

Analysis: Performance Comparison of Various CNN Models on different Datasets

Ayush Kumar¹ and Dr. Ajay Kaul²

¹⁻²Shri Mata Vaishno Devi University/School of Computer Science and Engineering, Katra, India
Email: kumarayush554@gmail.com, ajay.kaul@smvdu.ac.in

Abstract—Within the field of image processing, the term "computer vision" is growing in use. There is a big need for automatic object recognition now that computer vision applications are starting to take off. The computer vision community has profited from the convolution neural network (CNN), which has excelled in many disciplines including speech recognition, object categorization, segmentation, video editing, and many more. This manuscript's key contribution is to contrast several architectural evolutions based on architectural development, advantages, and disadvantages. A breakdown of the elements of CNN, advantages, and disadvantages of different variations on CNN, knowledge gaps or unresolved problems, CNN applications, and future research directions are also included. When it comes to machine learning issues, CNN performs superbly. In this essay, we will outline and identify every component of CNN, its key components, and its concerns.

Index Terms— classification, convolutional neural networks, deep learning, performance evaluation.

I. INTRODUCTION

Since its inception, neural networks have been used to simulate human brain function as closely as feasible. The amount of data created and used has significantly expanded with the advancement of technology, they are made up of neurons that self-optimize through learning, and convolutional neural networks (CNNs) are comparable to conventional ANNs. Due to its deep feed-forward design, Convolution Neural Network (CNN) is remarkably more generalizable than networks with fully connected layers. The key benefit CNN has over its rivals is that it uses machine learning to find crucial facts swiftly and automatically, which makes it the most well-liked. CNN needs less preprocessing in contrast to other techniques. The best learning system for understanding visual content is hence a CNN. People outside of academics are interested in CNN's success. Among the businesses working to create and enhance CNN architecture are Microsoft, AT&T, Google, NEC, and Facebook. The fundamental CNN design has so been modified in several ways. Normal neural networks are reconfigured by CNN, but it has the enticing feature of neurons with adjustable weights and biases. Each neuron does a dot product after taking in many inputs, and optional nonlinearity can then occur.

II. RELATED WORK

Md. Rakibul et al.[1] For training and testing, they were able to obtain a dataset with 350 COVID-19 lung CT frames and 400 non-COVID-19 CT frames. Using the data augmentation strategy, they were able to get 1595 CT

frames for those who lack COVID-19 and 1750 CT frames for individuals with COVID-19. 20% of the lung CT frames for testing and 80% of the frames for training. Our architecture included 82,146 different trainable parameters in total. Making the model more practical and effective is our goal for the future.

Jing et al.[2] introduced a methodology using the Alex-Net model to classify scenes in images. This study introduces a revolutionary scene classification method.

Srikanth Tammina[3] utilizes a Deep Convolutional Neural Network with the VGG-16 model, one of the pre-trained models, to identify images. A validation accuracy of 72.40% was produced by the first model created using a convolutional neural network. With image augmentation, the model improved its performance and reached an accuracy of 79.20%. Finally, the dataset was trained on the pre-trained models (VGG-16) and was improved with image augmentation with an accuracy of 95.40%.

Kaiming He et al.[4] described a residual methodology that will support the training of deeper networks that were used before. On the ImageNet dataset, deep residual nets up to 152 layers are being analyzed which are 8 levels deeper than VGG networks but are still less complicated and these 13 residual nets achieve 3.57 error. This ranked first in the 2015 ILSVRC classification task.

Riaz Ullah Khan et al.[5] use a combination of GoogleNet and ResNet models for malware detection. To detect the unusual or novel type of malware in this work, he used two different models. The ResNet and GoogleNet models, which come from two separate platforms—ResNet is owned by Microsoft, while GoogleNet is a work of Google's intellectual property—are studied and tested. For the models' training and validation, two different kinds of datasets are used. One of the datasets—a combination of 10,868 items, all of which are binary records—was obtained from Microsoft. Additionally, nine different classes have been separated from these records. There are 3000 innocuous files in the second dataset. The datasets underwent an opcode conversion process from EXE files to images. We achieved testing precision on GoogleNet of 88.4% and ResNet of 74.49%.

Peng, Cheng & Liu, and Yikun[6] Introduced a systematic and comparative analysis of the structure of the classification of the model. An enhanced-ResNet-v2 model based on CNN was created to increase the accuracy of the image classification. Then, multi-scale depth-wise detachable inversion was introduced. This manuscript also introduced a channel filtering module for feature extraction that was based on global information comparison. This model outperforms the others in each dataset because of the data augmentation, batch normalization, and modification of the learning rate. In this, the accuracy percentage can reach 94.8%.

Tugrul et al.[7] In this research, the researcher provided an analysis of the identification of leaf disease in agricultural plants based on CNN over the last 5 years. At most 100 publications were reviewed for the preparation of datasets, detection methods, comparison of popular CNN models, its applications in agricultural settings, detailed descriptions of performance evaluation, and solutions.

III. DATASET

The datasets used for the above work are:

A. CIFAR-10

A dataset containing 60,000-32x32 colored images in ten classes, having 6,000 images within a single class. These classes are automobile, airplane, cat, dog, deer, cat, horse, frog, truck, and ship. This dataset is commonly used in the field of machine learning and computer vision for developing and evaluating image-classification models.

CIFAR-10 is a subset of the larger CIFAR-100 dataset, consisting of 100 classes with 600 images per class. The CIFAR-10 dataset created by the research community at the University of Toronto was released at first during 2009. This dataset has since become a widely-used benchmark for image classification tasks and has been used in many research papers and competitions.

Some popular approaches for classifying images in the CIFAR-10 dataset contain models from deep learning like CNN-convolutional neural networks and residual neural networks (ResNets). State-of-the-art performance in CIFAR-10 as of 2021 is achieved by using a combination of techniques, including data augmentation, mixup, and adversarial training.

B. IMAGENET

ImageNet is a collection of 1000 classes having 1.28 million training images and 50,000 validation images. Hundreds of thousands of images are used to illustrate each node of the WordNet hierarchy in ImageNet, a collection of images that is structured by the WordNet hierarchy (currently just the nouns). Deep learning and computer vision research have benefited greatly from the initiative. For study purposes only and non-commercial

usage, the data is freely available. ILSVRC 2012, sometimes referred to as in ImageNet, is a large-size image dataset for image categorization.

IV. CNN COMPETITIVE ARCHITECTURES

As per the computer science bibliography, 367 papers show an architectural change in the last seven years. These models have become so complex and large that analyzing them has become quite difficult. There were approximately 100 and 60 articles published in the year 2000 and 2021 respectively. In this paper, several CNN versions are highlighted.

A. Mask R-CNN (2017)

It is an extension of the Faster R-CNN algorithm. The key innovation of Mask R-CNN is its ability to generate segmentation masks for each object detected in an image, in addition to classifying and localizing objects with bounding boxes. This is achieved by adding a small fully convolutional network on top of the existing Faster R-CNN architecture, which predicts a binary mask for each object instance.

Mask R-CNN has attained advanced outcomes on different benchmarks for object detection and segmentation tasks and has been widely adopted in computer vision research and applications. It is a powerful tool for tasks such as object recognition, instance segmentation, and object tracking in images and videos.

B. YOLOv4 (2020)

YOLOv4 (You Only Look Once version 4) is an object detection model developed by Joseph Redmon and his team at the University of Washington, and later maintained by Alexey Bochkovskiy and his team at the AI research company, YOLO AI.

YOLOv4 is the fourth iteration of the YOLO model family and features several improvements over its predecessors, including improved accuracy, speed, and robustness. Some of the key features of YOLOv4 include:

- 1 Advanced architecture: YOLOv4 uses a complex architecture that combines features from different detection algorithms to improve accuracy and robustness.
- 2 Backbone networks: YOLOv4 includes a range of backbone networks, such as CSPDarknet53 and Scaled-YOLOv4, that provide a solid foundation for feature extraction.
- 3 Training techniques: YOLOv4 uses advanced training techniques such as CutMix and Mosaic, which help prevent overfitting and improve generalization.
- 4 Post-processing: YOLOv4 employs a range of post-processing techniques, such as IoU-aware and DIOU-NMS, to refine object detection results.

Overall, YOLOv4 is considered to be one of the most accurate and efficient object detection models available, with state-of-the-art performance on several benchmark datasets.

C. GRCNN

Graph-based Residual Convolutional Neural Network[8], a type of neural network architecture is used for graph-based data. Graph-based data refers to data that is represented as a graph, where nodes represent entities, and edges represent relationships between them. Examples of graph-based data include social networks, molecular structures, and traffic networks.

The GRCNN architecture uses residual connections, which are connections that bypass some of the layers in a neural network, allowing the network to learn more complex representations.

GRCNN also uses convolutional layers, which are a type of layer commonly used in image recognition tasks. In GRCNN, convolutional layers are adapted to work with graph-based data, where they operate on local neighborhoods of nodes rather than image pixels.

The use of residual connections and convolutional layers in GRCNN allows it to learn hierarchical representations of graph-based data, which can be used for tasks such as node classification and graph classification.

D. MFRNet

MFRNet(Multi-Level Feature Review Network)[9] are deep neural network designed for image classification tasks. They were introduced in a research paper by Zhang et al. in 2018. The basic idea behind MFRNs is to incorporate both global and local information from different scales of an image to improve classification accuracy. The network architecture consists of multiple branches, each of which extracts features at a different

scale. The feature maps from each branch are then combined and reviewed to obtain a final classification decision.

The feature review process involves passing the combined feature maps through a set of convolutional layers and pooling layers to obtain a compact representation. This representation is then fed into a fully connected layer for the final classification. Overall, MFRNs are a powerful tool for image classification tasks that require the use of both global and local information from different scales of an image.

V. CNN CHALLENGES

CNNs-Convolutional Neural Networks have become a widespread deep learning model for solving a wide range of image and video recognition tasks. However, CNNs also comes with some challenges. Here are some of the key challenges in CNNs:

A. Overfitting

CNNs have a tendency to overfit the training data, which means they become too specialized in recognizing the patterns.

B. Data Augmentation

To avoid overfitting, CNNs require a large amount of diverse training data. However, collecting and annotating such data is often time-consuming and expensive.

C. Choosing the right architecture

CNNs have a complex architecture with many hyperparameters to be tuned, such as the number of filters, filter sizes, number of filter layers, pooling sizes, and activation functions. Choosing the right architecture requires a deep understanding of the problem and experimentation.

D. Computational complexity

CNNs require many computational resources, especially for training large-scale models on high-resolution images. This can limit their practicality for some applications.

E. Transfer learning

CNNs trained on one task may not be directly applicable to another task, as the learned features may not be transferable. Transfer learning techniques can help to overcome this challenge by leveraging pre-trained CNNs for other related tasks.

TABLE I. DIFFERENT CNN ARCHITECTURES

Architecture	Year of publication	Category	Parameter	Error Rate
LENET	1998	Spatial exploitation	0.060M	MNIST:0.952
ALEXNET	2014	Spatial exploitation	60M	IMAGENET:17.49
VGG	2014	Spatial exploitation	138 M	ImageNet: 7.32
GoogleNet	2015	Spatial exploitation	4M	ImageNet:6.69
ResNet	2016	Depth + multi-Path	25.6 M	ImageNet: 3.61 CIFAR-10: 6.432
Convolutional Block Attention Module (ResNeXt101 (32x4d) + CBAM)	2018	It exploited both spatial and feature map information	48.96 M	ImageNet: 5.6

VI. CONCLUSION

CNN, find widespread usage in several applications, containing face recognition, object detection, medical image analysis, and self-driving cars. They have shown state-of-the-art performance in various image recognition tasks, surpassing human-level performance in some cases. They have also been extended to various domains, including speech recognition and natural language processing, using recurrent and attention-based architectures. Despite their success, CNNs still have some limitations, such as their need for large amounts of labeled data and their lack of interpretability. However, ongoing research is addressing these issues and improving the effectiveness and interpretability of CNNs. Overall, CNNs are a powerful and versatile tool for

solving a wide range of computer vision tasks, and they will continue to play a significant role in the development of AI and machine learning in the years to come.

REFERENCES

- [1] Islam, Md & Matin, Abdul. (2020). Detection of COVID-19 from CT Image by The Novel LeNet-5 CNN Architecture. 10.1109/ICCIT51783.2020.9392723.
- [2] Sun, Jing & Cai, Xibiao & Sun, Fuming & Zhang, Jianguo. (2016). Scene image classification method based on the Alex-Net model. 363-367. 10.1109/ICCS.2016.7586482.
- [3] Tammina, Srikanth. (2019). Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. International Journal of Scientific and Research Publications (IJSRP). 9. p9420. 10.29322/IJSRP.9.10.2019.p9420.
- [4] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.
- [5] Khan, Riaz & Zhang, Xiaosong & Kumar, Rajesh. (2019). Analysis of ResNet and GoogleNet models for malware detection. Journal of Computer Virology and Hacking Techniques. 15. 10.1007/s11416-018-0324-z.
- [6] Peng, Cheng & Liu, Yikun & Yuan, Xinpan & Chen, Qing. (2022). Research of image recognition method based on enhanced inception-ResNet-V2. Multimedia Tools and Applications. 81. 1-21. 10.1007/s11042-022-12387-0.
- [7] Tugrul, B.; Elhoucine, E.; Eryigit, R. Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review. Agriculture 2022, 12, 1192. <https://doi.org/10.3390/agriculture12081192>
- [8] Lin, C. GRCNN: Graph Recognition Convolutional Neural Network for Synthesizing Programs from Flow Charts. arXiv 2020, arXiv:2011.05980
- [9] Ma, D. MFRNet: A New CNN Architecture for Post-Processing and In-loop Filtering. arXiv 2020, arXiv:2007.07099v2.
- [10] Zhang, W. Disentangled Dynamic Graph Deep Generation. arXiv 2021, arXiv:2010.07276v2.
- [11] Aex, B. Net2Vis—A Visual Grammar for Automatically Generating Publication-Tailored CNN Architecture Visualizations. arXiv 2021, arXiv:1902.04394v6.
- [12] Zou, C.; Zheng, Y.; Su, Q.; Fu, H. Chiew-Lan Tai Sketch-R2CNN: An Attentive Network for Vector Sketch Recognition. arXiv 2018, arXiv:1811.08170v1.
- [13] Haque, W.A. DeepThin: A novel lightweight CNN architecture for traffic sign recognition without GPU requirements. In Expert Systems with Applications; Elsevier: Amsterdam, The Netherlands, 2021; Volume 168.
- [14] Frizzi, S. Convolutional neural network for video fire and smoke detection. In Proceedings of the IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016; pp. 877–882.
- [15] Hinton, G. Matrix capsules with EM routing. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, 30 April–3 May 2018.
- [16] Justus, D. Predicting the Computational Cost of Deep Learning Models. In Proceedings of the 2018 IEEE International Conference on Big Data Big Data, Seattle, WA, USA, 10–13 December 2018.
- [17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.
- [18] Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25. 10.1145/3065386.
- [19] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.
- [20] C. Szegedy et al., "Going deeper with convolutions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1-9, doi: 10.1109/CVPR.2015.7298594.
- [21] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.
- [22] Woo, Sanghyun & Park, JongChan & Lee, Joon-Young & Kweon, Inso. (2018). CBAM: Convolutional Block Attention Module.