CISTER
Research Centre in
Real-Time & Embedded
Computing Systems

Conference Paper

A Hybrid Deep Learning Model for UAVs Detection in Day and Night Dual Visions

Alam Noor*
Kai Li*
Adel Ammar
Anis Koubâa*
Bilel Benjdira
Eduardo Tovar*

*CISTER Research Centre

CISTER-TR-211103

2021

A Hybrid Deep Learning Model for UAVs Detection in Day and Night Dual Visions

Alam Noor*, Kai Li*, Adel Ammar, Anis Koubâa*, Bilel Benjdira, Eduardo Tovar*

*CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: alamn@isep.ipp.pt, kai@isep.ipp.pt, aammar@psu.edu.sa, aska@isep.ipp.pt, emt@isep.ipp.pt

https://www.cister-labs.pt

Abstract

Unmanned Aerial Vehicle (UAV) detection for public safety protection is becoming a critical issue in non-fly zones. There are plenty of attempts of the UAV detection using single stream (day or night vision). In this paper, we propose a new hybrid deep learning model to detect the UAVs in day and night visions with a high detection precision and accurate bounding box localization. The proposed hybrid deep learning model is developed with cosine annealing and rethinking transformation to improve the detection precision and accelerate the training convergence. To validate the hybrid deep learning model, real-world experiments are conducted outdoor in daytime and nighttime, where a surveillance video camera on the ground is set up for capturing the UAV. In addition, the UAV-Catch open database is adopted for offline training of the proposed hybrid model, which enriches training datasets and improves the detection precision. The experimental results show that the proposed hybrid deep learning model achieves 65% in terms of the mean average detection precision given the input videos in day and night visions.

A Hybrid Deep Learning Model for UAVs Detection in Day and Night Dual Visions

1st Alam Noor
CISTER Research Center
Porto, Portugal
alamn@isep.ipp.pt

2nd Kai Li
CISTER Research Center
Porto, Portugal
kai@isep.ipp.pt

3rd Adel Ammar
Prince Sultan University
Riyadh , Saudi Arabia
aammar@psu.edu.sa

4th Anis Koubaa
Prince Sultan University
Riyadh , Saudi Arabia
CISTER Research Center
Porto, Portugal
akoubaa@psu.edu.sa

5th Bilel Benjdira
Prince Sultan University
Riyadh , Saudi Arabia
bilel.benjdira@gmail.com

6th Eduardo Tovar
CISTER Research Center
Porto, Portugal
emt@isep.ipp.pt

***Abstract*—Unmanned Aerial Vehicle (UAV) detection for public safety protection is becoming a critical issue in non-fly zones. There are plenty of attempts of the UAV detection using single stream (day or night vision). In this paper, we propose a new hybrid deep learning model to detect the UAVs in day and night visions with a high detection precision and accurate bounding box localization. The proposed hybrid deep learning model is developed with cosine annealing and rethinking transformation to improve the detection precision and accelerate the training convergence. To validate the hybrid deep learning model, real-world experiments are conducted outdoor in daytime and nighttime, where a surveillance video camera on the ground is set up for capturing the UAV. In addition, the UAV-Catch open database is adopted for offline training of the proposed hybrid model, which enriches training datasets and improves the detection precision. The experimental results show that the proposed hybrid deep learning model achieves 65% in terms of the mean average detection precision given the input videos in day and night visions.**

***Index Terms*—UAV Detection, IR Stream, RGB Stream, Convolutional Neural Networks, Rethinking Transformation, Cosine Annealing**

I. Introduction

Unmanned Aerial Vehicles (UAVs) are commonly used for commercial purposes due to flexible deployment and versatility, such as public surveillance, cartography, search and rescue, as shown in Figure 1. It is critical to automatically detect and locate the UAVs in the non-fly zone to ensure aviation safety or efficient air traffic control in non-fly zones. Specifically, it is essential to distinguish a UAV from an object with a similar shape, such as birds or aircraft. Most of deep learning models, e.g., [34] [35] [41] [36] [38], are developed for the UAV detection based on either RGB (the day vision camera) or IR (the night vision one). However, the UAV detection rate based on RGB is low when the camera has insufficient light in the daytime, e.g., cloudy or stormy weather. The UAV detection based on IR is affected when the UAV overlaps with an object (e.g., buildings or trees) in the background. Due to the effect of the light condition or similar shape of the UAV and other objects, false detection of the UAV and results in a low training accuracy of the deep learning with RGB or IR. Developing a hybrid deep learning model for processing RGB and IR videos is non-trivial since the RGB and IR video frames are trained independently with day or night vision features. As a result, the training for RGB and IR videos can not be directly combined for the UAV detection in a dual vision mode. Moreover, developing a hybrid model can suffer from a high complexity due to feature vanishing problems on the training of the RGB and IR videos.

The UAV needs to be located in the video and differentiated from other objects. The position of the UAV can be located by using bounding box localization (BBL) which determines the height, width, and X and Y-coordinates in annotations form [49]. Specifically, annotation is typically used to determine the UAV position with annotating of the upper left corner and the lower right corner of the UAV. Since the background color can be similar to the color of the UAV, the annotation with BBL experiences annotation errors, which results in classification and regression losses in the deep learning of BBL. In particular, the classification loss is due to the color difference between the UAV and the background, while the regression loss defines the difference between the actual bounding box and the predicted one of the training. In addition, classification and regression loss of the BBL training depends on the position of the UAV in the video and the annotation error.

In this paper, we propose a new hybrid deep learning model for the UAV detection with day and night dual visions. The contributions of our proposed work are as follows:

- We propose a new hybrid deep learning model to improve the detection accuracy of the UAV in the day and night dual visions. The hybrid deep learning model adopts rethinking transformation to accelerate the training model's convergence and reduce the classification and regression

Identify applicable funding agency here. If none, delete this.

losses. Moreover, the proposed hybrid model develops cosing annealing, which freezes stepwise the initial layers of the deep learning model to reduce the training time.
- The proposed hybrid model is developed to enhance the training of the BBL by using the information of coordinates dimensions (height and width) to determine the actual location of the UAV and reduce the annotation errors of the UAV detection.
- To evaluate the proposed hybrid model in the real-world, experiments are conducted to detect the UAV in daytime and nighttime. The experimental results show that the proposed hybrid deep learning model achieves 65% higher detection accuracy than the benchmark EfficientDet. In addition, the convergence of the hybrid model is 10% faster than the EfficientDet.

The paper is organized as follows. Section II presents the related works. In Section III, we investigate the proposed hybrid deep learning model and its implementation. We present the experimental setup and performance evaluation in Section IV. Finally, we conclude the paper in Section V.

II. Related Works

Several detection approaches with RGB or IR are studied in the literature to detect objects, such as people or vehicles. In [1] and [2], micro-Doppler signatures, thermal position intensity histogram of the oriented gradient are adopted for detecting one stream. In [3] and [4], aggregated channel features and shape context descriptor-based pedestrian detection are presented for the detection in day or night environment, respectively. Guan et al. [5] combine extracted features of two imaging sensors by using illumination and coefficients correction of the day and night estimated illumination. Many algorithms in the literature are used for the UAV detection, infrared-based systems, such as background subtraction, thermal visible video fusion, and robust multi-stage approaches detect objects with cameras [6]–[8]. Lin et al. [9] used Hidden Markov models to extract features and detect UAVs in noisy environments. Moreover, Li et al. [10] used a histogram of an oriented gradient to adjust parameters and geometric characteristics and support vector machine for the object detection in thermal images. Teutsch et al. [11], presented a two-stage person recognition model that adopts maximally stable extremal regions, and discrete cosine transform with random naïve bayes. Video-based object detection models are developed to detect objects based on extracting discriminant features such as distinctive invariant features [12], histograms of oriented gradients [13], and SURF (scale- and rotation-invariant detector and descriptor) [14].

Radar-based techniques are also developed in the literature [15]–[18]. Jahanger and Baker developed a holographic radar using Extended dwell Doppler characteristics to detect UAVs and distinguish them from birds [15]–[17]. Drozdowicz et al. [18], adopted an approach that uses multiple radars through the directional transmission to monitor a restricted area and used a geofence to detect UAVs. The mmWave radars detect UAVs at lower altitudes by using the Doppler spectrum. The mmWave radars detect moving UAVs concerning radars and prevent flight collisions of UAVs [19]–[21]. However, mmWave radars cannot differentiate the UAV from objects with a similar shape, like birds. Thus, the UAV detection accuracy using mmWave radars is low.

Convolutional Neural Networks (CNN) are developed for the UAV detection [22]–[24]. Muhammad et al. study a transfer learning method using the combination of VGG16 and Faster-RCNN for the detection of UAVs [22]. Al-Emadi et al. [23] applied CNN, RNN, and RCNN architectures on audio recorded samples to exploit the unique acoustic fingerprints of flying UAVs. Jihun et al. [24], present a method using Pan-Tilt-Zoom (PTZ) with a combination of camera and Faster R-CNN Inception Resnet algorithm for UAV detection with pan, tilt, and zoom actions. Byunggil and Daegun [42] used Short-time Fourier Transform and Wigner-Ville distribution to transform micro-Doppler signatures from diverse UAVs to images. The optimal hyperparameters for the CNN were determined using a heuristic search. Several advanced trainable object detection models have been proposed to compete with these problems.

Nevertheless, the above approaches cannot match the performance and precision to detect UAVs in day and night vision with different environmental changes and if the background similarity exists. Recently, CNN methods used for object detection, owing to large-scale datasets and advances in deep learning technologies. The EfficientNet-B3 [28], is used for the classification by using the technique of scaling up network width, depth, and resolution. Recently, scalable EfficientDet architecture for one-stage object detection was presented by the Facebook research community [29]. It uses a weighted bi-directional pyramid feature network with EfficientNet-B3 as the backbone. We propose a hybrid deep learning model using the EfficientDet for training with rethinking transformation, cosine annealing, and focal loss to increase the model accuracy and efficiency and avoid the high cost of computational power for detecting UAVs in videos from dual streams. Most previous works used a single stream for the detection of UAVs, as can be seen in Table II. In contrast, we used a single model (Hybrid) shown in the figure 2 for both day and night visions. In the next section, we present the detailed methodology of the proposed hybrid deep learning model.

III. The Hybrid Deep Learning Model

A. Dataset Preparation and Preprocessing

In this paper, we utilize open-source Anti-UAV catch video datasets [43]. This dataset contains 160 HD dual streams (RGB and IR videos), each with 100 validations and 60 training videos. Each video has multiple UAV models with three different sizes (large, medium, and tiny). Although the video frames are labeled, the labeling undergoes high annotation errors due to the poor UAV detection performance [44]. The video frames contain different backgrounds: day-night, cloudy/clear, similar objects (building, birds), scale variations, and occlusion. The UAVs move at different speeds and abruptly stop at very high speeds (100 mph). We used the 100 validation videos from each stream for the training of our model. 90 of the videos

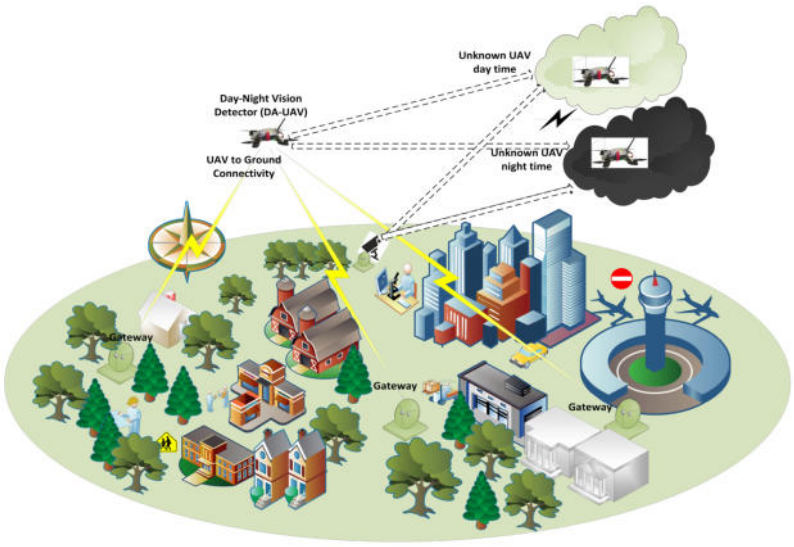

Fig. 1. Applications and framework of hybrid deep learning model for critical areas.

TABLE I
COMPARISON OF THE PRESENT WORK WITH RELATED WORKS ON UAV DETECTION

References	Dataset	Methods	Vision
Rozantsev et al., 2017 [34]	UAV/Aircraft Database	CNN and Boosted trees methods	Single Stream
Saqib et al., 2017 [35]	UAV/Bird Database	VGG and ZF Fine-tuning	Single Stream
Dong et al., 2017 [41]	Local Dataset	Feature Classifier	Single Stream
Peng et al., 2018 [36]	Synthetic Database	Faster-RCNN and ResNet-101 fine-tuning	Single Stream
Nalamati et al , 2019 [38]	Bird vs UAV	Faster R-CNN with ResNet-101	Single Stream
Lee et al., 2019 [39]	Web Data	Deep CNN	Single Stream
Hu et al., 2019 [40]	Local Dataset	Improved YOLO v3	Single Stream
Behera et al., 2020 [37]	Local Dataset	YOLOv3	Single Stream
Proposed Work	**UAV Catch Database**	**Hybrid**	**Dual Stream**

are used for training, and 10 videos are for validation/testing of the proposed model.

1) UAV Localization: The dataset annotation needs the proper conversion by using the information of coordinates dimensions. During the flight process, the location of the UAV can be determined by the X-coordinates and Y-coordinates with height (H) and width (W) of the UAV to reduce annotation errors in the detection. Therefore, we detect four ($x_{\max}$, $x_{\min}$ and $y_{\max}$, $y_{\min}$) translation parameters, which detects the UAV with the top-left corner and bottom-right corner. The X-axis position of the UAV presents the width ($x_{\max}$-$x_{\min}$) of the UAV, while Y-axis represents the height ($y_{\max}$-$y_{\min}$) of the UAV. The UAV's position varies by altering its rotation by global coordinate ($x_{\max}$, $x_{\min}$ and $y_{\max}$, $y_{\min}$ pivoted) and not fixed horizontally (non-pivoted), so the width and height change according to the sum of the X-axis ($x_{\max}$+$x_{\min}$) and Y-axis ($y_{\max}$+$y_{\min}$) respectively. In our work, the UAV is correctly located in RGB video frames and thermal (IR) video frames. To decide whether the operating procedure is in RGB or IR, we used the brightness information provided by the surveillance camera. The proposed UAV localization traces the movement of the UAV multiple times for the localization. Table II lists the features of the proposed UAV localization, as compared with the localization techniques in the literature.

TABLE II
UAV DATASET STATISTICAL APPROACH OF THE RGB AND IR FRAMES.

Categories	Training set	Validation set	Testing set	Total
Number of Frames	162k	19k	19k	200k
Percentage (%)	81.0%	9.5%	9.5%	100%
Number of RGB frames	81k	9.5k	9.5k	100k
Number of IR frames	81k	9.5k	9.5k	100k

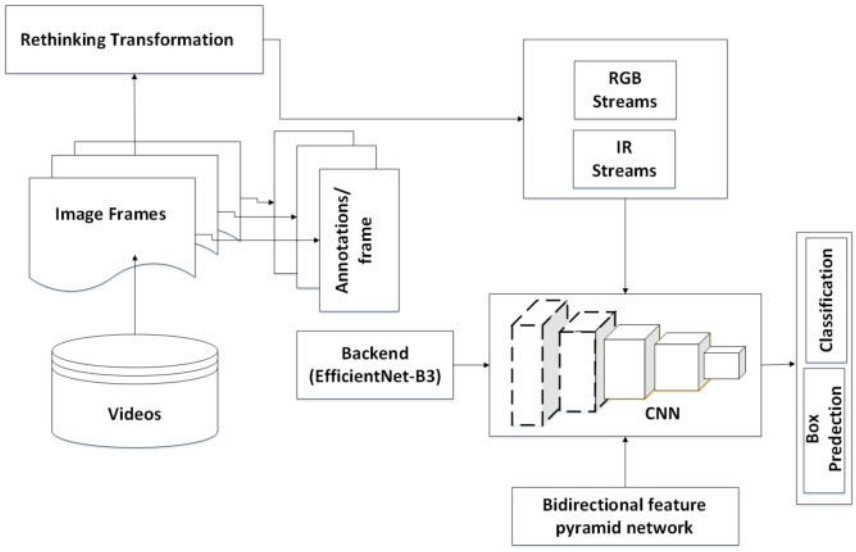

Fig. 2. The proposed architecture for the UAV detection using dual streams.

B. Backbone of the UAV detection

Pixels to features translation is developed with the proposed hybrid deep learning model. For scaling detection of the UAV, our hybrid model automatically adjusts the scaling of the width, depth, and resolution of CNN consistently to provide different sizes of the UAV. Moreover, the hybrid model extends EfficientNet-B3 to enable feature extractor for the UAV detection to achieve high accuracy and inference speed.

C. Dual-stream video processing

Using a hybrid deep learning model to process dual-stream videos is a challenging problem for UAV detection. Most UAV detection techniques are based on single-stream video, i.e., either RGB or IR. Moreover, dual-stream deep learning structures are used in cross-modality to process both RGB and IR inputs. The RGB features dominate the IR features in terms of vanishing during training the model. Designing a hybrid deep learning model based on the EfficientDet can overcome dual streams vanishing features during training the hybrid model of loss function approaches to zero. RGB and IR dominate each other during training which creates features vanishing either for RGB or IR.

The proposed hybrid model investigates Bi-Directional Feature Pyramid Network (BiFPN), which is an extension to EfficientDet. Specifically, BiFPN combines the ideas from Feature Pyramid Network, Path Aggregation Network, and NAS-FPN in the form of multi-level feature fusion, which efficiently helps the proposed hybrid deep learning model extract features for both RGB and IR streaming. The proposed hybrid model enables processing the input of dual streams, minimizing high computational cost and reducing vanishing features during training the hybrid model. The proposed hybrid model takes advantage of the BiFPN to minimize the computational cost by removing nodes with a single input edge and adding extra borders from input to output.

Moreover, the hybrid model utilizes BiFPN to fuse high-level features of RGB and IR efficiently. The BiFPN configures additional weight to the input features, enabling the hybrid model to learn the feature importance of each input.

D. Rethinking Transformation

The rethinking transformation shown in Figure 3 aims to enhance the accuracy and efficiency of the proposed hybrid model. Since the CNN architectures are prone to colour, rotation, an axis-aligned bounding box transformation, we study the rethinking transformation in color background transformation, i.e., contrast, flapping, equalizing, solarizing, and sharpness, to apply the visual effect video frames. In the rethinking transformation, the random probability for all parameters is set to 0.5, except 0.1 for solarization with a threshold of 128. Solarization tonal the values of the video frames in which darks areas appear bright and vice versa. The factor for adjusting the colour is $0 \leq x \leq 2$ to maintain the originality and quality of the data. The rethinking transformation rotation of the UAV performed 45-degree clockwise and anti-clockwise can vary according to the rotation degree from the center (width and height by 2) of the UAV. The new bounding box dimensions of the video frames have been computed after getting sine and cosine from the rethinking transformation rotation matrix. The bounding box annotation of the rethinking transformation is tackled by an array of $Ox5$, where O is an object (i.e., the UAV) in the frame, and 5 gives the attributes. The attributes include top-left corner coordinates, bottom-right corner coordinates, and a class of UAVs. In addition, the new boundary box of the rotation has been translated to the integer scaling using the rethinking transformation to avoid

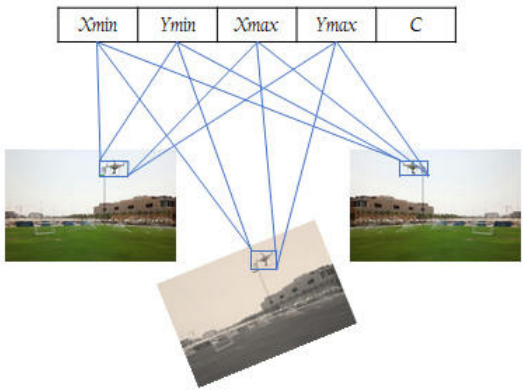

Fig. 3. Rethinking transformation of UAV with boundary box for precise bounding box localization.

the missing colour problems. The missing colour is due to rotation in which new pixels are added to the boundary of the video frames, and the new ones replace the original pixels. The rethinking transformation determines the nearest neighbour (integer scaling) is an up-scaling method of the coordinate mapping to the nearest pixel value.

E. Cosine Annealing

The proposed hybrid deep learning model reduces the training time using cosine annealing. Initial layers of the CNN represent low-level features. These early layers take up most of the allocation but have updates of very few parameters and converge to straightforward configurations influences compared to later layers (high-level features layers), where most of the parameters are updated during training. The initial layers do not require as much fine-tuning as influenced by this consideration. Cosine annealing is developed with the hybrid model to freeze the layers' weights, which preserves the pre-trained features for accelerating the learning. Cosine annealing is used for freezing out stochastic depth [32], [33]. In particular, cosine annealing trains each layer for several runs, gradually "freezes out" layers using equation 1, and excludes them from backward passes to accelerate the hybrid model.

$$LearningRate(\alpha_i) = 0.5 * \frac{\alpha}{t_i}(1 + cos(\frac{\pi * t}{t_i})) \quad (1)$$

where α defines the learning rate with the initial learning rate zero. t_i is the number of iterations between a user-selected t_0 and the total iteration during training. Iteration (t) depends on the rule of linear scheduling to reduce the training time.

F. Focal Loss

The two-stage object detection is computationally expensive to find the candidate object box [50]–[52]. First, the two-stage object detectors use the Region Proposal Network to predict candidate bounding boxes. In the second stage, features are pooled from each candidate box for classification and bounding box regression tasks using RoI (Region of Interest) operations. One-stage object detectors (i.e., EfficientDet) can detect the UAV without creating proposal drawing (region proposals) [29]. However, the problems of instances on new data mainly exist in the one-stage neural networks detectors. The prediction of the deep learning model is far from the correct prediction if not used the focal loss. Focal loss is specially designed for the one-stage detection scenario. It can tackle the significant imbalance between foregrounding and background UAV´s for the proposed training hybrid model to reduce the weight toward correct prediction. Cross-Entropy is frequently adopted loss function due to high precision to compare the approximate models.

$$Cross\ Entropy(_{p,t}) = \begin{cases} -\alpha_t\ log(p), \text{ if } t = 1 \\ -\alpha_t\ log(1-p), \text{ otherwise} \\ \text{When, } t \in \{\pm 1\}, p \in [0,1] \end{cases} \quad (2)$$

In 2, the t is the target value as ground truth, and the p is the probabilistic estimated model value for the UAV detection. Where α_t is the balanced parameters for positive and negative examples, but it can not distinguish between easy and hard examples. The modulated factor $(1-p_t)^\gamma$ of the focal loss is necessary for the numerical stability by using down weight methodology. The 2 is updated in the focal loss as: beginequation

$$Focal\ Loss(_{p,t}) = -\alpha_t\ (1-p_t)^\gamma\ log(p_t) \quad (3)$$

The focal loss down-weighted tuning process is dependent on the γ, and it varies from 0 to 2 to adjust the rate of easy examples. If γ is 0, then the focal loss equals the cross-entropy. The UAV detection scenarios increase the γ up to 2 to obtain the best training result.

IV. Experiments and Results

The experiments of the UAV detection are performed on the UAV dual-stream dataset of the Anti-UAV CVPR workshop. The dataset is divided into 162k frames for training and 19k frames for testing. Each frame is labeled with boundary boxes for UAVs and correctly localized by the UAV localization

technique. GPU used a workstation equipped with a GeForce RTX 2070 to train the model using Keras with the backend of Tensorflow. We opted for EfficientDet compatible with training on the GeForce RTX 2070 with a batch size of 8. Adam optimizer is used for its fast and easy convergence to the optimal point concisely compared to other optimizers. The learning rate decay set to 0.001 with $\beta 1$ of 0.9 and $\beta 2$ equal to 0.999. The model trained for 100 epochs, which took 97 hours. The focal loss kept the same as in the original EfficientDet paper α=0.25 and γ=1.5 with an aspect ratio between 1/2 and 2.

A. Results

We checked our hybrid deep learning model on the ground and with test videos to emphasize efficiency in various UAV positions. First of all, let us evaluate the identification accuracy of the proposed hybrid model with many traditional assessment methods. The video used in the experimentation is not used for testing or validation. We have also attempted to test the hybrid model for all different backgrounds and UAVs. During experimental results, different parameters are used to evaluate the performance of the training hybrid model. Focal loss is one of those parameters which is designed to differentiate between foreground and background for the one-stage UAV detection during training. We obtained 0.164 focal loss for classification and 0.15 with rethinking transformation and boundary box augmentation at the end of 100 epochs to train the hybrid model. The regression loss is obtained by using mean absolute error and the regression loss end up with 0.99 without augmentation and 0.97 with augmentation. Both Classification and regression losses are given in Figure 4. The diversity of the aerial scene creates variations in intra-class. These variations affected the classification and regression loss and degraded the scene classification performance. Batch loss (BL) is studied to eliminate the large variations of intra-class features with in-depth features in the batch and batch center of the given UAV. The equation 4 represents the batch loss which is given as;

$$BL = \frac{1}{2}\sum_{i=1}^{N}(x_i - B_c)^2 \tag{4}$$

Where B_c is a center of the corresponding batch and x_i denotes the deep learning features in the batch. We achieved a classification batch loss of 0.163 with simple training and 0.04 with rethinking transformation. Moreover, the regression batch loss is 0.99 and 0.5, respectively.

Mean average precision (mAP) is the essential evaluation metric to measure hybrid model UAV´s detection performance. The mAP is the mean of all corresponding average precision for a given recall. It defines the model's accuracy. A higher mAP denotes a high precision for the detection of objects. The mAP is defined as follows:

$$mAP = \frac{\sum_{q=1}^{Q} AP_q}{Q} \tag{5}$$

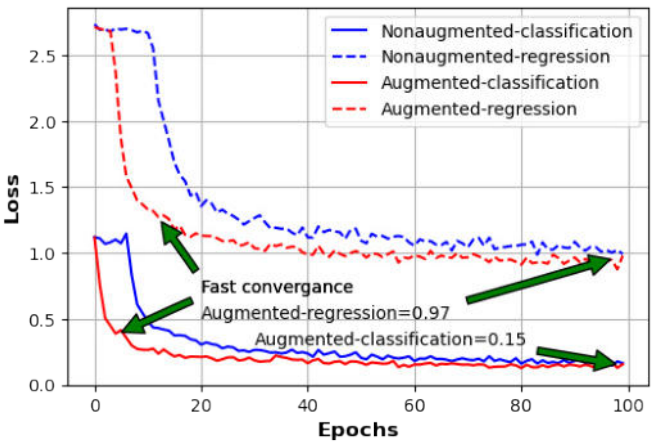

Fig. 4. Classification and regression Loss using dual stream with the augmented fast convergence and non-augmented classification and regression loss has been shown in this figure.

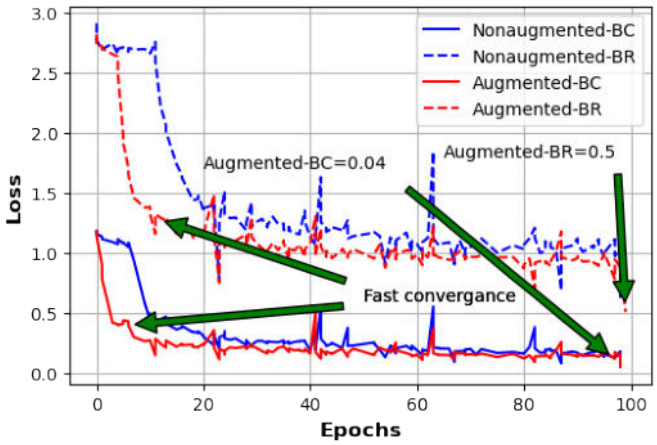

Fig. 5. In this figure the batch classification (BC) and regression (BR) losses for both augmentation and non-augmentation using boundary box re-thinking techniques is shown for the dual stream training.

As shown in Figure 5, we got a mean average precision of 0.658 by using dual streams (day and night vision) of the combination of RGB and IR frames. We got 0.64 mAP with training the Hybrid deep learning model with augmentation. The augmented data training converged very fast as compared to the non-augmented training data. The proposed hybrid deep learning model is trained using ImageNet pre-trained weights. We achieved the training and validation loss to 1.156 and 1.57 before rethinking transformation and 1.13 with 1.5 after boundary box augmentation. Also, faster convergence is a key point to train the model with rethinking transformation. The training and validation loss is depicted in Figure 7, and the validation base losses for the classification, regression with and without batch-wise is given in Figure 8. We achieved validation classification loss of 0.39 and 0.4 for validation data. The validation regression loss is 1.17 and 1.11, respectively, which is acceptable for detecting UAVs in videos with high

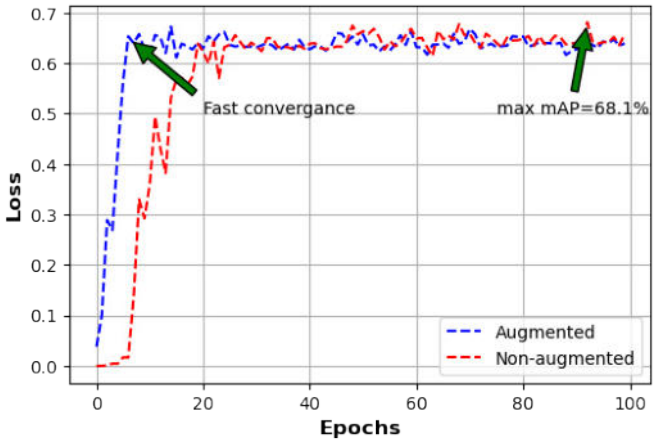

Fig. 6. Mean average precision (mAP) from the training for both streams with and without rethinking transformation and bounding box augmentation.

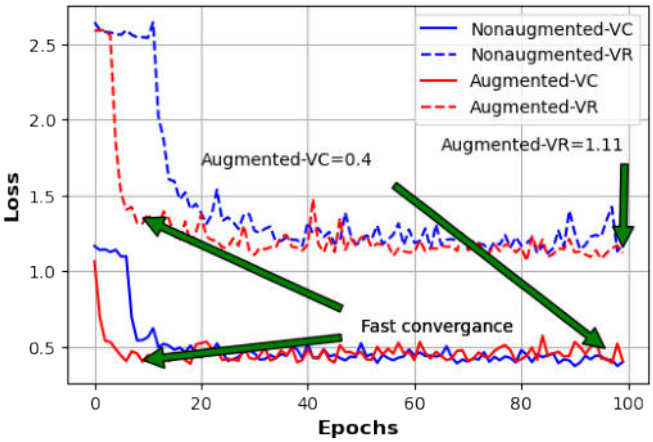

Fig. 8. Validation classification (VC) and regression (VR) losses of proposed model for both augmented data and non-augmented boundary box shown in this figure.

precision.

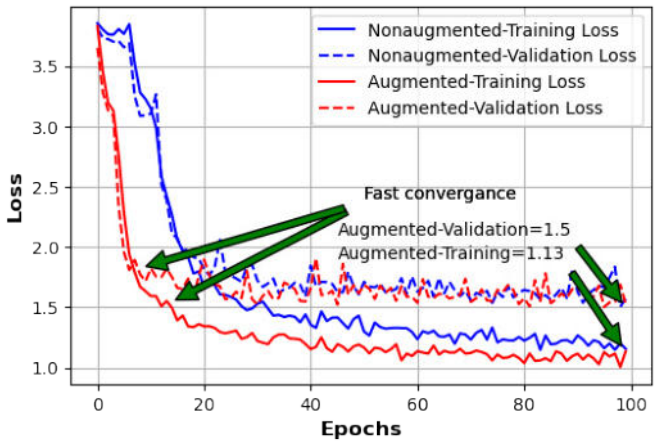

Fig. 7. Training and validation loss of the hybrid deep learning model with dual stream UAV detection using 100 epochs with and without rethinking transformation.

We used three different videos from diverse backgrounds to check the trained hybrid deep learning model on seen and unseen data. The testing videos are from RGB, IR, and local videos captured of the flying UAV. The trained hybrid model performed very well with RGB videos of the UAV Catch dataset. The detected samples of the RGB video are shown in Figure 9. Each frame contains the detected UAV with the correct label box, and precision is very high even with rotation of different positions of the UAV. The UAVs detection in night vision is a challenging problem, we tested the proposed hybrid model for the night vision of the IR stream. Our hybrid model performed best and accurately to identify the UAV and differentiate between UAV and birds; Figure 10 shows the IR-based UAV detection.

Furthermore, It is essential to test on the local data for the real-time application. The proposed hybrid deep learning model is used for locally collected flying UAV datasets and achieved the best accuracy, which is shown in Figure 11.

In last, we are presenting the miss-classification of the defined hybrid model. Our hybrid deep learning model still needs improvement in terms of precision. The hybrid model sometimes misses the UAV when the background and UAV have the same colour. UAV Catch dataset is used to test other objects with similar or similar UAV shapes detected by the proposed hybrid deep learning model. Similarly, It also misinterpreted the new local dataset background objects. The problem can be eliminated by adding similar data for the training because machine learning needs the same distribution dataset for the correct prediction. The samples of the misclassification are available in Figure 12.

V. Conclusion

In this study, we presented a new hybrid deep learning model for the UAV detection based on one stream structure for day-night dual visions. The proposed hybrid model highly improves the accuracy of the UAV detection with the RGB and IR videos. Moreover, rethinking transformation, BBL, and cosine annealing are developed with the proposed hybrid model to enhance the learning convergence and reduce the classification and regression losses. The experimental performance and results show that the proposed hybrid deep learning model achieved 65% higher detection accuracy than the benchmark EfficientDet. In addition, the convergence of the hybrid model is 10% faster than the EfficientDet.

VI. Acknowledgements

This work was partially supported by National Funds through FCT/MCTES (Portuguese Foundation for Science and Technology), within the CISTER Research Unit (UIDP/UIDB/04234/2020); also by national funds through the FCT, under CMU Portugal partnership, within project CMU/TIC/0022/2019 (CRUAV).

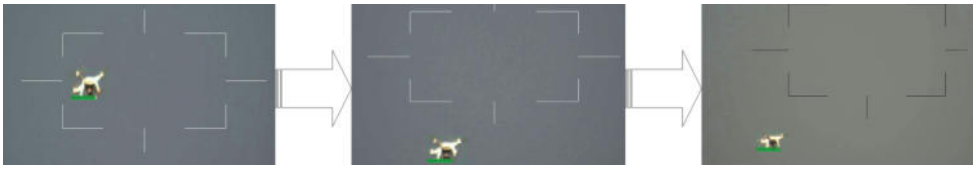

Fig. 9. UAV detection from RGB sample of day vision from test data.

Fig. 10. UAV detection from infrared (IR) videos of night vision samples from test data.

Fig. 11. UAV detection from real-time videos of on-campus samples from test data.

Fig. 12. Miss-classification samples of UAV from different testing videos of RGB, IR videos of both day-night vision of UAV Catch dataset, and on-Campus testing.

REFERENCES

[1] Molchanov, R. I. A. Harmanny, J. J. M. de Wit and K. Egiazarian and andJ. Astola,"Classification of small UAVs and birds by micro-Doppler sig-natures",Int. J. Microw. Wireless Technol., vol. 6, nos. 3–4,pp.435—444,Jun 2014.

[2] Baek. J, Hong. S, Kim. J and Kim. E,"Efficient Pedestrian Detection at Nighttime Using a Thermal Camera",Sensors 2017, vol.17, 1850.

[3] P. Dollár, R. Appel, S. Belongie and P. Perona,"Fast Feature Pyramids for Object Detection",in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 8, pp. 1532–1545, August 2014.

[4] W. Wang, J. Zhang and C. Shen,"Improved human detection and classification in thermal images",2010 IEEE International Conference on Image Processing, Hong Kong, pp. 2313–2316, 2010.

[5] ,Dayan Guan, Yanpeng Cao, Jiangxin Yang, Yanlong Cao and Michael Ying Yang,"Fusion of multispectral data through illumination-aware deep neural networks for pedestrian detection",Information Fusion, Vol.50, pp. 148–157, 2019.

[6] James W. Davis and Vinay Sharma,"Background-subtraction using contour-based fusion of thermal and visible imagery",Computer Vision and Image Understanding, Vol.106, Issues.2–3, pp. 162–182, 2007.

[7] A. Leykin, Y. Ran and R. Hammoud,"Thermal-Visible Video Fusion for Moving Target Tracking and Pedestrian Classification",2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, pp. 1–8, 2007.

[8] Bingwen Chen and Wenwei Wang and Qianqing Qin,"Robust multi-stage approach for the detection of moving target from infrared imagery",Opt. Eng. 51(6) 067006, June 2012.

[9] L. Shi and I. Ahmad and Y. He and K. Chang,"Hidden Markov model based drone sound recognition using MFCC technique in practical noisy environments",in Journal of Communications and Networks, vol. 20, no. 5, pp. 509–518, Oct 2018.

[10] W. Li and D. Zheng and T. Zhao and M. Yang,"An effective approach to pedestrian detection in thermal imagery",2012 8th International Conference on Natural Computation, Chongqing, pp. 325–329, 2012.

[11] M. Teutsch and T. Mueller and M. Huber and J. Beyerer,"Low Resolution Person Detection with a Moving Thermal Infrared Camera by Hot Spot

Classification",2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, pp. 209–216, 2014.

[12] Lowe David.G,"Distinctive Image Features from Scale-Invariant Keypoints",International Journal of Computer Vision 60, pp.91—110, 2004.

[13] N. Dalal and B. Triggs,"Histograms of oriented gradients for human detection,2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, pp. 886–893 vol. 1, 2005.

[14] Herbert Bay and Andreas Ess and Tinne Tuytelaars and Luc Van Gool,"Speeded-Up Robust Features (SURF)",Computer Vision and Image Understanding,Vol.110, pp.346–359,Issue.3,2008.

[15] M. Jahangir and C. Baker,"Robust Detection of Micro-UAS Drones with L-Band 3-D Holographic Rada",2016 Sensor Signal Processing for Defence (SSPD), Edinburgh, pp. 1–5,2016.

[16] M. Jahangir and C. Baker,"Extended dwell Doppler characteristics of birds and micro-UAS at l-band",2017 18th International Radar Symposium (IRS), Prague, pp. 1–10, 2017.

[17] M. Jahangir, C. J. Baker and G. A. Oswald,"Doppler characteristics of micro-drones with L-Band multibeam staring rada",2017 IEEE Radar Conference (RadarConf), Seattle, WA, pp. 1052–1057, 2017.

[18] J. Drozdowicz et al.,"35 GHz FMCW drone detection system",2016 17th International Radar Symposium (IRS), Krakow, pp. 1–4, 2016.

[19] F. Fioranelli and M. Ritchie and H. Griffiths and H. Borrion,"Classification of loaded/unloaded micro-drones using multistatic radar",in Electronics Letters, vol. 51, no. 22, pp. 1813–1815, Sep 2015.

[20] J. A. Nanzer and V. C. Chen,"Microwave interferometric and Doppler radar measurements of a UAV",2017 IEEE Radar Conference (RadarConf), Seattle, WA, pp. 1628–1633, 2017.

[21] M. U. de Haag and C. G. Bartone and M. S. Braasch,"Flight-test evaluation of small form-factor LiDAR and radar sensors for sUAS detect-and-avoid applications",2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, pp. 1–11, 2016.

[22] M. Saqib, S. Daud Khan, N. Sharma and M. Blumenstein, "A study on detecting drones using deep convolutional neural networks",2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Lecce, pp. 1–5, 2017.

[23] S. Al-Emadi, A. Al-Ali, A. Mohammad and A. Al-Ali,"Audio Based Drone Detection and Identification using Deep Learning",2019 15th International Wireless Communications Mobile Computing Conference (IWCMC), Tangier, Morocco, pp. 459–464, 2019.

[24] J. Park, D. H. Kim and Y. S. Shin and S. Lee,"A comparison of convolutional object detectors for real-time drone tracking using a PTZ camera",2017 17th International Conference on Control, Automation and Systems (ICCAS), Jeju, pp. 696–699, 2017.

[25] S. Ren, K. He and R. Girshick and J. Sun,"Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks",in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, June 2017.

[26] R. Girshick,"Fast R-CNN",2015 IEEE International Conference on Computer Vision (ICCV), Santiago, pp. 1440–1448, 2015.

[27] R. Girshick, J. Donahue and T. Darrell and J. Malik,"Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation",2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, pp. 580–587, 2014.

[28] Mingxing Tan and Quoc V. Le,"EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks",Iternational conference on machine learning ICLR, pp. 6105–6114, 2019.

[29] Mingxing Tan and Ruoming Pang and Quoc V. Le,"EfficientDet: Scalable and Efficient Object Detection",arXiv:1911.09070v4.

[30] Jansen, B,"Drone Crash at White House Reveals Security Risks",USA Today. 26 January 2015,https://www.usatoday.com/story/news/2015/01/26/drone-crash-secret-service-faa/22352857/.

[31] Pham, S,"Drone Hits Passenger Plane in Canada", https://money.cnn.com/2017/10/16/technology/drone-passenger-plane-canada/index.html, October 2017.

[32] C. Xu-hui, E. U. Haq, Z. Chengyu,"Efficient Technique to Accelerate Neural Network Training by Freezing Hidden Layers",2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), pp.542–546, 2019.

[33] Huang G, Sun Y, Liu Z, Sedra D. and Weinberger K.Q,"Deep Networks with Stochastic Depth",2016.

[34] A. Rozantsev, V. Lepetit and P. Fua,"Detecting Flying Objects Using a Single Moving Camera",IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.39, no.5, pp.879–892, 2017.

[35] M. Saqib, S. Daud Khan, N. Sharma and M. Blumenstein,"A study on detecting drones using deep convolutional neural networks", 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–5, 2017.

[36] Junkai Peng, Changwen Zheng, Pin Lv, Tianyu Cui, Ye Cheng and Si Lingyu,"Using images rendered by PBRT to train faster R-CNN for UAV detection", 2018.

[37] D. K. Behera and A. Bazil Raj,"Drone Detection and Classification using Deep Learning", 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1012–1016, 2020.

[38] M. Nalamati, A. Kapoor, M. Saqib, N. Sharma and M. Blumenstein,"Drone Detection in Long-Range Surveillance Videos", 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6, 2019.

[39] D. Lee, W. Gyu La and H. Kim,"Drone Detection and Identification System using Artificial Intelligence", 2018 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1131–1133, 2018.

[40] Y. Hu, X. Wu, G. Zheng and X. Liu, "Object Detection of UAV for Anti-UAV Based on Improved YOLO v3", 2019 Chinese Control Conference (CCC), pp. 8386–8390, 2019.

[41] Q. Dong and Q. Zou, "Visual UAV detection method with online feature classification", 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), pp. 429–432, 2017.

[42] Choi, Byunggil and Oh, Daegun, "Classification of drone type using deep convolutional neural networks based on micro-Doppler simulation", 2018 IEEE International Symposium on Antennas and Propagation (ISAP), pp.1–2, 2018.

[43] "UAV Dataset", https://anti-uav.github.io/dataset/,2020,CVPR.

[44] Koksal. Aybora, Ince. Kutalmis Gokalp, Aydin Alatan. A., "Effect of Annotation Errors on Drone Detection with YOLOv3", 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 4439–4447, 2020.

[45] Park, Seongjoon, Kim. Hyeong Tae, Lee. Sangmin, Joo. Hyeontae and Kim. Hwangnam, "Survey on Anti-Drone Systems: Components, Designs, and Challenges", IEEE Access, vol. 9, pp. 42635–42659, 2021.

[46] ,Zhang. H, Cao. C, Xu. L and Gulliver T. Aaron, "A UAV Detection Algorithm Based on an Artificial Neural Network", IEEE Access, vol.6, pp. 24720–24728, 2018.

[47] Kang. H, Joung. J, Kim. J, Kang. J and Cho. Y, "Protect Your Sky: A Survey of Counter Unmanned Aerial Vehicle System", IEEE Access, vol.8, pp.168671–168710, 2020.

[48] Shakhatreh Hazim, Sawalmeh Ahmad H., Al-Fuqaha Ala, Dou Zuochao, Almaita Eyad, Khalil Issa, Othman Noor Shamsiah, Khreishah Abdallah and Guizani Mohsen,"Unmanned Aerial Vehicles (UAVs): A Survey on Civil Applications and Key Research Challenges", IEEE Access, vol.7, pp. 48572–48634, 2019.

[49] He Yihui, Zhu Chenchen, Wang Jianren, Savvides Marios and Zhang Xiangyu, "Bounding Box Regression With Uncertainty for Accurate Object Detection", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June, 4, 2019.

[50] Lin Tsung-Yi, Dollár Piotr, Girshick Ross, He Kaiming, Hariharan Bharath, Belongie Serge, "Feature Pyramid Networks for Object Detection", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017.

[51] Q Hou, M Cheng, X Hu, A Borji, Z Tu and P Torr, "Deeply Supervised Salient Object Detection with Short Connections", in IEEE Transactions on Pattern Analysis and Machine Intelligence, 1 April 2019.

[52] Liu Shu, Qi Lu, Qin Haifang, Shi Jianping and Jia Jiaya, "Path Aggregation Network for Instance Segmentation", 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18-23 June 2018.