Application and analysis of landscape recognition based on efficient net for natural scene

Sisheng Jin

Department of Science and Engineering, University of Nottingham Ningbo China, Ningbo, 315100, China

scysj4@nottingham.edu.cn

Abstract. One significant assessing criteria of climate change is geometric evolution. The rate of evolution reveals the speed that environment worsens. Advanced space mirror monitors that and generates images timely. However, it might be difficult for human to deal with collected numerous image-related data. In previous research, convolutional neural network is regarded to have specific advantage in resolving image recognition tasks. Hence, a new type of convolutional neural network model is applied to identify different kinds of landscape. Virtually, this model is called Efficient Net which based on landscape recognition dataset with 5 classes of landscapes. The study also introduces the fine-tuning to further improve the performance of the model. To evaluate the model, the precision, recall, F1 score, accuracy and loss are adopted as assessing criteria. The results shows that the model predicts the target dataset to a great extent. However, it has been tested that the class of mountain might not be suitable for predicting because of vague criterion. That is helpful in real-condition geographical applications and environmental governance.

Keywords: Landscape Recognition, Efficient Net, Fine-tuning, Environmental Governance.

1. Introduction

There is an increasing concern in the fields regarding climate change especially for landscape evolution since 1980s [1]. Due to frequent human activities, the reaction of weather became more dynamic and unstable [1]. That was followed by the higher or lower probability of rainfall, which results in the unpredictable change of local landform. According to a study conducted by UN, dry land occupied an increased portion of 41.3% in whole land area by 2022 [2]. To mitigate the negative implications of abnormal weather change, quantities of remote-sensing images are collected daily for being analysed [3]. That had effects on the process of monitoring the concrete landform erosion. However, it seems that technology and methods that most scientists have invented and raised are not satisfied for the demand of practical use [3].

With the characteristic advantage of dealing with numerous image-related data, convolution neural network has shown the ability of addressing those problems [4]. For instance, the development of image recognition has provided practical experience and technology support for urban and rural cultural heritage [5]. Hence, supervised deep-learning had been widely used in remote-sensing areas such as geospatial target detection [3]. Cowls claimed that efficiency to train an artificial model to a relatively high level had improved by nearly 44 times compared to the beginning [6]. That provided more

probability for resolving great numbers of image-recognition targets. For instance, it seems easier for supervisors to make appropriate decisions about resource management and natural hazard assessment [7]. In addition to this, the utilization of transfer learning reinforces the structure of convolution neural network, which has been identified in categorizing images [8].

The purpose of this study is to introduce a convolutional neural network to construct a model to recognize given landscape images. Specifically, a convolutional neural network model called EfficientNet is used in this paper to accomplish this task. EfficientNet is a network structure based on automated model scaling. Compared with ordinary convolutional neural network models, EfficientNet has higher accuracy and smaller number of params [9]. In addition, the adopted method can effectively improve the efficiency, scalability and robustness of the model. Eventually, a great number of geometric images are assigned to specific categories indicating the effectiveness of the proposed model. The predicted results may contribute to better environmental management measures and facilitate the resolution of environmental crises, especially geohazards. In further applications, it can also be predicted that Efficient Net may have an impact on inferring local geomorphology using integrated and multiple views of the local geologic environment.

2. Methodology

2.1. Dataset description and preprocessing

The Landscape Recognition dataset from Kaggle contains 12,000 jpeg-format images which are categorized to 5 classes [10]. These classes are Coast, Desert, Forest, Glacier, Mountain respectively. Each of the image has been tagged by consistent label. The whole image files have been divided into 3 sub-datasets that are training set, validation set and testing set with a ratio of 20:3:1. The address of directory are obtained to read image files. These images have been resized to 224 x 224 pixels figure for easier processing. A dataset object is built to store the image data with their transformed int labels. Next, an EfficientNetB0 model is introduced with pre-trained weighting functions. It could be noticed that the collection of data within training set and validation set has been given the order of shuffle. That contributes to preventing the overfitting by randomly distributing same-type pictures while testing the model.

2.2. Proposed approach

This study's fundamental goal is to employ Efficient Net model to precisely predict the landscape of the given images. Besides, the different operation is also expanded to original EfficientNetB0 model to improve the accuracy. First, EfficientNetB0 is built up after importing the data. Second, As shown in Figure 1, the initial Efficient Net model is trained through the image data in training set. Meanwhile, the appropriate hyper-parameters and related function such as loss function is selected. This step is followed by the evaluation of model after training and testing the given image data. That process is plotted through using computer-vision libraries. Next, the image data is changed stochastically for better generalization. For instance, the image is rotated by an angle randomly within a given domain. Ultimately, fine-tuning is used to change hyper-parameters and other defaults for further improvement.
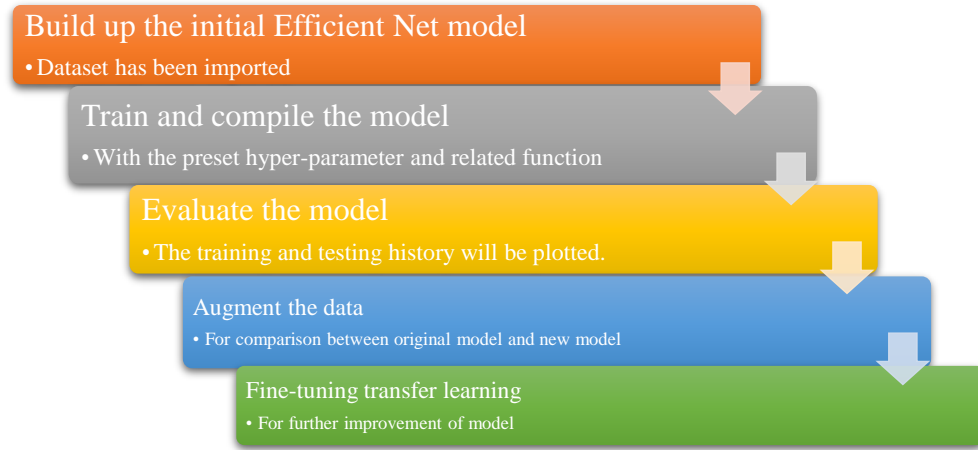
Figure 1. Flow Chart Graph.

2.2.1. EfficientNet. The EfficientNetB0 is the basic model of a series of Efficient net models. Compared to classic convolutional neural network (CNN) model, it has deeper convolution layers, lower stride and smaller scale of convolution kernels while the number of that have increased. Meanwhile, it also results in the increase of nonlinear factors combined with RELU function. All these improvements are highly likely to better the behaviour of model in accomplishing classification targets. However, it seems that blindly resetting only one of them is not useful in advancing model. To effectively utilize the params in the network, the author has used a formula to constrain the relationship between width, depth, resolution ratio of pictures. That is:

$$\alpha^{\Phi} \cdot \beta^{\Phi} \cdot \gamma^{\Phi} = 2 \tag{1}$$

The sum amount of computing increases to:

$$(\alpha \cdot \beta^2 \cdot \gamma^2)^{\Phi} \tag{2}$$

As shown above, φ represents combined scaling factor. α, β and γ represent the scaling cardinal number of width, depth, resolution ratio of pictures severally. If these cardinal numbers uniformly become φ times as big as original one, the sum amount of computing becomes 2 to the φ-th power compared to original amount accordingly. In EfficientNetB0, the φ is set to 1. Thus, the α, β, γ is set to 1.2, 1.1, 1.15 accordingly after NAS. In addition to this, the model consists of over fifty convolution layer and a full-connected layer. It could be indicated that the feature of image data has been extracted and expanded repeatedly. This study first applies EfficientNetB0 to landscape recognition dataset. Furthermore, data augmentation and fine-tuning are employed to improve the accuracy of model.

2.2.2. Evaluation Metrics and Visualization Tools. Evaluation of this model includes metrics about accuracy and tools to visualize the effectiveness. By applying the following metrics and visualization, the ability of model in predicting the given images is clearly and easily displayed.

Confusion Matrix with visualization: The confusion matrix is a visualization tool, which briefly checks the corresponding degree between images and their labels. It outputs the number and proportion relationship between prediction and true labels in samples. Besides, it also provides numerical ratio normalized over true values or predicted values. Accuracy with visualization: Accuracy is a significant criterion of assessing model in recognizing multi-class targets. The proportion of correctly identified samples to all samples is clearly shown by accuracy. The formula to compute accuracy is:

$$Accuracy = \frac{Number\ of\ correctly\ predicted\ cases}{Total\ number\ of\ cases} \tag{3}$$

Recall: It represents the ratio of the number of samples that the model correctly predicts as positive cases to the number of samples that are truly positive cases. It shows the capability of model in grabbing real positive cases, which is visualized with bar graphs in different landform classes. The formula is that:

$$Recall = \frac{Number\ of\ truly\ predicted\ positive\ cases}{Total\ number\ of\ positive\ cases} \tag{4}$$

Precision: It represents the ratio of the number of samples that the model correctly predicts as positive cases to the number of samples that are all predicted as positive cases. The formula is that:

$$Precision = \frac{Number\ of\ truly\ predicted\ positive\ cases}{Total\ number\ of\ predicted\ postive\ cases} \tag{5}$$

F1 score: It represents the harmonic mean of precision and recall, which is useful in comprehensively understanding the performance of model in predicting both positive cases and negative cases. High F1 score means that there is a good balance between precision and recall of model. The formula is that:

$$F1\ score = \frac{2*Precision*Recall}{Precision+Recall} \tag{6}$$

2.2.3. Loss function and optimizer. Sparse Categorical Cross-entropy function is oftentimes being employed in computing the cross-entropy of multi-classes problems. That measures the difference between the probability distribution of the model's output label and that of real label. Adaptive Moment estimation (Adam) optimizer combines Momentum with RMSProp, which is widely used in deep learning fields. This optimizer accumulates momentum to provide a shorter training runtime. It also optimizes the amplitude of swing in the process of updating of gradient.

2.2.4. Improvement of model. The performance of model is improved by data augmentation and fine-tuning. It includes data augmentation and fine-tuning. First, if the accuracy (or other criteria) were changed to a great extent after reasonable data augmentation, that means the model used is not appropriate for the given dataset. That is because operations include translation, rotation and rescaling do not change the type of the class of an image file. In this model applying in landform recognition dataset, the given images are processed by a random flip in horizontal direction, a random rotation ranged in 0.2 times as great as maximum and a random rescaling in 0.2 times as great as original scale. The mentioned methods above improve generalization of the model. In this process, a new model is created based on EfficientNetB0 with above data augmentation. This new model catches and saves the best weighting function while training with data augmentation. Besides, the weighting function is be frozen except for that in full connected layer. That avoids the repeat training in same dataset, thus lessening the runtime. Second, in the fine-tuning process, all layers are unfrozen again. In addition to this, the learning rate of optimizer is also set to 10-4 (original is 10-3) to modify weighting function. The training epoch is set to 10 to attain a higher accuracy.

2.3. Implemented details
The study is based on Python 3.10. Pandas and Matplotlib libraries have been employed for computer vision. Moreover, this study is conducted at Windows 10 environment with an 11th generation Intel i7 processor and Intel Iris Xe Graphics card. The EfficientNetB0 has the following configuration: batch size is set to 32, training epoch is set to 5, validation step is set to 0.25 times as great as length of validation dataset, activation function is SoftMax.

3. Result and discussion
This section displays and discusses the outcome of execution. First, the corresponding degree between true value and predicted class is shown by confusion matrix. Second, the value of final precision, recall and F1 score is shown by bar graphs. Third, the change of accuracy and loss of model is displayed. Ultimately, the knowledge of contrast between different models is discussed. The diagonals represent the matched classes, while the other represent confused classes. The left sub-graph is in the form of

number, the right two are in the form of normalized values. As shown in Figure 2, the basic model predicts the class of images to a relatively high level overall. In particular, the model has a high predicting rate in the desert class. That attributes to the distinctive differences between desert and other landscapes. However, the mountain landscape is confused on many samples. It seems that the mountain is difficult to classify because the most significant criterion of mountain is not surface but altitude. In this case, the identification of mountain should combine with other factors.
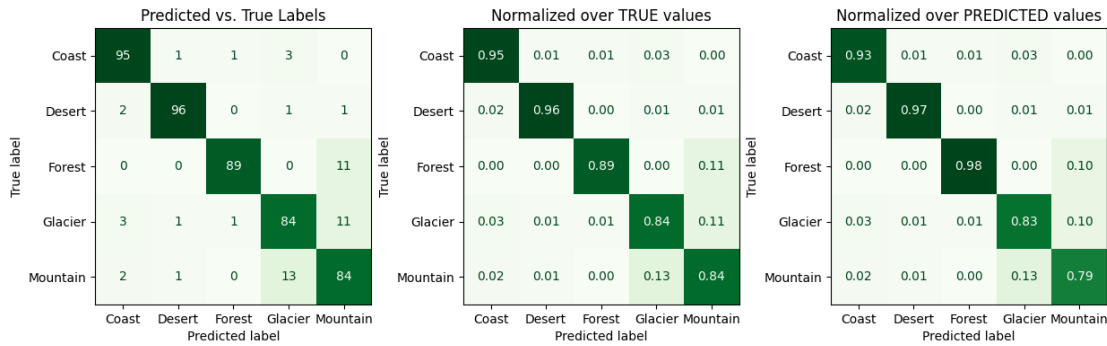

Figure 2. Confusion Matrix of basic EfficientNetB0 in the given training dataset.

As shown in Figure 3, the precision of forest is over 0.950. It could be claimed that other landscapes are rarely mistaken for this category. The recall of 5 classes is all higher than 0.80. This means model has a strong ability to catch positive accuracy samples. For F1 score, the model is conducted best in desert class while worst in mountain. That means there is a good balance in the precision and recall in that of desert class but not in that of mountain class.
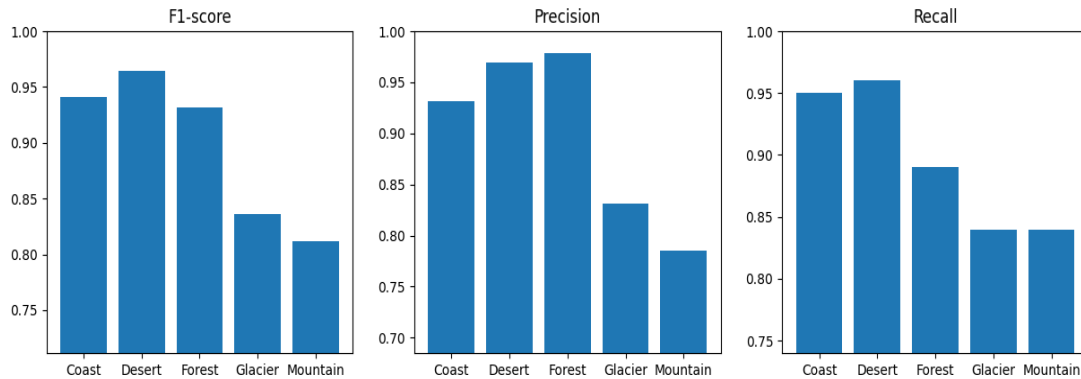

Figure 3. Final precision, recall and F1 score of the basic model (unmodified EfficientNetB0).

The Figure 4 shows that there is a strong overfitting condition. The training accuracy increased rapidly with the process of epochs, and there is a strong descending trend in the change of training loss. However, the validation loss is fluctuated from 0.42 to 0.50 while the validation accuracy is gradually decreasing during the validation process. That means the trained model predicts the landscape class to a high degree. Meanwhile, there also exists overfitting problem in the validation process. It could be attributed to nontypical image files in validation set (too many useless factors), or the number of given image files is not complex and great enough to train the EfficientNetB0 model. That is improved by import of regularization of model, adding greater scale of image data to train the model.
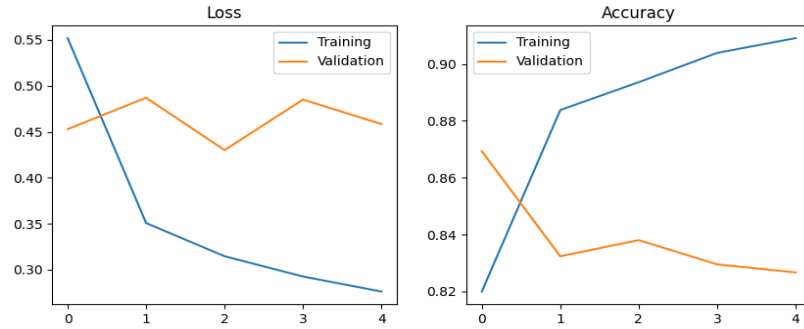
Figure 4. The change of accuracy and loss (unmodified EfficientNetB0) (0 represents the first epoch).

As shown in Figure 5 and 6, after data augmentation and fine-tuning, the difference in predicted rate between given classes has been expanded. The mountain and glacier have lower predicted rate, mountain class among which has been confused most. It seems that mountain is not appropriate for landscape recognition due to its distinctive judgement standard.
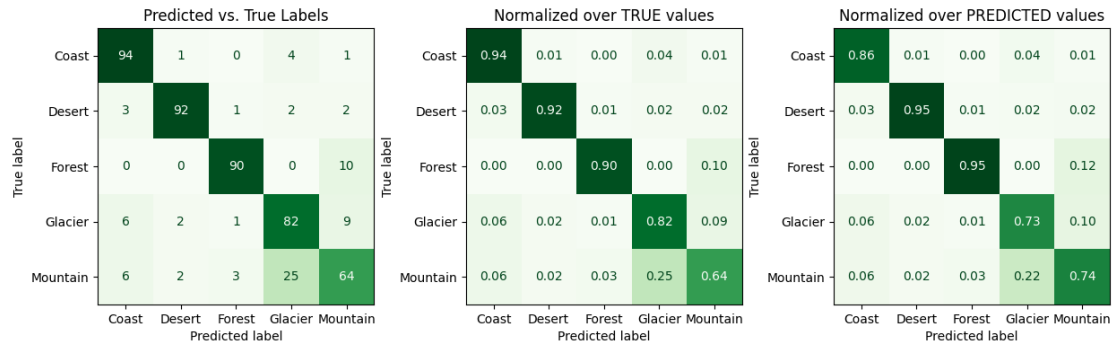

**Figure 5.** Confusion Matrix with data augmentation.
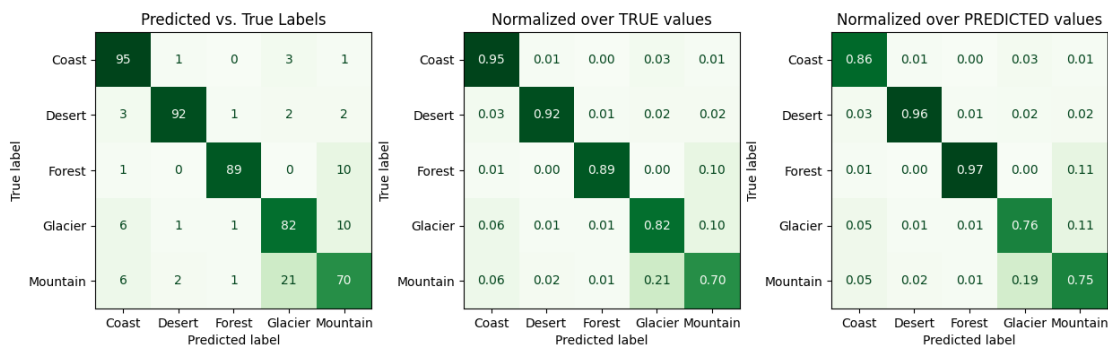
Figure 6. Confusion Matrix with fine-tuning.

As shown in Figure 7, the final training loss becomes about 0.45, which is higher than that of the basic model. The final accuracy becomes 0.82, which is lower than that of the basic model. It could be claimed that data augmentation has a negative implication on the improvement of basic model. The operations including rotation, rescaling, over flip seems to add more needless factors in landscape recognition. In conclusion, data augmentation might not be suitable for the model with the given dataset.
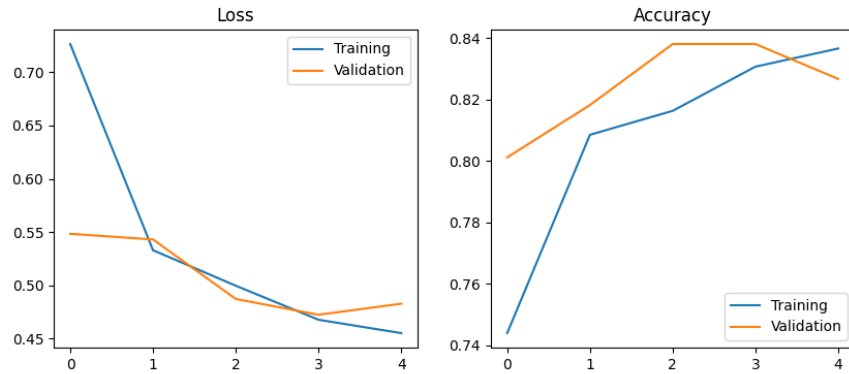
Figure 7. The change of loss and accuracy after data augmentation.

As shown in Figure 8, the loss and accuracy reach their limitations in the epoch of 5. From epoch 6-10, the learning rate of optimizer has shrunken to 10-4 and the whole layers have been unfrozen, the overfitting condition gain improvement to some extent compared to the model after data augmentation. Meanwhile, the training loss becomes stable in about 0.45. Overall, the fine-tuning implement has a beneficial effect on the improvement of model. In testing set, the final accuracy of EfficientNetB0 is 89.6%. After data augmentation, this has fallen to 84.4%. In the fine-tuning section, that has increased to 85.6% anew. It seems that models predict the class of landscape to a relatively high level.
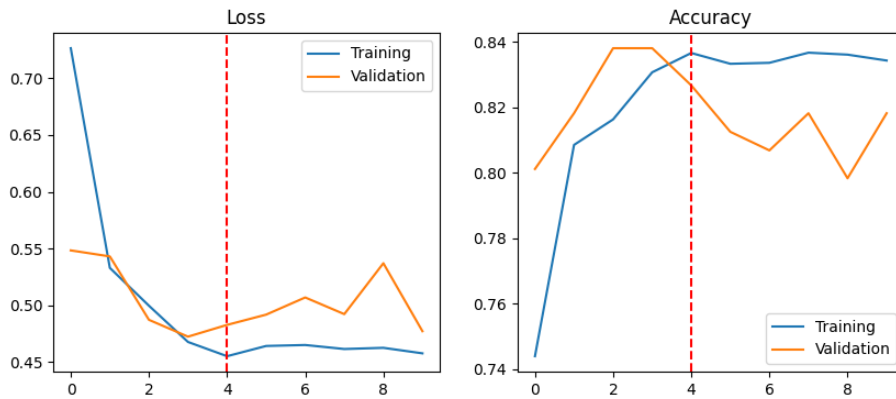
Figure 8. The change of loss and accuracy after fine-tuning.

4. Conclusion

This study employs EfficientNetB0 models to predict the landscape. The whole process includes initialization of the model, data augmentation, fine-tuning and evaluation of these models. The result of this study shows that the model predicts the type of given landscape images to a relatively high degree, except for mountain class. Besides, the data augmentation is likely to weaken the ability of recognition while fine-tuning improves this condition. With gradual upgrades and adjustments of scaling factor, the advanced model is applied in more complicated and numerous conditions. It decreases the overfitting rate after boosting the scale of the given dataset, which includes multiple angels of the same landscape for more frequent prediction. In addition, it could also be realized that transfer models to a higher-level configuration environment. That significantly optimizes the running time of the whole programme for programmers to adjust the hyper-parameters.

References

[1]   Goudie A 2020 The human impact in geomorphology–50 years of change Geomorpholog 366: p 106601

[2]   United Nations 2023 United Nations Decade for Deserts and the Fight Against Desertification https://www.un.org/en/events/desertification_decade/whynow.shtml.

[3]   Shafaey M A Salem M A M Ebied H M et al. 2018 Deep learning for satellite image classification//International Conference on Advanced Intelligent Systems and Informatics Cham: Springer International Publishing pp 383-391

[4]   Li Z Liu F Yang W et al. 2021 A survey of convolutional neural networks: analysis, applications, and prospects IEEE transactions on neural networks and learning system

[5]   Wang T Chen J Liu L et al. 2023 A Review: How Deep Learning Technology Impacts the Evaluation of Traditional Village Landscapes Buildings 13(2): p 525

[6]   Cowls J Tsamados A Taddeo M et al. 2021 The AI gambit: leveraging artificial intelligence to combat climate change—opportunities, challenges, and recommendations AI & Society pp 1-25

[7]   Sefrin O Riese F M Keller S 2020 Deep learning for land cover change detection Remote Sensing 13(1): p 78

[8]   Gupta J Pathak S Kumar G 2022 Deep learning (CNN) and transfer learning: a review//Journal of Physics: Conference Series. IOP Publishing 2273(1): p 012029

[9]   Tan M Le Q V 2019 Efficientnet: Rethinking model scaling for convolutional neural networks arXiv. arXiv preprint arXiv:1905.11946

[10]   Deepnets 2023 Landscape Recognition | Image Dataset | 12k Images Kaggle https://www.kaggle.com/datasets/utkarshsaxenadn/landscape-recognition-image-dataset-12k-images.