Integrating EfficientNet, Cosine Annealing, and Advanced Data Augmentation for Enhanced Aircraft Detection in Satellite Imagery

Lianzheng He [1,*,a], Junxin Wang [2,b] ,Rui Wang [3,c]

[1]School of North China Electric Power University(BaoDing),BaoDing,China
[2,3]School of Beijing Normal University at Zhuhai, Zhuhai, China

[a]helianzheng@foxmail.coma*,
[b]202111079160@mail.bnu.edu.cnb, [c]wangrui0921666@gmail.comc

Abstract. This study introduces a groundbreaking approach to aircraft detection in satellite imagery, featuring an integrated suite of advanced methodologies that include the EfficientNet architecture, sophisticated data augmentation, mixed precision training, and cosine annealing learning rate optimization. We demonstrate how the synergy of these innovations significantly enhances model performance, providing a nuanced solution to the challenges of small target recognition and environmental variability in satellite imagery analysis. EfficientNet, renowned for its balance between computational efficiency and accuracy, is meticulously fine-tuned with a comprehensive set of data augmentation techniques such as RandomResizedCrop and RandomHorizontalFlip, enriching the training dataset and bolstering the model's generalization capacity. The incorporation of mixed precision training facilitates faster computation and reduced memory usage, while the cosine annealing scheduler adeptly modulates the learning rate, fostering improved model convergence and robustness. The empirical outcomes underscore the superior detection capabilities of our model, marked by a high accuracy close to 95%, high precision, recall, and F1 scores, thereby establishing a new standard in satellite-based aircraft detection. This research not only propels the domain of remote sensing forward but also offers a scalable and efficient framework for real-time aerial surveillance and monitoring.

Keywords: EfficientNet;Image Recognition Optimization;Aircraft Detection;Data Augmentation;Mixed Precision Training;Cosine Annealing

1 Introduction

The advent of satellite technology has revolutionized various sectors, with satellite-based aircraft detection emerging as a critical application in national security, air traffic management, and surveillance. This technology, however, faces several challenges, such as the small target size of aircraft in vast landscapes, leading to detection difficulties. Environmental variability further complicates this task, as changes in weather,

T. Yao et al. (eds.), *Proceedings of the 2024 3rd International Conference on Engineering Management and Information Science (EMIS 2024)*, Advances in Computer Science Research 111,
https://doi.org/10.2991/978-94-6463-447-1_25

lighting, and seasonal conditions affect the visibility and recognition of aircraft. Moreover, the issue of class imbalance, where non-aircraft elements vastly outnumber actual aircraft in images, poses a significant challenge in training effective detection models, often leading to biased predictions toward the more frequent classes [1].

To address these challenges, this paper aims to develop a high-accuracy, computationally efficient model for real-time aircraft identification and categorization in satellite imagery. The proposed model leverages the EfficientNet architecture, known for its balance between accuracy and computational efficiency [2]. By integrating advanced data augmentation techniques, such as RandomResizedCrop and RandomHorizontalFlip, the model aims to improve generalization across diverse environmental conditions and mitigate the effects of class imbalance [3]. Mixed precision training methods will be employed to enhance training speed and reduce memory usage, while a cosine annealing learning rate scheduler will be used to optimize the learning process, adjusting the rate in a cyclic pattern to better converge on the model's parameters [4].

2 Methodological Framework

2.1 Comprehensive Data Augmentation for Robustness

To enhance the robustness of the aircraft detection model, a comprehensive data augmentation strategy will be employed, utilizing the following techniques:

1.RandomResizedCrop: The operation is defined by:

$$I' = crop(I, s, r) \tag{1}$$

where I is the original image, I' is the cropped image, s is the scale of the crop relative to the original image, and r is the aspect ratio of the crop [5].

2.RandomHorizontalFlip: This operation flips the image horizontally with a probability p :

$$I' = \begin{cases} flip(I) & with\,probability \\ I & otherwise \end{cases} \tag{2}$$

3.RandomRotation: The image is rotated by an angle θ, randomly chosen from a predefined range:

$$I' = rotate(I, \theta) \tag{3}$$

where θ is typically within $[-\theta_{max}, \theta_{max}]$.

4.ColorJitter: Adjusts the brightness, contrast, saturation, and hue of the image by factors b,c,s and h respectively:

$$I' = jitter(I, b, c, s, h) \tag{4}$$

5.RandomAffine: Applies affine transformations including translation, scale, rotation, and shear:

$$I' = affine(I, t_x, t_y, \theta, s, shear) \tag{5}$$

Where t_x , t_y are translations, θ is the rotation angle, s is the scale, and shear refers to the affine shear parameters.

6.RandomPerspective: Warps the perspective of the image with a distortion scale d :

$$I' = prespective(I, d) \tag{6}$$

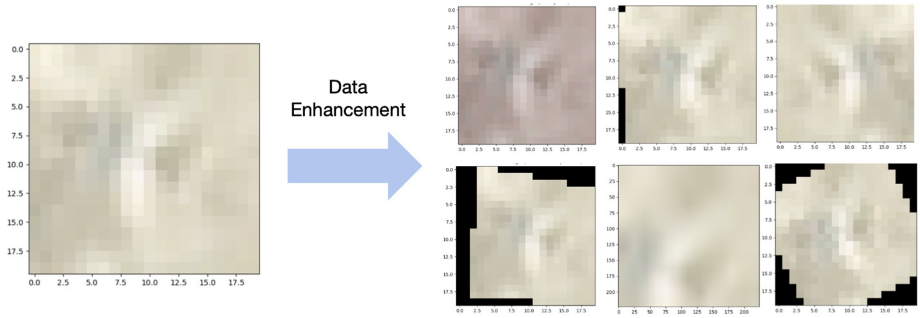

Fig. 1. Six examples of advanced data enhancement

These data augmentation techniques collectively enhance the diversity of the training dataset, enabling the model to learn robust features invariant to scale, orientation, lighting, and perspective (Fig 1). By presenting the model with a broader range of potential scenarios, these methods prevent overfitting and improve the model's ability to generalize to new, unseen satellite images, ultimately lead.

2.2 Mixed Precision Training with Gradient Scaling

Mixed precision training utilizes both single-precision (float32) and half-precision (float16) formats to accelerate training and reduce memory consumption. This process is facilitated by torch.cuda.amp in PyTorch, which includes GradScaler and autocast. The methodology involves:

1.Autocast: Automatically casts variables to float16 where beneficial, preserving float32 where necessary for numerical stability:

$$with\ torch.cuda.amp.autocast(\square): output = mode(I) \tag{7}$$

2.Gradient Scaling: Scales the loss before backpropagation to prevent gradient underflow in float16:

$$L' = GradScaler.scale(L) \tag{8}$$

where L is is the original loss, and L' is the scaled loss.

3.Backward Pass and Step:

$$GradScaler.\mathrm{backward}(L) \tag{9}$$

$$GradScaler.\mathrm{step}(optimizer) \tag{10}$$

$$GradScaler.\mathrm{update}(\square) \tag{11}$$

Mixed precision training, especially with gradient scaling, allows for a reduction in memory usage, enabling larger batches or more complex models to be trained on the same hardware. This approach can significantly accelerate training speed due to faster computation in float16, while maintaining the accuracy provided by float32 calculations where needed, thus optimizing the training process of the aircraft detection model in satellite imagery efficiently.

2.3 Learning Rate Optimization with Cosine Annealing

Cosine Annealing is a learning rate scheduling technique that adjusts the learning rate η following a cosine curve between an upper bound η_{max} and a lower bound η_{max} The learning rate at epoch t is computed as:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})\left(1 + \cos\left(\frac{t}{T_{max}}\pi\right)\right) \tag{12}$$

where T_{max} is the maximum number of epochs. This method allows for large learning rates initially for rapid descent towards minima, followed by smaller learning rates to fine-tune the approach to the minima, simulating a restart mechanism.

Cosine Annealing helps in avoiding local minima and ensures more robust convergence in training deep neural networks. For satellite-based aircraft detection, this adaptive learning rate strategy fine-tunes the EfficientNet-based model to effectively learn from diverse and complex datasets, improving generalization and reducing the risk of overfitting. By dynamically adjusting the learning rate, Cosine Annealing aids in stabilizing and optimizing the training phase, especially beneficial for nuanced tasks like detecting small objects (aircraft) in large satellite images.

3 Experimental Setup and Evaluation

3.1 Dataset

The "Planes in Satellite Imagery" [6] dataset is employed for this study, consisting of 32,000 20x20 pixel RGB images labeled as 'plane' (8,000 images) and 'no-plane' (24,000 images). These images are derived from Planet satellite imagery, specifically over California airports, and are intended for the development and evaluation of machine learning models capable of detecting aircraft.

The preprocessing steps for this dataset are as follows:

1.Resizing: Images are upscaled to 256x256 pixels to provide a more detailed input for the neural network.

2.Augmentation: To enhance model robustness and generalization, several augmentations are applied:

- RandomResizedCrop to 224x224 pixels, adapting to EfficientNet's input requirements.
- RandomHorizontalFlip, RandomRotation, ColorJitter, RandomAffine, and RandomPerspective are used to introduce variability and simulate different viewing conditions.

3.2 Model Training and Testing

The model of choice is EfficientNet-B0 (Fig 2), known for its balance between efficiency and accuracy. The training setup is structured as follows:

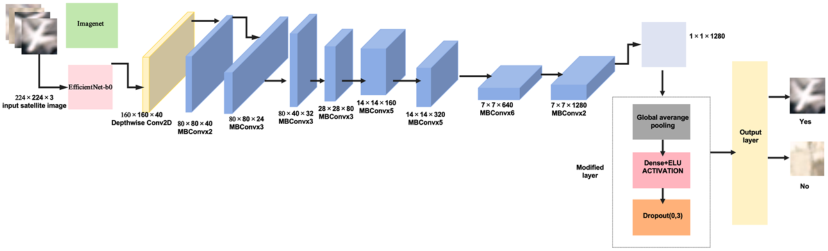

Fig. 2. Image of Improved EfficientNet network architecture

Loading the Model:

- EfficientNet-B0 is initialized with pre-trained weights from ImageNet to leverage transfer learning for feature extraction relevant to aerial image classification.

Training Environment:

- The training is conducted on a GPU-enabled environment to leverage CUDA for accelerated computing, ensuring rapid model iterations and evaluations.

Training Details:

- **Optimizer**: AdamW is used with an initial learning rate of 0.001 and weight decay set at 1e-3, optimizing for both speed and regularization.
- **Loss Function**: CrossEntropyLoss is employed to differentiate between the 'plane' and 'no-plane' classes effectively.
- **Batch Size**: Set to 32, balancing the trade-off between memory usage and speed.
- **Epochs**: The model is trained for 100 epochs, with periodic evaluation to monitor progress and performance.
- **Learning Rate Scheduler**: CosineAnnealingLR is utilized, decreasing the learning rate periodically in a cosine pattern to refine training stability and convergence.

Mixed Precision Training:

- Implemented using torch.cuda.amp for GradScaler and autocast, enhancing the training efficiency by allowing faster computation and reduced memory usage.

Training Loop:

- Each epoch consists of training and validation phases, where model parameters are updated, and performance is assessed, respectively.
- The best model state is saved based on the highest accuracy achieved on the validation set.

Performance Metrics:

- After training, the model is evaluated on the test set, calculating accuracy, precision, recall, and F1 score to assess its classification efficacy.

The parameters and superparameters designed are as follows (Table 1):

Table 1. Statistical table of experimental specific parameters and superparameters

Parameter/Hyperparameter	Value	Parameter/Hyperparameter	Value
Dataset Image Resize	256x256 pixels	Random Resized Crop Size	224x224 pixels
Random Horizontal Flip	Enabled	Random Rotation	45 degrees
Color Jitter (Brightness, Contrast, Saturation, Hue)	0.3, 0.3, 0.3, 0.2	Random Affine (Degrees, Translate, Scale)	20 degrees, (0.1, 0.1) translate, (0.8, 1.2) scale
Random Perspective Distortion Scale	0.3	Normalization Mean	[0.485, 0.456, 0.406]
Normalization Standard Deviation	[0.229, 0.224, 0.225]	Batch Size	32
Number of Epochs	100	Loss Function	CrossEntropyLoss
Optimizer	AdamW	Learning Rate (LR)	0.001
Weight Decay	1e-3	LR Scheduler	CosineAnnealingLR
T_max for CosineAnnealingLR	20	Gradient Scaling (for Mixed Precision)	Enabled with GradScaler

3.3 Results Discussion and Evaluation

In the airplane satellite image recognition project, we employed the improved EfficientNet model for classification (Fig 3), yielding impressive results: an accuracy of 94.94%, precision of 93.22%, recall of 93.39%, and an F1 score of 93.31% (Table 2). These metrics underscore EfficientNet's robust capability in distinguishing airplanes from non-airplanes within satellite imagery. Notably, the high precision and recall rates reflect the model's efficacy in accurately detecting airplanes and minimizing false negatives. The F1 score corroborates the model's consistent and reliable performance.

Table 2. The classification results of the EfficientNet model.

Accuracy	Precision	Recall	F1 Score
0.9494	0.9322	0.9339	0.9331

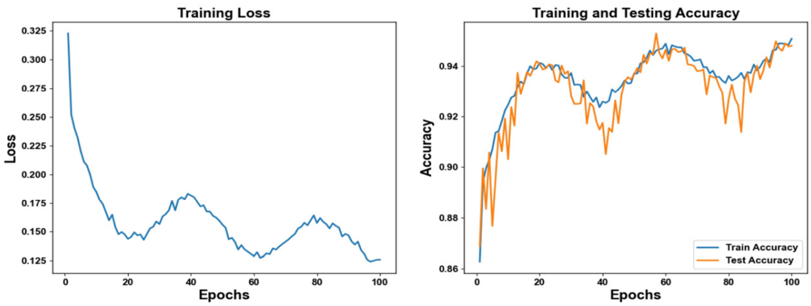

Fig. 3. Loss in training process and the changing image of accuracy

In advancing the field of aircraft detection from satellite images, our study distinguishes itself by utilizing a vast dataset of 32,000 images and incorporating the EfficientNet model, significantly enhancing model robustness and accuracy with an mAP of 0.9494. In contrast, previous efforts, such as Chen et al. (2013) utilizing Deep Belief Networks (DBN) achieved an mAP of 0.7954 [7], and Wu et al. (2015) applying Convolutional Neural Networks (CNN) reached 0.8425 [8]. Zhao & Ren (2019) improved detection with YOLOv3, securing an mAP of 0.93 [9], while Maher, Gu, & Zhang (2018) combined YOLO with the Deep-patch Orientation Network (DON) for an mAP of 0.85 [10] (Table 3). These studies, though innovative, were limited by smaller, less diverse datasets and methods that offered narrower applicability. Our approach not only addresses these limitations by employing a broader dataset but also sets a new benchmark by being the first to apply EfficientNet in this context, thus achieving superior accuracy and model generalization across complex scenarios.

Table 3. Overview of various aircraft detection approaches.

	Author	Sample size	Resolution	Source	Technique	mAP
1	Chen	25	-	Google	DBN	0.7954
2	Wu	26	565*369	earth	BING+CNN	0.8425
3	Zhao	350	600*600	Google	YOLOv3	0.93
4	Ali	600	511*511	earth	YOLO+DON	0.85
5	Our	32000	20*20	Google earth UCAS-AOD Kaggle	EfficientNet	0.9494

4 Conclusions

Our investigation has elucidated the efficacy of the EfficientNet architecture within the domain of satellite-based aircraft detection, achieving notable metrics in accuracy, precision, recall, and F1 score. The employment of a vast and heterogeneous dataset, combined with this sophisticated convolutional neural network, has culminated in a model with enhanced generalizability and resilience against diverse imaging conditions. This advancement in methodological approach signifies a substantial progression in the analytical capabilities applied to satellite imagery for aircraft recognition.

Future endeavors will concentrate on the exploration of cutting-edge neural network architectures that promise further enhancements in computational efficiency and analytical precision. Delving into more intricate data augmentation methodologies will also be pivotal, aiming to simulate an expansive array of operational scenarios to amplify the model's adaptability and performance consistency across a spectrum of aerial surveillance challenges.

References

1. Vergara, V. M., Wojcik, J. J., Kain, E. T., & Lovelly, T. M. (2023). Comparing Deep Learning Performance for Aircraft Detection in Satellite Imagery. NAECON 2023 - IEEE National Aerospace and Electronics Conference. Comparing Deep Learning Performance for Aircraft Detection in Satellite Imagery
2. You, H., Lu, Y., & Tang, H. (2023). Plant Disease Classification and Adversarial Attack Using SimAM-EfficientNet and GP-MI-FGSM. Sustainability. Plant Disease Classification and Adversarial Attack Using SimAM-EfficientNet and GP-MI-FGSM
3. Dai, H., Liu, Z., Liao, W., Huang, X., Wu, Z., Zhao, L., Liu, W., Zhu, D., Cai, H., Li, Q., Shen, D., Liu, T., & Li, X. (2023). ChatAug: Leveraging ChatGPT for Text Data Augmentation. ArXiv. ChatAug: Leveraging ChatGPT for Text Data Augmentation
4. Wang, Z., Hou, F., Qiu, Y., Ma, Z., Singh, S., & Wang, R. (2022). CyclicAugment: Speech Data Random Augmentation with Cosine Annealing Scheduler for Automatic Speech Recognition. Interspeech. CyclicAugment: Speech Data Random Augmentation with Cosine Annealing Scheduler for Automatic Speech Recognition
5. Zhang, J., Meng, Y., & Chen, Z. (2021). A Small Target Detection Method Based on Deep Learning With Considerate Feature and Effectively Expanded Sample Size. IEEE Access. A Small Target Detection Method
6. Planes in Satellite Imagery. (2018). Detect aircraft in Planet satellite image chips. Retrieved from https://www.kaggle.com/datasets/rhammell/planesnet
7. Chen, X., Xiang, S., Liu, C. L., & Pan, C. H. (2013, November). Aircraft detection by deep belief nets. In 2013 2nd IAPR Asian Conference on Pattern Recognition (pp. 54-58). IEEE.
8. Wu, H., Zhang, H., Zhang, J., & Xu, F. (2015, September). Fast aircraft detection in satellite images based on convolutional neural networks. In 2015 IEEE International Conference on Image Processing (ICIP) (pp. 4210-4214). IEEE.
9. Zhao, K., & Ren, X. (2019, May). Small aircraft detection in remote sensing images based on YOLOv3. In IOP Conference Series: Materials Science and Engineering (Vol. 533, No. 1, p. 012056). IOP Publishing.

10. Maher, A., Gu, J., & Zhang, B. (2018). Deep-patch orientation network for aircraft detection in aerial images. In Advances in Image and Graphics Technologies: 12th Chinese conference, IGTA 2017, Beijing, China, June 30–July 1, 2017, Revised Selected Papers 12 (pp. 178-188). Springer Singapore.