baseline Models

The study utilizes two models, Faster R-CNN (R50-FPN) and Mask R-CNN (R50-FPN), which are both based on the Detectron2 platform developed by Facebook AI Research (FAIR) [22]. This platform provides a flexible environment for developing and deploying computer vision algorithms and includes various object detection techniques, such as Mask R-CNN [23], RetinaNet [24], Faster R-CNN [25], and RPN. The Faster R-CNN (R50-FPN) [26] is a Faster R-CNN model with a ResNet50+FPN backbone, and the Mask R-CNN (R50-FPN) [27] is a Mask R-CNN model with a ResNet50+FPN backbone. These models will be further explained in Sections 3.3.1 and 3.3.2.

3.3.1. Faster R-CNN (R50-FPN) Architecture

The architecture of Faster R-CNN (R50-FPN) [28] is depicted in Figure 4 and is made up of three main parts: the Backbone Network, the Region Proposal Network, and the Box Head. The Backbone Network is a ResNet+FPN backbone that extracts feature maps from the input image. The ResNet part of the backbone consists of residual blocks stacked on top of one another, which are simpler to optimize and improve accuracy compared to traditional deep networks. The Feature Pyramid Network (FPN) [29] part of the backbone creates proportionally scaled feature maps from a single-scale input image of any size.

The Region Proposal Network (RPN) [28] is a deep learning network used for object detection that generates rectangular object proposals with corresponding objectness scores from the input image. It shares its convolutional layers with the Fast R-CNN object identification network to save computation time. The Box Head is a type of region of interest head that uses fully connected layers to refine box placements and classify objects. It takes the feature maps and region proposals generated by the RPN and performs computations on each region, cutting and warping the feature maps with proposal boxes to create multiple fixed-size features. The final output is limited to 100 boxes after non-maximum suppression.

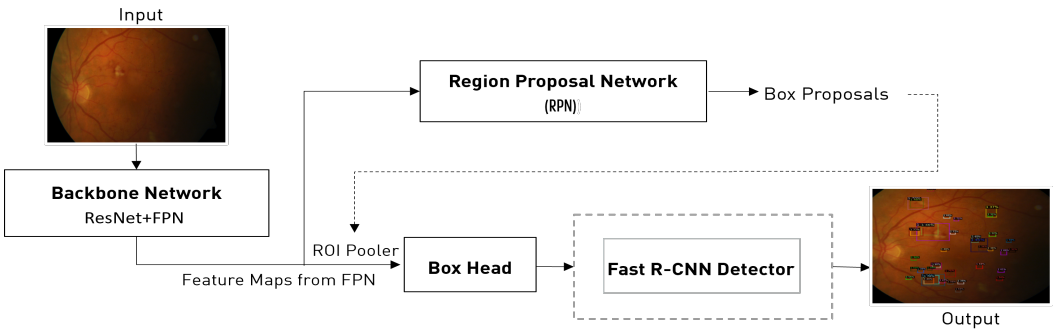


Figure 4. The architecture of Faster R-CNN (R50-FPN) consists of 3 stages: (1) Backbone Network, (2) Region Proposal Network (RPN), (3) Box Head, and Fast R-CNN for object identification. Mask R-CNN (R50-FPN) employs the same 3-step procedure as Faster R-CNN (R50-FPN), but also includes a binary mask produced for each ROI in addition to class and box offset predictions in the third stage.

3.3.2. Mask R-CNN (R50-FPN) Architecture

Mask R-CNN [23] is an extension of Faster R-CNN, with the main difference being the inclusion of an additional output branch for generating object masks. While Faster R-CNN outputs class labels and bounding-box offsets, Mask R-CNN also generates binary masks for each region of interest (ROI). The procedure for Mask R-CNN (R50-FPN) [28] is similar

to that of Faster R-CNN (R50-FPN), with the first two stages being the backbone network and RPN. In the third stage, in addition to class and box offset predictions, a binary mask is generated for each ROI. This allows for more precise spatial arrangement of objects, as it involves pixel-to-pixel alignment.

Mask Representation: Mask R-CNN [23] predicts binary masks for objects in an image by using a fully convolutional network (FCN). The FCN generates an m x m mask for each region of interest (ROI), preserving the pixel-to-pixel correspondence through convolutions. This allows for precise extraction of the object's spatial structure. The RoIAlign layer in Mask R-CNN helps maintain the accuracy of the small RoI features by aligning them with the input, leading to better mask prediction performance. The RoIAlign layer is crucial for accurate mask prediction and ensures per-pixel spatial correspondence.

RoIAlign: The RoIAlign layer in Mask R-CNN improves the accuracy of the features extracted from regions of interest (RoIs) compared to the standard RoIPool operation. RoIPool quantizes the RoI to the granularity of the feature map and divides it into spatial bins, which can result in inaccuracies. RoIAlign, on the other hand, aligns the retrieved features with the input, removing the harsh quantization introduced by RoIPool and providing improved alignment. This improved alignment allows for more accurate bounding box regression, resulting in better object detection performance compared to models like Fast R-CNN.

3.4. Training and Loss Function

The RPN is trained to classify the anchor boxes as objects or not objects by applying back-propagation and Stochastic Gradient Descent (SGD) [30]. The RPN training uses the "image-centric" sampling method, where each mini-batch originates from a single image with a mix of good and bad example anchors. The loss function is computed by randomly selecting 256 anchors from the image, with a positive to negative anchor ratio of up to 1:1. If an image has fewer than 128 positive samples, the mini-batch is boosted with negative samples.

The entire architecture is trained using a four stage alternating training method [23]. First, the backbone CNN network is initialized with ImageNet weights and region proposals are generated through fine-tuning these weights. Then, the RPN is trained and used as a proposal for the object detection network. The backbone network is again initialized with ImageNet weights, but it is not yet connected to the RPN network. The RPN and Fast R-CNN detector are then fine-tuned by fixing the common layer weights and tweaking only the layers specific to the detector network.

Both machine learning models were trained on an Intel CPU with 4GB of RAM, using Python 3.7, PyTorch 1.8 or later, TorchVision, and TensorBoard version 1.6.0, on a Linux system. With a learning rate of 0.00025, the models were trained for 30,000 iterations, using data augmentations such as Re-size-Shortest-Edge to maintain image aspect ratios [31]. This can help prevent the image from being distorted, which is especially important for image classification and object detection tasks where the shape and size of objects are critical features for determining their class or location.

4. Results

4.1. Evaluation Metrics

The performance of the detector is evaluated using sensitivity, specificity, and accuracy [32]. These performance measures are commonly used in medical diagnosis and are defined as follows:

Sensitivity is the ratio of true positive (TP) cases to the sum of true positive and false negative (FN) cases. Specificity is the ratio of true negative (TN) cases to the sum of true negative and false positive (FP) cases [33]. Accuracy is the ratio of the sum of true positive and true negative cases to the overall sample.

$$Sensitivity = TP/(TP + FN) \quad (1)$$

$$Specificity = TN/(TN + FP) \quad (2)$$

$$Accuracy = TP + TN/(TN + TP + FN + FP) \quad (3)$$

In the evaluation of object detection and instance segmentation models in Section 4.3, the key metrics utilized to assess the performance are Intersection over Union (IoU) and the standard metrics provided by the Common Objects in Context (COCO) dataset [17]. IoU measures the overlap between the predicted and ground-truth bounding boxes and is calculated as the ratio of their intersection to their union. A high IoU score indicates a good match between the two bounding boxes. An example of IoU scores is shown pictorially in a Figure 5.

In addition to IoU, COCO provides standard metrics, including Average Precision (AP) and Average Recall (AR) [34]. These metrics are listed in Table 2. AP measures the accuracy of the model's object detection, while AR measures the model's ability to recall objects. AP is calculated across different scales and at different IoU thresholds (IoU = 0.50, 0.75) and is divided into categories based on object size (small, medium, large). These categories are represented as APsmall (for small objects with an area < 322), APmedium (for medium objects with an area between 322 and 962), and APlarge (for large objects with an area > 962) [35].

Similarly, average recall will be calculated at different levels of maximum detections per image (ARmax = 1, 10, 100) and divided into categories based on object size (ARsmall, ARmedium, ARlarge). These metrics provide a comprehensive evaluation of the performance of object detection models and help assess the strengths and weaknesses of the approach, specifically with respect to object size and detection count.

Table 2. Evaluation metrics from COCO [35] used to assess the performance of object detection models on the COCO dataset. These evaluation metrics include Average Precision (AP) and Average Recall (AR) for bounding box detection, as well as AP and AR for different object sizes and IoU thresholds.

Average Precision (AP)		**AP Across Scales**	
AP	% AP at IoU = 0.50:0.05:0.95	APmedium	% AP for medium objects: 322 < area < 962
APIoU = 0.50	% AP at IoU = 0.50	APsmall	% AP for small objects: area < 322
APIoU = 0.75	% AP at IoU = 0.75	APlarge	% AP for large objects: area > 962
Average Recall (AR)		**AR Across Scales**	
ARmax = 1	% AR given 1 detection per image	ARsmall	% AR for small objects: area < 322
ARmax = 10	% AR given 10 detections per image	ARmedium	% AR for medium objects: 322 < area < 962
ARmax = 100	% AR given 100 detections per image	ARlarge	% AR for large objects: area > 962

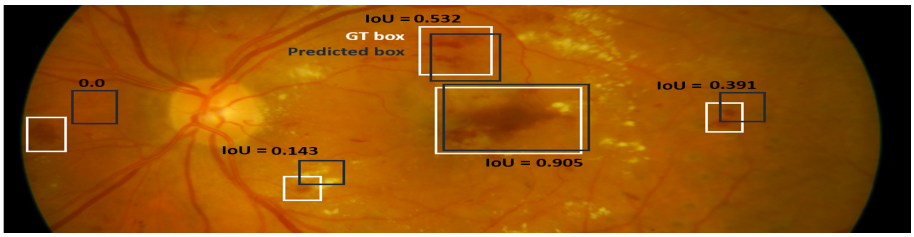


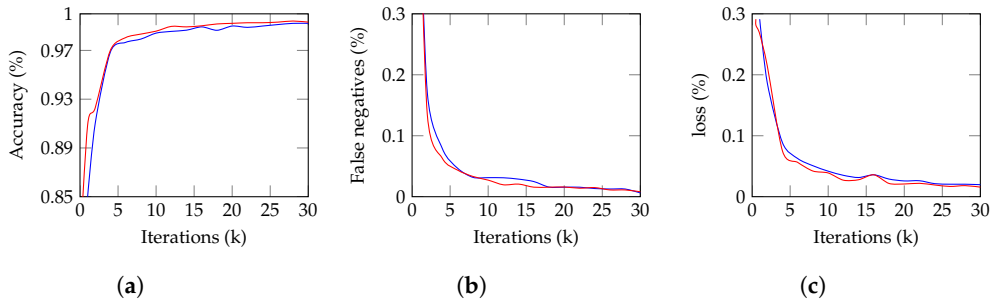
Figure 5. Example IoU scores for the detected bounding box.

4.2. Analysis of Accuracy, False Negative Rate, and Convergence Speed

Figure 6a illustrates the accuracy of the Mask R-CNN (R50-FPN), which was 99.34% with a sensitivity of 97.5% and a specificity of 96.6%. This indicates that the model accurately classified the majority of positive and negative cases, with a low number of false positive and false negative results. For the Faster R-CNN (R50-FPN), the accuracy was 99.22%

with a sensitivity of 97.37% and a specificity of 96.49%. Although the accuracy is slightly lower compared to the Mask R-CNN (R50-FPN), the sensitivity and specificity are still high, which means the model still has a good performance in classifying the positive and negative cases.

Figure 6b shows the comparison of the false negative rate of the two models. The results indicate that Mask R-CNN (R50-FPN) performs better, with a lower false negative rate of approximately 0.8% compared to Faster R-CNN (R50-FPN)'s rate of 0.6%. Figure 6c illustrates the convergence of the loss value for both models during the training process. Both models converge to a minimum value, with Faster R-CNN (R50-FPN) demonstrating faster convergence and a lower minimum loss value compared to Mask R-CNN (R50-FPN).



(**a**) (**b**) (**c**)

Figure 6. Faster R-CNN (R50-FPN) in red and Mask R-CNN (R50-FPN) in blue over 30k iterations. (**a**) Illustrates accuracy, with Mask R-CNN (R50-FPN) having a higher overall accuracy of 99.34%. (**b**) Compares false negative rate, with Mask R-CNN (R50-FPN) performing better with a final false negative rate of 0.8%. (**c**) Demonstrates convergence of loss value, with Faster R-CNN (R50-FPN) demonstrating faster convergence and a lower minimum loss value.

4.3. Analysis of Baseline Models in Object Detection and Instance Segmentation

In the context of this work, 'detection' refers specifically to the identification and localization of hemorrhages in retinal images, and should not be confused with the more general concept of object detection. For Faster R-CNN (R50-FPN), Table 3 reveals that the highest Average Precision (AP) is obtained at an Intersection over Union (IoU) threshold of 0.50:0.95 with all maxdets set to 100, with a value of 0.477. However, the Average Recall (AR) is low at 0.032 under the same setting. The model performs well for large objects, with an AP of 0.812 and an AR of 0.830 at IoU 0.50:0.95 with all maxdets set to 100. But it performs poorly on small objects, with an AP of 0.180 and an AR of 0.182 under the same setting.

Table 4 indicates that Mask R-CNN (R50-FPN) has superior performance in both object detection and instance segmentation tasks, with a slightly better performance in object detection. The AP for object detection is 0.475 at IoU 0.50:0.95 with all maxdets set to 100, while the AR is 0.031 at the same setting. The AP for instance segmentation is 0.424 at the same setting. The AP is highest for medium-sized objects, with a value of 0.612 in object detection and 0.543 in instance segmentation.

The objective evaluation of the accuracy and sensitivity of a Faster R-CNN (R-50-FPN) model can be facilitated by visually comparing the model's output to the ground truth annotations, as demonstrated in Figure 7. This comparison enables the detection of discrepancies or errors in the model's predictions and provides insights into the degree of agreement with expert annotations.

In the context of hemorrhage detection and segmentation, baseline models were evaluated, and the results are presented in Figure 8. The first image demonstrates early stages of DR without DME, with a marked area on the top right that indicates a failure in identification. The second image shows successful hemorrhage detection in a case with both DR and DME. These results showcase the potential of baseline models for detecting and segmenting hemorrhages, which can aid in the diagnosis and monitoring of DR in patients.

In conclusion, the summary of results shown in Table 5 indicate that both models have a relatively high average precision, with Mask R-CNN (R50-FPN) performing slightly better overall compared to Faster R-CNN (R50-FPN). However, the performance of both models can be improved by adjusting the IoU and maximum detection settings.

Table 3. COCO evaluation metrics for object detection using Faster R-CNN (R50-FPN) baseline. The metric maxdets specifies the maximum number of detections per image for calculating the average precision (AP) and average recall (AR).

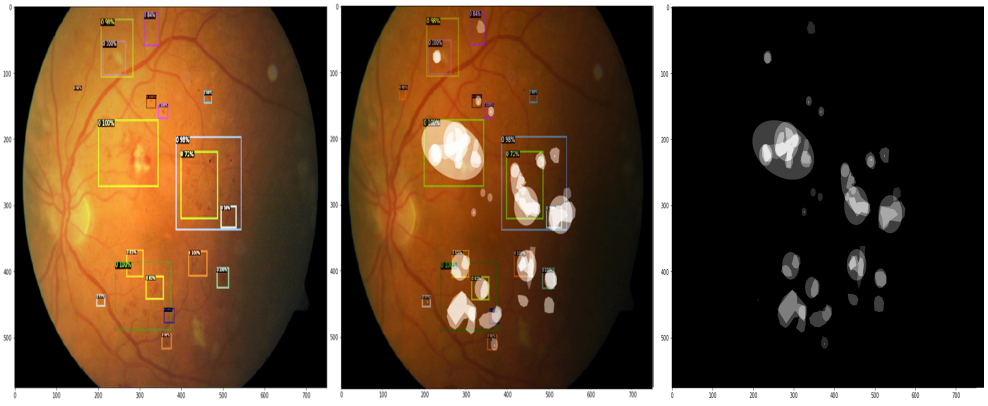
Backbone: Faster R-CNN(R50-FPN)							
IoU	**Area**	**Maxdets**	**AP**	**IoU**	**Area**	**Maxdets**	**AR**
0.50:0.95	all	100	0.477	0.50:0.95	all	1	0.032
0.50	all	100	0.590	0.50:0.95	all	10	0.317
0.75	all	100	0.478	0.50:0.95	all	100	0.514
0.50:0.95	small	100	0.180	0.50:0.95	small	100	0.182
0.50:0.95	medium	100	0.608	0.50:0.95	medium	100	0.662
0.50:0.95	large	100	0.812	0.50:0.95	large	100	0.830

Table 4. COCO Evaluation Metrics for object detection and instance segmentation using Mask R-CNN (R50-FPN) baseline.

Backbone: Mask R-CNN(R50-FPN)							
Object Detection							
IoU	**Area**	**Maxdets**	**AP**	**IoU**	**Area**	**Maxdets**	**AR**
0.50:0.95	all	100	0.475	0.50:0.95	all	1	0.031
0.50	all	100	0.584	0.50:0.95	all	10	0.303
0.75	all	100	0.470	0.50:0.95	all	100	0.514
0.50:0.95	small	100	0.205	0.50:0.95	small	100	0.204
0.50:0.95	medium	100	0.612	0.50:0.95	medium	100	0.677
0.50:0.95	large	100	0.734	0.50:0.95	large	100	0.747
Instance Segmentation							
0.50:0.95	all	100	0.424	0.50:0.95	all	1	0.030
0.50	all	100	0.545	0.50:0.95	all	10	0.274
0.75	all	100	0.441	0.50:0.95	all	100	0.459
0.50:0.95	small	100	0.160	0.50:0.95	small	100	0.171
0.50:0.95	medium	100	0.543	0.50:0.95	medium	100	0.603
0.50:0.95	large	100	0.686	0.50:0.95	large	100	0.700

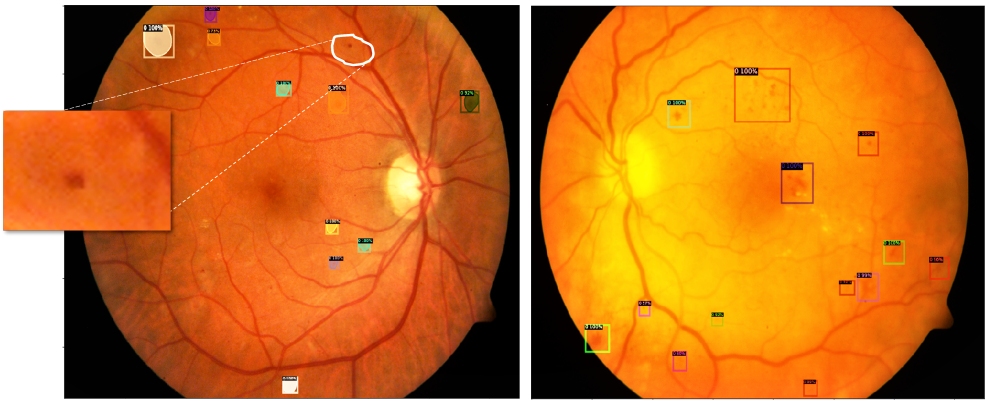
Table 5. Summary of object detection performance reported as Average Precision (AP) at different IoU thresholds, using different backbones."BB" denotes the bounding box-based approach, while "IS" denotes the instance segmentation-based approach.

Backbone	**AP**	**AP50**	**AP75**	**APs**	**APm**	**APl**
Faster rcnn: BB	47.720	58.988	47.782	17.975	60.761	81.220
Mask rcnn: BB	47.492	58.421	47.032	20.525	61.198	73.415
Mask rcnn: IS	42.378	54.517	44.141	16.031	54.304	68.574



(a) (b) (c)

Figure 7. The results of Faster R-CNN (R-50-FPN) can be visually compared to the ground truth to assess the performance of the model. (**a**) Shows the output image from Faster R-CNN (R-50-FPN), (**b**) Shows the ground truth image overlayed on the output image, and (**c**) Shows the ground truth image marked by experts. The visual comparison of the model's predictions to expert annotations offers an objective evaluation of the model's accuracy and sensitivity. It highlights any discrepancies or errors in the model's predictions and provides insight into the degree of agreement between the model's predictions and the expert annotations.



(a) (b)

Figure 8. Results of hemorrhage detection and segmentation from baseline models. (**a**) Hemorrhage detection indicating early stages of DR without DME. The marked area on the top right represents a failure in identification. (**b**) Hemorrhage detection indicating both DR and a clear case of DME.

5. Discussion and Future Work

In recent years, deep learning models have been utilized to analyze retinal images and detect the presence of diabetic retinopathy and diabetic macular edema. Previous work in this field has relied on multiple levels of representation to make predictions, without explicitly displaying the diabetic retinopathy lesions such as microaneurysms and retinal hemorrhages. This black-box approach, while effective, has raised concerns about its acceptability for clinical use among physicians. These studies have primarily focused on detecting diabetic retinopathy and diabetic macular edema at later stages, such as individuals with referable DR or advanced DR, which indicates that these patients require closer follow-up or treatment from ophthalmologists. However, early-stage detection is essential to achieving the best possible outcomes for patients with diabetic retinopathy and diabetic macular edema. Research shows that with appropriate care at an early stage, it may be possible to significantly slow the development of DR and even reverse moderate NPDR to a DR-free stage in order to achieve optimal control of blood pressure, glucose levels, and lipid profiles. In addition, advanced DR is incurable, which underscores the importance of early detection.

The results of the evaluation of the deep learning models, Faster R-CNN (R50-FPN) and Mask R-CNN (R50-FPN), in detecting early stages of DR and DME by focusing on the presence of hemorrhages in the retinal fundus show promising results. Both models achieved a high accuracy, with the Mask R-CNN (R50-FPN) having a higher accuracy of 99.34% compared to the Faster R-CNN (R50-FPN) which had an accuracy of 99.22%. The Mask R-CNN (R50-FPN) also had a higher sensitivity of 97.5% and a higher specificity of 96.6%, while the Faster R-CNN (R50-FPN) had a sensitivity of 97.37% and a specificity of 96.49%. This indicates that both models accurately classified the majority of positive and negative cases, with a low number of false positive and false negative results.

In the future, improvement in the accuracy of these models can be achieved through fine-tuning with a larger and more diverse dataset that includes small objects. The impact of alternate backbone architectures, such as InceptionNet and DenseNet, on the models' performance should also be investigated. The ultimate goal is to make the detection of DR and DME fully automatic by training the model to identify all pathologies, the macula, and the optic disc based on distinctive features. It is important to note that most of the publicly available datasets which are commonly used for diabetic retinopathy and diabetic macular edema detection, have limitations in terms of annotations and lesion localization. These datasets only provide a severity score for each image on a scale of 0 to 4, without specific annotations of the lesions on the images, which can make it challenging to use them for tasks that require lesion localization. Thus, it is crucial to develop and use datasets with more specific annotations for lesion detection and localization in the training of deep learning models for accurate and efficient detection of diabetic retinopathy and diabetic macular edema.

Author Contributions: Conceptualization, F.C.; software, F.C.; data curation, F.C.; writing—original draft preparation, F.C.; writing—review and editing, F.C.; supervision, H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: A publicly available dataset is used in this work which is available at https://www.it.lut.fi/project/imageret/diaretdb1/ (accessed on 1 July 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Diabetic Retinopathy: Cdc.gov-visionhealth-factsheet. Available online: https://www.cdc.gov/visionhealth/pdf/factsheet.pdf (accessed on 1 September 2022).
2. Zhang, X.; Saaddine, J.B.; Chou, C.F.; Cotch, M.F.; Cheng, Y.J.; Geiss, L.S.; Gregg, E.W.; Albright, A.L.; Klein, B.E.K.; Klein, R. Prevalence of Diabetic Retinopathy in the United States, 2005–2008. *JAMA* **2010**, *304*, 649–656.
3. Zachariah, S.E.A. Grading diabetic retinopathy (DR) using the Scottish grading protocol. *Community Eye Health* **2015**, *28*, 72–73. [PubMed]
4. Diabetes Retinal Screening, Grading and Management Guideline. Available online: https://www.worlddiabetesfoundation.org/sites/default/files/WDF08-386%20Pacific%20Island%20Ret%20Screen%20Guidelines.pdf (accessed on 1 September 2022).
5. Ting, D.S.W.; Cheung, C.Y.L.; Lim, G.; Tan, G.S.W.; Quang, N.D.; Gan, A.; Hamzah, H.; Garcia-Franco, R.; San Yeo, I.Y.; Lee, S.Y.; et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA* **2017**, *318*, 2211–2223. [CrossRef] [PubMed]
6. Dai, L.; Wu, L.; Li, H.; Cai, C.; Wu, Q.; Kong, H.; Liu, R.; Wang, X.; Hou, X.; Liu, Y.; et al. A deep learning system for detecting diabetic retinopathy across the disease spectrum. *Nat. Commun.* **2021**, *12*, 3242. [CrossRef] [PubMed]
7. Pragathi, P.; Rao, A.N. An effective integrated machine learning approach for detecting diabetic retinopathy. *Open Comput. Sci.* **2022**, *12*, 83–91. [CrossRef]
8. Reddy, G.T.; Bhattacharya, S.; Ramakrishnan, S.S.; Chowdhary, C.L.; Hakak, S.; Kaluri, R.; Reddy, M.P.K. An ensemble based machine learning model for diabetic retinopathy classification. In Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 24–25 February 2020; pp. 1–6.

9. Alsaih, K.; Lemaitre, G.; Rastgoo, M.; Massich, J.; Sidibé, D.; Meriaudeau, F. Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images. *Biomed. Eng. Online* **2017**, *16*, 1–12. [CrossRef] [PubMed]
10. Nguyen, Q.H.; Muthuraman, R.; Singh, L.; Sen, G.; Tran, A.C.; Nguyen, B.P.; Chua, M. Diabetic retinopathy detection using deep learning. In Proceedings of the 4th International Conference on Machine Learning and Soft Computing, New York, NY, USA, 17–19 January 2020; pp. 103–107.
11. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
12. Zhao, S.W.; Li, C.D.; Zhang, W.; Wang, X.H.; Wang, L.M.; Li, H.Y. Deep learning-based approach for grading diabetic macular edema. *J. Ophthalmol.* **2018**, *2018*, 8415759.
13. Xu, W.; Wang, Y.; Zhang, J.; Wang, W. An end-to-end deep learning approach for detecting and grading diabetic macular edema in digital fundus images. *Med. Biol. Eng. Comput.* **2020**, *58*, 2191–2199.
14. Chen, W.; Lu, J.; Li, R.; Guo, S.; Lu, Z.; Liu, M.; Yang, X. Deep transfer learning-based detection of diabetic macular edema in digital fundus images. *J. Med. Syst.* **2019**, *43*, 418.
15. Alyoubi, W.L.; Shalash, W.M.; Abulkhair, M.F. Diabetic retinopathy detection through deep learning techniques: A review. *Inf. Med. Unlocked* **2020**, *20*, 100377. [CrossRef]
16. Kauppi, T.; Kalesnykiene, V.; Kamarainen, J.K.; Lensu, L.; Sorri, I.; Raninen, A.; Voutilainen, R.; Uusitalo, H.; Kalviainen, H.; Pietila, J. The DIARETDB1 diabetic retinopathy database and evaluation protocol. In Proceedings of the British Machine Vision Conference, Warwick, UK, 10–13 September 2007; Volume 2007, pp. 1–10.
17. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part V 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
18. Roboflow. Available online: https://roboflow.com/ (accessed on 1 July 2022).
19. [Messidor—ADCIS]. Available online: https://http://messidor.crihan.fr (accessed on 1 June 2022).
20. Hadid, A.; Pietikainen, M.; Martinkauppi, B. Color-based face detection using skin locus model and hierarchical filtering. In Proceedings of the 2002 International Conference on Pattern Recognition, Quebec City, QC, Canada, 11–15 August 2002; pp. 196–200.
21. Kaggle [Online]. Available online: https://https://kaggle.com/c/diabetic-retinopathy-detection (accessed on 1 June 2022).
22. Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.Y.; Girshick, R. Detectron2. 2019. Available online: https://github.com/facebookresearch/detectron2 (accessed on 1 July 2022).
23. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2017; pp. 2961–2969.
24. Zhang, H.; Chang, H.; Ma, B.; Shan, S.; Chen, X. Cascade retinanet: Maintaining consistency for single-stage object detection. *arXiv* **2019**, arXiv:1907.06881.
25. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
26. COCO-Detection: Faster rcnn R50 FPN. Available online: https://github.com/facebookresearch/detectron2/blob/main/configs/COCO-Detection/faster_rcnn_R_50_FPN_3x.yaml (accessed on 1 July 2022).
27. COCO-InstanceSegmentation: Mask rcnn R50 FPN. Available online: https://github.com/facebookresearch/detectron2/blob/main/configs/COCO-InstanceSegmentation/mask_rcnn_R_50_FPN_3x.yaml (accessed on 1 July 2022).
28. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 3–7. [CrossRef] [PubMed]
29. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
30. Zheng, S.; Meng, Q.; Wang, T.; Chen, W.; Yu, N.; Ma, Z.M.; Liu, T.Y. Asynchronous Stochastic Gradient Descent with Delay Compensation. *arXiv* **2016**, arXiv:1609.08326.
31. Detectron2.Data.Transforms. Available online: https://detectron2.readthedocs.io/en/latest/modules/data_transforms.html#detectron2.data.transforms.ResizeShortestEdge (accessed on 1 June 2022).
32. Van Stralen, K.J.; Stel, V.S.; Reitsma, J.B.; Dekker, F.W.; Zoccali, C.; Jager, K.J. Diagnostic methods I: Sensitivity, specificity, and other measures of accuracy. *Kidney Int.* **2009**, *75*, 1257–1263. [CrossRef] [PubMed]
33. Lalkhen, A.G.; McCluskey, A. Clinical tests: Sensitivity and specificity. *Contin. Educ. Anaesth. Crit. Care Pain* **2008**, *8*, 221–223. [CrossRef]
34. Goutte, C.; Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In Proceedings of the Advances in Information Retrieval: 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, 21–23 March 2005; Proceedings 27; Springer: Berlin/Heidelberg, Germany, 2005; pp. 345–359.
35. COCO. Available online: https://cocodataset.org/#detection-eval (accessed on 1 September 2022).