


Article

Computer-Assisted Fine-Needle Aspiration Cytology of Thyroid Using Two-Stage Refined Convolutional Neural Network

Wensi Duan ^{1,†}, Lili Gao ^{2,†}, Juan Liu ^{1,*} , Cheng Li ³, Peng Jiang ¹, Lang Wang ¹, Hua Chen ¹, Xiaorong Sun ³, Dehua Cao ³, Baochuan Pang ³, Rong Li ³ and Sai Liu ³

¹ Institute of Artificial Intelligence, School of Computer Science, Wuhan University, Wuhan 430072, China

² Department of Pathology, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China

³ Landing Artificial Intelligence Center for Pathological Diagnosis, Wuhan 430070, China

* Correspondence: liujuan@whu.edu.cn

† These authors contributed equally to this work.

Abstract: Fine-needle aspiration cytology (FNAC) is regarded as one of the most important preoperative diagnostic tests for thyroid nodules. However, the traditional diagnostic process of FNAC is time-consuming, and its accuracy is highly related to the experience of the cytopathologist. Computer-aided diagnostic (CAD) systems are rapidly evolving to provide objective diagnostic recommendations. So far, most studies have used fixed-size patches and usually hand-select patches for model training. In this study, we develop a CAD system to address these challenges. In order to be consistent with the diagnostic working mode of cytopathologists, the system is mainly composed of two task modules: the detecting module that is responsible for detecting the regions of interest (ROIs) from the whole slide image of the FNAC, and the classification module that identifies ROIs having positive lesions. The system can then output the top-k ROIs with the highest positive probabilities for the cytopathologists to review. In order to obtain the overall good performance of the system, we compared different object detection and classification models, and used a combination of the YOLOV4 and EfficientNet networks in our system.

Keywords: two-stage CAD system; fine-needle aspiration cytology (FNAC); thyroid cytopathology; deep learning; object detection



Citation: Duan, W.; Gao, L.; Liu, J.; Li, C.; Jiang, P.; Wang, L.; Chen, H.; Sun, X.; Cao, D.; Pang, B.; et al. Computer-Assisted Fine-Needle Aspiration Cytology of Thyroid Using Two-Stage Refined Convolutional Neural Network. *Electronics* **2022**, *11*, 4089. <https://doi.org/10.3390/electronics11244089>

Academic Editor: Donghyeon Cho

Received: 4 November 2022

Accepted: 5 December 2022

Published: 8 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the popularization of ultrasound technology and the improvement in people's health awareness, an increasing number of thyroid nodules have been found in health examinations, and the incidence of thyroid cancer is also increasing worldwide [1]. In 2020, about 586,000 new cases of thyroid cancer were diagnosed, and more than 43,000 people died [2]. Fine-needle aspiration biopsy (FNAB), which is a minimally invasive examination guided by ultrasound, is regarded as one of the most important preoperative evaluations in the diagnosis of thyroid cancer [3]. The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) outlines the criteria for the cytological classification of thyroid FNAB specimens [4].

Traditionally, after sample preparation using conventional smear methods or liquid-based cytology, cytopathologists examine the specimen under an optical microscope to determine the risk of malignancy according to the morphology of thyroid follicular cell groups. With the development of high-performance scanners and digital pathology, it has become common practice to perform a diagnosis on a whole-slide image (WSI) [5]. However, the manual screening process is time-consuming and laborious. Such diagnostic methods are also highly subjective and are related to cytopathologists' experience. Thus,

computer-aided diagnostic (CAD) systems have been developing rapidly to accelerate the diagnostic process and provide objective diagnostic suggestions [6].

With the advancement of artificial intelligence (AI) and the advent of digital pathology, AI solutions in thyroid cytopathology have been a research hot spot for over a decade. Hand-crafted feature extraction has worked great in many fields [7,8]. Previous CAD systems generally adopted conventional machine-learning algorithms that required complex image-preprocessing and feature-extraction steps. In 2006, Cochand-Priollet et al. [9] extrapolated morphometric nuclear data from May Grunwald–Giemsa-stained smears and used statistical classifiers to categorize them into either benign or malignant. Gopinath et al. [10] developed an automatic CAD system to classify thyroid nodules into benign or malignant by extracting statistical features on the basis of the textural characteristics of the thyroid cells, and integrated four different classifiers' results using majority-voting and linear-combination rules. Chain et al. [6] revealed that differences in nuclear area and elongation could objectively distinguish papillary thyroid carcinoma (PTC) from benign nodules and noninvasive follicular thyroid neoplasm with papillar features (NIFTP).

In recent years, deep learning has achieved great breakthroughs in the field of computer vision and image processing [11,12]. End-to-end image recognition methods have been widely applied in various tasks in medical image analysis [13]. Deep learning brings a promising solution for automatic thyroid cancer diagnosis, and an increasing number of scholars have studied deep-learning-based diagnostic methods for thyroid cancer in the past five years [14]. More recent studies focused on CNN-based methods for thyroid cancer screening in FNAB cytological images. Sanyal et al. [15] developed a simple five-layer network to differentiate between PTCs and non-PTCs on 186 microphotographs from Romanowsky/Pap-stained smears of papillary carcinoma and 184 microphotographs from the smears of other thyroid lesions (at $\times 10$ and $\times 40$ magnification). A subsequent study [16] used the VGG-16 [17] and Inception-v3 models [18] to differentiate PTCs from benign thyroid nodules using 279 cytological images. An improved two-stage multiple instance learning (MIL) algorithm proposed in [19] achieved thyroid malignancy prediction from cytopathology WSIs. Duc et al. [20] utilized stain normalization and ensemble deep learning methods to improve performance for the automatic classification of malignant PTC cell clusters from FNAB images. We summarize the works mentioned above and list them in Table 1.

Table 1. General summary of recent works.

Reference	Dataset	Method
Cochand-Priollet et al. [9]	157 cases of thyroid FNA	Nuclear features extraction + parametric classifiers
Gopinath et al. [10]	35 cases of thyroid FNA	Textural feature extraction + four traditional classifiers
Chain et al. [6]	35 cases of thyroid FNA	Calculating nuclear area and elongation as classification criteria
Sanyal et al. [15]	370 cases of PTC and non-PTC	Simple five-layer network
Guan et al. [16]	279 cases of PTCs and non-PTCs	VGG-16 [17] and Inception-v3 model [18]
Dov et al. [19]	908 WSIs from 659 patients	Improved two-stage multiple instance learning (MIL) algorithm
Duc et al. [20]	367 hematoxylin–eosin (H&E)-stained images	Utilizing stain normalization and ensemble deep learning methods

However, these studies have several limitations. First, most of these studies classified fixed-size patches rather than using WSI-level data. Second, patch-level cytological images used for model training are often manually selected rather than automatically generated. Since the discriminative features that are helpful in diagnosis occupy only a small part of the entire WSI, it is crucial in clinical practice to find the regions of interest (ROIs) among much useless background information. These issues greatly impede the application of automatic end-to-end CAD systems in clinical practice.

Thus, to address the above challenges, this paper proposes a two-stage refined convolutional neural network (CNN) to realize automatic screening for thyroid cancer in FNAB

cytology. To meet actual clinical application needs, we collected 360 samples from multiple medical institutions to construct the thyroid cytopathology dataset. Inspired by the diagnostic workflow of cytopathologists, namely, locating ROIs and then carefully screening, we built our system so that it (1) detected ROIs with object detection framework YOLOV4 [21] and (2) further refined those ROIs into benign and malignant using the EfficientNet [22] model. The main contributions of this paper are summarized as follows:

- To better migrate and apply it to clinical programs, we built a thyroid cytopathology dataset generated from 360 FNAB specimens in practical medical institutions. This dataset is well-annotated for both detection and classification tasks.
- Mimicking the diagnostic experiences of pathologists, we utilized an object detection algorithm to search suspected target areas in WSI and then leverage another network to refine the classification result. To the best of our knowledge, this is the first work to apply an object detection algorithm for the automatic discovery of ROIs in thyroid cytology screening.
- Extensive experiments with promising results on the built thyroid cytopathology dataset validated the effectiveness of the proposed method for thyroid cancer diagnosis. The proposed refined two-stage network provides a novel solution for automatic CAD systems in practical thyroid cancer screening.

The remainder of this paper is organized as follows: Section 2 describes the collection process of the thyroid cytology dataset in detail. Section 3 introduces the whole two-stage diagnostic procedure of thyroid cancer, including the detection of suspicious areas and further precise identification. In Section 4, experimental results are presented and analyzed to evaluate the performance of the proposed approach. Lastly, this paper is summarized in Section 5.

2. Dataset Generation

2.1. Overview

Clinical image data are key to constructing a deep-learning-based diagnostic system. A sufficient amount of high-quality data can contribute to the success of a study. However, there are few public thyroid cytological datasets due to the difficulty of data collection and the high cost involved to some extent. Therefore, at the beginning of this work, we built a real-world clinical dataset of thyroid FNAB cytology. We collected 360 WSIs, and both WSI-level labels and the malignant thyroid regions of each WSI were carefully annotated by expert pathologists.

As our system consisted of two main parts, (1) an object detection network and (2) a classification network, two different datasets for the two tasks were generated. The process is illustrated in Figure 1. First, we collected thyroid FNAB samples from the participants. Then, using liquid-based cytologic preparation, we obtained 360 WSIs, removing the ones that did not meet the requirements. Next, we developed one dataset for detection and another for classification.

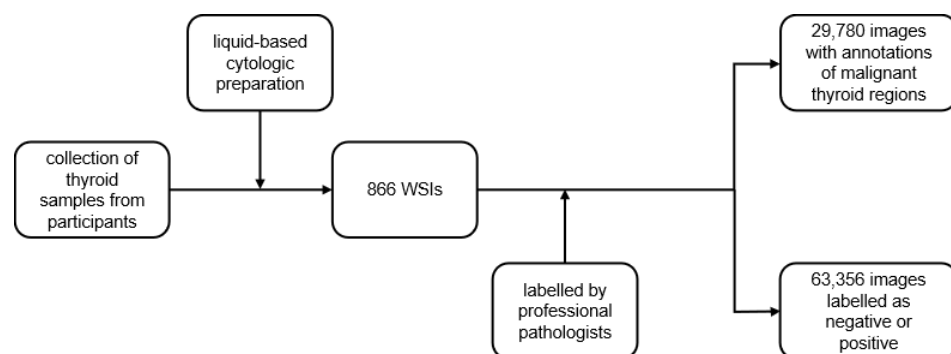


Figure 1. The process of the dataset generation.

This study abides by the ethical standards of the institutional research committee and the tenets of the Helsinki Declaration. Since this work is anonymous and retrospective, the need for informed consent has been waived.

2.2. WSI Collection

We conducted a retrospective study and collected thyroid cell samples from participants aged 25–64 years from multiple healthcare institutions in 2020 and 2021. Table 2 shows the distribution of the 360 WSIs, each of which corresponds to a participant, with liquid-based cytologic preparation. The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC), a standardized system used for the diagnosis of thyroid FNAB, comprises 6 diagnostic categories: nondiagnostic, benign, atypia of undetermined significance or follicular lesion of undetermined significance, follicular neoplasm, suspicious for malignancy, and malignant [4]. The nondiagnostic category was obviously not considered, and the other five categories were classified as positive or negative for the present study.

Table 2. The distribution of WSIs.

Class	TBSRTC Diagnostic Category	WSI Image Count	Type
0	Nondiagnostic or unsatisfactory	-	-
1	Benign	222	negative
2	Atypia of undetermined significance or follicular lesion of undetermined significance	10	positive
3	Follicular neoplasm or suspicious for a follicular	2	positive
4	Suspicious for malignancy	36	positive
5	Malignant	90	positive

2.3. Image Labeling

We built two datasets from the existing WSIs, one for the object detection module and one for the classification module. Expert pathologists provided the ground-truth annotations of the malignant thyroid regions for the detection model, and Figure 2a shows an annotated image. Images in the dataset of the classification model were labeled as negative or positive. Figure 2b,c give examples of each of the two categories. We obtained a total of 29,780 images and 63,356 images for the two datasets, and divided them for training and testing, respectively. The training set was for fitting and optimizing the parameters of the model, while the independent test set was for evaluating the performance of the model. Table 3 lists the outline.

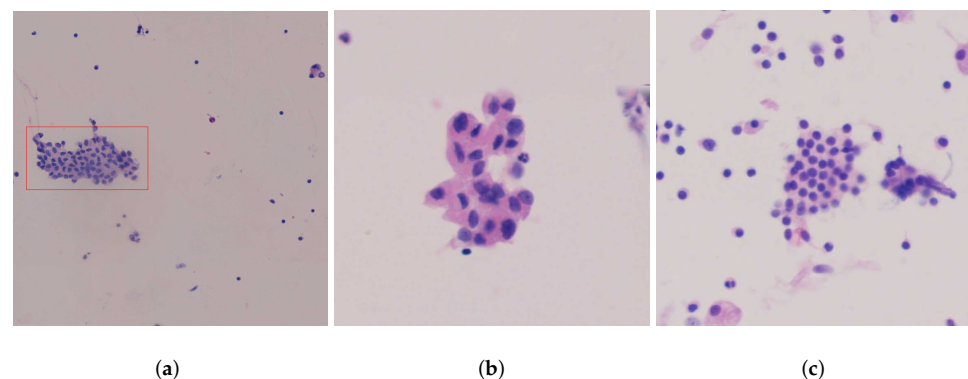


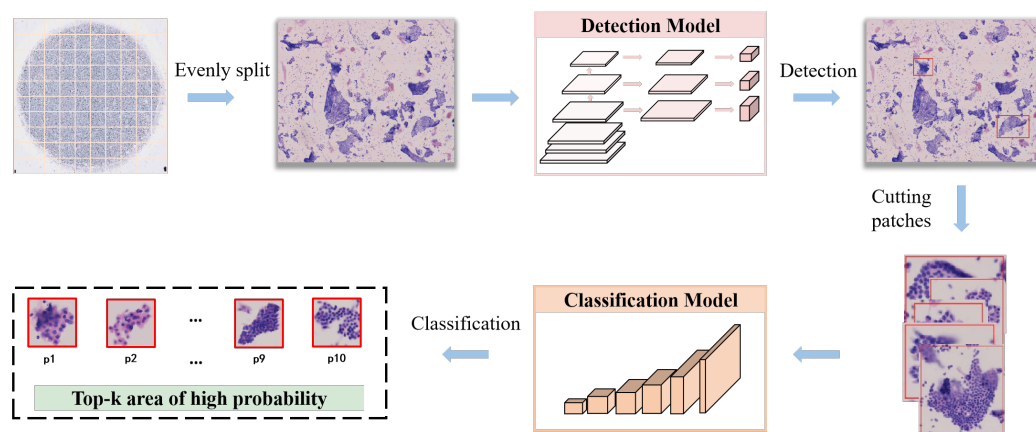
Figure 2. Examples of the datasets: (a) annotated image; (b) an image labelled as positive; (c) an image labelled as negative.

Table 3. Number of images in the developed datasets.

Dataset for the Detection Model			
Set of boxing images		positive	
Training set		26,802	
Testing set		2978	
total		29,780	
Dataset for the Classification Model			
Set of labeled images	Positive	Negative	total
Training set	30,726	31,085	61,811
Testing set	840	705	1545
Total	31,566	31,790	63,356

3. System Architecture

As shown in Figure 3, our CAD system was designed on the basis of two artificial intelligence models. The input of the system was a WSI that was later split evenly into patches.

**Figure 3.** The workflow of our proposed system.

Then, the detection model detects the suspected positive areas and sends these areas to the classification model, which predicts the final pathology. After lesion detection and classification, the system displays the 10 images with the highest probability of positivity to the doctor.

3.1. Suspicious Area Detection

The YOLO series [21,23–26] is one of the most classical and effective one-stage object detection frameworks. It can handle both classification and location tasks at the same time. Considering the efficiency of clinical screening, we used the YOLO V4 framework in this work, which has fast detection speed while maintaining high accuracy, to realize the real-time detection of suspicious thyroid cancer areas. YOLO V4 is a fourth-generation derivative network of the YOLO network aiming for the optimal balance between the number of convolutional layers and the number of parameters. As illustrated in Figure 4, the overall structure of YOLO V4 consists of three parts: a backbone used for extracting base features, a detection neck for better feature fusion, and a detection head for final target classification and location regression.

Backbone. To reduce the amount of computation while maintaining high-accuracy performance, YOLO V4 utilizes CSPDarkNet-53 as the backbone, which adopts the network modeling strategy in a cross-stage partial network (CSPNet) [11]. CSPNet achieves better gradient interaction and combination via cross-stage hierarchal feature partitioning and concatenation. Due to the efficient division of feature maps, CSPNet becomes more lightweight and boosts the detection speed of YOLO V4.

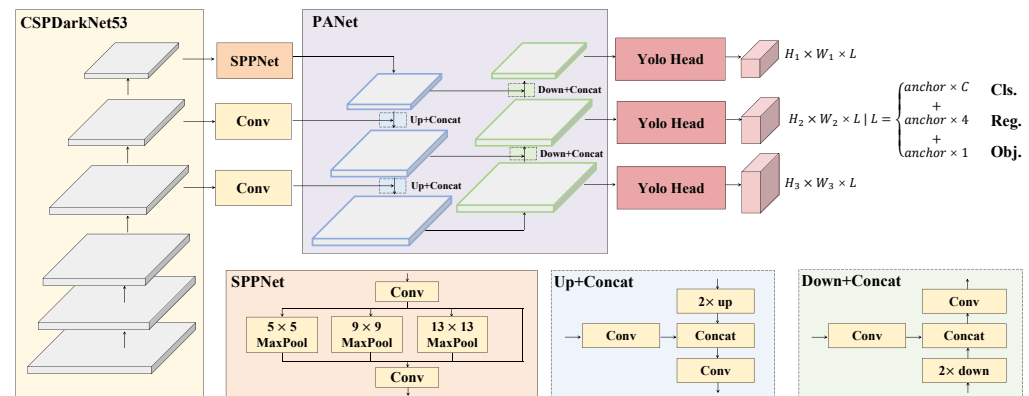


Figure 4. The overall structure of YOLO V4.

Neck. High-level features in deep layers strongly respond to entire objects and are thereby more discriminative for a classification task. In contrast, low-level features in shallow layers are more likely to be activated by local patterns and edges, and are helpful in location tasks. Thus, the appropriate fusion of low-level and high-level features is critical for integrating contextual information and realizing comprehensive detection. FPN [27] augments a top-down path and horizontal connections to propagate semantically strong features and effectively fuse multilevel features. Inspired by FPN, YOLO V4 has adopted PANet [28] to further leverage multilevel features by introducing top-down and bottom-up paths. Moreover, YOLO V4 utilizes spatial pyramid pooling (SPP) [29] to enhance the discriminative capability of multiscale features in which different kernel sizes of pooling operations are implemented to generate different receptive fields.

Head. YOLO V4 is an anchor-based network that utilizes a single path to complete the construction of the detection head. Specifically, via cascade 3×3 and 1×1 convolutional layers, YOLO V4 transforms input feature maps generated from the PANet into a tensor $U \in \mathbb{R}^{H \times W \times L}$, where $L = \text{anchors} \times (k + 4 + 1)$ represents the values for k category classification, 4 coordinate regression, and 1 object probabilistic output. There are also 3 preset anchors for each feature map.

After the construction of the network, a CIoU-loss and a NMS are applied to refine the object localization performance and filter those predicted boxes with less confidence. On the whole, YOLO V4 realizes real-time detection with high accuracy and is ideal for the detection of thyroid lesions.

Anchor cluster analysis. The anchors of the original YOLO V4 were designed for the detection of natural images, which are significantly different from cytological images. Thus, we adjusted the size of anchors to allow them to fit better in our thyroid cytological dataset. Specifically, we first counted the size of all labeled boxes, and presented the width and height distribution of these boxes as shown in Figure 5. Then we used the k-means algorithm to cluster 9 points that could be assigned to the corresponding anchors. Through the above adjustment, the anchor size is more suitable for our dataset and is helpful for performance.

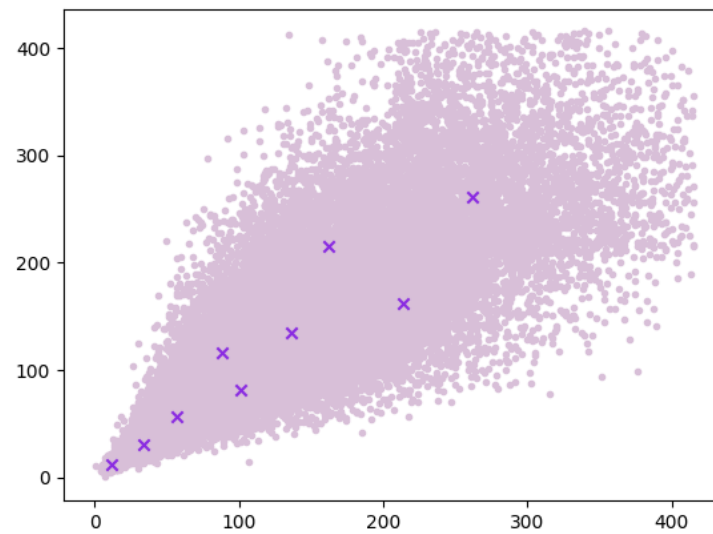


Figure 5. Width and height distribution of the built thyroid cytological dataset for ROI detection. (Horizontal coordinates indicate the width, and vertical coordinates indicate the height).

3.2. Lesion Classification

The classification model was trained on the dataset mentioned above to avoid missing genuinely negative cases that were similar to the positive ones. As shown in Table 3, 63,356 images were labeled as positive or negative by expert pathologists and divided into a training set and a testing set. The EfficientNet model provides higher accuracy values than those of other models due to the use of scaling at depth, width, and resolution. EfficientNet uses mobile inverted bottleneck MBConv [30,31] as its backbone, and the squeeze-and-excitation method [32] to optimize the network structure. It scales from B0 to B7 with different parameters from 5.3 to 66 M. We used EfficientNet-B0 as our lesion classifier, as it is suitable for the proposed CAD system. Its framework is shown in Figure 6. Each MBConv is followed by a 1 or a 6, where 1 or 6 is the multiplicity factor, i.e., the first 1×1 convolutional layer in the MBConv expands the channels of the input feature matrix by a factor of n , where $k3 \times 3$ or $k5 \times 5$ indicates the size of the convolutional kernel used in Depthwise Conv in the MBConv.

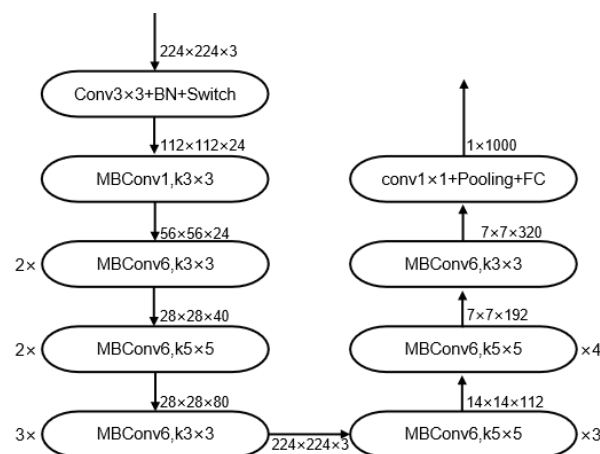


Figure 6. Illustration of the EfficientNet-B0 framework.

4. Experiment

4.1. Experimental Setup and Evaluation Measures

All models were trained, validated, and tested using four NVIDIA GeForce GTX 3090Ti GPUs. For lesion detection, the model experiment was completed on the MMDetection platform [33] of the deep-learning object-detection toolbox based on PyTorch; we

adopted the $1\times$ training schedule of MMDetection. For a fair comparison, we adopted the same training and testing strategies as those of the detection models, and they were not pretrained.

For lesion classification, we built all the models on the basis of *timm* 0.5.4 [34] and PyTorch. We used the same hyperparameters for all of the models for fairness, i.e., all experiments followed most of the settings in Deit [35]. In detail, we trained all the classification models in 200 epochs with a batch size of 512 (128 images per GPU). The initial learning rate was set to be 5×10^{-4} , and for the first 5 epochs, we used a linear warm-up with a learning rate starting from 1×10^{-6} . For model optimization, we used the AdaW [36] optimizer with a cosine decay learning scheduler. We set the weight decay to 0.05. To enlarge the dataset, we also used most of the augmentation strategies in Deit, including Cutmix [37], Mixup [38], and RandAugment [39].

In this work, we adopted *recall*, mean average accuracy (mAP), *accuracy*, *precision*, and *F1 – score* to evaluate the performance of the models, and the scores were calculated according to the following formulas:

$$\begin{aligned} recall &= \frac{TP}{TP + FN}, \\ accuracy &= \frac{TP + TN}{TP + FN + FP + TN}, \\ precision &= \frac{TP}{TP + FP}, \\ F1 - score &= 2 \frac{precision \times recall}{precision + recall} \\ &= \frac{2TP}{2TP + FP + FN}, \end{aligned}$$

where *TP* (true positive) denotes the number of correctly detected positive samples, *FN* (false negative) is the number of positive samples that were incorrectly identified as negative, *TN* (true negative) denotes the number of correctly detected negative samples, and *FP* (false positive) is the number of negative samples that were incorrectly identified as positive.

4.2. Lesion Detection Performance

Taking the application and deployment of our CAD system into consideration, we narrowed down the selection of models to one-stage object detection models, as one-stage models are faster than two-stage models. Used as the lesion detection model to localize suspicious areas, the YOLO V4 framework has fast detection speed while maintaining high accuracy. Figure 7 gives some examples of the detection results on the testing set by YOLO V4. YOLOX [26] and Single Shot MultiBox Detector (SSD) [40] are both fast one-shot object detectors. We evaluated the chosen detector YOLO V4 with these two. *mAP* and *recall* were the main evaluation indicators. Utilizing CSPDarkNet-53 as the backbone, YOLO V4 adopts PANet [28] and SSP [29] to better leverage and discriminate multilevel features. Owing to these improvements, YOLO V4 demonstrated greater ability to improve detector accuracy. Figure 8, where SSD300 means the input size of SSD is 300×300 , lists the final results. Compared with YOLOX [26] and SSD [40], YOLO V4 performed better in both mAP and recall.

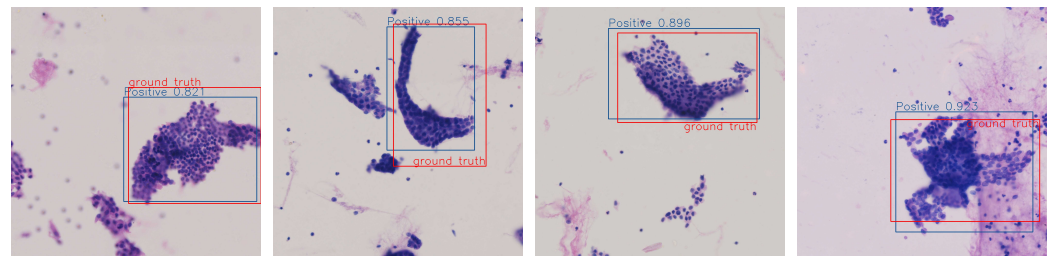


Figure 7. Some examples of the detection results.

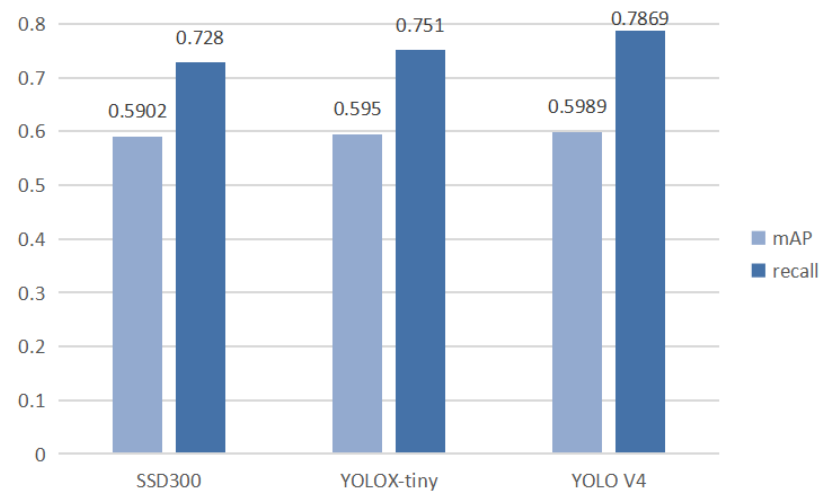


Figure 8. Detection performance comparison results.

4.3. Lesion Classification Performance

The lesion classification model, EfficientNet, was trained and evaluated on the dataset mentioned in Section 2.

As shown in Figure 9, the *accuracy* and *loss* curves were plotted with the number of training iterations on the X axis, and the corresponding training accuracy and loss on the Y axis. Moreover, we compared EfficientNet with GhostNet [41], MobileNet V3 [42], Edge-Vit [43], and Light-Vit [12].

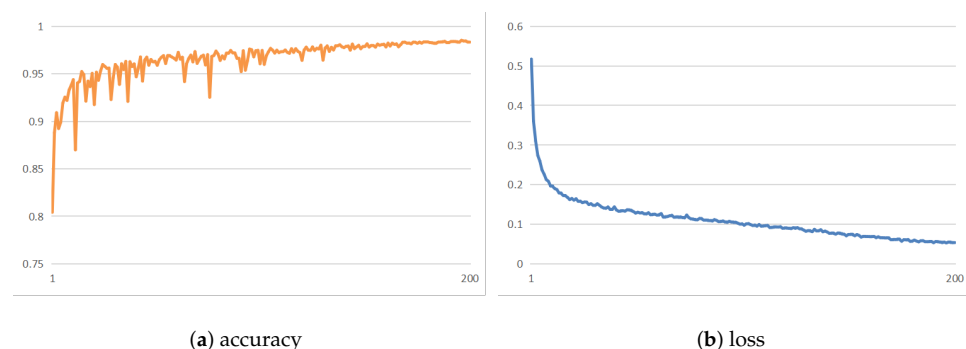


Figure 9. The training accuracy and loss curves of EfficientNet.

A confusion matrix was used to see how the model performed on each class and which class was not easily distinguishable. Figure 10 shows the confusion matrices of the final results, and *accuracy*, *recall*, *precision*, and *F1 – score* values are listed in Table 4.

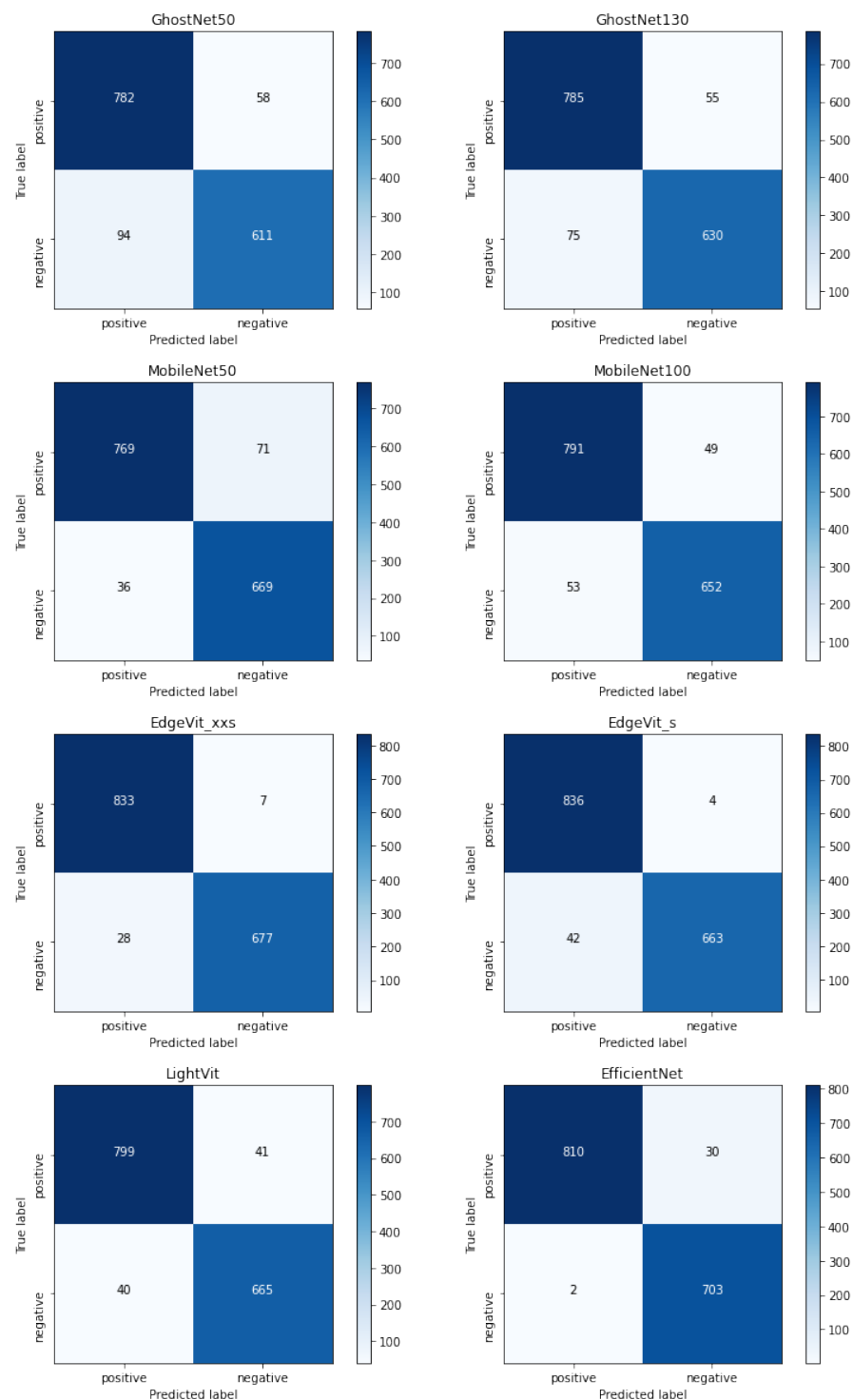


Figure 10. Confusion matrices of the classifiers.

On the basis of these results, the performance of the dataset that we built of different classifiers could be visually evaluated. According to the confusion matrices, EfficientNet was more likely to avoid identifying a negative as a positive. Reaching 97.93% in *accuracy*, 97.92% in *F1 – score*, 98.08% in *recall*, and 97.83% in *precision*, EfficientNet also achieved more balanced performance compared to the other classifiers due to the use of the highly effective compound scaling method, which balances network width, depth, and resolution.

Table 4. Classification performance comparison results (%).

Model	Accuracy	F1-Score	Precision	Recall
Ghost50	90.16	90.04	90.30	89.88
Ghost130	91.59	91.50	91.62	91.41
Mobile50	93.07	93.04	92.97	93.21
Mobile100	93.40	93.34	93.37	93.32
EdgeVit_xxs	97.74	97.71	97.86	97.60
EdgeVit_s	97.02	96.98	97.31	96.78
LightVit	94.76	94.72	94.71	94.72
EfficientNet	97.93	97.92	97.83	98.08

4.4. Two-Stage Network Performance

Lastly, we designed experiments to demonstrate the validity of the proposed two-stage refined network. Specifically, we used the detection results of the previously mentioned trained YOLOv4 object detection network on the test set as input to the EfficientNet classification network. The experimental results show that our proposed system, with the introduction of the classification module (81.84%), achieved a 3.16% improvement in *precision* compared to the single detection network (78.68%). *Precision* describes how many of the predicted positive cases are true-positive cases from the perspective of the predicted outcome. This result shows our proposed network's capability of avoiding missing genuinely negative cases that are similar to positive ones.

5. Conclusions

In this paper, we proposed a two-stage CAD system for thyroid pathology detection and classification using AI technology. We established a dataset of 360 FNAB WSIs, of which 222 were negative and 138 were positive. On the basis of the dataset, we generated two datasets for the training and evaluation of our object detector and classifier separately. When given a WSI, our system detects the suspected positive areas and send them to the classifier. After classification, the system displays the 10 images with the highest probability of positivity to the doctor. The proposed two-stage refined network provides a novel solution for automatic CAD systems in practical thyroid cancer screening. The experimental results indicate that the proposed method can assist in the diagnosis of thyroid cancer and possesses a certain clinical value. For further study, we aim to focus on the classification of multiple categories, such as TBSRTC diagnosis and the segmentation of suspicious areas at the pixel level to render our system more practical in the clinic.

Author Contributions: Conceptualization, W.D., J.L. and C.L.; methodology, W.D.; validation, W.D., P.J. and L.W.; investigation, W.D., L.G., B.P. and D.C.; resources, L.G., R.L. and S.L.; writing—original draft preparation, W.D.; writing—review and editing, W.D., J.L. and H.C.; supervision, X.S.; project administration, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Major Projects of Technological Innovation in Hubei Province (2019AEA170), the Frontier Projects of Wuhan for Application Foundation (2019010701011381), and the Translational Medicine and Interdisciplinary Research Joint Fund of Zhongnan Hospital of Wuhan University (ZNJC202226).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki.

Informed Consent Statement: Patient consent was waived due to that this work is anonymous and retrospective.

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

References

1. Cabanillas, M.E.; McFadden, D.G.; Durante, C. Thyroid cancer. *Lancet* **2016**, *388*, 2783–2795. [\[CrossRef\]](#)
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [\[CrossRef\]](#)
3. Haugen, B.R.; Alexander, E.K.; Bible, K.C.; Doherty, G.M.; Mandel, S.J.; Nikiforov, Y.E.; Pacini, F.; Randolph, G.W.; Sawka, A.M.; Schlumberger, M.; et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* **2016**, *26*, 1–133. [\[CrossRef\]](#)
4. Cibas, E.S.; Ali, S.Z. The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* **2017**, *27*, 1341–1346. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Kumar, N.; Gupta, R.; Gupta, S. Whole slide imaging (WSI) in pathology: Current perspectives and future directions. *J. Digit. Imaging* **2020**, *33*, 1034–1040. [\[CrossRef\]](#)
6. Chain, K.; Legesse, T.; Heath, J.E.; Staats, P.N. Digital image-assisted quantitative nuclear analysis improves diagnostic accuracy of thyroid fine-needle aspiration cytology. *Cancer Cytopathol.* **2019**, *127*, 501–513. [\[CrossRef\]](#)
7. Gadekallu, T.R.; Rajput, D.S.; Reddy, M.; Lakshmana, K.; Bhattacharya, S.; Singh, S.; Jolfaei, A.; Alazab, M. A novel PCA—Whale optimization-based deep neural network model for classification of tomato plant diseases using GPU. *J. Real-Time Image Process.* **2021**, *18*, 1383–1396. [\[CrossRef\]](#)
8. Nagaraj, P.; Deepalakshmi, P.; Mansour, R.F.; Almazroa, A. Artificial flora algorithm-based feature selection with gradient boosted tree model for diabetes classification. *Diabetes Metab. Syndr. Obes. Targets Ther.* **2021**, *14*, 2789.
9. Cochand-Priollet, B.; Koutroumbas, K.; Megalopoulou, T.M.; Pouliakis, A.; Sivolapenko, G.; Karakitsos, P. Discriminating benign from malignant thyroid lesions using artificial intelligence and statistical selection of morphometric features. *Oncol. Rep.* **2006**, *15*, 1023–1026. [\[CrossRef\]](#)
10. Gopinath, B.; Shanthi, N. Development of an automated medical diagnosis system for classifying thyroid tumor cells using multiple classifier fusion. *Technol. Cancer Res. Treat.* **2015**, *14*, 653–662. [\[CrossRef\]](#)
11. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
12. Huang, T.; Huang, L.; You, S.; Wang, F.; Qian, C.; Xu, C. LightViT: Towards Light-Weight Convolution-Free Vision Transformers. *arXiv* **2022**, arXiv:2207.05557.
13. Li, T.; Bo, W.; Hu, C.; Kang, H.; Liu, H.; Wang, K.; Fu, H. Applications of deep learning in fundus images: A review. *Med. Image Anal.* **2021**, *69*, 101971. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Kezlarian, B.; Lin, O. Artificial intelligence in thyroid fine needle aspiration biopsies. *Acta Cytol.* **2021**, *65*, 324–329. [\[CrossRef\]](#)
15. Sanyal, P.; Mukherjee, T.; Barui, S.; Das, A.; Gangopadhyay, P. Artificial intelligence in cytopathology: A neural network to identify papillary carcinoma on thyroid fine-needle aspiration cytology smears. *J. Pathol. Inform.* **2018**, *9*, 43. [\[CrossRef\]](#)
16. Guan, Q.; Wang, Y.; Ping, B.; Li, D.; Du, J.; Qin, Y.; Lu, H.; Wan, X.; Xiang, J. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study. *J. Cancer* **2019**, *10*, 4876. [\[CrossRef\]](#)
17. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
18. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
19. Dov, D.; Kovalsky, S.Z.; Assaad, S.; Cohen, J.; Range, D.E.; Pendse, A.A.; Henao, R.; Carin, L. Weakly supervised instance learning for thyroid malignancy prediction from whole slide cytopathology images. *Med. Image Anal.* **2021**, *67*, 101814. [\[CrossRef\]](#)
20. Duc, N.T.; Lee, Y.M.; Park, J.H.; Lee, B. An ensemble deep learning for automatic prediction of papillary thyroid carcinoma using fine needle aspiration cytology. *Expert Syst. Appl.* **2022**, *188*, 115927. [\[CrossRef\]](#)
21. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolo4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
22. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
23. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
24. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
25. Redmon, J.; Farhadi, A. Yolo3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
26. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YoloX: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
27. 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.
28. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [\[CrossRef\]](#) [\[PubMed\]](#)

30. Tan, M.; Chen, B.; Pang, R.; Vasudevan, V.; Sandler, M.; Howard, A.; Le, Q.V. Mnasnet: Platform-aware neural architecture search for mobile. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2820–2828.
31. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
32. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
33. Chen, K.; Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Xu, J.; et al. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv* **2019**, arXiv:1906.07155.
34. Wightman, R. PyTorch Image Models. 2019. Available online: <https://github.com/rwightman/pytorch-image-models> (accessed on 28 September 2022). [CrossRef]
35. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jégou, H. Training data-efficient image transformers & distillation through attention. *arXiv* **2021**, arXiv:2012.12877v2.
36. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
37. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 6023–6032.
38. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond empirical risk minimization. *arXiv* **2017**, arXiv:1710.09412.
39. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. Randaugment: Practical automated data augmentation with a reduced search space. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 702–703.
40. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016.
41. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1580–1589.
42. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
43. Pan, J.; Bulat, A.; Tan, F.; Zhu, X.; Dudziak, L.; Li, H.; Tzimiropoulos, G.; Martinez, B. EdgeViTs: Competing Light-weight CNNs on Mobile Devices with Vision Transformers. *arXiv* **2022**, arXiv:2205.03436.