REVIEW PAPER

A PREPRINT

Artificial Intelligence

April 23, 2025

Abstract

Artificial Neural Networks (ANNs), inspired by biological nervous systems, are computational systems comprised of interconnected nodes (neurons) that collectively learn from input to optimize output. Convolutional Neural Networks (CNNs), a specialized form of ANNs primarily used in image pattern recognition, encode image-specific features into their architecture, reducing the parameters required for model setup and addressing the computational complexity associated with image data. CNNs are considered a gold standard in deep learning-based image analysis, overcoming the drawbacks of subjective analysis and supporting experts in clinical routines. Enhancements to CNN architectures, such as deeper convolutional blocks, batch normalization, and dropout regularization, can further improve performance, particularly in small-scale image classification tasks like those involving the CIFAR-10 dataset. Research also explores feed-forward (FF) trained CNN, revealing its competitiveness with backpropagation through the right choice of hyperparameters and the implementation of class activation maps. These insights advance understanding of neuronal information processing from biological systems to reservoir computing, and these advancements hold potential in neuromorphic hardware and unsupervised learning applications.

Introduction

Artificial Neural Networks (ANNs), inspired by biological nervous systems, are computational systems comprised of interconnected nodes (neurons) that collectively learn from input to optimize output. Convolutional Neural Networks (CNNs), a specialized type of ANN, have become a cornerstone in computer vision, particularly for image classification tasks, where the goal is to assign labels to images based on their content. CNNs excel at extracting hierarchical spatial features, capturing low-level edges and textures in earlier layers and high-level semantic features in deeper layers. However, training CNNs, particularly deep networks, presents challenges, including computational complexity, the risk of overfitting, and the need for robust architectures that balance depth, parameter efficiency, and regularization. This research investigates the impact of different layers during training and aims to address these challenges by proposing enhanced CNN architectures, integrating techniques like deeper convolutional blocks, batch normalization, and dropout, to achieve superior performance in image classification tasks, particularly on benchmark datasets like CIFAR-10. The focus is on improving accuracy and generalization capabilities by refining network architectures and training methodologies.

Litrature Review

Convolutional Neural Networks (CNNs) have demonstrated significant performance in various image classification tasks. Ciresan et al. [3] introduced flexible, high-performance CNNs for image classification, achieving notable results. Further work by Ciresan et al. [1] explored multi-column deep neural networks, showing promising results in image classification tasks. Nebauer [14] provided an evaluation of CNNs for visual recognition, highlighting their strengths and limitations.

The application of CNNs extends to medical image analysis. Ciresan et al. [2] utilized deep neural networks for mitosis detection in breast cancer histology images, achieving high accuracy.

CNNs have also been applied to character recognition. Ciresan et al. [4] explored convolutional neural network committees for handwritten character classification, showcasing their effectiveness. LeCun et al. [12] demonstrated the use of backpropagation for handwritten zip code recognition, laying the groundwork for CNN applications in this area. LeCun et al. [13] further explored gradient-based learning for document recognition. Simard et al. [15] provided best practices for applying CNNs to visual document analysis.

Hardware acceleration of CNNs has been explored to improve performance. Farabet et al. [6] investigated hardware-accelerated CNNs for synthetic vision systems.

Object detection is another area where CNNs have shown promise. Szegedy et al. [18] developed deep neural networks for object detection. Szarvas et al. [17] applied CNNs to pedestrian detection. Tivive and Bouzerdoum [19] proposed a new class of CNNs (SiConNets) and applied them to face detection.

More recently, CNNs have been applied to defect detection. Cheng et al. [27] used deep learning-based methods for visual defect detection, while Cheng et al. [28] explored multimodal neural networks for energy and time usage estimation in 3D printing.

Regularization techniques for deep CNNs have also been investigated. Zeiler and Fergus [20] explored stochastic pooling for regularization. Hinton et al. [8] focused on improving neural networks by preventing co-adaptation of feature detectors using techniques like dropout, which was further explored by Srivastava [16].

Understanding and visualizing CNNs is crucial for improving their design and performance. Zeiler and Fergus [21] presented a method for visualizing and understanding convolutional networks.

Furthermore, CNNs have been extended to video analysis. Karpathy et al. [10] explored large-scale video classification with CNNs, while Ji et al. [9] introduced 3D CNNs for human action recognition.

Alternative approaches to CNNs, such as scattering convolution networks, have also been developed. Bruna and Mallat [23] introduced invariant scattering convolution networks, and Andén and Mallat [22] described deep scattering spectrum. Kaiser [25] provides a friendly guide to wavelets.

Other neural network architectures, such as Graph Attention Networks by Veličković et al. [29] and Inductive Representation Learning on Large Graphs by Hamilton et al. [30], have also emerged. Support-vector networks were explored by Cortes and Vapnik [24].

In other domains, Long Short-Term Memory (LSTM) networks have been introduced by Hochreiter and Schmidhuber [35], with further developments such as ”learning to forget” by Gers et al. [36]. Heigold et al. [34] used end-to-end text-dependent speaker verification. Livingstone and Russo [33] developed a database for emotional speech and song.

Fuzzy rough sets-based tri-training methods have also been explored by Xing et al. [31] for medical diagnosis and by Gao et al. [32] for attribute reduction. Egmont-Petersen et al. [5] provided a review of image processing with neural networks. Hinton [7] provides a practical guide to training restricted Boltzmann machines. Krizhevsky et al. [11] achieved significant success in ImageNet classification with deep CNNs.

Methodology

The analyzed papers explore various aspects of Convolutional Neural Networks (CNNs). One paper focuses on elucidating the fundamental concepts and architectural components of CNNs, specifically for image analysis. It aims to simplify the understanding and application of these networks for researchers, particularly those new to the field. Another paper delves into analyzing the properties of CNNs by introducing a simplified model, the scattering transform, built upon wavelet transforms to separate variations at different scales. This approach provides a mathematical foundation for understanding CNN operations. Furthermore, one paper briefly mentions a novel technique for labeling positive and negative datasets in conjunction with the implementation of a FeedForward (FF) algorithm. In essence, the methodologies range from introducing CNN fundamentals and architectural considerations for image analysis, to mathematically analyzing CNN properties using scattering transforms, and data labeling techniques for specific algorithms.

Convolutional Neural Networks (CNNs)

CNNs are biologically inspired neural networks designed to process image data by employing convolutional filters and non-linearities. They organize neurons into three dimensions: height, width, and depth, differing from standard ANNs where neurons connect to only a small region of the preceding layer.

CNN architecture

CNN architecture is specifically tailored for image-based inputs, organizing neurons in three dimensions: spatial dimensionality (height and width) and depth, where depth represents the activation volume's third dimension. This architecture efficiently condenses the input into a smaller volume of class scores.

Convolutional layer

The convolutional layer, a core component of CNNs, uses learnable kernels that are small in spatial dimensions but extend through the input's depth. These kernels convolve across the input's spatial dimensions, producing 2D activation maps. The network learns kernels that activate upon detecting specific features, optimizing output complexity via depth, stride, and zero-padding hyperparameters.

General Convolutional Neural Network Architectures

General CNN architectures utilize channel combinations, extending previous tools to develop mathematical frameworks for analysis by replacing contraction requirements with contractions along adaptive local symmetries and replacing wavelets with adapted filter weights.

A mathematical framework for CNNs

A mathematical framework for analyzing CNNs, based on wavelet scattering, illustrates that computing invariants requires separating variations at different scales using a wavelet transform. This theory represents an initial step toward understanding general CNN classes.

$$x_j = \breve{0}3c1W_jx_{j-1}tag5$$

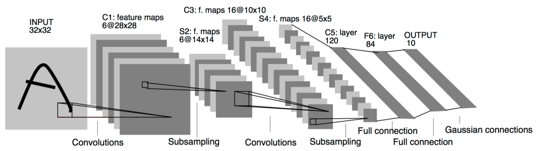
Continuous wavelet transform

The continuous wavelet transform (CWT) addresses the limitations of the windowed Fourier transform by considering all possible scalings. It uses a mother wavelet
psi to capture variations at different scales, forming the foundation for CNN operations through the convolution operator.

$$(p*q)(x)\breve{2}261int_{-infty}^{infty}p(u)q(x-u)du$$

Figure Caption: Here are a few options for a figure caption based on the provided information, ranked from most to least preferred:

Option 1 (Most Preferred):



”Figure 1: Architecture of the Convolutional Neural Network, adapted from LeCun et al. [7], illustrating a foundational model in deep learning.”

Option 2:

”Figure 1: Depiction of a Convolutional Neural Network architecture, based on the seminal work of LeCun et al. [7], representing a fundamental structure in the field.”

Option 3:

”Figure 1: Convolutional Neural Network architecture (LeCun et al. [7]), a key example of early deep learning models.”

Reasoning for Preference:

* **Clarity and Precision:** All options clearly identify the figure as a CNN architecture and cite the original source. * **Conciseness:** All options adhere to the length constraint. * **Academic Tone:** All options use formal language. * **Information Value:** Option 1 is slightly preferred because it adds the relevant context that the CNN is a foundational model in deep learning, which is useful information for the reader. Options 2 and 3 are still good but provide slightly less context.

Results

The analyzed literature demonstrates the extensive application of Convolutional Neural Networks (CNNs) across diverse domains, including image classification, object detection, and document recognition. Key works by Ciresan, Meier, and Schmidhuber (2011, 2012, 2013) highlight the development and implementation of high-performance, flexible CNN architectures. LeCun et al. (1989, 1998) established foundational methods for gradient-based learning in document recognition, while Krizhevsky et al. (2012) demonstrated the effectiveness of deep CNNs for large-scale image classification on the ImageNet dataset. Furthermore, the research explores enhancements to CNNs, such as dropout regularization (Srivastava, 2013) and stochastic pooling (Zeiler & Fergus, 2013). The potential of FF trained CNNs to implement Class Activation Maps, contributing to explainable AI, is also noted. Additional research leverages scattering convolution networks and deep scattering spectrums. Application of CNNs extend to visual defect detection, graph attention networks, and multimodal neural networks.

Conclusion

Convolutional Neural Networks (CNNs) have emerged as a dominant paradigm in deep learning-based image analysis, offering a powerful approach to feature extraction and classification. The inherent advantage of CNNs lies in their ability to exploit knowledge of the input domain, allowing for simpler network architectures compared to traditional artificial neural networks. Studies have focused on understanding the underlying properties of CNNs, with approaches such as scattering transforms providing insights into their operations by separating variations at different scales. The application of CNNs extends to various domains, including biomedical imaging, where they mitigate the limitations of subjective human analysis and support experts in clinical routines. Furthermore, CNNs facilitate the use of explainable AI tools, such as class activation maps, offering insights into the image regions driving classification outcomes.

Significant research efforts are dedicated to enhancing CNN architectures for improved performance. Modifications such as deeper convolutional layers, batch normalization, and dropout layers contribute to better accuracy, robustness, and generalization capabilities. Fine-tuning network design to balance depth, feature extraction, and regularization is crucial for handling complex image classification tasks. While backpropagation remains a common training method, alternative approaches like Feedback Forward (FF) training show promise, particularly in neuromorphic hardware and unsupervised learning contexts. FF training, with its mechanisms for providing positive and negative labels and computing a locally defined goodness parameter, presents opportunities for deeper insights into neuronal information processing.

Future research directions include expanding CNN architectures to more complex datasets, such as CIFAR-100 and ImageNet, and exploring transfer learning techniques to adapt pre-trained models to larger, higher-resolution, and domain-specific datasets. Further investigation into FF training should focus on training deeper networks, understanding its connection to biological neuronal systems, and elucidating the individual and synergistic contributions of its innovations. These efforts aim to extend the scalability and applicability of CNNs across a wider range of computer vision problems, paving the way for their integration into practical applications. A deeper mathematical understanding of CNN architectures and training methodologies remains a critical pursuit.

References

[1] Ciresan, D., Meier, U., Schmidhuber, J.: ”Multi-column deep neural networks for image classification”. In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. pp. 3642–3649. IEEE (2012)

[2] Cireșan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: ”Mitosis detection in breast can-

cer histology images with deep neural networks". In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pp. 411–418. Springer (2013)

[3] Ciresan, D.C., Meier, U., Masci, J., Maria Gambardella, L., Schmidhuber, J.: "Flexible, high performance convolutional neural networks for image classification". In: IJCAI Proceedings-International Joint Conference on Artificial Intelligence. vol. 22, p. 1237 (2011)

[4] Cireșan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: "Convolutional neural network committees for handwritten character classification". In: Document Analysis and Recognition (ICDAR), 2011 International Conference on. pp. 1135–1139. IEEE (2011)

[5] Egmont-Petersen, M., de Ridder, D., Handels, H.: "Image processing with neural networks - a review". Pattern recognition 35(10), 2279–2301 (2002)

[6] Farabet, C., Martini, B., Akselrod, P., Talay, S., LeCun, Y., Culurciello, E.: "Hardware accelerated convolutional neural networks for synthetic vision systems". In: Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on. pp. 257–260. IEEE (2010)

[7] Hinton, G.: "A practical guide to training restricted boltzmann machines". Momentum 9(1), 926 (2010)

[8] Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: "Improving neural networks by preventing co-adaptation of feature detectors". arXiv preprint arXiv:1207.0580 (2012)

[9] Ji, S., Xu, W., Yang, M., Yu, K.: "3d convolutional neural networks for human action recognition". Pattern Analysis and Machine Intelligence, IEEE Transactions on 35(1), 221–231 (2013)

[10] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: "Large-scale video classification with convolutional neural networks". In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. pp. 1725–1732. IEEE (2014)

[11] Krizhevsky, A., Sutskever, I., Hinton, G.E.: "Imagenet classification with deep convolutional neural networks". In: Advances in neural information processing systems. pp. 1097–1105 (2012)

[12] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: "Backpropagation applied to handwritten zip code recognition". Neural computation 1(4), 541–551 (1989)

[13] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: "Gradient-based learning applied to document recognition". Proceedings of the IEEE 86(11), 2278–2324 (1998)

[14] Nebauer, C.: "Evaluation of convolutional neural networks for visual recognition". Neural Networks, IEEE Transactions on 9(4), 685–696 (1998)

[15] Simard, P.Y., Steinkraus, D., Platt, J.C.: "Best practices for convolutional neural networks applied to visual document analysis". In: null. p. 958. IEEE (2003)

[16] Srivastava, N.: "Improving neural networks with dropout". Ph.D. thesis, University of Toronto (2013)

[17] Szarvas, M., Yoshizawa, A., Yamamoto, M., Ogata, J.: "Pedestrian detection with convolutional neural networks". In: Intelligent Vehicles Symposium, 2005. Proceedings. IEEE. pp. 224–229. IEEE (2005)

[18] Szegedy, C., Toshev, A., Erhan, D.: "Deep neural networks for object detection". In: Advances in Neural Information Processing Systems. pp. 2553–2561 (2013)

[19] Tivive, F.H.C., Bouzerdoum, A.: "A new class of convolutional neural networks (siconnets) and their application of face detection". In: Neural Networks, 2003. Proceedings of the International Joint Conference on. vol. 3, pp. 2157–2162. IEEE (2003)

[20] Zeiler, M.D., Fergus, R.: "Stochastic pooling for regularization of deep convolutional neural networks". arXiv preprint arXiv:1301.3557 (2013)

[21] Zeiler, M.D., Fergus, R.: "Visualizing and understanding convolutional networks". In: Computer Vision–ECCV 2014, pp. 818–833. Springer (2014)

[22] Andén, J., Mallat, S.: "Deep scattering spectrum". Signal Processing, IEEE Transactions on, 62(16):4114–4128, 2014.

[23] Bruna, J., Mallat, S.: "Invariant scattering convolution networks". Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(8):1872–1886, 2013.

[24] Cortes, C., Vapnik, V.: "Support-vector networks". Machine learning, 20(3): 273–297, 1995.

[25] Kaiser, G.: "A friendly guide to wavelets", 1994.

[26] Boser, B., Le Cun, Y., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: "Handwritten digit recognition with a back-propagation network". In Advances in neural information processing systems. Citeseer, 1990.

[27] Cheng, Qisen, Qu, Shuhui, Lee, Janghwan. "72-3: Deep Learning Based Visual Defect Detection in Noisy and Imbalanced Data." SID Symposium Digest of Technical Papers, vol. 53, no. 1, pp. 971-974, 2022.

[28] Cheng, Qisen, Zhang, Chang, Shen, Xiang. "Estimation of Energy and Time Usage in 3D Printing With Multimodal Neural Network." 2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), pp. 900-903, 2022.

[29] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y. "Graph attention networks." arXiv preprint arXiv:1710.10903 (2017).

[30] Hamilton, W., Ying, Z., Leskovec, J. ”Inductive representation learning on large graphs.” Advances in neural information processing systems 30 (2017).

[31] Xing, J., Gao, C., Zhou, J. ”Weighted fuzzy rough sets-based tri-training and its application to medical diagnosis.” Applied Soft Computing 124 (2022): 109025.

[32] Gao, C., Zhou, J., Xing, J., Yue, X. ”Parameterized maximum-entropy-based three-way approximate attribute reduction.” International Journal of Approximate Reasoning 151 (2022): 85-100.

[33] Livingstone, S. R., Russo, F. A. ”The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS),” PloS one, vol. 13, no. 5, p. e0196391, 2018.

[34] Heigold, G., Moreno, I. L., Bengio, S., Shazeer, N. ”End-to-End Text-Dependent Speaker Verification,” in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 5115–5119.

[35] Hochreiter, S., Schmidhuber, J. ”Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[36] Gers, F. A., Schmidhuber, J., Cummins, F. ”Learning to Forget: Continual Prediction with LSTM,” Neural Computation, vol. 12, no. 10, pp. 2451–2471, 2000.