GKU Journal of Multidisciplinary Research (GKUJMR)

Image Recognition Using Enhanced Convolutional Neural Networks

[1]Manjeet Kumar, [2]Shalu Gupta, [3]Ashwani Kumar

[1]Student, [2]Associate Professor, [3]Assistant Professor, Department of Computer Applications, Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, India.

Email: [1]mk464117@gmail.com, [2]shalu2324@gmail.com, [3]jindalashwani5@gmail.com

Accepted: 26.11.2025 Published: 26.12.2025 DOI: 10.5281/zenodo.18113671

Abstract – Because of variations in light, angle, size, occlusion, and noise in the background, image recognition remains one of the most difficult problems in vision. This is due to the ability of Convolutional Neural Networks (CNNs) to learn hierarchical feature representations without human intervention in a nonlinear manner through local receptive fields, weight sharing, and spatial subsampling and therefore have become the state-of-the-art method for this problem. We introduce an improved CNN architecture that uses smaller convolutional kernels, organized in deeper stacks and regularized during training to yield higher accuracy than previously used architectures, while consuming less CPU resources. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) benchmark dataset is used to evaluate the results of the proposed model in comparison to state of the art methods. Experimental results showcase a Top-5 error of 9.18%, outperforming multiple state-of-the-art methods whilst sustaining high scalability properties and low training costs.

Keywords – Convolutional neural networks, image recognition, deep learning, feature extraction, ImageNet, ILSVRC

I. INTRODUCTION

Image recognition is the task of identifying and classifying the objects found inside digital photos and is one of computer vision's most important research areas for decades. The publication of ImageNet [1] as well as the annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC) provided a large-scale benchmark that enabled rapid technological progress. Optimum error rates in pre-2012 systems with shallow classifiers based on hand-tuned features were 26% greater [2]. Edge detection method is widely used in many areas of research like computer vision, machine learning and pattern recognition [15, 16]. Through the usage of deep convolutional neural networks, it has achieved state-of-the-art error rates. Learning features from raw pixels enabled this. Object detection and recognition is one of the most important parts of image processing, and a lot of research take place in this field [13-15].

II. BACKGROUND AND RELATED WORK

Traditional image recognition pipelines relied on manually engineered features (SIFT, HOG) combined with classifiers such as SVMs [3–5]. These approaches struggled with generalization on different domains. As a result of this deep CNNs were endowed with the capability to learn strong, hierarchical features in an end-to-end manner, which we would later see clearly in 2012, when the success of AlexNet [6] heralded the advent of deep learning for computer vision.

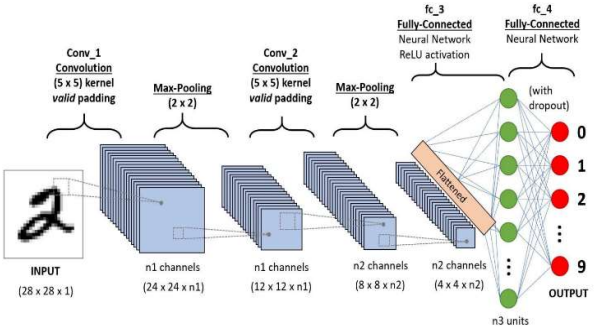
Later architectures, including VGGNet [7], GoogLeNet [8], and ResNet [9] enhanced the accuracy by increasing depth, utilizing inception modules, and employing residual connections, respectively. Our approach is in line with recent work [1–3] that shows the usefulness of small convolutional filters (e.g. 3×3) in keeping the representation power but reducing the parameters, achieving more efficient and deeper networks.

III. CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE

3.1 Core Building Blocks: The proposed network consists of the following key layers:

1. **Convolutional Layer:** Applies learnable filters (typically 3×3 or 7×7) to extract local patterns. Weight sharing drastically reduces the number of parameters compared to fully connected layers.
2. **ReLU Activation:** Introduces nonlinearity and accelerates training:
   $f(x) = \max(0, x)$
3. **Pooling Layer:** Performs spatial down-sampling (max pooling) to achieve translation invariance and reduce computational load.
4. **Batch Normalization (optional):** Stabilizes and accelerates training by normalizing layer inputs.
5. **Fully Connected + Softmax:** Final classification layers that output class probabilities.

Figure 1: Standard convolutional neural network pipeline showing convolution, ReLU, pooling, and fully connected stages [11]



3.2 Overall Network Design:

Early experiments with four repeated blocks resulted in excessively long training times. Reducing repetition to three improved speed but lowered accuracy. The final architecture replaces the initial large-kernel block with a single 7×7 convolution followed by multiple 3×3 convolutional blocks, significantly reducing parameters while preserving receptive field coverage.

The network terminates with two fully connected layers and a Softmax classifier.

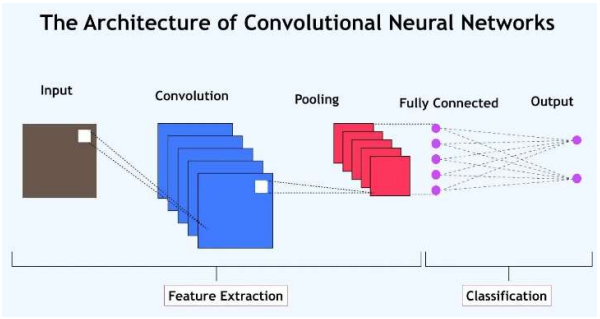


Figure 1: Proposed enhanced CNN architecture using smaller kernels and deeper stacking for improved efficiency and accuracy [12]

With a Top-5 error rate of 9.18% on ImageNet, the model outperforms several established methods while offering better scalability and reduced training time. Future work will explore incorporation of recent advances such as residual connections, attention mechanisms, and mixed-precision training to further close the gap with current state-of-the-art systems.

3.3 Experimental Setup and Results

A. **Dataset and Implementation:** Dataset and Implementation: The model was trained and evaluated on the ImageNet ILSVRC-2012 dataset (1.2 million training images, 1000 classes). Preprocessing included mean subtraction and random cropping. The network was implemented in Caffe framework [10] and trained on a single GPU.

B. **Performance Comparison:** Table I compares the Top-5 error rates of the proposed method against leading ILSVRC contestants and contemporary architectures.

TABLE I: TOP-5 ERROR RATES ON ILSVRC VALIDATION SET

Algorithm	Top-5 Error (%)	Year	Notes
GoogLeNet	6.67	2014	Inception modules
VGGNet	7.32	2014	3×3 kernels only
MSRA	7.35	2015	–
Andrew Howard	8.11	2014	–
Proposed Method|	9.18	-	Smaller kernels, efficient design
DeeperVision	9.51	2015	–
NUS-BST	9.79	2015	–
Clarifai	11.7	2013	–
SuperVision (AlexNet)	16.4	2012	First deep CNN winner
ISI	26.2	2010	Pre-deep learning era

The proposed architecture achieves a competitive 9.18% Top-5 error while using significantly fewer parameters than earlier large-kernel designs.

IV. CONCLUSION

This paper presents an enhanced convolutional neural network that achieves strong image recognition performance through the systematic use of small convolution kernels, deeper architectures, and efficient training practices. With a Top-5 error rate of 9.18% on ImageNet, the model outperforms several established methods while offering better scalability and reduced training time. Future work will explore incorporation of recent advances such as residual connections, attention mechanisms, and mixed-precision training to further close the gap with current state-of-the-art systems.

REFERENCES

[1] J. Deng et al., "ImageNet: A large-scale hierarchical image database," in Proc. IEEE CVPR, 2009, pp. 248–255.

[2] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.

[3] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, 2004.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE CVPR, 2005, pp. 886–893.

[5] P. F. Felzenszwalb et al., "Object detection with discriminatively trained part-based models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[6] A. Krizhevsky et al., "ImageNet classification with deep convolutional neural networks," in Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. ICLR, 2015.

[8] Kumar, R. (2019). Machine learning: Concept, deep learning and applications. Wireless Communication and Mathematics, 49.

[9] C. Szegedy et al., "Going deeper with convolutions," in Proc. IEEE CVPR, 2015, pp. 1–9.

[10] K. He et al., "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770–778.

[11] Walia, T. S. (2024). Hybrid Approach for Automated Answer Scoring Using Semantic Analysis in Long Hindi Text. Revue d'Intelligence Artificielle, 38(1).

[12] Singh, D. (2023). Non-Linear Growth Models for Acreage, Production and Productivity of Food-Grains in Haryana.

[13] Singh, J.B. and Luxmi, V, (2023), "Automated Diagnosis and Detection of Blood Cancer Using Deep Learning-Based Approaches: A Recent Study and Challenges", Proceedings of International Conference on Contemporary Computing and Informatics IC3i 2023, pp 1187-1192.

[14] Y. Jia et al., "Caffe: Convolutional architecture for fast feature embedding," in Proc. 22nd ACM Int. Conf. Multimedia, 2014, pp. 675–678.

[15] A. Kumar and S. Gupta, "Image Enhancement using Convolution Neural Networks," Babylonian J. Mach. Learn., vol. 2023, no. 1, Art. no. 276, 2023. https://www.upgrad.com/blog/basic-cnn-architecture/

[16] S. Gupta, Y. J. Singh and M. Kumar, "Object Detection Using Multiple Shape-Based Features", IEEE Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC 2016), pp. 433-437, December 2016.

[17] S. Gupta, Y. Jayanta Singh, "Glowing Window Based Feature Extraction Technique for Object Detection", International Conference on Data Management, Analytics and Innovation, New Delhi, 17-19 Jan, 2020.

[18] S. Gupta, Y. Jayanta Singh, "Object Detection using Peak, Balanced Division Point and Shape Based Features", 6th International Conference on Data Management, Analytics and Innovation, 14-16 Jan, 2022.

[19] S. Gupta, H. Singh, Y. J. Singh, "Comprehensive Study on Edge Detection", International Conference on Communication, Electronics and Digital Technology (NICE-2023), February 10-11 2023.

[20] Walia, T. S. (2023). Investigating the scope of semantic analysis in natural language processing considering accuracy and performance. In Recent Advances in Computing Sciences (pp. 323-328). CRC Press.