Efficient Image Classification Using ReXNet: Distinguishing AI-Generated Images from Real Ones

Yanxi Liu[a]
School of Information Management, Shanghai Lixin University of Accounting and Finance, Shanghai, China

Keywords: ReXNet, Convolutional Neural Network, Image Classification, AI-Generated.

Abstract: Image classification, a topic of growing interest in recent years, holds significant applications in computer vision and medical domains. This paper introduces an image classifier built on the Rank eXpandsion Networks (ReXNet) model, specifically designed to effectively distinguish between Artificial Intelligence (AI)-generated and real images. Leveraging Convolutional Neural Network (CNN) architecture, this method utilizes deep separable convolution layers to minimize parameters and computational complexity. It also features a compact network structure and optimized hyperparameters for efficient feature extraction and classification. Experimental results demonstrate the model's high classification accuracy across various image types, showcasing its efficiency. This experiment underscores the ReXNet model's potential in image classification and offers valuable insights for future research directions. This study not only validates the accuracy and generalization capabilities of lightweight models but also lays a solid groundwork for more intricate image classification studies. The findings highlight the importance of efficient model design in addressing real-world image classification challenges, particularly in distinguishing between AI-generated and authentic images, with implications for advancing both theoretical understanding and practical applications in the field.

1 INTRODUCTION

With the continuous development of Artificial Intelligence (AI), an increasing number of intelligent technologies are being applied to various aspects of life, constantly enriching lives and bringing great convenience (Goyal, 2020). Image classification, as the core problem of computer vision, is a classic subject of research in recent years, and it is also the foundation of the integration of visual recognition with other domains (Chen, 2021). The synthesis of images by artificial intelligence is one of the important areas in image classification (Bird, 2024). Efficiently distinguishing between AI-generated images and real images is crucial to ensuring the authenticity and effectiveness of image data (Bird, 2024).

Currently, an increasing number of deep learning methods are being applied to image classification. Neural networks, decision tree classifiers, and remote sensing classification methods have achieved certain effectiveness in image classification (Lu, 2007). Compared to traditional image features, which heavily rely on manual settings, the development of deep learning enables image features to be hierarchically represented by computers nowadays (Suzuki, 2017). Among them, Residual Neural Network (RESNET) is one of the classic methods for image classification. It first concretizes the neural network (RESNET) through pre-training, then extracts image features from it, and finally uses these features to train a machine learning Support Vector Machine (SVM) classifier to achieve the ultimate goal of image classification (Mahajan, 2019). Despite the emergence of numerous methods and some level of success achieved, due to the difficulty in representing features in images, image classification remains one of the most challenging topics in the field of computer vision (Wang, 2022). In 2023, a report on the underwater image enhancement network, Rank eXpandsion Networks (ReXNet), demonstrated the practicality of ReXNet in visual tasks through validation on multiple datasets, providing new insights for image classification (Zhang, 2023). The

[a] https://orcid.org/0009-0003-6440-3167

Liu, Y.
Efficient Image Classification Using ReXNet: Distinguishing AI- Generated Images from Real Ones.
DOI: 10.5220/0012925200004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 234-238
ISBN: 978-989-758-713-9

Recursive Aggregation Operator (ReX) in this model reduces peak RAM and average latency of various adaptive models on the device by bypassing large early activations and local representation methods, greatly improving the efficiency of image classification (Qian, 2022). At the same time, the layer channel setting principle derived through the progressive increase in the number of channels has effectively resolved the bottleneck problem of the image classification layer, leading to a significant improvement in the accuracy of image classification (Han, 2021).

To enhance image classification accuracy, this study utilizes the ReXNet model as the classifier for both AI-generated and real images. The image dataset is primarily sourced through web scraping, providing real images for the model. Additionally, AI is employed to generate synthetic data, mimicking the characteristics of real images, for model training and evaluation, available on Kaggle. Data preprocessing involves standardization and normalization, ensuring consistent dimensions and magnitudes, which is crucial for subsequent modeling. The ReXNet model is employed for feature extraction and classification, undergoing cyclic training with a specified number of epochs. Backpropagation is used to update model parameters based on calculated loss for different data batches. The study incorporates early stopping techniques during training to prevent overfitting, maximizing training efficiency. The experiment indicates that this research is able to effectively extract image features. As an image classifier, this model can accurately and efficiently differentiate between real images and AI-generated images.

2 METHODOLOGIES

2.1 Dataset Description and Preprocessing

The dataset for this study is primarily obtained through random web crawling and AI generation, and can be accessed on Kaggle (Kaggle, 2024). This dataset contains 538 AI-generated images and 435 real images randomly collected from the web, providing a basis for further analysis of their similarities. Each image has a size of 224x224 pixels, encompassing a variety of themes, with special emphasis on people, animals, landscapes, and psychedelia. The dataset has been divided into training set and test set. Simultaneously, resizing and normalizing the images eliminates differences in dimensions and scales, making the model easier to converge and laying the foundation for subsequent modeling. Figure 1 and Figure 2 respectively illustrate partial AI-generated images and real images.

Figure 1: AI generated image (Photo/Picture credit: Original).

Figure 2: Real image (Photo/Picture credit: Original).

2.2 Proposed Approach

This study primarily utilizes the ReXNet model as the classifier for AI-generated images and real images. Before establishing the model, the data needs to be normalized for further modeling and analysis. This model is further optimized based on Convolutional Neural Network (CNN), and to some extent, it solves the bottleneck problem in the representation of the shrinking layer through progressive increase in the number of channels. When these techniques are combined, they can better extract the feature information in the images, effectively improve the training efficiency of the model, and achieve better performance in image classification tasks. The following Figure 3 illustrates the structure of the system.

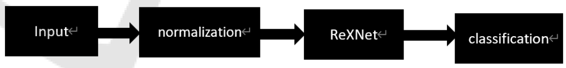

Figure 3: Model flow chart (Photo/Picture credit: Original).

2.2.1 ReXNet

RexNet is a lightweight neural network architecture designed to strike a balance between model accuracy and computational efficiency, enabling efficient image classification tasks in resource-constrained environments. Its primary goal is to maintain a high level of accuracy while optimizing computational resources. In comparison to the traditional network architecture paradigm, there may be expression bottleneck issues, which in turn affect model performance. This model undergoes slight adjustments on the benchmark network, adopts a progressively increasing approach to channel count design, and replaces more expansion layers to address this issue. While the model is lightweight, ReXNet performs exceptionally well on image classification tasks, demonstrating high accuracy and generalization ability. At the same time, ReXNet, as the foundation model of this study, focuses especially on enhancing channel interaction, further optimizing efficiency while ensuring accuracy.

The ReXNet model used in this study is implemented based on the PyTorch framework, and undergoes continuous training and validation through multiple epochs of iteration. The model reduces the number of parameters and computational complexity by using depthwise separable convolutional layers, while also maintaining the model's expressive power. During the training phase, the model calculates and stores metrics such as loss and accuracy based on the training data, in order to update the model parameters for better performance. During the validation phase, the model computes the corresponding metrics obtained in the training set, further storing and printing the validation metrics. In addition, the final validation loss will also be checked to see if there is any improvement. If there is improvement, the model will be saved as the best model; if there is no improvement, the number of periods without improvement will continue to be tracked. If the specified time still does not show any improvement, the training will be stopped. The following Figure 4 illustrates the training and validation process of the model.

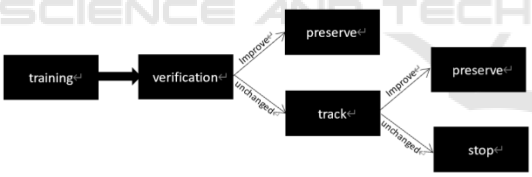

Figure 4: Training and inspection process (Photo/Picture credit: Original).

2.2.2 Loss Function

Choosing the right loss function plays a crucial role in the training process of the model. This research selects the cross-entropy loss function for AI-generated image and real image classifiers. As a common loss function in classification tasks, the cross-entropy loss function can calculate the negative log-likelihood loss of the classification problem, where the model outputs the probability of each class. During the training loop, the loss function computes the loss value for each batch of data and accumulates it over the entire epoch. After completing the cycle, divide the loss value by the total number of batches to obtain the average loss for that cycle. By training the model using loss values, the parameters can be adjusted during backpropagation. In addition, the validation loss of the model is also calculated in a similar manner to monitor the model's performance on unseen data. The following formula represents the cross-entropy loss function, as:

$$H(p,q) = -\sum_x p(i) \log q(i) \tag{1}$$

where $p(i)$ represents the actual probability of the model and $q(i)$ represents the predicted probability of the model.

2.3 Implementation Details

In the process of building the model, this research needs to emphasize the following aspects. Firstly, this study employs the Stochastic Gradient Descent (SGD) optimizer to update the model parameters. This method primarily updates parameters based on each training sample and label, and SGD avoids redundant calculations in each update, resulting in faster execution speed. In addition, the model also utilizes a learning rate scheduler to dynamically adjust the learning rate. After multiple experiments, 3e-4 was finally selected as the learning rate for the model to improve training stability and convergence speed. Finally, the model stops training early by setting a patience value to prevent overfitting or unstable training. If the loss value does not decrease for multiple consecutive epochs, the training is stopped prematurely.

3 RESULTS AND DISCUSSION

In this study, the ReXNet model is utilized as an AI image generator and real image classifier. This model classifies images of different themes that are randomly captured to ensure its accuracy. Figures 5, 6, and 7 respectively demonstrate the performance of

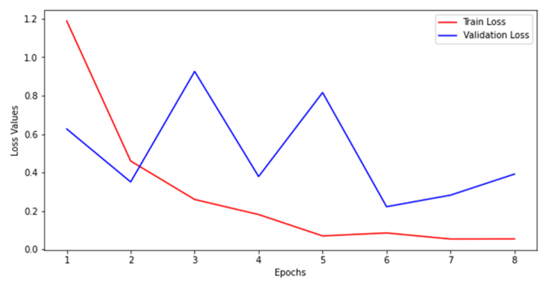

Figure 5: Loss Values Image (Photo/ Picture credit: Original).

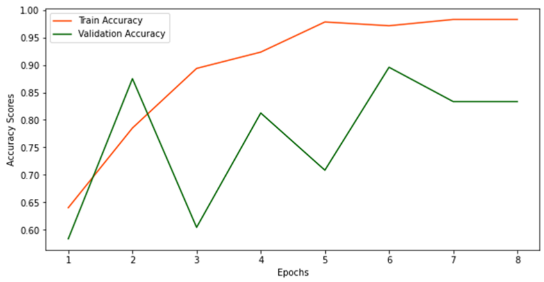

Figure 6: Accuracy Scores Image (Photo/Picture credit: Original).

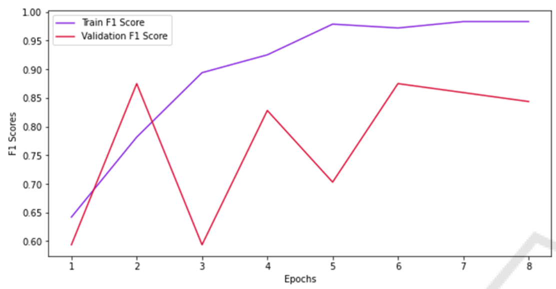

Figure 7: F1 Scores Image (Photo/Picture credit: Original).

the training set and validation set based on the Loss Values, Accuracy Scores, and F1 Scores of this model.

As shown in Figures 5 and 6 that the model experiences a gradual decrease in loss value and a continuous improvement in accuracy score over the first five epochs. The training set in the ReXNet model achieved an impressive 97% accuracy after just 5 epochs of training. Even though there was a slight decrease in test set accuracy at the same time, it showed an upward trend after a brief adjustment, and maintained a consistently high accuracy rate with the training set in subsequent training cycles until the training stopped. As shown in Figure 7, the frequently used F1 score for measuring the performance of binary classification models also nicely illustrates this point. The efficiency of this model is closely related to its lightweight characteristics. The model adopts an efficient network structure and parameter optimization, which enables the model to have a small model size and computational complexity, greatly improving the model's classification efficiency. In addition, the carefully designed network architecture also enables the model to have strong feature extraction capability and generalization ability during the image classification process.

Figure 8: Image classification display (Photo/Picture credit: Original).

As shown in Figure 8, the first and third images are generated by AI, while the second image is a real image. This image shows the result of the image after being processed by the classifier, where different colors differentiate the visualized model's different focus areas on the image. This model mainly visualizes the model's predictions for these images using the Class Activation Map (CAM) technique, and displays the ground truth label and predicted label for each class name. This technology mainly utilizes the form of superimposing heat maps and original images for visualization, with the red highlighted areas serving as the primary basis for its analysis. When comparing AI-generated images to real ones, there may be some inconsistencies or irrational details in terms of fine details, such as blurry edges, unnatural colors, while real images tend to be more authentic and clear. The AI image classifier, generated by the ReXNet model, precisely utilizes these details to classify the images.

Overall, the ReXNet model generated by this study significantly improves the accuracy of AI image classification compared to real image classification. The images in this classification involve various aspects including humans, animals, landscapes, etc., laying the groundwork for future applications of image classification in multiple domains. The future optimized model can be applied to industrial quality inspection, medical imaging, and various other fields, further benefiting mankind.

4 CONCLUSIONS

This study aims to develop a classifier model that can effectively distinguish between AI-generated images and real images. The research further explores and optimizes the basis of CNN, ultimately selecting the ReXNet model to generate the AI image and real image classifier. Experimentation with the image dataset showcased the classifier's high accuracy in categorizing images across various topics. This commendable performance is attributed to the model's lightweight design and high-performance characteristics, significantly enhancing the efficiency and accuracy of AI-generated and real image classification. Future research endeavors will explore data augmentation strategies and refine model architectures to further elevate classification accuracy, particularly under challenging conditions such as varying angles and lighting. Integration of transfer learning strategies will continue to bolster the model's classification and generalization capabilities. Additionally, methods including integrated learning and hyperparameter tuning will be explored to optimize model performance. The ReXNet model holds promise in effectively distinguishing complex images, thus advancing the application of deep learning in image classification.

REFERENCES

Bird J J, Lotfi A. (2024). Cifake: Image classification and explainable identification of ai-generated synthetic images.

Chen, L.; Li, S.; Bai, Q.; Yang, J.; Jiang, S.; Miao, Y. (2021). Review of Image Classification Algorithms Based on Convolutional Neural Networks. Remote Sens. vol. 13, p: 4712.

Goyal M, Knackstedt T, Yan S, et al. (2020). Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. Computers in biology and medicine, vol. 127, p: 104065.

Han D, Yun S, Heo B, et al. (2021). Rethinking channel dimensions for efficient model design. Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition.pp: 732-741.

Kaggle. (2024). AI Generated Images vs Real Images. https://www.kaggle.com/datasets/cashbowman/ai-generated-images-vs-real-images

Lu D, Weng Q. (2007). A survey of image classification methods and techniques for improving classification performance. International journal of Remote sensing, vol. 28(5), pp: 823-870.

Mahajan A, Chaudhary S. (2019). Categorical image classification based on representational deep network (RESNET). International conference on Electronics, Communication and Aerospace Technology (ICECA). pp: 327-330.

Qian X, Hang R, Liu Q. (2022). ReX: an efficient approach to reducing memory cost in image classification. Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36(2), pp: 2099-2107.

Suzuki K. (2017). Overview of deep learning in medical imaging. Radiological physics and technology, vol. 10(3), pp: 257-273.

Wang R, Lei T, Cui R, et al. (2022). Medical image segmentation using deep learning: A survey. IET Image Processing, vol.16(5), pp: 1243-1267.

Zhang D, Zhou J, Zhang W, et al. (2023). ReX-Net: A reflectance-guided underwater image enhancement network for extreme scenarios. Expert Systems with Applications, vol. 231, p: 120842.