Parameter Investigation in Low Computing Cost Model-Based EfficientDet for Unmanned Aerial Vehicles Object Detection

1st Iga Narendra Pramawijaya
School of Electrical Engineering
Telkom University
Bandung, Indonesia
iganarendra@student.telkomuniversity.ac.id

2nd Suryo Adhi Wibowo*
School of Electrical Engineering
Telkom University
Bandung, Indonesia
suryoadhiwibowo@telkomuniversity.ac.id
*Corresponding Author

3rd Koredianto Usman
School of Electrical Engineering
Telkom University
Bandung, Indonesia
korediantousman@telkomuniversity.ac.id

***Abstract*—Unmanned Aerial Vehicles (UAV) are equipped with high-resolution cameras. Previous researches novel object detection models for UAV images, but available models took high computational cost to retrain. Models that not trained using the UAV image will result in poor performance. EfficientDet object detection model is significantly lower in computing cost. In this paper, the author wants to investigate epochs and optimisation function impact to the performance of the low computing cost model-based EfficientDet on UAV images. An object detection system on UAV imagery is trained using conventional computer will be designed to detect 10 classes object in Visdrone dataset using the D0 version of the EfficientDet model. After the data is obtained then preprocessing will be carried out in the form of annotation conversion. Next, model training process will be trained from 10 to 50 epochs. In each training, the model will be tested so as to produce a validation value. The last validation value will be analysed as performance benchmark. As a result, our proposed model surpassed state-of-the-art models in Average Recall score with 2.1% ARmax1 for Visdrone by only using 50 epochs or less.**

***Index Terms*—EfficientDet, UAV Images, Low Computing Cost.**

I. INTRODUCTION

Unmanned Aerial Vehicles (UAV) is a type of aerial vehicle (or a type of vehicle that operates at certain height above the surface) that does not require human pilot to be physically flying inside the vehicle [1]. UAV can be controlled remotely using radio control, or artificially controlled by computers or with minimum human interference. UAVs, or also to be called as drones, have tremendous contribution in military back after the First World War. Earliest drone that fits its modern definition launched back in 1922, from a British aircraft carrier, HMS Argus [2]. A century later, UAV expanded its service scope from military to commercial and research. Modern UAV is equipped with a high-resolution camera to capture UAV images and video. Data captured using UAV has become more common since the demand for data in remote and rural areas is increasing. Since UAV cameras are able to take capture high resolution data, we can use images and videos from this kind of camera for many use case.

Wu, et al. utilise UAV's high resolution data and remote control capability to deploy automatic rice seeding counting system using the help of deep learning [3]. Oleire-Oltmanns, et al. also utilising UAV's reliability in remote areas data capturing (or remote sensing) to monitor soil erosion rate in Morocco [4]. Panda, et al. designed and deployed a post-disaster emergency network using help of the UAVs [5]. Neto, et al proposed UAV usage in surveillance task under natural disaster scenario [6]. UAV images are having unique characteristics such as (i) ultra-high spatial resolution [7], (ii) affected by weather, and (iii) affected by altitude.

Unlike in another images, objects in UAV images will appear smaller and from a bird's eye point of view. Ultra-high spatial resolution, and altitude effect from UAV images will increase detection complexity. Also, since UAV images quality affected by weather such as fog and visibility, certain weather condition like rain, fog, haze, and sunlight could affect object detection quality. Pikun, et, al. implemented an improved dark channel haze removal (DCHR) algorithm to remove haze and sea fog for naval military implementation using UAV [8]. These three factors invites more researchers to tackle object detection-based UAV challenges, as UAV implementation is proven to be effective in many implementation scenarios.

Artificial Neural Network Model can help implement the use UAV image [9]. Artificial neural network model can be trained to recognise objects and locate the object in an image, known as object detection model. Mittal, et, al. reviewed well-known state-of-the-art object detection model performances on images taken from a UAV flying in low altitude. Faster RCNN, Cascade RCNN, R-FCN, two staged model such as YOLO variants, SSD, and one staged model like CornerNet were reviewed by previous research [10]. Available Artificial Neural Network model performance can be improved by modifying layers. Wiranata, et, al. investigate padding scheme impact to AlexNet performance in vehicle detection model [11]. These previous researches prove object detection model also able to process traffic UAV images, and being modified for another implementations such as visual tracking [12] or performance upgrade [11]. By default, the available pretrained object detection models are not trained using the UAV image, and will result in poor performance when detecting objects from UAV images. Retraining the model using UAV images could increase model detection performance. Model retraining or transfer learning is proven to increases model performance for vehicle detection [13]. When existing object detection models are retrained using UAV images, it will take up a high computational cost due to high resolution images from UAV. Vaddi et, al. proposed deep feature pyramid architecture (FPN) based model to reduce computational cost. The focal loss function inside the FPN is also modified to reduce class imbalance issue. Computational cost reduce proposal made real-time object detection application in drones possible [14].

In the previous research, Vaddi et, al. used use high computing hardware such as high performance computing (HPC) device to train the model. Unfortunately, HPC devices are not available for everyone. Since HPC's limited availability, there is a demand to tackle high computational cost problem using more accessible computing devices. Previous researches mentioned in this paper also uses high performance computing devices for training proposes. Despite an architecture already proposed to deliver lower computing cost, this proposed method is still difficult to replicate and modified in specific use-case due to its hardware dependency. On the other hand, Tan, et al developed EfficientDet, an even more efficient object detection model capable to deliver state-of-the-art capability [15]. Tan, et al. proposed EfficientDet with flexibility in mind. EfficientDet has eight different pyramid network depth, that can be used as needed to reduce unnecessary computation. Unfortunately, we are lacking of literature that shares EfficientDet as low computing cost model and in limited environment, away from HPCs. This is important because UAV based object detection model needed to be modified easily, deployed easily and work in real-time regardless how strong computing power available. Also to mention UAV usage as a remote data-capturing device will face a situation where high performance computing device is simply not available and unreachable.

Hence, in this research we proposed utilisation of the smallest version of EfficientDet (D0) and exploit its performance when faced with UAV dataset to highlight the most of its low computing capability under strict limited resources such in UAV implementation. As an addition to highlight its capability as low computing cost model, the EfficientDet D0 model will be trained for maximum 50 epochs only on a personal-grade laptop. This proposed method will contribute exploration results of low-computing cost EfficientDet for UAV images. Unlike previous researches that focuses only on EfficientDet performance and not taking computing complexity into account. Also, this proposed method will contribute low-computing cost EfficientDet for UAV implementation.

Section II of the paper briefs the implementation of Object Detection-Based UAV in previous researches and EfficientDet object detection network. Section III of the paper explains our training process, data preparation and the data used in this research. Section IV of the paper gives the research's result in detail when Stochastic Gradient Descend and Adam optimiser function is being used. Also in section IV briefs how our model could be an ideal implementation in object detection-based UAV. Finally, section V of the paper discuss about the conclusion and future scope of the research.

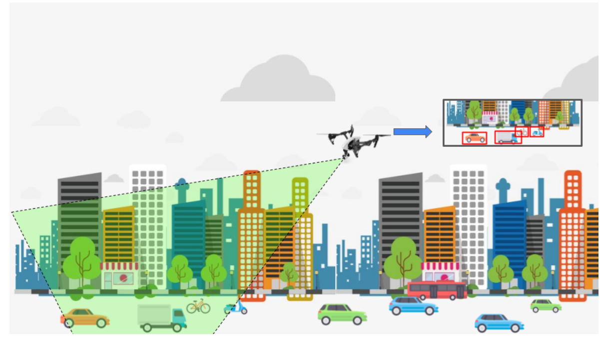

Fig. 1: Illustration of Object Detection-Based UAV in urban scenario.

II. Theory

A. Object Detection-Based UAV

In area of computer vision research, object detection is a branch branch that recognise, then pinpoint objects relative position in an image, as illustrated in figure 1. Deep learning approach shows incredible result solving this problem [16]. Szegedy, et al [17] proofed that deep learning network not only learning image's features for classification, but also geometric information that will make the model easier to detect objects. Mittal, et al discussed about several deep learning algorithms in UAV images object detection model, such as RCNN and its variants like Cascade RCNN and Faster RCNN, YOLO and its modifications, RetinaNet, RetinaNet, SSD, and more [18]. Mittal, et al. discussed mean average precision result from various network on VisDrone dataset, with the novel of PENet structure has the highest mean average precision value (41.1 mAP) [19]. In the previous research [18], outside of high computing cost model that even trained around 2000 epochs such as in [20], it is also discussed about others low computing cost model. With tiny-YOLO used as criterion, [21] tested several low computing cost model such as TinyYoloVoc, TinyYoloNet, SmallYoloV3 and DroNet. Amongs other models, DroNet shows best accuracy around 5 to 18 fps. In all literatures, low computing cost based EfficientDet performance result is not available.

B. EfficientDet

EfficientDet is a object detection model based on EfficientNet [22] developed by Tan, et al [15]. This object detection model focuses on delivering top-performing object detection quality and accuracy even limited by less computational cost. The largest version of EfficientDet (D7) achieved 52.2% AP on COCO test-dev. EfficientDet reached this level with only 325 million FLOPs, 4-9 times smaller in size and 13-42 times fewer FLOPs between its variations than its predecessor state-of-the-art model, such as AmoebaNet-based NAS-FPN detector [15]. This model uses a two-way pyramidal network as in figure 2 to feature extraction tool. This pyramid model consists of 8 levels, where the Bi-directional pyramid network

(BiFPN) adapted from Chen, et al [23] applied from third to seventh level. This pyramid network is intended for making smaller feature maps, but with more information in each features. One tier pyramid to another is connected using the convolution operation, so that when moving levels, the feature map matrix becomes smaller in dimension, but increasingly dense information. In addition, when moving from one level to another is given a weight as width difference, as in equation 1:

$$Width_{BiFPN} = 64 \times (1.35^n) \quad (1)$$

where n is the EfficientDet level. This represents during the learning process, EfficientDet network will learn which levels have high and low weights. Each level has its own BiFPN layer's depth formulated in equation 2:

$$Depth_{BiFPN} = 3 + n \quad (2)$$

By using this adaptability, it can increase precision by assigning a higher weight to the level which produces a feature map that is representative of the predicted results. In this research, we exploited the performance of lightest EfficientDet with $n = 0$, or the least level available in the detection model.

C. optimisation Function

optimisation Function is for readjust parameters such as bias and weight in order to reduce loss and improve model performance. In this research, Adam and Stochastic Gradient Descend (SGD) is used to improve model performances.

1) Adam optimisation: Adam optimisation function is derived from adaptive moment estimation which has advantages, especially in memory usages and computational efficiency due to this optimisation process is a combination between Adaptive Gradient Algorithm or simply called momentum propagation and Root Mean Square Propagation [24]. Adam gains fast gradient descent rate like in momentum propagation algorithm and great at control gradient descent, especially in suppressing oscillation when it reaches global minimum. Since Adam can take big enough step in learning rate, it wont stuck in local minimum. The formula of Adam optimisation can be seen in:

$$W_{i+1} = W_i - \hat{m}_i\left(\frac{\alpha_i}{\sqrt{v_i + \epsilon}}\right) \quad (3)$$

W_{i+1} stands for result on future iterations, W_i for current iterations and $\hat{m}_i(\frac{\alpha_i}{\sqrt{v_i+\epsilon}})$ stands for bias correction on Adam optimisation for improving function to get its global minimum.

2) Stochastic Gradient Descent: Stochastic Gradient Descent (SGD) is an optimisation function based on gradient descent associated with random probabilities. Since it is randomised and selected for each iteration process, SGD tends to get louder oscillations like zigzagging paths, however high oscillations are tolerable problem due to SGD has shorter time in finding global minimum. Due to its oscillating result, SGD needs a higher iterations to find global minimum and prone to get stuck in local minimum result.

D. VisDrone Dataset

This research is using VisDrone dataset. Visdrone dataset is a collection of UAV photos from AISKYEYE team. This team created in the laboratory of Machine Learning and Data Mining in Tianjin University. Tianjin University is located in mainland China, making the dataset consisting of Chinese urban city setup [25]. The dataset created consist of many scenarios. Object detection scenario part of the dataset containing 10,209 UAV areal photos. These photos split into three parts. 548 of the photos meant for validation; 6,471 train images; and leaving the rest 3,190 photos as model tester. The UAV areal photos was captured using real drone, not a dolly-crane platform. The ASKYEYE team gave heights and places variations across the dataset. All ground-truth annotations are created using human aid. This dataset including train, valid, and test sets are publicly accessible. VisDrone dataset has 10 different classes to detect (such as tricycle, a not moving person, van, car, truck, passenger bus, motorcycle, awning-tricycle,bicycle, and pedestrian [moving person]), and rare or uncommon-proposed vehicles (for example like forklift, tanker, fire-engine) are ignored. Detected objects in Visdrone dataset visualised in figure 3.

E. Training Process and Data Preparation

Before being trained, we transformed the annotation format from PASCAL VOC to tfrecord, in order to accommodate the EfficientDet D0 model trained using automl framework. In each epoch, we randomly picked 5696 from 6471 images for training, and the random picking process will be repeated each epoch. We used EfficientDet D0 from ImageNet checkpoint as backbone. Training process will be limited to maximum 50 epochs. Training will be analysed every 10 epochs. We trained the model not using any high performance computing device GPUs to represent lacking of access to high performance computing devices. Instead we used a laptop-grade GPU. We used a laptop with NVIDIA GeForce GPU of GTX 1660 Ti mobile edition or Max-Q Design with 6GB of GDDR6 dedicated GPU memory, 9th generation 2.6 GHz (up to 4.5 GHz performance turbo) Intel Core i7-9750H CPU and 16 GB, DDR4 2933 MHz dual channel RAM. EfficientDet D0 trained on two different optimiser function, stochastic gradient descend [26], and Adam [24]. As in previous research [25], we measured the model performance with six values, AP50, AP75, APl, ARmax1, ARmax10, and ARmax100.

III. Experiment Result

A. Using Stochastic Gradient Descend as Optimiser Function

The EfficientDet D0 model using stochastic gradient descend optimiser function training result available at Table I Results in **bold** shows performance exceeded previous research. Results in **<u>bold and underline</u>** shows performance exceeded state-of-the-art models in previous research. As expected, the average precision values are small. This is due to average precision needs hundreds or even thousands of epoch to get its optimal value. Especially when we are using smaller level of the EfficientDet object detection model. Beside of

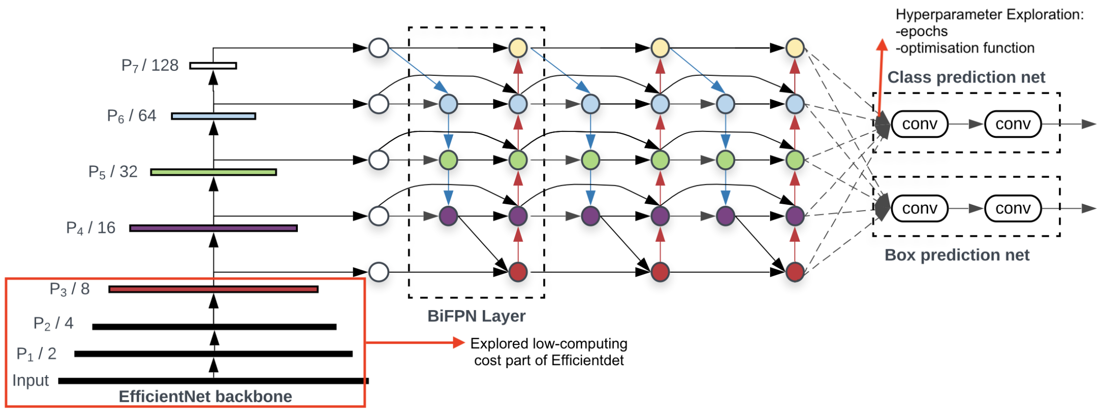

Fig. 2: proposed EfficientDet Exploration Methods, based on Tan, et. al. [15]

TABLE I: EfficientDet D0 Result with SGD and Adam optimiser

epochs	optimiser	AP50	AP75	AP	ARmax1	ARmax10	ARmax100
10	SGD	0.018	0.006	0.046	**0.009**	**0.026**	0.033
	Adam	0.025	0.01	0.058	**0.009**	**0.031**	0.039
20	SGD	0.012	0.002	0.024	**0.003**	**0.01**	0.014
	Adam	0.039	0.02	0.105	**0.018**	**0.043**	0.052
30	SGD	0.037	0.017	0.09	**0.015**	**0.04**	0.05
	Adam	0.043	0.02	0.093	**0.018**	**0.05**	**0.059**
40	SGD	0.041	0.018	0.098	**0.017**	**0.045**	**0.055**
	Adam	0.049	0.024	**0.122**	**0.021**	**0.05**	**0.059**
50	SGD	0.041	0.022	**0.116**	**0.018**	**0.045**	**0.056**
	Adam	0.052	0.026	**0.11**	**0.021**	**0.054**	**0.066**

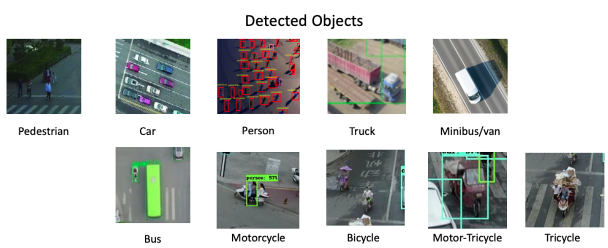

Fig. 3: Detected objects in Visdrone dataset

average precision, average recall actually revealed impressive values, especially in ARmax1. Bold values are the training results that exceeded previous models in the category [25]. By only 30 epochs, EfficientDet d0 succeeded to be the best ARmax1 model, beating other vigorously trained object detection model. Adding up to 50 epochs, this result is getting better, with 0.018 ARmax1 score worth 50 epochs of training.

B. Using Adam as Optimiser Function

The EfficientDet D0 model using Adam optimiser function training result also available at Table I, with SGD optimiser value for comparison. Unlike the average precision value, average recall value shows a promising result. By only 20

TABLE II: AR Comparison with Previous Model

Method	AR1[%]	AR10[%]	AR100[%]
Proposed EfficientDet D0	**2.1**	**5.4**	**6.6**
TridentNet [27]	1.17	8.3	28.98
ConstraintNet [28] (A.10)	0.44	3.97	21.23
EnDet (A.13) [29]	0.31	2.49	24.47
RetinaNet	0.21	1.21	5.31

training iteration, the EfficientDet D0 able to position its performance as the best ARmax1 value [25]. Only adding 10 epochs to the training process, the EfficientDet D0 able to surpass TridentNet, the current top performing model for ARmax1. TridentNet only able to reach 1.17% ARmax1 value, and worth to mention that this model typically trained in hundreds of epochs [25]. EfficientDet D0 able to reach 1.8% ARmax1 value by only 20 epochs. This result making EfficientDet easily reproducible, and deployed across systems. This result is getting better the more epochs added. Finally after 50 epochs, EfficientDet D0 scored 2.1% ARmax1 value, almost twice as much than TridentNet [25].

C. Model Implementation as Object Detection-Based UAV

As seen on both of the results, average precision score value of EfficientDet D0 is far below another state-of-the-art model

[25]. This result is tolerable because the importance of average recall value in object detection-based UAV implementation scenario. With better average recall value, the model will detect more object and focuses on localisation accuracy [30]. When an object detection model has high recall but low precision value, model will classify more object, but some of them will be classified as a different object, for example a truck got detected as a car [31]. In object detection-based UAV implementation scenario, a higher recall value is needed for live inference when the UAV is on mobile. Recall is being prioritised over precision because detecting there is an object is more important than correctly detect the class of an object. After detecting object presence, detection speed is also important in object detection-based UAV implementation scenario, especially when the UAV is on mobile. In many scenario, decreases of the detection quality to prioritise detection speed is common and tolerable [30]. Adam's optimisation function is generally superior to SGD. This advantage is because in this case, the appearance of objects in the image is not always in the same or adjacent place. Not only the location in the image, the size of the target object for detection in each image also varies. The diversity in this image makes Adam an optimisation function with a superior adaptive learning rate to find the smallest loss compared to the stochastic gradient descend function which will be more easily trapped in local minima [32] [33]. However, with Adam's superiority, average precision scores are not high enough to be equivalent to the model in previous studies. This is caused by many factors, one of which is the training iteration which is still relatively low compared to the model in previous studies. Various models in previous studies were trained with high training iterations, more than 100 times of training. Proposed method's result compared to results from previous research shown in Table II.

test test test

IV. Conclusion and Future Research

Looking up to the simulation and model performance analysis, EfficientDet d0 model able outperform in terms of ARmax1 score of the previous research model in 20 iterations or more, even though it trained under limited hardware computing capabilities. The EfficientDet d0 model is suitable as a low-cost object detection model for UAV implementations because it has a high value of recall. Due to the lower computational costs required compared to another UAV-based object detection models, it will simplify the training process, eliminate high-end hardware requirements, increase the detection speed of the UAV and the scalability of object detection in the future. Suggestions that can be applied to further research related to the topic of the research is that there is potential to increase the detection quality of EfficientDet models. Not only by improving recall values, but also to increase precision. Future research can improve the overall model reliability by applying more preprocessing to simplify the input image, as well as modifying the pooling layer on the model to increase the model's capabilities when faced with high-resolution images from UAV.

V. Acknowledgement

This research would not be possible without supports from the Directorat of Research and Community Services Telkom University and PT Epindo Prima Solusi as an industrial partner.

References

[1] A. C. Watts, V. G. Ambrosia, and E. A. Hinkley, "Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use," *Remote Sensing*, vol. 4, no. 6, pp. 1671–1692, 2012. [Online]. Available: https://www.mdpi.com/2072-4292/4/6/1671

[2] S. Agostino, M. Arjomandi, M. Mammone, M. Nelson, and T. Zhou, *CLASSIFICATION OF UNMANNED AERIAL VEHICLES*, 2007. [Online]. Available: https://d1wqtxts1xzle7.cloudfront.net/29666442/group9-with-cover-page-v2.pdf

[3] J. Wu, G. Yang, X. Yang, B. Xu, L. Han, and Y. Zhu, "Automatic counting of in situ rice seedlings from uav images based on a deep fully convolutional neural network," *Remote Sensing*, vol. 11, no. 6, 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/6/691

[4] S. Oleire-Oltmanns, I. Marzolff, K. D. Peter, and J. B. Ries, "Unmanned aerial vehicle (uav) for monitoring soil erosion in morocco," *Remote Sensing*, vol. 4, no. 11, pp. 3390–3416, 2012. [Online]. Available: https://www.mdpi.com/2072-4292/4/11/3390

[5] K. G. Panda, S. Das, D. Sen, and W. Arif, "Design and deployment of uav-aided post-disaster emergency network," *IEEE Access*, vol. 7, pp. 102 985–102 999, 2019.

[6] J. M. M. Neto, R. A. da Paixão, L. R. L. Rodrigues, E. M. Moreira, J. C. J. dos Santos, and P. F. F. Rosa, "A surveillance task for a uav in a natural disaster scenario," in *2012 IEEE International Symposium on Industrial Electronics*, 2012, pp. 1516–1522.

[7] G.-H. Kwak and N.-W. Park, "Impact of texture information on crop classification with machine learning and uav images," *Applied Sciences*, vol. 9, no. 4, 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/4/643

[8] W. Pikun, W. Ling, Q. Jiangxin, and D. Jiashuai, "Unmanned aerial vehicles object detection based on image haze removal under sea fog conditions," *IET Image Processing*, 2022.

[9] C. Kyrkou, G. Plastiras, T. Theocharides, S. I. Venieris, and C.-S. Bouganis, "Dronet: Efficient convolutional neural network detector for real-time uav applications," in *2018 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2018, pp. 967–972.

[10] P. Mittal, R. Singh, and A. Sharma, "Deep learning-based object detection in low-altitude uav datasets: A survey," *Image and Vision Computing*, vol. 104, p. 104046, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885620301785

[11] A. Wiranata, S. A. Wibowo, R. Patmasari, R. Rahmania, and R. Mayasari, "Investigation of padding schemes for faster r-cnn on vehicle detection," in *2018 International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC)*, 2018, pp. 208–212.

[12] S. A. Wibowo, H. Lee, E. Kim, and S. Kim, "Visual tracking based on complementary learners with distractor handling," *Mathematical Problems in Engineering*, vol. 2017, 04 2017.

[13] A. Wiranata, S. A. Wibowo, and R. Patmasari, "Analysis of transfer learning on faster r-cnn for vehicle detection," in *Symposium of Future Telecommunication and Technologies (SOFTT)*, no. 2, 2018.

[14] S. Vaddi, D. Kim, C. Kumar, S. Shad, and A. Jannesari, "Efficient object detection model for real-time uav application," *Computer and Information Science*, vol. 14, p. 45, 01 2021.

[15] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10 781–10 790.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[17] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," 2013.

[18] P. Mittal, R. Singh, and A. Sharma, "Deep learning-based object detection in low-altitude uav datasets: A survey," *Image and Vision Computing*, vol. 104, p. 104046, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0262885620301785

[19] Z. Tang, X. Liu, G. Shen, and B. Yang, "Penet: Object detection using points estimation in aerial images," *CoRR*, vol. abs/2001.08247, 2020. [Online]. Available: https://arxiv.org/abs/2001.08247

[20] C. Iuga, P. Drăgan, and L. Busoniu, "Fall monitoring and detection for at-risk persons using a uav," *IFAC-PapersOnLine*, vol. 51, pp. 199–204, 01 2018.

[21] C. Kyrkou, G. Plastiras, T. Theocharides, S. I. Venieris, and C.-S. Bouganis, "Dronet: Efficient convolutional neural network detector for real-time uav applications," in *2018 Design, Automation Test in Europe Conference Exhibition (DATE)*, 2018, pp. 967–972.

[22] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019. [Online]. Available: http://arxiv.org/abs/1905.11946

[23] J. Chen, H. Mai, L. Luo, X. Chen, and K. Wu, "Effective feature fusion network in bifpn for small object detection," in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 699–703.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[25] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," 2021.

[26] C. Lemaréchal, "Cauchy and the gradient method," 2012.

[27] Y. Li, Y. Chen, N. Wang, and Z. Zhang, "Scale-aware trident networks for object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6054–6063.

[28] M. Brosowsky, O. Dünkel, D. Slieter, and M. Zöllner, "Sample-specific output constraints for neural networks," *CoRR*, vol. abs/2003.10258, 2020. [Online]. Available: https://arxiv.org/abs/2003.10258

[29] P. Zhu, L. Wen, D. Du, X. Bian, H. Ling, Q. Hu, Q. Nie, H. Cheng, C. Liu, X. Liu *et al.*, "Visdrone-det2018: The vision meets drone object detection in image challenge results," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0–0.

[30] J. H. Hosang, R. Benenson, P. Dollár, and B. Schiele, "What makes for effective detection proposals?" *CoRR*, vol. abs/1502.05082, 2015. [Online]. Available: http://arxiv.org/abs/1502.05082

[31] A. Gad, "Evaluating object detection models using mean average precision," Mar 2022. [Online]. Available: https://www.kdnuggets.com/2021/03/evaluating-object-detection-models-using-mean-average-precision.html

[32] A. Lydia and S. Francis, "Adagrad - an optimizer for stochastic gradient descent," vol. Volume 6, pp. 566–568, 05 2019.

[33] J. Zhang, S. P. Karimireddy, A. Veit, S. Kim, S. J. Reddi, S. Kumar, and S. Sra, "Why are adaptive methods good for attention models?" 2019. [Online]. Available: https://arxiv.org/abs/1912.03194