UNDERSTANDING CNNS FOR IMAGE & AUDIO CLASSIFICATION

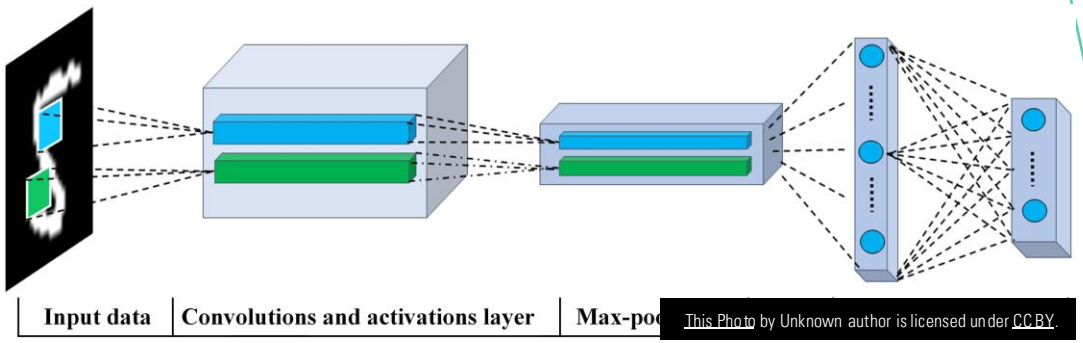
DEEP LEARING PROJECT

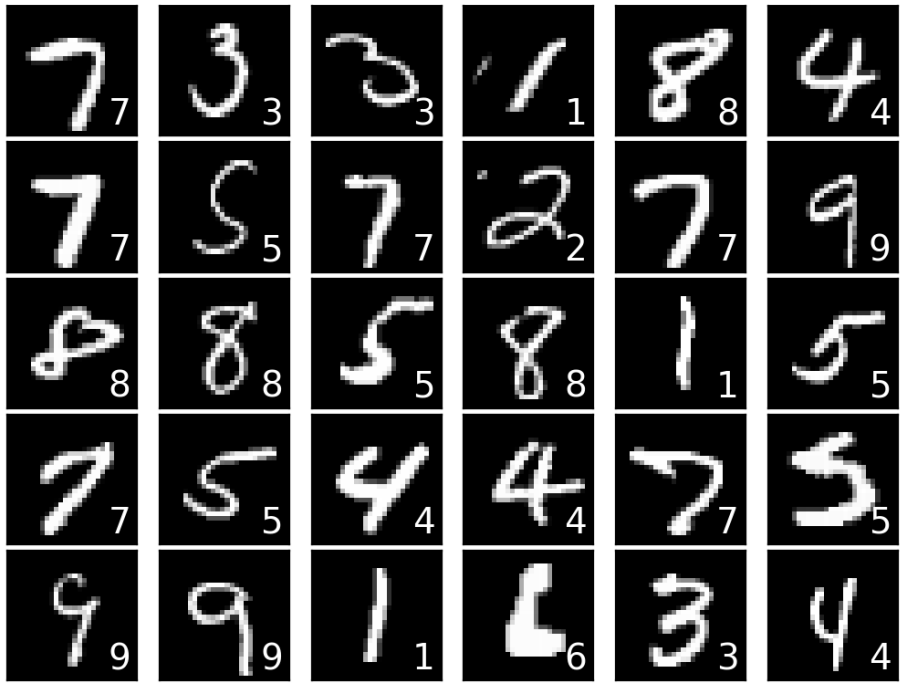
CONVOLUTIONAL NEURAL NETWORKS

- For computer vision applications, Convolutional Neural Networks (CNNs) are deep learning models.
- They consist of fully connected layers, pooling layers, and convolutional layers.
- CNNs excel at learning hierarchical representations from visual input, making them ideal for tasks such as object identification, segmentation, and image recognition.

CONVOLUTIONAL NEURAL NETWORK (CNN) ARCHITECTURE

- Input Layer:** Receives raw input data, typically images.
- - Convolutional Layers:** Apply filters to extract features such as edges and textures.
- - Activation Function:** Introduces non-linearity to the network.
- - Pooling Layers:** Reduce spatial dimensions of feature maps, controlling complexity.
- - Fully Connected Layers:** Process flattened feature maps for classification or regression.
- - Output Layer:** Produces final prediction or output.





UNDERSTANDING MINST AND CIFAR-10 DATASETS

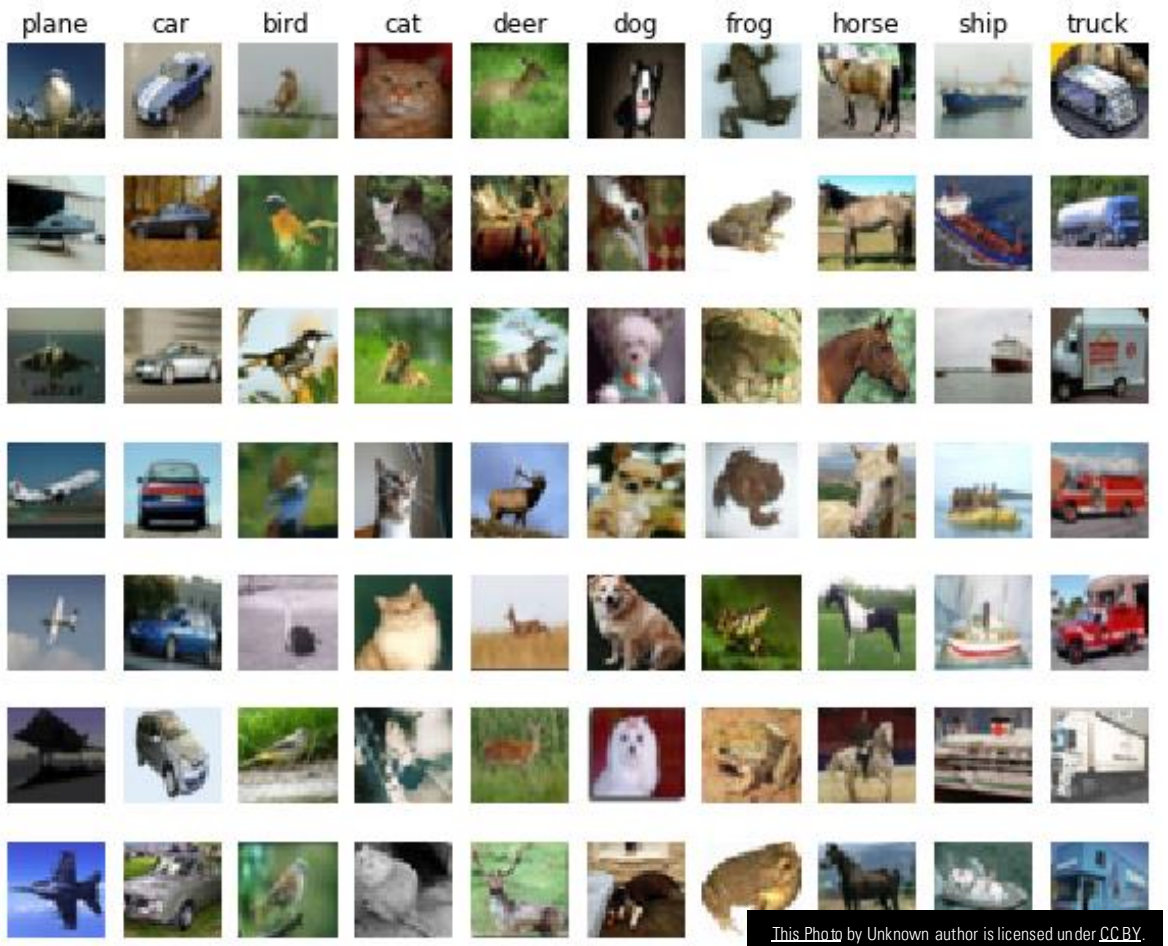
MINST:

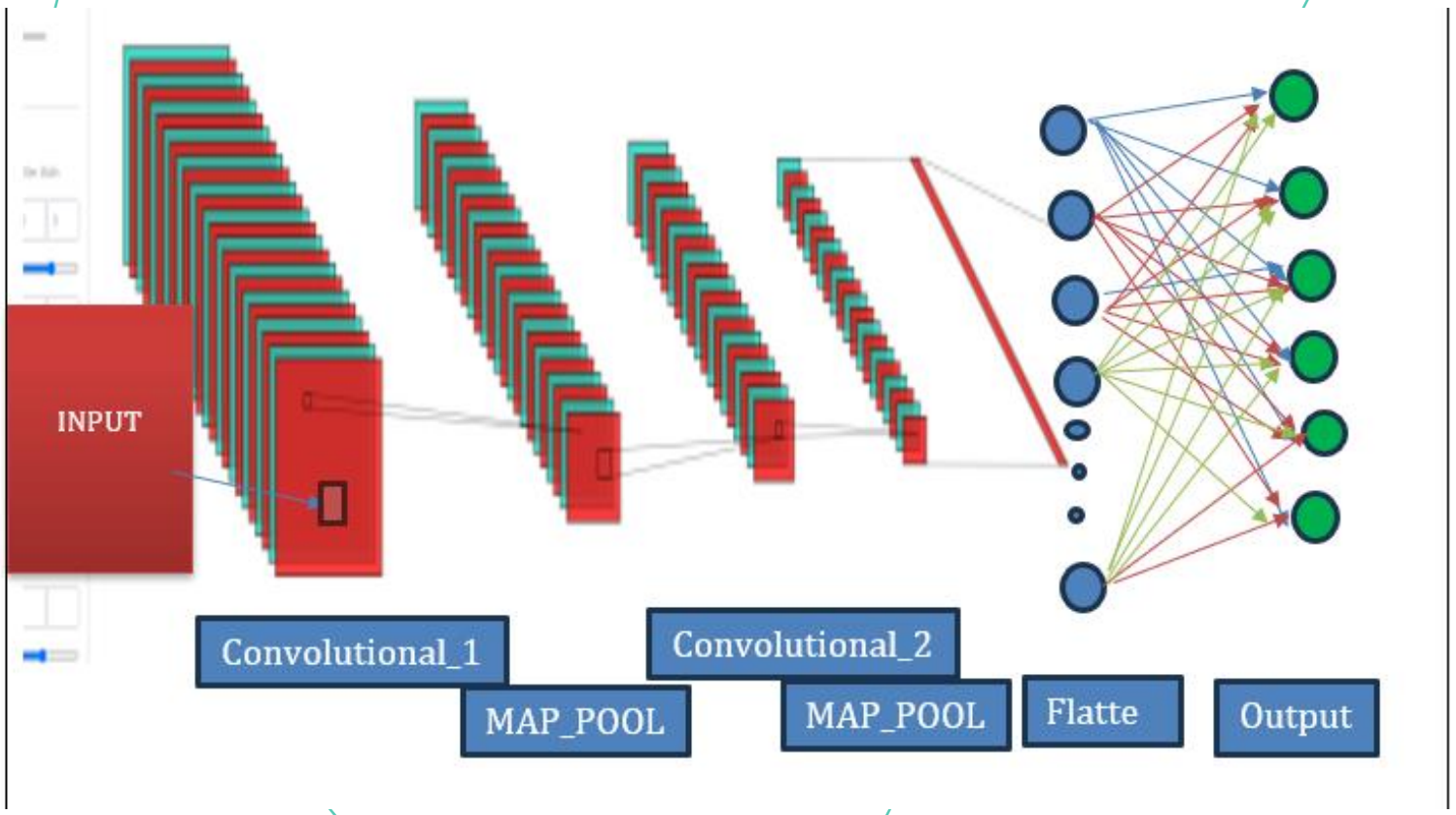
contains 70,000 28x28 pixel grayscale pictures.

- Handwritten numbers 0–9, frequently utilized in challenges involving digit recognition.
- An introductory dataset for deep learning and computer vision education.

CIFAR_10

- CIFAR-10
- includes 60,000 32 x 32 pixel color pictures.
- Ten classes: vehicles; birds; cats; deer; dogs; frogs; horses; ships; trucks; and aircraft.
- popular benchmark for problems involving picture categorization.





CNN ARCHITECTURE ON CIFAR_10

RESULT

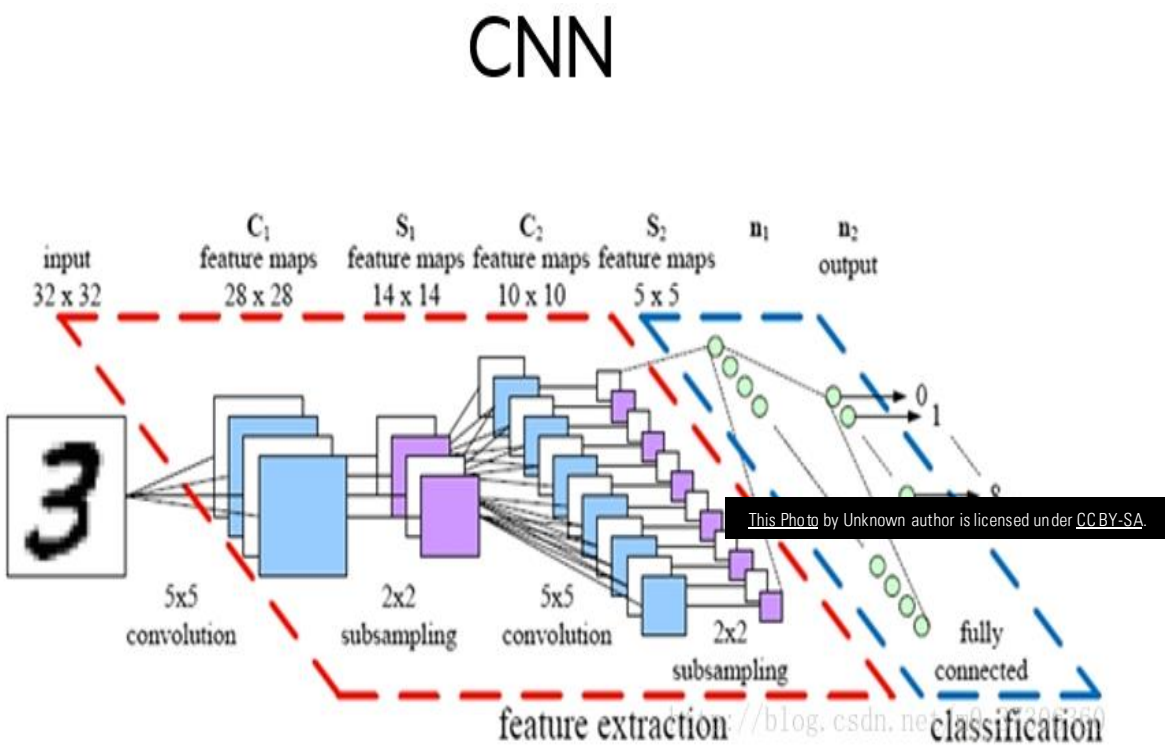
Calculation Method as the Accuracy Metric

- To continue evaluating the model's performance on the CIFAR-10 dataset, accuracy should be the main evaluation parameter used.

Initial CNN Model Accuracy on CIFAR-10 Dataset: 0.7113

- To assess the efficacy of further enhancements, take into consideration the baseline accuracy result of 0.7113 from the original CNN model.

PERFORMANCE COMPARISON OF BASIC CNN ON MNIST AND CIFAR-10



Accuracy Results:

- MNIST: 0.9894
- CIFAR-10: 0.7113

Observations:

- MNIST, a simpler dataset, achieves significantly higher accuracy compared to CIFAR-10.
- CIFAR-10's complexity, with colored images of various objects, poses greater challenges for classification.

Possible Reasons for Performance Gap:

- CIFAR-10's diverse images with color variations and multiple objects demand more sophisticated feature extraction.
- MNIST comprises grayscale images of single digits, making it easier for the CNN to learn distinctive patterns.
- Deeper networks or additional layers may be required to extract relevant features from CIFAR-10 images.

IMPROVEMENT OF CNN FOR CIFAR-10

Addition of Convolutional Layer:

- Increase model depth to capture more intricate features.

Integration of Max-Pooling Layer:

- Downsample feature maps to reduce computational complexity and enhance spatial invariance.

Introduction of Fully Connected Layer:

- Improve classification accuracy by adding a dense layer for higher-level feature representation.

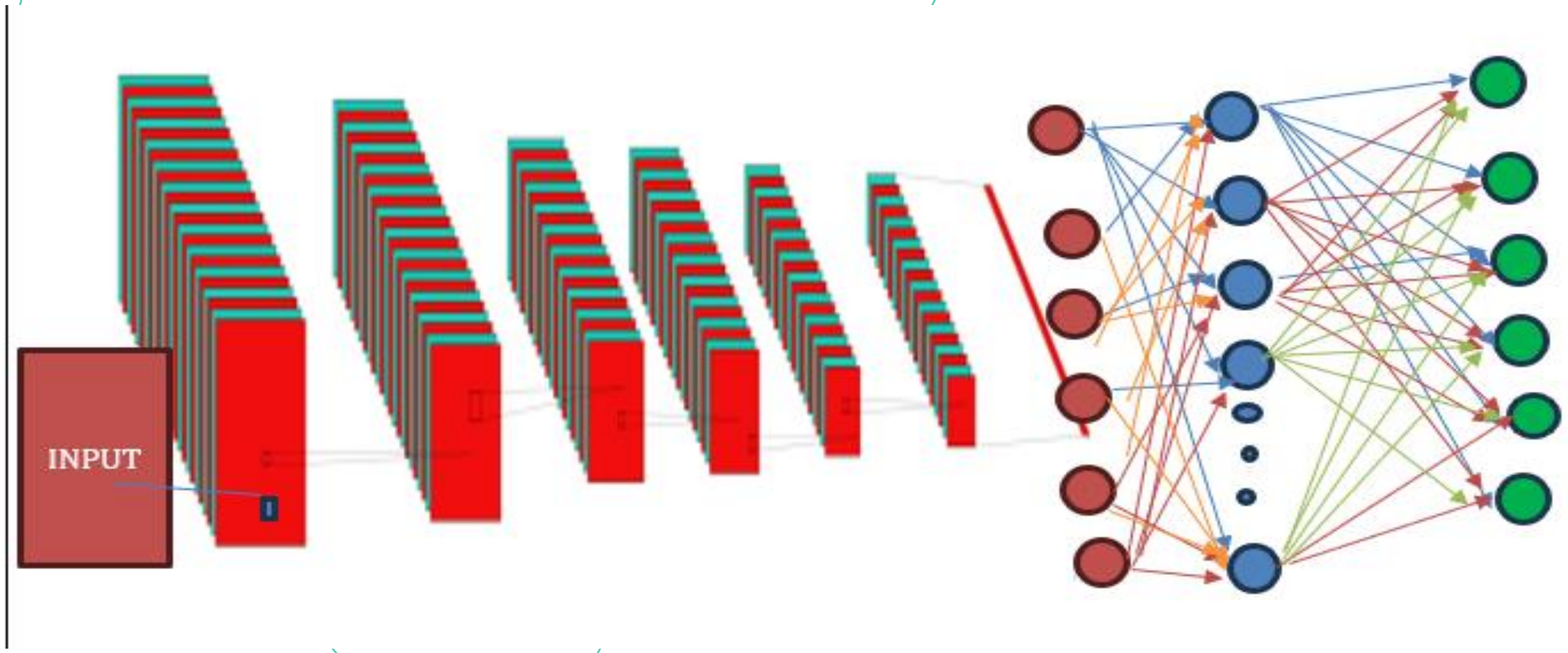
Modified Architecture:

- Incorporated proposed changes into the existing CNN architecture.

Hyperparameter Tuning:

- Adjusted learning rate, batch size, and other parameters for optimal training performance.

IMPROVED CNN



CONCLUSION

- A synopsis of the modifications:

The CNN architecture was expanded to include more convolutional, max-pooling, and fully linked layers.
To guarantee improved feature learning and convergence, the total amount of training steps was increased.
Impact on Precision:

Better feature extraction & representation were made possible by the inclusion of more layers, which improved classification accuracy.
By taking incremental training stages, the model was able to converge more successfully and identify smaller patterns in the information.
Accuracy Improvement:

Through architectural and training process optimization, the CIFAR-10 CNN model's accuracy increased from 0.7113 into 0.7285.