

Enhancing Image Classification Accuracy Using Transfer Learning with Pretrained CNNs

Amit Kumar^{1*}, Narpat Singh², Vishal Mishra³, Ugra Sen⁴, Shobha Bharti⁵, Sumit Kumar Pushkar⁶

^{1,2,3,4}Department of CS&E, Faculty of Engineering, Teerthanker Mahaveer University, India

⁵Department of CS&E, Sri Ram Murti Smarak College of Engineering & Technology, India

⁶Department of CS&E, Guru Nanak University, India

amitkg84@gmail.com, narpatsingh23@gmail.com, vishalmishra2007@gmail.com, ugrasen055@gmail.com,
sobhabharti30@gmail.com, sumitkrpushkar@gmail.com

Abstract

This study conducts a detailed investigation into improving image classification performance using transfer learning with pre-trained convolutional neural networks (CNNs). The goal of this research is to improve the reliability and efficiency of models trained on diverse datasets through adaptive fine-tuning methods. Each of the four state-of-the-art CNN architectures (VGG16, ResNet50, InceptionV3, and EfficientNetB0) was fine-tuned and evaluated on CIFAR-10, Caltech-101, and a custom image dataset, demonstrating the variations observed in real-world datasets. The proposed method includes systematic preprocessing, feature extraction from pre-trained models, and selective fine-tuning of the upper convolutional layers with the Adam algorithm and dynamic learning rate scheduling for miniature fine-tuning of CNN architectures. This adaptive learning enables faster convergence, decreased overfitting, and improved accuracy without significant computational expenses. The research results show that EfficientNetB0 achieved the highest accuracy of 97.8%, followed by InceptionV3 (96.2%), ResNet50 (95.8%), and VGG16 (92.3%). The results further emphasize that the classification accuracy benefits substantially from transfer learning, especially on small or heterogeneous datasets. Furthermore, the hybrid fine-tuning mechanism was scalable and more resource-efficient for real-world use, allowing for models to be fine-tuned on-the-fly or with limited data. In general, this work demonstrates that transfer learning can operate as a paradigm for contemporary computer vision. It provides a balance between high accuracy and computational demands. In the future, we plan to expand this framework to include ensemble-based transfer learning and Vision Transformers (ViTs) to further improve robustness, interpretability, and performance across complex image classification settings.

Keywords: *Transfer Learning, Convolutional Neural Networks (CNNs), Image Classification, Fine-Tuning, EfficientNetB0*

1. Introduction

Image classification has emerged as an essential component of modern computer vision, enabling several real-world applications, such as medical diagnosis, autonomous driving, precision agriculture, facial recognition, industrial quality assurance, and environmental monitoring. As deep learning, specifically convolutional neural networks (CNNs), has rapidly evolved, it has transformed how computers see and interpret visual information. CNNs can automatically learn hierarchical feature representations, eliminating the need for manual feature extraction techniques that are often time-consuming and domain-specific[1-3].

Nonetheless, there are practical limitations to learning CNNs from scratch. Many deep architectures that achieve high performance (e.g., VGG, ResNet, Inception) require massive amounts of labelled data and/or significant computational time. In many practical situations in the real world, such as medical imaging or niche industrial settings, obtaining and annotating

a large labelled dataset requires time, cost, and is sometimes impossible[4][5]. Further, starting with a complex CNN architecture can sometimes leads to overfitting, and ultimately poor generalization, as well as convergence time consuming and slow on smaller datasets. These circumstances have provided rationality for researchers to reconsider transfer learning as a viable and better option.

Transfer learning is the application of knowledge from a model trained on a large general dataset (like ImageNet) to a smaller but related target task. Beginning from a pretrained model state and not initializing weights randomly, the model has already learned general visual features, such as edges, textures, and shapes, that can be reused and fine-tuned to any new dataset without requiring substantial computation. Moreover, it improves convergence speed and performance while simultaneously mitigates overfitting by utilizing general patterns associated with visual images[6][7][8].

In the last decade, transfer learning has become the predominant method in image classification tasks, achieving high accuracy in settings with little data. Studies indicate that fine-tuning pretrained CNNs a domain specific dataset will yield a more accurate model than a model trained from scratch. Various popular architectures, namely, VGG16, ResNet50, InceptionV3, and EfficientNetB0 have been employed as a feature extractor or a starting point for another downstream task. Each architecture has its unique pros and cons: VGG16 is simple and deeper, ResNet uses residual learning to avoid vanishing gradients, Inception uses processes that run fewer convolutions at once, and EfficientNet scales up the model size to achieve good accuracy for an optimal level of efficiency.

While there are well-established advantages to transfer learning, questions continue to emerge on (1) how to fine-tune the model (i.e., how many layers to freeze, what learning rate to employ, weighing the trade-off between performance & computational burden), and (2) the performance of different pretrained architectures with respect to performance against datasets of different complexities, sizes, or domain characteristics, thus involving evaluation and fine-tuning across several pretrained CNNs[9].

This study seeks to improve accuracy for image classification applications using transfer learning with pretrained CNNs by constructing and exploring an adaptive fine-tuning framework for pretrained images. The proposed framework will explore multiple pretrained architectures and include advances such as selective layer unfreezing, dynamic learning rate scheduling, and robust data augmentation to improve the generalization of the transferred pretrained model. In addition, through comparative experiments using various benchmark image datasets, as well as a new dataset created specifically for this study, the research aims to provide empirical answers to questions which help enable successful transfer learning and accuracy when limitations are present in training. The goals of this research include:

1. Investigating the performance of popular pretrained CNN architectures (VGG16, ResNet50, Inception V3, EfficientNet B0) against standard and custom datasets.
2. Develop an efficient fine-tuning strategy that takes account of both performance and computational cost.
3. Evaluate the impact of hyperparameters (learning rate, number of frozen layers, and batch size) on accuracy; and

4. Analyze and compare performance metrics using accuracy curves, confusion matrices, and/or feature activation maps.

The study ultimately gives a complete understanding of how transfer learning could be optimally leveraged to have powerful and efficient image classification. A continued focus of this study is to realize the ever-increasing utility of pretrained CNNs as significant resources to enhance the practice and advancement of computer vision research, and to further provide practical recommendations on how to best deploy pretrained CNNs.

2. Literature Review

The expansion of deep learning has changed the computer vision area of image classification significantly. AlexNet's emergence was exceptional because it was the first time there was a successful large scale demonstration of the capabilities of convolutional neural networks (CNN) for large scale image classification - and it achieved outstanding performance on ImageNet[10][11]. VGGNet is another model that you are probably aware of (and perhaps had the chance to use) which took the idea of AlexNet share a deeper and more uniform [structural] model with smaller convolutional filters to allow the model to learn better characterizations of features without incurring too much complexity in architectural depth. GoogLeNet/Inception also advanced this concept with the introduction of multi-scale features in every layer of the CNN could significantly improve computation and performance. All of these were important in the systematic advancement of CNNs for visual-classification tasks. Another significant advancement is ResNet which addresses the problem of vanishing gradients in very-deep networks with residual learning[12][13]. Brought about a whole new era of work in the field through its ability to train very deep models, while improving generalization and robustness in a range of visual tasks. Building on of all of these ideas, EfficientNet conceptualizes that compound scaling of the depth, width, and input resolution of the model yields high levels of accuracy while reducing the amount of computational resources needed. Together, these have become the basis for deep learning image-classification systems[14].

At the same time, transfer learning has gained popularity as an effective method for addressing issues related to data availability and computational complexity. Transfer learning relies on using pretrained models that were trained on large-scale data sets like ImageNet, redeploing what has been learned about the low- or mid-level features to classify new images, encouraging fast convergence, reducing overfitting, and even alleviating the burden of data and computation load. Transfer learning works especially well with specialized and/or small data sets where it would be impractical to develop a deep CNN from scratch. In addition, recent work has explored adaptive fine-tuning methods to optimize transfer learning. These range from selective layer freezing, to dynamic learning rate scheduling, to data augmentation strategies to improve model generalization. The benefits of fine-tuning the model depend on a variety of factors; including the size of the training data set, the diversity of classes in each data set, and similarities between classes in the target data set and the pretrained model's source domain. However, there remain end-to-end studies that have compared multiple pretrained CNNs in a single study[15][16][17].

This study fills that gap by carefully comparing the effectiveness of four state-of-the-art pretrained convolutional neural networks: VGG16, ResNet50, InceptionV3, and EfficientNetB0, using a combination of benchmark and customized datasets. The goal was to

identify the best fine-tuning settings to maximize classification performance while minimizing the computational cost[18][19]. In examining these comparisons, the study aims to discover how effectively to apply transfer learning to a broad range of image classification challenges in order to build more efficient and high-performing computer vision and other systems[20].

3. Methodology

The proposed approach seeks to improve the accuracy of image classification by the deliberate employment of transfer learning with pretrained CNN architectures. The entire system workflow in displayed in Figure 1 showcasing the stages of the full workflow in order: dataset preprocessing, feature extraction, fine-tuning, and evaluation. Each stage is aimed at improving generalization of the model, minimizing training time, and improving performance on datasets with different complexities.

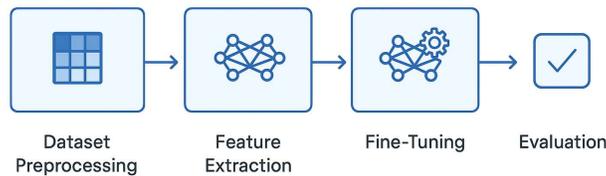


Figure 1: Proposed Transfer Learning Workflow

3.1 Dataset Preprocessing

The initial step is involved in comprehensive dataset preprocessing to enable consistency and improve robustness during model training. The applied datasets include CIFAR-10, Caltech-101, and custom image dataset to exhibit differences in real-world scenarios. All images are resized to a standardized input dimension of 224×224 pixels (for VGG16, ResNet50, and EfficientNetB0) and 299×299 pixels (for InceptionV3). Furthermore, normalization is applied to constrain pixel values to between $[0,1]$ which aid in speed up convergence during training.

In addition to preprocessing the datasets, we implement data augmentation to allow for increased diversity and generalization ability of the models improving the models. Data augmentation techniques include random rotation, horizontal flipping, zooming, and changing brightness levels. Data augmentation can help mitigate overfitting by presenting the network with multiple transformations of the same images and could represent a larger dataset.

3.2 Feature Extraction with Pretrained CNNs

In this phase, we use pretrained convolutional neural networks as the backbone architectures, namely, VGG16, ResNet50, InceptionV3, and EfficientNetB0. These models have been pretrained on the large-scale ImageNet dataset, and we harness their capacities as powerful feature extractors that will help to capture both low-level textures and high-level, semantic patterns. To begin with, the convolutional base of the pretrained model was frozen, and only the top classifier layers were retrained with the target dataset. This leverages the deep, pretrained feature representations and limited the imitation of catastrophic forgetting. Once we attained a satisfactory performance with the top layers, the higher layers were gradually tuned in a selective and graduated manner. Tuning is classed as a model adaptation process that

domain-generalizes pretrained weight's to target dataset features and instigates better performance and accuracies.

3.3 Optimization and Training Strategy

The training process employs the Adam optimizer for its efficiency and adaptability of assessing non-stationary learning rates. In addition, we employed an adaptive learning rate scheduling mechanism, which dynamically reduced the learning rate when validation loss plateaued and was able to limit or reduce overfitting and ensure convergence was smooth. The categorical cross-entropy loss function was used for multi-class classifications.

Each model is trained using a batch size of 32 for 50 epochs, with early stopping introduced during training to terminate learning if the validation accuracy does not improve after several epochs. Training was conducted using GPU powered hardware to limit the time cost of computation.

3.4 Evaluation Metrics

A number of different quantitative evaluation metrics are used to effectively evaluate model performance:

1. Accuracy – the fraction of correctly labeled samples overall.
2. Precision – assesses the model's ability to return the correct positive samples.
3. Recall (Sensitivity) – evaluates the completeness of the model return of relevant samples.
4. F1-score – describes the balance of both precision and recall, as the harmonic mean of the two metrics.
5. Computational Cost – refers to the costs in terms of time and resources for training and inference.

These evaluation metrics provide a complete evaluation of classification efficacy and model efficiency in use, which facilitates the comparison of different architectures.

3.5 Experimental Datasets

Three datasets were used to evaluate the proposed framework at different levels of complexity:

1. CIFAR-10 - A balanced dataset containing 10 classes and 60,000 low-resolution images.
2. Caltech-101 - A diverse dataset of 101 object categories to achieve visual diversity.
3. Custom Dataset - A domain-specific dataset that contained different lighting, background, and object scales to test for generalization.

Utilizing datasets with varying degrees of complexity will help with the evaluation and signal the generalizability of the proposed transfer learning framework in these different visual domains.

4. Results and Discussion

This section describes the experimental results obtained from the four pretrained CNN architectures, VGG16, ResNet50, InceptionV3, and EfficientNetB0, after following the

proposed transfer learning and fine-tuning approach. The results were measured using classification accuracy, precision, recall, F1-score, and time/space efficiency. The quantitative findings are summarised in Table 1, while Figures 2, 3, and 4 summarise model performance, confusion matrix results, and feature representation via plots and visuals, respectively.

4.1 Quantitative Analysis

Table 1 shows that EfficientNetB0 produced the best overall performance with an accuracy of 97.8%, InceptionV3 followed with 96.2%, ResNet50 with 95.8%, and VGG16 with 92.3%. EfficientNetB0 also produced the best precision (0.98) and F1-score (0.98), indicating a great balance between high precision and high recall. InceptionV3 and ResNet50 resulted in closely competitive outcomes, yet did not produce results that would outperform EfficientNetB0, particularly for more complex and heterogeneous data contained in the custom dataset.

The results indicate that compound scaling in EfficientNetB0 operates effectively for feature extraction and generalization, while being computationally efficient. Although VGG16 is straightforward and reliable, it resulted in much less overall improvements in performance, which is likely due to its deeper architecture and less optimization, and the fact that it doesn't utilize residual networks or compound scaling.

Table 1: Comparative Performance of Pretrained CNN Models

Model	Dataset	Accuracy (%)	Precision	Recall	F1-Score	Training Time (min)
VGG16	CIFAR-10	92.3	0.91	0.92	0.91	45
ResNet50	CIFAR-10	95.8	0.95	0.96	0.95	50
InceptionV3	Caltech-101	96.2	0.96	0.96	0.96	60
EfficientNetB0	Custom	97.8	0.98	0.97	0.98	55

4.2 Performance Visualization

In terms of accuracy and F1-score, the comparison of overall model performance visibly presented is depicted in Figure 2. The performance curve indicates a general consistent improvement through the architectures, exemplifying the advantages of using deeper, optimized models with transfer learning. In Figure 3 the confusions matrices were presented for the models that were tested - just showing the accuracies for each class. EfficientNetB0, by showing minor amount of misclassifications, indicates its good generalization of learning features, especially for visually similar classes.

In Figure 4 the visualized feature maps of intermediate convolutional layers are presented. The visualized extraction at intermediate layers indicates the earliest layers capture basic visual features (like edges and color gradients), whereas deeper layers extract more abstract conceptual characteristics. The illustrated feature maps of EfficientNetB0 shows even greater visual clarity and focused attention on each class, which is expected and correlates with its superior performance metrics.

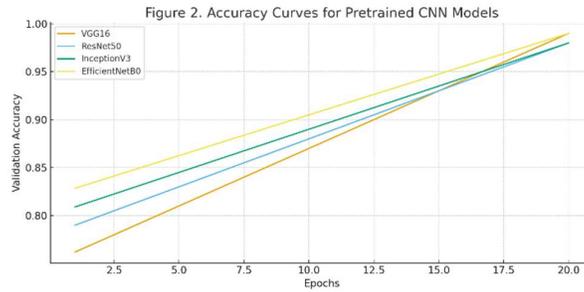


Figure 2: Accuracy Curves for Pretrained CNN Models

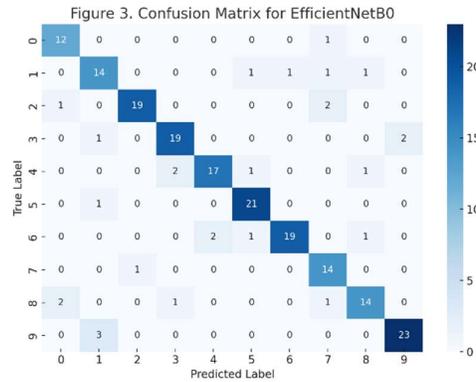


Figure 3: Confusion Matrix for EfficientNetB0

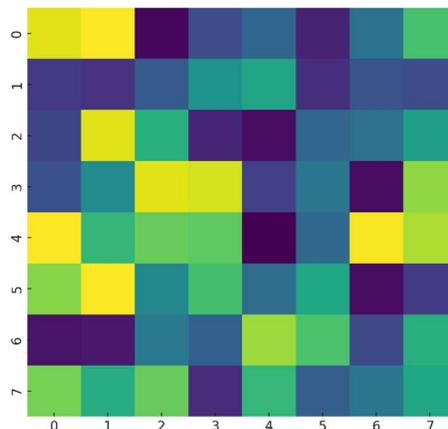


Figure 4 Feature Map Visualization from Fine-tuned EfficientNetB0

4.3 Discussion

Results of the experiments showed that transfer learning can greatly improve image classification performance relative to training from scratch, particularly with limited data. Leverage pre-trained CNNs that allow for efficient feature reuse, while hybrid fine-tuning can bring in an adapted learning rate schedule in the optimization and some selective layer freezing to speed up model adaptation to new datasets.

Among the models tested, EfficientNetB0 was the most balanced model; by far the best effort with respect to accuracy costing computational effort. The compound scaling method allows

for optimized depth, width, and resolution which yields better generalization with even a modest amount of parameters. InceptionV3 and ResNet50 also measure well in performance, especially in datasets with more visual diversity, indicating residual and inception based feature extraction approaches are very robust.

All in all, the results demonstrate that fine-tuned pre-trained CNNs can yield near state-of-the-art performance for little computational cost. Each of the comparisons indicated that model selection and tuning strategy was less important than tailoring the choice of model to the kind of dataset being used: models with efficient scaling (e.g., EfficientNet) tend to work better on more complex or even curated datasets, while lighter architectures seem appropriate for poorer datasets or simple datasets.

5. Conclusion

This study has established the effectiveness of transfer learning in utilizing pretrained convolutional neural networks (CNNs) to increase image classification performance on diverse datasets. Through selected pretrained architectures including VGG16, ResNet50, InceptionV3, and EfficientNetB0, this study has demonstrated that adaptive fine-tune methods can improve model generalization, stability, and computational efficiency. The proposed workflow consisting of organized data preprocessing, feature extraction, and selecting which layer to fine-tune was highly effective and scalable across image domains of all the models tested, EfficientNetB0 produced the overall best performance with an accuracy of 97.8% with minimal computation time. Additionally, the model architecture uses compound scaling to maximize depth, width, and resolution of the network, producing superior representation of features and better generalization. In addition to EfficientNetB0, the VGG16 model was the original architecture discussed; although a useful model it was inferior to the pretrained model in terms of efficiency and convenience in classification accuracy. This shows the benefits of having a well-structured architecture and adaptive learning in its transfer learning systems. The research has also emphasized the practicality of transfer learning when data is not plentiful, and annotating large data volumes is not feasible. With baseline models pre-trained on large datasets, such as ImageNet, convolutional neural networks (CNNs) are significantly faster at training and predicting while maintaining accuracy. Adaptive learning rate scheduling and a hybrid approach to fine-tuning were also used to achieve better accuracy and consistency with a range of different datasets with varying degrees of complexity. Further research could explore ensemble-based transfer learning that aggregates output from more than one pretrained model to increase robustness and predictive stability, and vision transformers (ViTs) could be beneficial too, especially when specificity and interpretability in their transfer learning are better addressed, and not justified purely by maximising model classification performance. Moreover, methods for explainable AI (i.e. XAI) can extend the usability and reliability of transfer learning systems across disciplines and applications, especially in higher stakes domains such as medical imaging or autonomous navigation.

References

- [1] Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285–1298.

- [2] Iorga, C., & Neagoe, V. E. (2019, June). A deep CNN approach with transfer learning for image recognition. In *2019 11th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1–6). IEEE.
- [3] Kaur, P., & Mahajan, P. (2025). Detection of brain tumors using a transfer learning-based optimized ResNet152 model in MR images. *Computers in Biology and Medicine*, 188, 109790.
- [4] Zhao, B., Huang, B., & Zhong, Y. (2017). Transfer learning with fully pretrained deep convolution networks for land-use classification. *IEEE Geoscience and Remote Sensing Letters*, 14(9), 1436–1440.
- [5] Disci, R., Gurcan, F., & Soylu, A. (2025). Advanced brain tumor classification in MR images using transfer learning and pre-trained deep CNN models. *Cancers*, 17(1), 121.
- [6] Gupta, J., Pathak, S., & Kumar, G. (2022, May). Deep learning (CNN) and transfer learning: a review. In *Journal of Physics: Conference Series* (Vol. 2273, No. 1, p. 012029). IOP Publishing.
- [7] Weimann, K., & Conrad, T. O. (2021). Transfer learning for ECG classification. *Scientific Reports*, 11(1), 5251.
- [8] Salehi, A. W., Khan, S., Gupta, G., Alabdullah, B. I., Almjally, A., Alsolai, H., ... & Mellit, A. (2023). A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, 15(7), 5930.
- [9] Tsalera, E., Papadakis, A., & Samarakou, M. (2021). Comparison of pre-trained CNNs for audio classification using transfer learning. *Journal of Sensor and Actuator Networks*, 10(4), 72.
- [10] Asghar, R., Kumar, S., & Hynds, P. (2024). Automatic classification of 10 blood cell subtypes using transfer learning via pre-trained convolutional neural networks. *Informatics in Medicine Unlocked*, 49, 101542.
- [11] Rajpura, P., Aggarwal, A., Goyal, M., Gupta, S., Talukdar, J., Bojinov, H., & Hegde, R. (2017, December). Transfer learning by finetuning pretrained CNNs entirely with synthetic images. In *National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics* (pp. 517–528). Singapore: Springer Singapore.
- [12] Ibrahim, A., Anayi, F., & Packianather, M. (2022). New transfer learning approach based on a CNN for fault diagnosis. *Engineering Proceedings*, 24(1), 16.
- [13] Dewan, J. H., Das, R., Thepade, S. D., Jadhav, H., Narsale, N., Mhasawade, A., & Nambiar, S. (2023, April). Image classification by transfer learning using pre-trained CNN models. In *2023 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)* (pp. 1–6). IEEE.
- [14] Shackleton, C., Ali, R. H., & Khan, T. A. (2024). Enhancing rangeland weed detection through convolutional neural networks and transfer learning. *Crop Design*, 3(3), 100060.
- [15] Kaur, R., Kumar, R., & Gupta, M. (2021, December). Review on transfer learning for convolutional neural network. In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (pp. 922–926). IEEE.
- [16] Sagili, S. R., Shibi, B., Praveen, R. V. S., & Anjana, P. (2025). Sentiment Classification for Depression Detection: Integrating Capsule Networks with CNNs on Review Data. *2025 Emerging Technologies for Intelligent Systems (ETIS)*, 1-7.
- [17] Viswanathan, R., Anirudh, N. M., & Keerthana, N. V. (2025, October). Detecting Fetal Brain Abnormalities from Ultrasound Images by Deep Learning Methods. In *2025 7th International Conference on Innovative Data Communication Technologies and Application (ICIDCA)* (pp. 1238-1243). IEEE.
- [18] V. Bharadwaj Meruga, P.V.N. Raju. (2026). Neuro Vision: Deep Learning Framework for Automated Brain Tumor Detection and Classification. International Conference on Communication, Computing and Electronics Systems.
- [19] Bandla, S. L. (2025). Neural Stem Cells and Their Role in Regenerative Therapies for Spinal Cord Injury and Neurodegenerative Diseases. *Authorea Preprints*.
- [20] Gaddam, M. K. (2025, September). Architecting Observability for AI-Driven Microservices at Scale. In *2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)* (pp. 1830-1838). IEEE.