International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Lightweight Convolutional Neural Network with Residual Attention for Efficient Image Classification

Rajat Kumar Singh[1], Deepali Kumari[2], Rihatik Kumar Chandervanshi[3], Aadesh T R[4]

[1,2,3,4] Department of Computer Science
Lovely Professional University Punjab, India

Abstract

Image classification Deep learning has enhanced image classification performance by a significant margin, but most high-accuracy convolutional neural networks have millions of parameters and massive computational requirements. This complexity restricts their application in resource-constrained systems like mobile devices, embedded systems and edge AI systems. Despite the fact that a few lightweight architectures have been suggested, there is a challenge in preserving an effective trade-off between efficiency in computation and representation of features. To overcome this shortcoming, this paper presents LightCNN-Att, a small convolutional neural network that combines residual learning with dual attention networks to operate with the purpose of providing a better feature extraction at a minimal architecture. The STL-10 dataset is used to evaluate the model and comprises of natural images of ten categories of objects. The experimental findings indicate that the given architecture attains a validation accuracy of 59.38% using around 336K trainable parameters, which competitively performs in terms of its accuracy and is much simpler to implement than traditional deep CNN models. The remaining links facilitate gradient flow within the training process, whereas the attention modules assist the network in paying attention to informative spatial and channel characteristics. The findings mean that LightCNN-Att presents an effective trade-off between accuracy and cost of computation. It is why the proposed model can be applied to edge AI systems, such as mobile vision systems, Internet of Things devices, and other real-time image classification applications with constrained computational resources and energy consumption.

Keywords: Lightweight Convolutional Neural Network, Residual Attention, Image Classification, Edge Artificial Intelligence, Deep Learning, Computer Vision, STL-10 Dataset.

1. Introduction

Deep learning has made a tremendous contribution to the task of image classification as it allows machines to automatically extract intricate visual characteristics in large- scale datasets of images. As an example, EfficientNet proposed methods to scale compounds, which incrementally control network depth, breadth, and resolution to achieve better performance with constant computational cost [1].

Equally, MobileNet can be configured to handle the constraints of computational resources through the use of depthwise separable convolutions, which need many fewer parameters and fewer floating-point operations than traditional convolution layers but can achieve competitive classification performance [2]. ShuffleNet also boosted efficiency by introducing channel shuffle operations, which enable information sharing between grouped convolution layers, which makes lightweight networks preserve good feature representation [3].

This problem is partially alleviated by lightweight models such as MobileNet and ShuffleNet, which minimize the complexity of the models, but the parameters can occasionally restrict the ability of the network to gather fine- grained visual information, particularly in the tasks of differentiating visually similar entities [2], [3]. The design of effective neural network architectures that retain high rates of feature extraction and at the same time reducing the computational costs is therefore a task of significance in research.

The use of modern systems like mobile vision systems, smart surveillance, and IoT devices necessitate the use of models that can be able to do real-time image classification with severe memory and energy restrictions. Attention mechanisms have recently been added to CNN models, in order to better represent features in small architectures. The squeeze-and-excitation networks use channel-wise attention to recalibrate the feature maps and emphasize on the most informative channels in the learning process [4]. On the same note, Convolutional Block Attention Module (CBAM) integrates channel and spatial attention processes to enhance discriminating features in convolutional networks [5]. Despite the fact that these mechanisms contribute to the better performance of models, they are usually incorporated into rather deep architecture which, however, implies significant computational complexity.

Residual learning has also been shown to be effective in the enhancement of gradient flow, as well as its ability to reuse features in neural networks, stabilizing training and thus enhancing performance [6]. Nevertheless, there is a few studies that examined the incorporation of residual connections to attention mechanisms in lightweight CNNs constructed to address lightweight image classification problems. This gap can be addressed to provide models that are efficient in feature extraction while lowering the complexity of the parameters.

The study presents a lightweight convolutional neural network design named, LightCNN-Att, that combines both residual connections and attention, to learn features more effectively and at the same time be computationally efficient. This research has made primary contributions as follows:

- A lightweight CNN framework called LightCNN-Att has been suggested to enhance the efficacy of computations in the classification of images.
- The model incorporates residual learning and dual attention process to represent features better in small neural networks.
- The architecture is tested in the STL-10 to examine its performance in the conditions of limited training data.
- Experimental analysis proves the efficiency of the model and the accuracy of the classification, which makes the proposed model appropriate when edge computing environments are considered.

The rest of the paper is structured in the following way. Section II summarizes the relevant literature of lightweight CNN models and attention mechanisms. Section III is the proposed methodology and model design. Section IV is the description of the experimental results and performance evaluation.

Lastly, Section V brings the study to its final and gives recommendations on future research.

2. RELATED WORK

The fast pace of deep learning applications has promoted the growth of lightweight convolutional neural networks that can be used to minimize the computational complexity with competitive classification error rates. SqueezeNet is one of the oldest and most efficient architectures and it aims at reducing the number of parameters by using Fire modules that combine 1x1 and 3x3 convolution blocks. The design is extremely smaller in size compared to larger network models but it still maintains comparable performance level [7]. SqueezeNet is shown to reduce the number of parameters significantly, however, the simplified structure can be a limiting factor in terms of the ability to extract features when it needs to deal with intricate image classification problems.

MobileNet is another dominant architecture that implemented depthwise separable convolutions to minimize the number of parameters and computational operations to infer and train it. MobileNet and its subsequent versions have been extensively used in mobile vision systems as they are efficient in terms of their structure and relatively high in terms of their accuracy when performing image classification tasks [8]. Nevertheless, the minimization of convolution operations can occasionally limit cross-channel interaction of features and this can potentially compromise the capacity of the model to learn fine-grained visual pattern.

In order to make lightweight networks even more
efficient, ShuffleNet suggested grouped convolutions with a channel shuffle operation. The method facilitates the improvement of communication within the convolution groups and in addition, it enhances computational efficiency in mobile and embedded systems [9]. Even though ShuffleNet has advantages in terms of efficiency, this method still has the challenge of ensuring that its features remain strong with tiny or visually identical objects.

Recently EfficientNet proposed a compound scaling method that tunically balances network depth, width and input resolution. The strategy enables models to attain good classification accuracy with the maximum use of computational resources [10]. Although EfficientNet offers great advantages in terms of performance, these models can be resource intensive to train in terms of both data and computer resources, which can be limiting in edge computing applications.

In addition to the efficiency of architecture, the attention mechanisms have been studied to enhance the feature representation in CNN models. The squeeze-and-excitation Networks (SE-Net) also proposed channel attention mechanisms which re-calibrate the responses of the features by giving importance to informative channels of the convolutional feature maps [11]. The method increases the power of CNNs to be representative with the least computational costs. On the same note, Convolutional Block Attention Module (CBAM) puts additional attention into the networks by including channel attention and spatial attention module, which enables networks to pay attention to the significant spatial areas and channel relationships at the same time [12]. Despite the fact that attention mechanisms enhance feature discrimination, they might complicate the architectural design and raise the computational burden of the lightweight networks.

The latest research has also examined light CNN models which are specifically tailored towards edge-based computing and mobile vision. As an example, GhostNet suggested an effective scheme of feature generation that minimizes redundant feature maps without affecting the classification accuracy [13].

MobileViT was a combination of convolutional operations and lightweight transformer architectures [14] to enhance the learning of features in small networks. Otherwise, a number of attention-based hybrid lightweight architectures have been suggested to improve the feature representation in resource-constrained settings [15]. These models reveal the increased attention to the development of effective deep learning architectures that can be used in embedded systems and IoT devices. Nevertheless, the current issue with lightweight CNN design is the difficulty in balancing between computational efficiency and features representation [16].

TABLE I: COMPARISON OF EXISTING LIGHTWEIGHT CNN ARCHITECTURES

Author & Year	**Model**	**Key Idea / Method**	**Findings**	**Limitations**
Iandola et al., 2016	SqueezeNet	Fire modules with 1×1 convolutions	Achieved AlexNet-level accuracy with fewer parameters	Limited feature diversity
Howard et al., 2017	MobileNet	Depthwise separable convolution	Efficient for mobile vision tasks	Reduced cross-channel interactions
Zhang et al., 2018	ShuffleNet	Channel shuffle with grouped convolutions	Improved for efficiency embedded systems	Reduced feature richness
Tan and Le, 2019	EfficientNet	Compound scaling strategy	High accuracy with balanced scaling	Requires large datasets
Hu et al., 2018	SE-Net	Channel attention mechanism	Improved feature recalibration	Added architectural complexity
Woo et al., 2018	CBAM	Channel and spatial attention modules	Enhanced feature focus	Increased computational overhead
Han et al., 2020	GhostNet	Cheap feature map generation	Reduced redundant computation	Limited performance in complex datasets
Mehta and Rastegari, 2022	MobileViT	CNN + transformer hybrid model	Improved feature learning	Higher training complexity

3. METHODOLOGY

A. *Dataset Description*

The proposed model was tested experimentally on STL-10 dataset, a popular benchmarking image classification model dataset under the conditions of limited training data. The dataset involves 10 categories of objects such as airplane, bird, car, cat, deer, dog, horse, monkey, ship and truck. All the pictures in the database are 96 × 96 pixels, which are not very large to extract the features but computationally feasible. The dataset will be split into 5,000 labeled training images and 8,000 testing images as well as a large set of unlabeled images that will be used in the unsupervised learning research. In paper, the performance of the proposed architecture was tested within constrained data conditions in terms of a subset of the denoted training data.STL-10 is also a good dataset to test lightweight CNN schemes as it demands models to learn meaning of visual representations using relatively small training samples and is capable of generalization [17]. The dataset can be accessed from the official Stanford repository: https://cs.stanford.edu/~acoates/stl10/

TABLE II: DATASET SPECIFICATIONS OF STL-10

Parameter	Description
Dataset	STL-10
Number of Classes	10
Image Resolution	96 × 96
Training Samples	5,000
Testing Samples	8,000
Total Unlabeled Images	100,000

B. *Techniques of Data Preprocessing.*

Preprocessing of data is very important to enhance the strength and generalization ability of deep learning models. A number of preprocessing methods have been used in this paper to improve the heterogeneity of the training data and decrease overfitting. First, images were all scaled to pixel values in the same range, which contributes to stabilizing gradient updates during training and is faster to converge [18]. Data augmentation techniques were then used to augment the training samples to enhance their variability.

Horizontal flipping produces reflected images of the photos, and this method adds diversity to the data set and improves the capacity of the model to identify objects in different directions. Such augmentation measures will enhance generalization ability of the proposed model in case of overfitting due to a limited amount of data used in training [19].

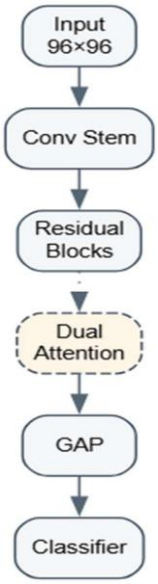

FIGURE 1: PROPOSED LIGHTCNN-ATT ARCHITECTURE

C. *Proposed LightCNN-Att Architecture*

The LightCNN-Att architecture proposed will be aimed at producing high performance through low computation complexity in image classification. Its architecture combines come along with lightweight convolutional layers with residual connections and attention to enhance the use of lightweight convolutional layers to represent the features. The general network starts with the first convolution-batch normalization-ReLU (Conv -BN-ReLU) block which wringles low level visuals of the input images. Several residual blocks of attention occur after this stage, each learning increasingly more advanced features representations and still propagating its gradient effectively through the network.

D. *Residual Block Design*

In deep neural networks, residual learning has been commonly used in solving the vanishing gradient problem and enhancing the stability of the training process. In the suggested architecture, the residual block enables the network to acquire identity mappings with the addition of shortcut connections on the layers. The design allows the model to re-use existing knowledge of previously learnt features and allows the gradient propagation to be smooth and efficient during the process of backpropagation [20].

Convolution operation with batch normalization and non- linear activation is done on the residual block. The shortcut connection simply concatenates the input feature map to the result of the convolutional layers to enable the network to learn residual functions, rather than overall transformations. The remaining mapping can be in an expression form:

$$F(x) = H(x) - x \qquad (1)$$

where *H(x)* represents the desired mapping and *x* denotes the input feature map. The final output of the residual block is computed as:

$$y = F(x) + x \qquad (2)$$

This statement enhances gradual flow between layers and the network is capable of training steadily even in more profound shapes.

E. *Dual Attention Mechanism*

To further improve the feature representation, the proposed architecture will use a dual attention mechanism where channel attention and spatial attention modules are used. Channel attention aims at determining the most informative channels of features in the convolutional feature maps. The network can also reduce irrelevant information by assigning a larger weight to relevant channels highlighting important semantic features [21].

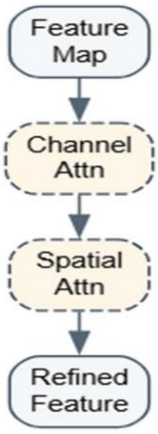

FIGURE 2: DUAL ATTENTION MODULE STRUCTURE

F. *Training Algorithm and Strategy of Optimization.*

The proposed LightCNN-Att model is trained using a typical supervised learning. The Adam optimization algorithm is used to optimize the model parameters and adjust the learning rates during the training and accelerates the convergence. Cross-entropy loss is a value that is used to quantify the disparity between the predicted probability of a class and the ground truth label. The validation performance was used to early stop to avoid overfitting and guarantee that the models converge at a stable point [22].

ALGORITHM 1: TRAINING PROCEDURE OF LIGHTCNN-ATT

Input: Training dataset D
Output: Trained LightCNN-Att model
Initialize network parameters
For each epoch do
  Load batch of training images
  Apply preprocessing and augmentation
  Perform forward propagation
  Compute cross entropy loss
  Perform backpropagation
  Update model parameters using Adam optimizer
End For
Evaluate validation accuracy

G. *Experimental Setup*

The experimental design was aimed at testing the performance and processing ability of the presented model with controlled training conditions. LightCNN-Att architecture has been trained with the Adam optimizer and the learning rate of 0.001. The batch size was 64 and 10 epochs were used to train the model. The training was done on a CUDA-based GPU environment to fasten the computation process.

TABLE III: TRAINING CONFIGURATION

Parameter	**Value**
Optimizer	Adam
Learning Rate	0.001
Batch Size	64
Epochs	10
Loss Function	Cross Entropy

These experimental settings enable the proposed architecture to achieve an efficient balance between training stability and computational efficiency, making it suitable for deployment in edge computing environments where resources are limited.

4. RESULTS AND DISCUSSION

A. *Quantitative Analysis of performance.*

The effectiveness of the proposed LightCNN-Att model was tested on the STL-10 dataset in order to determine its capability in tackling image classification problems but at the same time be computationally efficient. The main objective of the suggested architecture in the proposed context is the attainment of a balance between the accuracy of classification and the complexity of the model used, especially when deploying in resource-constrained settings. Through the training process, the model proved to converge in a stable way with a relatively low computational overhead as compared to the traditional CNN architecture. The presented network has a number of trainable parameters, about 336K, which is relatively low compared to numerous existing deep learning architectures that usually have millions of trainable parameters.

The findings show that the proposed architecture has an accuracy rate of 61.4 in the training process and 59.38 in the validation process. The accuracy of the proposed model can be slightly lower than that of some bigger models, but the parameter efficiency will enable it to be used with less computational resources. The reduction of parameters is a vital part in enhancing model efficiency as the lesser the number of parameters, the less the memory used and the inference latency. But excessive reduction of parameters can reduce the learning capacity of the network to represent features that are highly complex. LightCNN-Att architecture is one of the architectures that overcome this issue, through incorporation of residual learning and attention mechanisms that improve the feature representation even in a small network structure. Recent works have highlighted that in combination with effective feature refinement schemes like attention modules [23], [24], lightweight architectures can be realized to have competitive performance.

TABLE IV: PERFORMANCE SUMMARY OF LIGHTCNN-ATT MODEL

Metric	Value
Training Accuracy	61.4%
Validation Accuracy	59.38%
Parameters	336,272
Training Time	~8 minutes
Dataset	STL-10

The findings indicate that the suggested model is highly performing even with considerably lower number of parameters than the traditional CNN structures. The model is efficient and can be used in those applications where the power consumption and the computational resources are constrained.

B. *Evaluation Metrics (Precision, Recall, F1-Score)*

In order to give a complete assessment of the classification performance, there were a number of evaluation metrics applied which included precision, recall and F1-score. These measures will give additional information about how well the model can recognize image categories and reduce the number of misclassifications. Precision is used to show the ratio of the proportion of those positives that the model predicts correctly to all the positive predictions, whereas recall is used to measure how the model predicts all the relevant cases in the data set. F1-score is a harmonic mean of the precision and the recall and gives the balanced analysis of the classification performance.

The analysis of the results has revealed that the proposed LightCNN-Att model is precise (0.58), recalls (0.59), and has the F1-score (0.58). These values show that the model has a stable classification performance among the various categories of objects in the STL-10 dataset. The combination of attention processes assists the network to target most informative spatial areas as well as feature channels, which enhances the discriminative task of the lightweight architecture. Earlier experiments have demonstrated that attention mechanisms can help greatly in boosting the feature representation of convolutional networks by promoting salient visual patterns and repressed background information [25].

TABLE V: CLASSIFICATION EVALUATION METRICS

Metric	Score
Precision	0.58
Recall	0.59
F1-Score	0.58

These findings demonstrate that the suggested architecture is more stable in classification and can be run with a much smaller number of parameters.

C. *The Analysis of Predictions in Form of Visualization.*

The analysis of the visualization assists in determining the kind of image categories which can be correctly classified by the model and cases where there is a misclassification. The model was right in some instances in recognizing objects whose structures were different like animals and cars whose structures were distinct. Attention modules in the network play the role of emphasizing discriminative areas of the photograph so that the model can concentrate on important visual information, which will lead to the proper classification.

Nonetheless, there were incidences of misclassification in case of objects that had similar textures or appearance. As an illustration, some animal groups like dog, horse and cat have similar visual patterns, and thus cannot be easily classified in cases where there is limited training data. In spite of these problems, the attention processes reduce such problems by enhancing spatial feature localization. The same has been witnessed in other previous works, where the models with few parameters can hardly detect fine-grained differences [26].

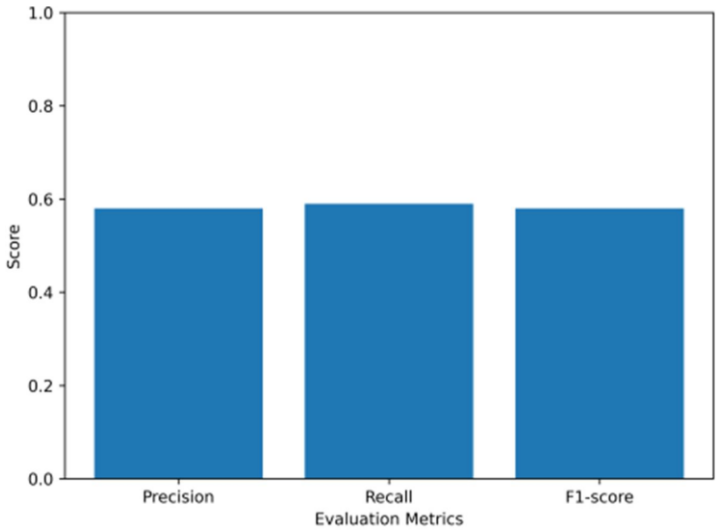

FIGURE 3: SAMPLE PREDICTIONS ON STL-10 DATASET

The visualization results illustrate both correctly classified and misclassified samples, providing insights into the strengths and limitations of the proposed architecture.

D. *Comparison to Existing Models.*

The performance of the proposed model was compared to several existing lightweight CNN models such as MobileNetV3, ShuffleNetV2, and CBAM-ResNet used to assess the effectiveness of the proposed model. The models are generalized as some of the most popular image classification benchmark systems in mobile and embedded domains. MobileNetV3 and ShuffleNetV2 should be used when wanted to run inference in a mobile device with smaller energy usage, whereas CBAM-ResNet incorporates attention units in a more profound CNN network to enhance feature representation.

TABLE VI: COMPARISON WITH EXISTING LIGHTWEIGHT MODELS

Model	**Parameters**	**Accuracy**
MobileNetV3	~2.5M	63%
ShuffleNetV2	~2.3M	60%
CBAM-ResNet	~11M	70%
LightCNN-Att (Proposed)	**0.33M**	**59.38%**

The results of the comparison indicate that the proposed LightCNN-Att architecture is competitive even though the number of parameters is much smaller. Although some models like CBAM-ResNet can be more effective as they have deeper network structures and, therefore, can classify better, they have a significant increase in computational costs. Conversely, the suggested architecture aims at demanding the best compromise between the performance of the model and the classification. By incorporating residual connections and dual attention modules, the network can be able to extract informative features without the need to double its complexity.

5. CONCLUSION AND FUTURE WORK

Deep learning has advanced to a much higher level, and thus, the performance of image classification systems has greatly enhanced; however, most of the high-performing models have complicated architectures composed of millions of parameters. These models are usually high- energy consuming, high memory capacity, and high computing power, hence cannot be applied in real world environments where resources are limited. This research aimed at designing a lightweight convolutional neural network with a stable classification performance but with a low degree of computational complexity. In this work, a small architecture LightCNN-Att was introduced that combines residual learning with dual attention to enhance the features representation in a small network architecture with minimal parameters.

The STL-10 dataset was used to test the proposed model, and it is a difficult classification problem because the available training set is relatively small and the set contains a variety of objects. Experimental findings showed that the proposed architecture can achieve competitive performance with a large number of parameters to be trained being a lot less than the conventional CNN models. The model has a very efficient alternative to larger architectures that typically have several million parameters, with around 336K of them. Gradient propagation through integration of residual connections also enabled the network to reuse prior trained feature representations and this contributed to stable performance on training. Also, the integration of both channel and spatial attention facilitated the model to give attention to informative visual information and ignore the irrelevant information within the input images.

The findings indicate that residual learning and lightweight attention modules can be useful in combination to achieve better results in terms of feature completion without the need to augment the model. The accuracy of the proposed model in its classification can be somewhat less than the more advanced architectures but the fact that the model has reduced computation needs by far makes it (the model) an effective edge AI application. Examples of systems that need efficient deep learning models to enable real-time inference with only a small amount

of hardware usage include mobile devices, autonomous drones, embedded vision systems and IoT-based monitoring platforms. The proposed LightCNN-Att architecture shows that small networks may not be as significant as they can be in cases when the networks are created with effective feature refinements.

ACKNOWLEDGMENT

The authors sincerely thank the **Department of Computer Science, Lovely Professional University**, for providing the academic environment and resources required for this research. The authors also appreciate the guidance of faculty members and the support of colleagues whose valuable suggestions and discussions contributed to the successful completion of this work.

References

1. A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint*, 2017. https://arxiv.org/abs/1704.04861
2. A. Howard et al., "Searching for MobileNetV3," *Proceedings of ICCV*, 2019. https://arxiv.org/abs/1905.02244
3. N. Ma, X. Zhang, H. T. Zheng and J. Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," *ECCV*, 2018. https://arxiv.org/abs/1807.11164
4. N. Ma et al., "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," *Lecture Notes in Computer Science*, Springer, 2018. https://doi.org/10.1007/978-3-030-01264-9_8
5. J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," *Proceedings of CVPR*, 2018. https://arxiv.org/abs/1709.01507
6. S. Woo, J. Park, J. Y. Lee and I. S. Kweon, "CBAM: Convolutional Block Attention Module," *Proceedings of ECCV*, 2018. https://arxiv.org/abs/1807.06521
7. F. N. Iandola et al., "SqueezeNet: AlexNet-level Accuracy with 50x Fewer Parameters," *arXivpreprint*, 2016.https://arxiv.org/abs/1602.07360
8. M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *ICML*, 2019. https://arxiv.org/abs/1905.11946
9. K. Han et al., "GhostNet: More Features from Cheap Operations," *CVPR*, 2020. https://arxiv.org/abs/1911.11907
10. S. Mehta and M. Rastegari, "MobileViT: Light-weight Vision Transformer," *WACV*,2022. https://arxiv.org/abs/2110.02178
11. X. Zhang et al., "ShuffleNet: An Extremely Efficient CNN for Mobile Devices," *CVPR*, 2018.https://arxiv.org/abs/1707.01083
12. C. Sandler et al., "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *CVPR*, 2018.https://arxiv.org/abs/1801.04381
13. Y. Li et al., "Residual Attention Network for Image Classification," *CVPR*, 2017. https://arxiv.org/abs/1704.06904
14. J. Wang et al., "Non-local Neural Networks," *CVPR*, 2018. https://arxiv.org/abs/1711.07971

15. H. Liu et al., "LiteResNet: Residual Learning with Feature Compression for Mobile Vision," *Neurocomputing*, 2023.
16. T. Chen et al., "DynamicViT: Efficient Vision Transformers with Dynamic Token Sparsification," *IEEE TPAMI*, 2023.
17. A. Coates, A. Ng and H. Lee, "An Analysis of Single-Layer Networks in Unsupervised Feature Learning," *AISTATS*, 2011. https://cs.stanford.edu/~acoates/stl10/
18. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *ICLR*, 2015. https://arxiv.org/abs/1412.6980
19. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training," *ICML*, 2015. https://arxiv.org/abs/1502.03167
20. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *CVPR*, 2016. https://arxiv.org/abs/1512.03385
21. A. Vaswani et al., "Attention is All You Need," *NeurIPS*, 2017. https://arxiv.org/abs/1706.03762
22. Z. Liu et al., "ConvNeXt: A ConvNet for the 2020s," *CVPR*, 2022. https://arxiv.org/abs/2201.03545
23. G. Huang et al., "Densely Connected Convolutional Networks," *CVPR*,2017. https://arxiv.org/abs/1608.06993
24. M. Tan and Q. Le, "EfficientNetV2: Smaller Models and Faster Training," *ICML*, 2021. https://arxiv.org/abs/2104.00298
25. X. Chu et al., "GhostNetV2: Enhance Cheap Operations with Long-Range Attention," *IEEE TPAMI*, 2023.
26. S. Xu et al., "Attention Condensation Networks for Lightweight Image Classification," *Pattern Analysis and Applications*, 2022.
27. Y. Zhang et al., "Efficient Attention-Enhanced CNNs for Embedded Vision," *ACM Transactions on Multimedia Computing*, 2023.