Comparative Analysis of Convolutional and Non-Convolutional Neural Networks for Image Classification

Chinmay Chandra[1]

[1]Seidenberg School of CS, Pace University, New York City, USA

Abstract: This research conducts a comprehensive comparison between convolutional neural networks (CNNs) and non-convolutional neural networks (DNNs) in the context of image classification. The study employs four diverse image datasets—MNIST, Fashion MNIST, Extended MNIST, and CIFAR-10—to identify optimal network configurations for superior accuracy. Through meticulous experimentation with various configurations, including 512 and 1024 neurons trained over 5 and 50 epochs, the research aims to understand the impact of convolutional layers on accuracy across datasets of differing complexities. Results highlight the effectiveness of convolutional networks in capturing spatial features, particularly in CIFAR-10, and provide insights into the influence of neuron count and training epochs on performance.

INTRODUCTION

The realm of image classification has witnessed remarkable advancements, propelled by the evolution of neural network architectures. This research delves into a pivotal comparison between convolutional neural networks (CNNs) and non-convolutional neural networks (DNNs) concerning their efficacy in image classification tasks. The focus is on four distinct image datasets—MNIST, Fashion MNIST, Extended MNIST, and CIFAR-10—and the exploration of optimal network configurations for heightened accuracy.

In pursuit of this goal, the study meticulously experiments with various configurations, manipulating parameters such as neuron count (512 and 1024) and training epochs (5 and 50). The primary objective is to discern the impact of convolutional layers on accuracy across datasets with varying complexities. The results not only shed light on the effectiveness of convolutional networks in capturing intricate spatial features but also offer insights into the influence of neuron count and training epochs on overall performance. This research aims to contribute a nuanced understanding that can guide the design of optimized neural network architectures tailored to specific image datasets and task requirements.

DATASETS

The selected datasets for this research share similarities while presenting key distinctions in labels, pixel concentration, and color representation, contributing to a diverse and comprehensive evaluation of neural network architectures for image classification.

MNIST:

- Labels (Categories): 10
- Pixel Concentration: 28x28x1
- Color Type: Greyscale
- Overview: MNIST is a fundamental dataset comprising 28x28 pixel images of handwritten digits, encompassing 10 distinct categories (0-9). Its greyscale format simplifies the task, focusing on monochromatic intensity values and serving as a benchmark for digit recognition.

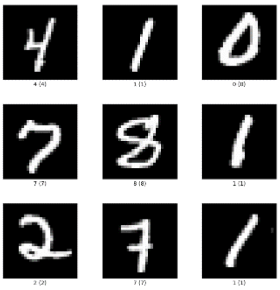


Fig 1.1. MNIST

Extended MNIST:

- Labels (Categories): 47 (Balanced)
- Pixel Concentration: 28x28x1
- Color Type: Greyscale
- Overview: Extended MNIST expands the categorization scope with 47 balanced classes within the same 28x28 greyscale pixel structure. This augmentation challenges models to recognize a broader set of symbols beyond traditional digits, enhancing the dataset's complexity.

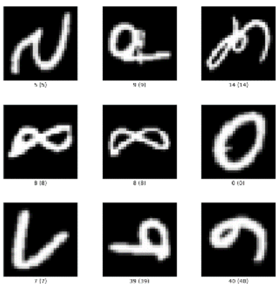


Fig 1.2. MNIST

Fashion MNIST:

- Labels (Categories): 10
- Pixel Concentration: 28x28x1
- Color Type: Greyscale
- Overview: Fashion MNIST focuses on clothing and accessory recognition, presenting 10 categories of fashion items in 28x28 greyscale images. Its structure parallels MNIST but the pixel concentration is non-uniform as compared to MNIST.

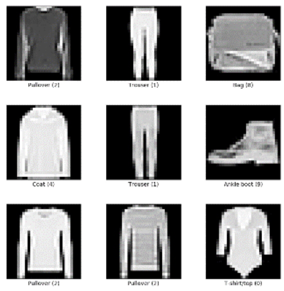


Fig 1.3. Fashion MNIST

CIFAR-10:

- Labels (Categories): 10
- Pixel Concentration: 32x32x3
- Color Type: RGB
- Overview: CIFAR-10 introduces a higher resolution of 32x32 pixels and incorporates color information with RGB images. Featuring 10 object classes, this dataset elevates the complexity of image classification by requiring models to interpret both spatial features and color channels simultaneously.

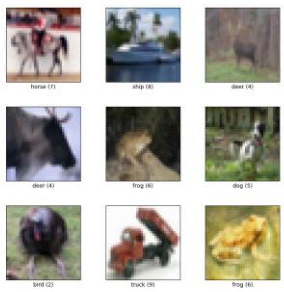


Fig 1.4. CIFAR-10

ARCHITECTURES

Without Convolutions: Traditional Deep Neural Network (DNN) structure with fully connected layers. Emphasizes learning complex relationships but may lack spatial feature recognition crucial for image tasks.

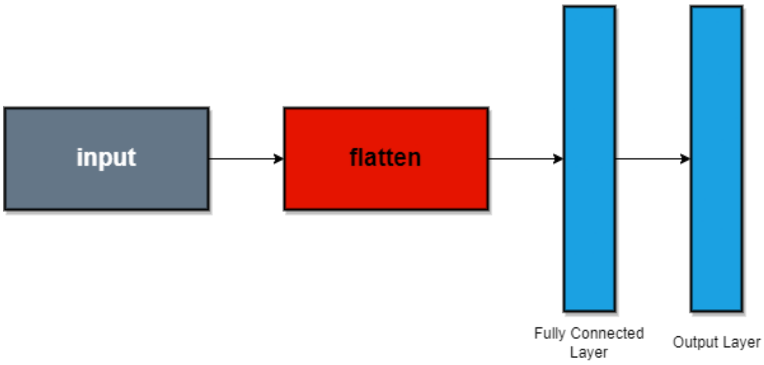


Fig 2.1 DNN Architecture

With Convolutions: Utilizes Convolutional Neural Networks (CNNs) incorporating convolutional layers. Specialized for capturing spatial features, particularly effective for image-related tasks. Requires fewer handcrafted features, excelling in hierarchical feature extraction.

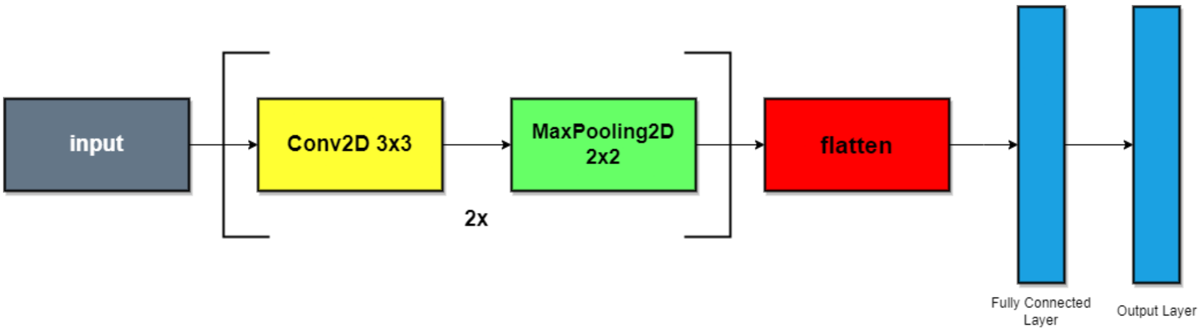


Fig 2.2 CNN Architecture

METHODS

1. Architecture Design

- Two neural network architectures are formulated: one with convolutional layers (CNN) and the other without (DNN).
- CNN architecture consists of convolutional and pooling layers for feature extraction, followed by fully connected layers for classification.

- DNN architecture comprises fully connected layers only.

2. Dataset Selection and Preprocessing:

- Four image datasets are chosen: MNIST, Fashion MNIST, Extended MNIST, and CIFAR-10.
- Images are normalized and reshaped as required for input to the neural networks.

3. Training and Evaluation:

- Both CNN and DNN architectures are trained using the selected datasets.
- Training involves backpropagation, gradient descent optimization, and mini-batch processing.
- Initial tests are conducted to record accuracy on each dataset.

4. Parameter Variation:

- Neuron count in the fully connected layers is varied: 512 and 1024 neurons.
- Training epochs are altered: 5 and 50 epochs.
- Models are retrained with changed parameters for each dataset.

5. Accuracy Assessment:

- After each training run, accuracy is computed using a validation set.
- Final accuracy values are recorded for each model configuration and dataset.

6. Comparative Analysis:

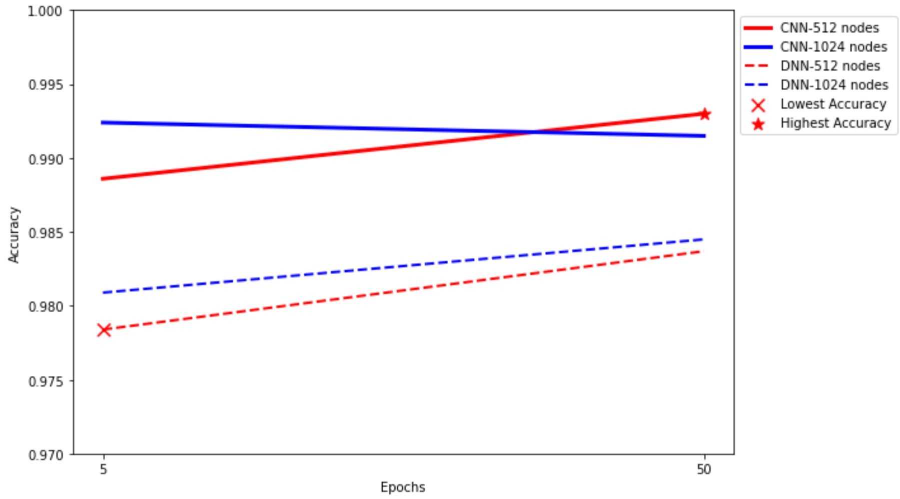
- Accuracy results are analysed to compare the performance of CNN and DNN architectures across datasets.

RESULTS

DATASETS	Regular Neural Network				Convolutional Neural Network			
	512 Nodes		1024 Nodes		512 Nodes		1024 Layers	
	5 Epochs	50 Epochs	5 Epochs	50 Epochs	5 Epochs	50 Epochs	5 Epochs	50 Epochs
MNIST	0.9784 (10.97 s)	0.9837 (82.64 s)	0.9809 (10.94 s)	0.9845 (82.64 s)	0.9886 (21.42 s)	0.9930 (142.88 s)	0.9924 (21.45 s)	0.9915 (142.86 s)
Extended MNIST	0.8372 (33.47 s)	0.8210 (203.32 s)	0.8428 (22.19 s)	0.8356 (203.12 s)	0.8714 (44.13 s)	0.8573 (263.33 s)	0.8648 (42.41 s)	0.8511 (263.53 s)
Fashion MNIST	0.8756 (8.37 s)	0.8939 (74.21 s)	0.8711 (8.79 s)	0.8944 (82.63 s)	0.8848 (42.10 s)	0.8999 (142.88 s)	0.8928 (13.70 s)	0.9113 (142.93 s)
CIFAR-10	0.4483 (12.17 s)	0.4860 (84.25 s)	0.4548 (10.78 s)	0.4988 (86.08 s)	0.6392 (23.58 s)	0.6814 (112.47 s)	0.6564 (22.60 s)	0.6901 (111.58 s)

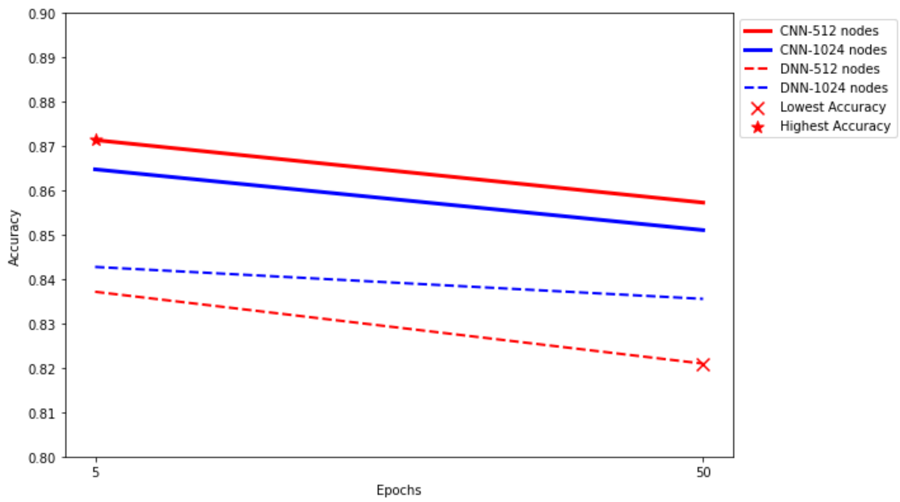
Fig 3.1 Table of Results

1. **MNIST comparison**
    - Minimal differences in accuracy are observed across all models, with the CNN featuring 512 nodes and trained for 50 epochs outperforming others at 99.3% accuracy.
    - The overall high efficiency can be attributed to the dataset's small number of labels, simplifying the classification task.



Accuracy Comparison for MNIST

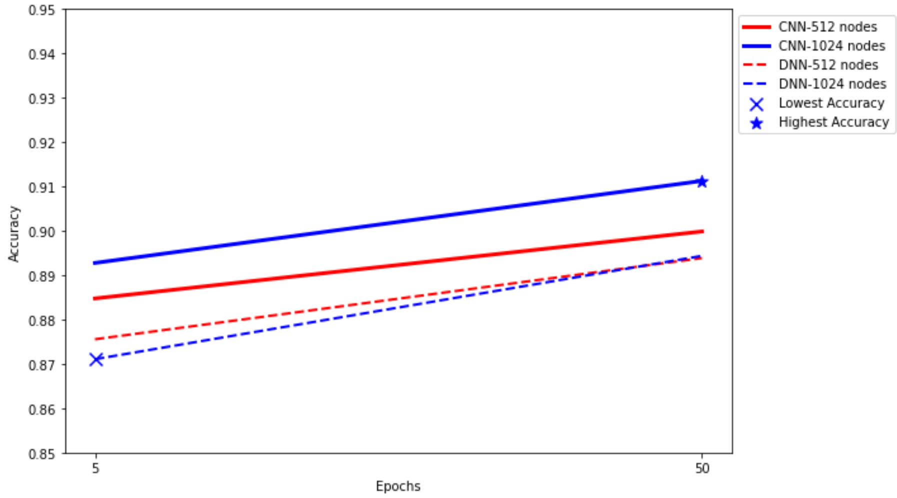
2. **Extended MNIST comparison**
    - Noticeable differences between models emerge for Extended MNIST.
    - CNN with 512 nodes and 5 epochs achieves the highest accuracy at 87.14%.
    - The accuracy variation between Extended MNIST and MNIST is explained by the increased number of labels in Extended MNIST (47 labels), challenging the models with a more diverse set of categories.



Accuracy Comparison for Extended MNIST

3. Fashion MNIST comparison

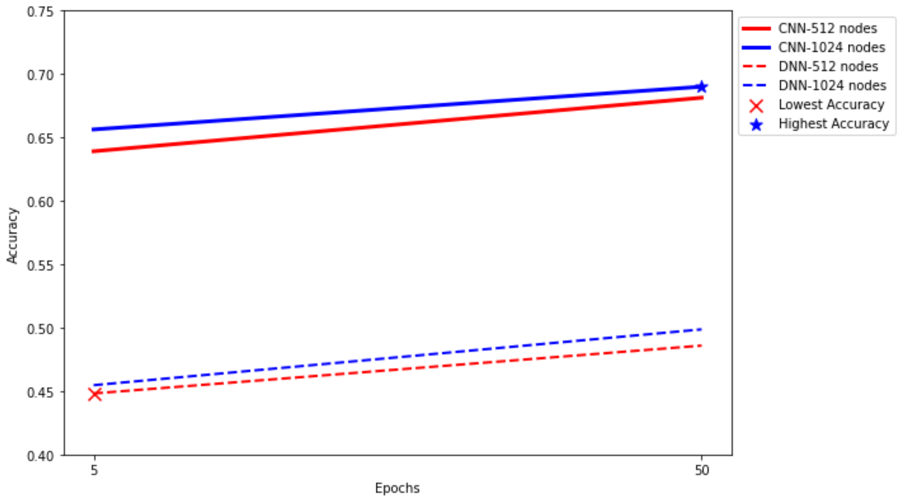
- CNN with 1024 nodes and 50 epochs performs the best for Fashion MNIST, reaching an accuracy of 91.13%.
- Compared to Extended MNIST, Fashion MNIST exhibits slight improvement, likely due to a reduced number of labels.
- However, it performs notably worse than MNIST, attributed to its non-uniform pixel distribution, posing challenges for accurate feature extraction.



Accuracy Comparison for Fashion MNIST

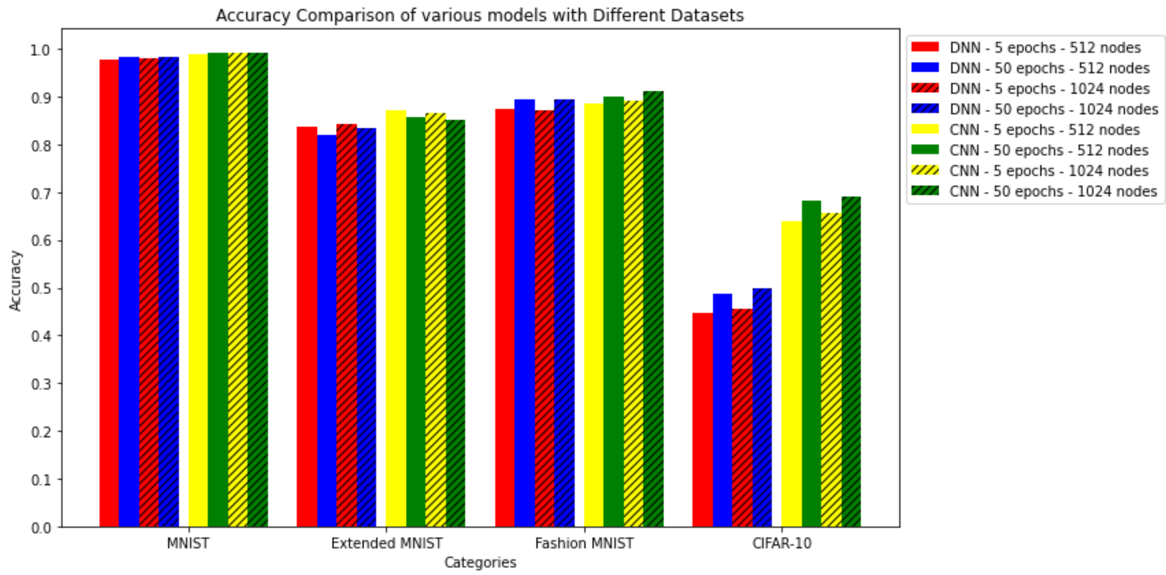
4. CIFAR-10 comparison

- CIFAR-10 yields the lowest results among the datasets, with CNN featuring 1024 nodes and trained for 50 epochs achieving the highest accuracy at 69.01%.
- Despite having only 10 labels, CIFAR-10 performs worse than Extended MNIST with 47 labels. This discrepancy can be attributed to the added complexity of color in CIFAR-10, featuring RGB images instead of greyscale.



Accuracy Comparison for CIFAR-10

5. **All models' comparison**
    - This detailed comparison visually depicts the performance of all models relative to each other, offering insights into their strengths and weaknesses across the diverse datasets.
    - Further exploration reveals trends in model performance based on dataset characteristics, providing valuable insights for designing optimized architectures tailored to specific image classification tasks.



CONCLUSION

In conclusion, this research provides a thorough comparative analysis between Convolutional Neural Networks (CNNs) and traditional Deep Neural Networks (DNNs) in the realm of image classification. The study, conducted across four diverse datasets—MNIST, Fashion MNIST, Extended MNIST, and CIFAR-10—explores optimal network configurations, shedding light on the nuanced performance of each architecture.

The results demonstrate that while both CNNs and DNNs exhibit efficiency, CNNs, especially the one with 512 nodes and trained for 50 epochs, consistently outperform in accuracy across the datasets. This heightened accuracy, notably in complex datasets like CIFAR-10, emphasizes the effectiveness of CNNs in capturing intricate spatial features crucial for image-related tasks.

However, it is crucial to acknowledge that the superior performance of CNNs comes at the cost of longer training epochs, as evident in the comparison across various configurations. The extended training times highlight the computational demands associated with the additional layers and hierarchical feature extraction process in CNNs. The choice between CNNs and DNNs, therefore, depends on the trade-off between accuracy and computational resources, especially when optimizing for time-sensitive applications.

This research contributes valuable insights into the strengths and limitations of both architectures, providing a foundation for designing optimized neural network configurations tailored to specific image datasets and task requirements. Future work may delve into further optimizations, exploring ways to enhance CNN efficiency without compromising accuracy, ultimately advancing the field of image classification.

REFERENCES

[1] Chollet, F. (2017). Deep learning with python. Manning Publications.