Dean&Francis

ISSN 2959-6157

Analyzing the Performance of Machine Learning Models on Aircraft Classification Tasks across Various Conditions

Yuchen Liu

Institute of Electronic Science and Engineering, Southeast University, Nanjing, Jiangsu, China
213222313@seu.edu.com

Abstract:

As more and more aircraft are used, using an efficient and accurate tool to identify aircraft types will help manage air safety. As a basic framework for machine learning, a Convolutional neural network is widely used to make image recognition. This paper presents a method to classify different types of military and civilian aircraft using neural networks and explore the performance differences by changing batch size and learning rate. Five models are selected to achieve classification for ten types of aircraft images from the Fine-Grained Visual Classification of Aircraft (FGVC)-Aircraft dataset. Except for the existing models like Visual Geometry Group 16-layer Network (VGG16), Residual Network with 152 Layers (Resnet152), Densely Connected Convolutional Networks (DenseNet), and Efficient Neural Network (EfficientNet), a customized Convolutional Neural Network (CNN) is also designed to complete the image classification task. The parameters used to demonstrate performance are latency, throughput, model convergence speed, accuracy, f1-score, and confusion matrix. By analyzing the results, the customized CNN performs best among these five models, having both good accuracy and latency. VGG16 and DenseNet perform well although each has certain deficiencies in terms of latency and accuracy respectively. The performance of EfficientNet and Resnet152 is not good, and their accuracy is relatively lower compared to other models. Therefore, the best model can be obtained for solving this task and make predictions on the performance of each model when classifying pictures of more types of aircraft.

Keywords: Machine learning; aircraft image classification; computer vision.

1. Introduction

Nowadays, the aircrafts are widely used in numerous fields. Classifying military and civilian aircraft is essential for the airspace safety and airports, which means estimate the danger of the invalid military aircraft or just help to calculate how many flights and what kinds of planes arrive at a certain airport during some period. Machine learning, as a powerful method, can automatically recognize the type of aircraft, which can be considered in this situation to improve classification efficiency significantly.

There has been a lot of research that analyzes the performance of different models on similar tasks. For instance, Alkharji et al. use Resnet50, VGG, and Convolutional Neural Network (CNN) to judge whether the aircraft is a military plane or a civil plane [1]. While Edhah et al. utilize the advantages of CNNs [2], this work presents a comprehensive tool to detect and analyze military aircraft logos. Besides, in [3], they design a high-accuracy convolutional neural network to detect and classify aircraft logos as either adequate or inadequate based on specified criteria. These researches all give many methods to use machine learning and artificial intelligence models to identify aircraft, but there are several deficiencies in how the batch size and learning rate affect the performance of these models. Therefore, this research will focus on the performance difference between five models: Customized CNN, Resnet152, VGG16, EfficientNet, and DenseNet in aircraft image classification, including accuracy, f1-score, latency, and throughput. All these parameters will be compared to finding out the most suitable model for this task.

Several kinds of images of military and civilian aircraft are chosen to be the dataset in this study. After the pre-processing of the dataset, these pre-trained models (except customized CNN) are able to predict the aircraft's brand and variant of the aircraft in the image, and they can be compared with the original labels to describe the accuracy of the prediction. In this research, to make a trade-off between execution speed and accuracy, the best batch size and learning rate of all these five models should be found. To be noted, EfficientNet is the most efficient model among these five models, it also has moderate accuracy when the batch size is more than 32. Customized CNN and DenseNet have less latency. While Resnet152 and VGG16 have very large latency, their accuracy is the highest, and the convergence speed of these two models is also the highest among these five models. So, the best-matched model to deal with this task can be chosen according to different requirements of latency, throughput, and accuracy.

2. Method

2.1 Dataset Preparation

This study chooses the FGVC-Aircraft as the dataset, which contains 10, 000 images of 100 kinds of aircraft [4, 5]. The download site is listed in reference [5]. In this study, 10 categories of aircraft are chosen: Cessna 172, BAE-125, DR-400, Boeing 747, SR-20, F-16, Tornado, Spitfire, Il-76, and Hawk T1, including both civilian and military planes. In the pre-processing part, this study randomly divided the dataset into two equal-sized train and test datasets. The images are resized so that the shorter side can be converted to 128 pixels with an original aspect ratio. Then a 128´128 pixels patch is cropped from the center, and the RGB channels are normalized to values from 0-255 to 0-1, helping training stability and convergence. After that, these processed images are sent to the models for further operation. Some sample images are provided in Fig. 1.

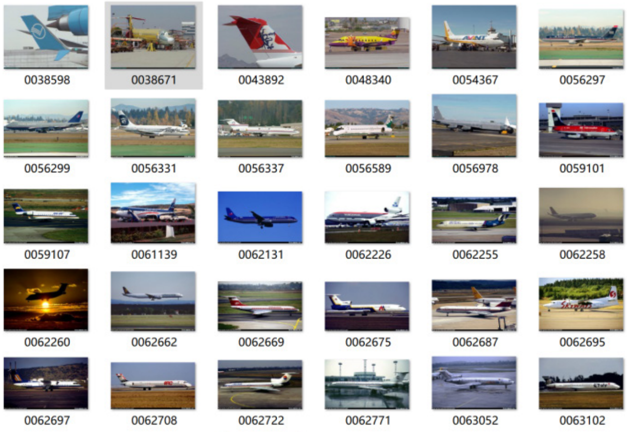

Fig. 1 FGVC-Aircraft dataset [5].

2.2 CNN-based Classification Models

In this study, all models used are transformed from CNN, which is the most popular model to achieve image recognition in machine learning. These models will be introduced in the following sections.

2.2.1 Customized CNN

In this work, a customized CNN is designed, consisting of four successive down-conv blocks each followed by a 2´2 max-pooling layer. This is followed by an adaptive max-pooling operation that reduces the feature map to a fixed 12´12 spatial size, and a fully connected layer. Firstly, the 3-channel RGB input is converted to 32 channels, pooled to half spatial dimensions, and then it is convolved to 64, 128, and 256 channels in the same way. At last, it passed through a linear layer to produce 10-class

ISSN 2959-6157

scores. The total number of learnable parameters is approximately 757,000.

2.2.2 DenseNet

Densenet is a convolution neural network architecture introduced by Gao Huang et al. in 2016 [6], which establishes L(L+1)/2 direct connections between layers so that each layer receives the feature-map outputs of all preceding layers and passes on its own outputs to all subsequent layers, thereby alleviating the vanishing-gradient problem, strengthening feature propagation and reuse, and substantially reducing parameter count. In this work, DenseNet-121 contains roughly 7.98 million parameters, and the classifier adds 10,250 parameters, for a total of about 7,990,250 learnable parameters.

2.2.3 EfficientNet

EfficientNet is a family of convolutional neural networks proposed by Google AI in 2019 [7]. The compound scaling method that uniformly scales network depth, width, and resolution with a single coefficient ϕ and employs neural architecture search to design efficient MB Conv blocks, yielding models from EfficientNet-B0 to B7 that achieve superior accuracy and efficiency on ImageNet and have been widely adopted for image classification, object detection, and segmentation. This study comprises about 4,049,000 parameters after removing the original 1,281,000-parameter head, and the new linear layer adds 12,810 parameters, yielding roughly 4,061,810 learnable parameters in total.

2.2.4 Resnet-152

Resnet-152 is a 152-layer variant of the residual network family introduced by Kaiming He, et al. in 2015 [8]. Bottleneck residual blocks are the key part of Resnet—each block first applies a 1×1 convolution for dimensionality reduction, then a 3×3 convolution for feature extraction, and finally a 1×1 convolution for dimensionality restoration, while a “skip connection” adds the block’s input directly to its output. This design both alleviates the vanishing-gradient and degradation problems that arise in very deep networks and enables lossless information propagation via identity mappings, resulting in more stable training and superhuman accuracy on large-scale image classification benchmarks. In this study, the final classifier leads into 2048×10 + 10 = 20,490 new parameters. In total, ResNet152 has about 58,143,808 + 20,490= 58,164,298 parameters, compared to the original 60.34 M.

2.2.5 VGG16

VGG16 is a classic deep convolutional neural network introduced by Karen Simonyan and Andrew Zisserman in 2014 [9], consisting of thirteen 3×3 convolutional layers and three fully connected layers, using small-sized filters throughout and deep stacking to increase network depth while keeping per-layer parameter counts low to manage overall model size. The three fully connected layers of sizes 4096, 4096, and 1000, and a final Softmax classifier for 10 ImageNet classes. The model comprises approximately 138 million parameters in total.

2.3 Implementation Details

In this study, pytorch is chosen to be the deep learning framework [10]. The optimizer is SGD and the loss function is Cross Entropy Loss, all imported from the torch. The default learning rate is 0.0005, and 50 for the default number of iterations, and 128 for default batch sizes. To evaluate the model performance under various conditions, accuracy, f1-score, latency, and throughput are recorded for comparison.

3. Results and Discussion

The following Table 1, Table 2, Table 3, Table 4 and Table 5 shows the results of these five models, including the latency, throughput, accuracy on the test dataset, f1-score, and number of incorrect predictions in the confusion matrix. The performance differences can be obviously seen in this chart. The five situations in the table are: (1) batch size is 128, learning rate is 0.001 (2) batch size is 64, learning rate is 0.001 (3) batch size is 32, learning rate is 0.001 (4) batch size is 32, learning rate is 0.0005 (5) batch size is 32, learning rate is 0.0001.

Table 1. CNN performance under different conditions

Parameters	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
Latency(ms/image)	4.80	4.53	5.54	5.65	5.65
Throughput(image/s)	208.26	224.71	180.45	177.11	176.91
Accuracy(percentage)	100	100	100	100	100
F1-score(percentage)	100	100	100	100	100
Wrong number in the confusion matrix	0	0	0	0	0
Number of iterations to 100% accuracy	13	13	46	12	5
Number of iterations to 100% f1-score	12	9	32	8	4

Table 2. VGG16 performance under different conditions

Parameters	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
Latency(ms/image)	25.2	17.78	24.33	24.28	24.11
Throughput(image/s)	208.26	224.71	180.45	41.21	41.49
Accuracy(percentage)	100	100	100	100	100
F1-score(percentage)	100	100	100	100	100
Wrong number in the confusion matrix	0	0	0	0	0
Number of iterations to 100% accuracy	9	6	9	5	6
Number of iterations to 100% f1-score	8	6	7	5	7

Table 3. ResNet performance under different conditions

Parameters	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
Latency(ms/image)	18.12	18.03	24.2	24.28	20.76
Throughput(image/s)	56.2	55.46	41.33	41.23	48.53
Accuracy(percentage)	90.1	97.0	99.6	96.0	80.0
F1-score(percentage)	94.5	97.6	100	94.2	76.4
Wrong number in the confusion matrix	25	15	0	12	79
Number of iterations to 100% f1-score	/	/	16	/	/
Latency(ms/image)	18.12	18.03	24.2	24.28	20.76

Table 4. DenseNet performance under different conditions

Parameters	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
Latency(ms/image)	8.40	7.45	10.06	9.99	10.03
Throughput(image/s)	119.13	134.41	99.5	100.18	99.75
Accuracy(percentage)	98.2	99.8	100	98.5	89.6
F1-score(percentage)	98.8	100	100	99.8	88.5
Wrong number in the confusion matrix	5	0	0	7	43
Number of iterations to 100% accuracy	/	/	13	/	/
Number of iterations to 100% f1-score	/	42	5	/	/

ISSN 2959-6157

Table 5. EfficientNet performance under different conditions

Parameters	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
	(1)	(2)	(3)	(4)	(5)
Latency(ms/image)	2.96	2.86	3.58	3.47	3.52
Throughput(image/s)	338.19	351.50	280.00	288.21	284.38
Accuracy(percentage)	83.0	79.2	92.0	99.8	46.6
F1-score(percentage)	87.8	83.2	97.5	94.2	46.6
Wrong number in the confusion matrix	55	27	12	40	202
	(1)	(2)	(3)	(4)	(5)

Among them, latency and throughput are the average values of 50 iterations. As parameters for evaluating the degree of accuracy, accuracy, and f1-score will become 100% after a certain number of iterations for some models that converge very fast. Therefore, the table also gives the number of iterations when accuracy and f1-score start to be stable at 100 percent.

It can be seen that from Table 1-5, EfficientNet has the smallest latency among the five models under any conditions, and correspondingly, the throughput is also the largest. Customized CNN and DenseNet also perform well in this regard. ResNet and VGG are slower when processing images. It is noted that, although EfficientNet is very fast, its accuracy is not very ideal, and prediction errors often occur. Customized CNN has advantages of both high accuracy and low latency. Similarly, DenseNet also has both advantages when the batch size is small. Although VGG has the largest latency among all models, its accuracy and convergence speed are the fastest, making it the most reliable model among these five. Resnet performed poorly in several tests, with low accuracy and latency second only to VGG. It is worth mentioning that Resnet can still maintain good accuracy when the number of aircraft types to be classified is relatively small, indicating that Resnet may not be suitable for multi-class classification.

In general, the increase in batch size means that the model can deal with more images at the same time, but the impact of intermediate results must be considered. In this study, all these five models have the smallest latency and the biggest throughput when the batch size is 64. Compared to 32 and 128, this batch size makes the number of images most suitable for GPU to accelerate the machine learning rate. As for the accuracy and f1-score, they increase as the number of batch sizes decreases. At the same time, the decrease in the learning rate has little impact on latency and throughput. Furthermore, except for customized CNN and EfficientNet, all the models perform even worse than the condition that the learning rate is 0.0001. So, a rule that applies to all five models is, that when the batch size is 64 and the learning rate is 0.0001, in the comprehensive consideration of accuracy, latency, and throughput, the performance is the best.

Since this study is to classify 10 types of aircraft, if these tasks need to classify more types of aircraft, the classification task will be more complicated, so the impact of accuracy is of great importance. In these tested models, it is seen that the accuracy of customized CNN and VGG16 is very high. It is believed that they will be applicable to the recognition process of more types of aircraft. DenseNet and Resnet may also be considered.

4. Conclusion

This study analyzed the performance of five different machine learning models when they are used to classify ten types of military and civil aircraft images, exploring the feasibility of using machine learning models for aircraft image classification. Considering latency and accuracy comprehensively, the customized CNN shown in this research is the best model to classify the types of aircraft given in this article, while VGG16 and DenseNet also have good performance in terms of accuracy and latency respectively. The future study will continue to explore the performance of more models and evaluate how these models perform with more kinds of aircraft images.

References

[1] Alkharji N, Almazrouei H, Alzaabi S, Nassif AB, Elsalhy M, Talib MA. Aircraft-type classification using deep learning algorithms. In2024 IEEE 12th International Conference on Intelligent Systems (IS) 2024 Aug 29 (pp. 1-6). IEEE.

[2] Li Y, Zhang S, Zhao J, Tan W. Aircraft detection in remote sensing images based on deep convolutional neural network. In2017 International Conference on Computer Technology, Electronics and Communication (ICCTEC) 2017 Dec 19 (pp. 942-945). IEEE.

[3] Edhah S, Awadallah A, Madboly M, Dawed H, Werghi

N. Image Classification and Text Identification in Inspecting Military Aircrafts Logos: Application of Convolutional Neural Network. In2022 IEEE International Symposium on Robotic and Sensors Environments (ROSE) 2022 Nov 14 (pp. 01-06). IEEE.
[4] Maji S, Rahtu E, Kannala J, Blaschko M, Vedaldi A. Fine-grained visual classification of aircraft. arXiv preprint arXiv:1306.5151. 2013 Jun 21.
[5] Kaggle. FGVC Aircraft. https://www.kaggle.com/datasets/seryouxblaster764/fgvc-aircraft. 2020.
[6] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4700-4708).
[7] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. InInternational conference on machine learning 2019 May 24 (pp. 6105-6114). PMLR.
[8] Zhang L, Li H, Zhu R, Du P. An infrared and visible image fusion algorithm based on ResNet-152. Multimedia Tools and Applications. 2022 Mar;81(7):9277-87.
[9] Theckedath D, Sedamkar RR. Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks. SN Computer Science. 2020 Mar;1(2):79.
[10] Ketkar N, Moolayil J, Ketkar N, Moolayil J. Introduction to pytorch. Deep learning with python: learn best practices of deep learning models with PyTorch. 2021:27-91.