Revolutionizing Image Recognition: Next-Generation CNN Architectures for Handwritten Digits and Objects

Md Nurul Absur
Department of Computer Science
City University of New York
New York, USA
mabsur@gradcenter.cuny.edu

Kazi Fahim Ahmad Nasif
College of Computing and Software Engineering
Kennesaw State University
Georgia, USA
knasif@students.kennesaw.edu

Sourya Saha
Department of Computer Science
City University of New York
New York, USA
ssah42@gradcenter.cuny.edu

Sifat Nawrin Nova
Department of Computer Science
Chalmers University of Technology
Gothenburg, Sweden
esifatn@chalmers.se

***Abstract*—This study addresses the pressing need for computer systems to interpret digital media images with a level of sophistication comparable to human visual perception. By leveraging Convolutional Neural Networks (CNNs), we introduce two innovative architectures tailored to distinct datasets: the MNIST handwritten digit dataset and the Fashion MNIST dataset. Unlike traditional machine learning methods such as Support Vector Machines (SVM) and Random Forests, our customized CNN models remarkably enhance image attribute comprehension and recognition accuracy. Specifically, the model developed for the MNIST dataset achieved an unprecedented accuracy of 98.71% without any bias, while the Fashion MNIST model reached 91.39%, marking significant advancements over conventional algorithms without any bias. This research showcases the superior efficiency of CNNs in processing and understanding digital images. It underscores the potential of deep learning technologies in bridging the gap between computational systems and human-like visual recognition. Through meticulous experimentation and analysis, we illustrate how deep CNNs require less preparatory work than other image-processing algorithms, setting a new benchmark in computer vision.**

***Index Terms*—Handwriting Recognition, Visual Object Recognition, Deep CNN, Performance Analysis, Deep Learning, Data Science**

I. INTRODUCTION

In today's digital landscape, information holds the highest value. Its accuracy and relevance provide a foundation for a wide range of applications. The process of turning raw data from various sources into usable insights is where information begins. Technological advancements have expanded the capacity for creating and sharing visual data, enriching the knowledge pool. It is crucial to analyze this vast array of visual data using computational methods to understand our environment better. In the academic field, data applications have experienced significant growth in recent years. Technologies such as Computer Vision [1], Machine Learning [2], Deep Learning [3], Reinforcement Learning [4], and Data Mining [5], etc., are being leveraged to collect and analyze data in real-life scenarios. These tools empower researchers to make groundbreaking discoveries and advancements within their respective fields.

A family of feed-forward neural networks known as convolutional neural networks (CNNs) has shown impressive results in resolving various machine-learning issues. A significant advancement in this domain is the Deep Convolutional Neural Network, which has multiple layers of architecture. Recent research has consistently shown that Deep CNNs deliver precise classification outcomes across various applications. However, utilizing Deep CNNs requires a training phase, where the network learns from a dataset and subsequently applies this knowledge to perform classifications based on the characteristics of the training data. This process underscores the critical need for sophisticated computational resources and algorithmic strategies to advance the field of image recognition and classification in artificial intelligence.

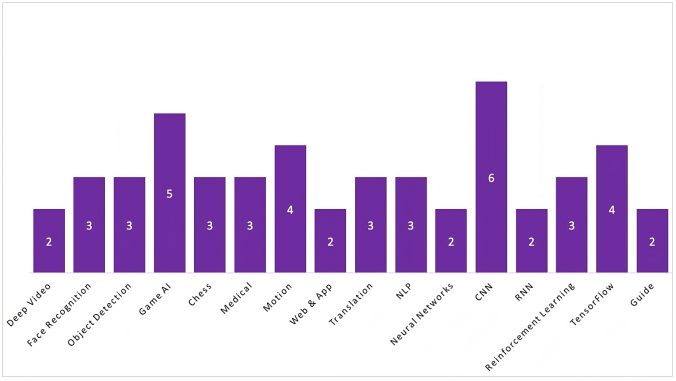

Fig. 1. Popularity of Deep Learning for Image-Related Tasks [21].

Convolutional Neural Networks (CNNs) are specialized neural networks used for image analysis. They break the image into smaller segments and use convolutional layers to identify lower-level features. The result is an output vector with solid responses to similar inputs, demonstrating the importance of filtering in CNNs for accurate classification. These days, deep learning is more famous for handling vast amounts of data, and deep CNN, in particular, is popular among researchers for object recognition and image classification, as demonstrated in Figure 1.

Our contributions in this paper are:

- Through meticulous efforts, we have successfully eliminated all inaccuracies within the datasets, resulting in a refined version that is now universally accessible.
- Our proposed model has demonstrated exceptional performance by accurately predicting numerous instances across these curated datasets.
- We have developed a three-layered convolutional neural network architecture to demonstrate our method's effectiveness. This sophisticated arrangement yielded more accurate results than those from earlier algorithms, and its effectiveness was thoroughly tested.

II. RELATED WORKS

Image property recognition methods have been an essential scientific sector in recent years. It is "the construction of explicit, meaningful descriptions of physical objects from images." The primary image-processing elements are understanding the properties of images, image classification, localization, and detection [6]. In the Convolutional Neural Network, the hierarchy of images is maintained [7]. Convolutional Neural Networks (CNNs) utilize a grid-based approach for processing images instead of feeding the entire image into the network. Rather than processing the image, it is divided into smaller parts, and each segment is passed to individual neural networks for specialized training. For a color image, a 2D array is used for each featured map and a 3D array for each featured map of video [8]. Menghan Sheng and other authors represented the process of using a Convolutional Neural Network for facial detection [9]. Before CNNs, Hybrid Neural Networks (HNNs) were extensively utilized for facial recognition, albeit with notable limitations such as inconsistent face identification, challenges in recognizing faces in various poses and low light, and difficulties in extracting unique features and accurately matching images to a dataset, transitioning to CNNs markedly reduced training errors, the exploration of the Karhunen-Loeve method yielded promising results, with error rates closely rivaling those of CNNs (5.3% vs. 3.8%). In comparison, Multi-Layer Perceptions underperformed significantly (40% errors vs. 3.8%). Furthermore, Farabet et al.'s approach to pixel-wise classification for scene labeling demonstrated the nuanced progress in neural network methodologies, emphasizing the shift towards more accurate and efficient facial recognition technologies [10].

Recurrent architecture is used in Convolutional Neural Network [11]. Adopting R-CNNs, facilitated by network sharing of sequential parameters, enhances error identification and correction as context size increases, significantly improving object detection and classification accuracy [12]. According to Cheng Liang's analysis, Deep Convolutional Neural Networks (DCNNs) have reached a performance plateau, falling short of the precision required for specific tasks. This is especially true in the work of George Papandreou and Liang-Chieh Chen, who note that the final layer of DCNNs needs more localization accuracy for effective object segmentation. A specialized Conditional Random Field (CRF) has been proposed to overcome this challenge. This approach has significantly improved segmentation accuracy, achieving a 71.6% Intersection Over Union (IOU) accuracy on the benchmark set and setting a new state-of-the-art performance on the PASCAL semantic image segmentation task. This progress highlights the ongoing need for innovative solutions to push the limits of current deep-learning architectures in complex image-processing tasks [13].

Dhananjay et al. said that CNN performs better due to its standard features and classifier learning [14]. Zhicheng et al. introduce a Hierarchical Deep Convolutional Neural Network (HDCNN) architecture premised on the observation that certain classes within image classification tasks present more significant ambiguity than others. This architecture innovates beyond traditional CNN models, which operate as N-way classifiers, by adopting a coarse-to-fine classification strategy coupled with a modular design approach. The implementation of HDCNN, utilizing the CIFAR100-NIN (Network In Network) building block, demonstrates a testing accuracy of 65.33%. This performance surpasses standard deep learning models and other HD-CNN frameworks tested on the CIFAR100 dataset, thereby underscoring the efficacy of hierarchical structuring in managing class-based ambiguities and enhancing model accuracy [15].

Xiao and colleagues have emphasized the importance of identifying foreground objects in creating effective image classification systems. Their innovative approach utilizes visual interest points and a Deep Neural Network (DNN) model that learns from CNN-generated eye movements to assign fine-grained classes. This technique achieves high accuracy with minimal supervision, relying solely on class labels and not requiring object-bound boxes or part landmarks. Notably, their approach achieves the highest accuracy on the CUB2000-2011 dataset within weakly supervised learning environments. Additionally, the team introduces modifications to the CNN architecture to address translation and distortion challenges, resulting in a structure that exhibits translation invariance and improved feature extraction capabilities. This advancement further improves state-of-the-art image classification [16].

Shuiwang et al. detail the development of a 3-D Convolutional Neural Network (3-D CNN) tailored for motion recognition, representing a significant advancement over traditional 2-D CNN models. This innovative approach allows for extracting features not just from spatial dimensions but also from the temporal dimension by implementing 3-D convolutions. This is attributed to its enhanced ability to analyze and interpret complex motion patterns within the video data, as exemplified

by its effectiveness in handling SPM (Spatial et al.) cube gray scenarios, where the dynamic nature of the content necessitates a more sophisticated model for accurate analysis and interpretation [17].

Addressing this challenge necessitates the exploration of diverse datasets to ensure the robustness and versatility of the model. This approach not only enhances the generalizability of the solution but also mitigates the risk of overfitting to specific data characteristics, leading to more accurate and reliable outcomes across varying contexts (e.g., MNIST handwritten digit dataset, Fashion MNIST dataset) [18] [19]. Over the years, many classifiers have been employed to categorize these datasets [20]. Our study introduces two distinct methodologies that leverage the Convolutional Neural Network (CNN) framework for enhanced classification efficacy, particularly in addressing visual recognition challenges.

III. DATASET DESCRIPTION

Our research focuses on assessing the efficacy of Convolutional Neural Networks (CNNs) in image understanding and classification for computer vision. We use a range of relevant datasets during training to ensure model stability. Rigorous testing and validation against these datasets enable us to evaluate performance comprehensively. Our research is supported by two distinct datasets, selected to represent diverse aspects of the subject matter, allowing for a thorough examination of CNN's effectiveness in this domain.

A. MNIST Handwritten Digit Dataset

This research utilizes the MNIST dataset [23], which contains handwritten English digits and serves as a benchmark for digit recognition systems. Comprising 28×28 grayscale images, processed as 28×1 vectors, each annotated with a class label, the MNIST dataset is crucial for training models to recognize the varied styles of human handwriting. It includes a training set of 60,000 images and a testing set of 10,000 images, totaling 70,000 images. These sets are carefully divided to ensure models are tested on unseen data, facilitating the evaluation of machine learning algorithms that have reached near-human accuracy in digit recognition. The distribution of images is detailed in an accompanying in Table I, offering a clear overview of the dataset's structure.

TABLE I
MNIST HANDWRITTEN DIGIT TRAINING VS. TESTING SET

Name	Total images	Training set	Testing set
MNIST-handwritten digit	70,000	60,000	10,000

B. Fashion MNIST Dataset

The Fashion MNIST dataset [24] features many article images from Zalando, an esteemed online fashion retail platform. This dataset encompasses ten distinct categories of fashion products and closely mirrors the structure of the MNIST handwritten digit dataset. A significant component of this dataset is the data per category (DPC), which serves as a crucial foundation for analysis. The dataset's relevant particulars and distribution are meticulously presented in the subsequent Table II for a comprehensive overview.

TABLE II
FASHION MNIST DATA TRAINING VS TESTING SET

Name	Total images	Training set	Testing set	DPC
MNIST fashion	70,000	60,000	10,000	10

IV. PROPOSED MODEL

This study introduces two divergent methodologies tailored to the unique characteristics of the MNIST handwritten digit and Fashion MNIST datasets. For the MNIST handwritten digit dataset, a model configuration employing three convolutional layers is utilized, whereas the classification of the Fashion MNIST dataset incorporates an architecture featuring three convolutional layers. A max-pooling layer of 2×2 is strategically placed between every pair of convolutional layers to enhance feature extraction and reduce dimensionality.

A. MNIST Handwritten Digit Dataset

Specifically, the model configuration designed for the MNIST handwritten digit dataset comprises three convolutional layers with an input specification of $28 \times 28 \times 1$. The following Table III meticulously delineates the names of the layers and the filters per node, ensuring a precise and data-driven exposition of the model architecture.

TABLE III
MODEL CONFIGURATION

Layer name	Filters per node
Input layer	$28 \times 28 \times 1$
Convolution layer-1	32 filters, 3×3
Max pooling layer-1	2×2
Convolution layer-2	64 filters, 3×3
Max pooling layer-2	2×2
Convolution layer-3	128 filters, 3×3
Max pooling layer-3	2×2
Fully connected layer	128
Output layer	10

Model proposed design for the MNIST handwritten digit dataset is defined with three convolution layers. The proposed model for this dataset is shown in Figure 2.

B. Fashion MNIST Dataset

Proposed model configuration for the Fashion MNIST dataset consists of three convolutional layers. $28 \times 28 \times 1$ input configuration is associated with this model. Table IV can determine the summary of the different layers with their associated filters.

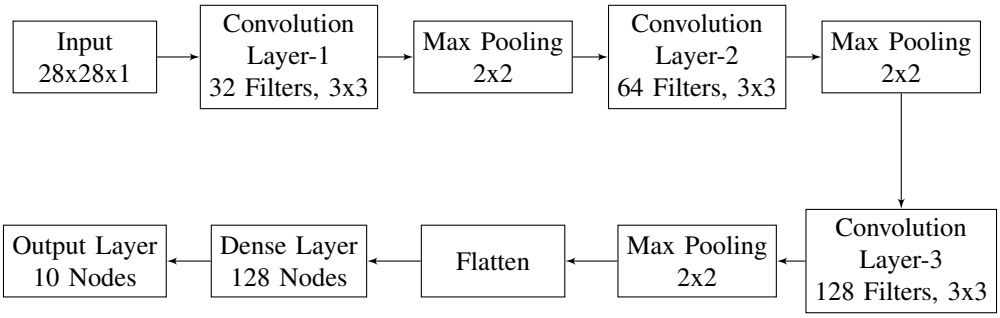

Fig. 2. Convolutional Neural Network Model Architecture for MNIST Handwritten Digit Dataset & Fashion MNIST Dataset

TABLE IV
MODEL CONFIGURATION

Layer name	Filters per node
Input layer	$28 \times 28 \times 1$
Convolution layer-1	32 filters, 3×3
Max pooling layer-1	2×2
Convolution layer-2	64 filters, 3×3
Max pooling layer-2	2×2
Convolution layer-3	128 filters, 3×3
Max pooling layer-3	2×2
Fully connected layer	128
Output layer	10

The proposed architecture for the Fashion MNIST dataset in Figure 2, characterized by its three convolutional layers, is precisely articulated in the subsequent figure.

Our algorithm expertly combines Convolutional Neural Networks (CNNs) with the MNIST Handwritten Digit and Fashion MNIST datasets, resulting in a standardized framework for image classification that promotes the advancement of CNN applications. The procedure is highly effective, offering a streamlined process from sequential model initialization to optimization and evaluation, thereby facilitating the creation of efficient, dataset-agnostic models. This broad applicability and future extensibility provide researchers with a valuable blueprint for harnessing the strengths of CNNs across various domains, leading to significant strides in deep learning and pattern recognition.

V. RESULT & PERFORMANCE ANALYSIS

This section will discuss conducted experiments, performance analysis, and comparison with other learning algorithms. Two different layer setups are used for two different datasets. Three convolution layers are used in MNIST handwritten digit dataset. On the other hand, the Fashion MNIST dataset is classified with three convolution layers.

A. MNIST Handwritten Digit Dataset

Training and validation accuracy versus epoch plot is shown in Figure 3. The graph depicts training and validation accuracy over epochs during a machine learning model's training process, with both accuracies starting below 0.94 and plateauing close to 0.98, indicating stable and high model performance.

Algorithm 1 CNN Training and Evaluation

0: **Inputs:**
1. Fashion_MNIST: $(X_{\text{fashion\_train}}, Y_{\text{fashion\_train}})$, $(X_{\text{fashion\_test}}, Y_{\text{fashion\_test}})$
1. Handwritten Digit_MNIST: $(X_{\text{handwritten\_train}}, Y_{\text{handwritten\_train}})$, $(X_{\text{handwritten\_test}}, Y_{\text{handwritten\_test}})$

0: **Procedure:**
1. Load Data: Fashion_MNIST and Handwritten Digit_MNIST.
2. Normalize Images: $X /= 255$ for both datasets.
3. Reshape Data: X.reshape(samples, $28, 28, 1$).
4. Encode Labels: $Y_{\text{one\_hot}} = \text{OneHot}(Y)$.
5. Partition: Split $(X_{\text{train}}, Y_{\text{train\_one\_hot}})$ into $(X_{\text{train}}, X_{\text{val}})$ and $(Y_{\text{train\_one\_hot}}, Y_{\text{val\_one\_hot}})$.
6. Initialize Model: $model = \text{Sequential}()$.
7. Define Layers for each layer l in $\{1, 2, \ldots, L\}$:
   model.add(Conv2D(filters_l, kernel_size=(3, 3), activation='linear', padding='same'))
   model.add(LeakyReLU(alpha=0.1))
   model.add(MaxPooling2D(pool_size=(2, 2), padding='same'))
   If Fashion_MNIST: model.add(Dropout(rate_l))
8. Flatten Output: model.add(Flatten()).
9. Add Dense Layers: model.add(Dense(128, activation='linear')), model.add(LeakyReLU(alpha=0.1)).
10. Output Layer: model.add(Dense(num_classes, activation='softmax')).
11. Compile Model: model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy']).
12. Train Model: model.fit(X_train, Y_train_one_hot, epochs=20, batch_size=64, validation_data=(X_val, Y_val_one_hot)).
13. Evaluate Model: model.evaluate(X_test, Y_test_one_hot).
14. Predict Classes: model.predict(X_test).
15. Compute Metrics: Calculate precision, recall, F1 score.

0: **Outputs:**
Model parameters, loss, accuracy, precision, recall, and F1 score.

=0

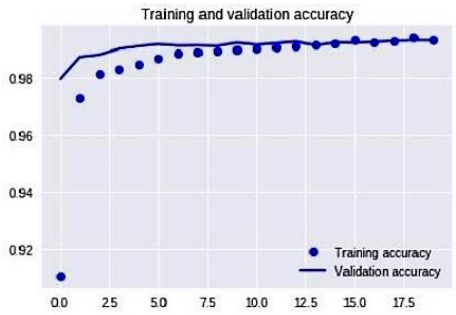

Fig. 3. Training and validation accuracy vs epoch

The implemented model for the MNIST Handwritten digit dataset achieved an impressive accuracy of approximately 98.71%. Out of 10,000 test labels, the model accurately predicted 9,915 labels. Illustrations of some correctly predicted labels are provided in Figure 4. We found 85 labels as incorrect from our experiment.

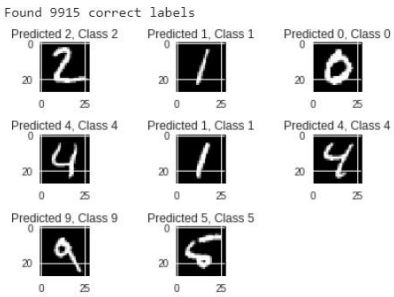

Fig. 4. Sample of correctly predicted labels For MNIST Handwritten Digit Dataset

The findings of this study showcase the outstanding capabilities of the suggested learning algorithm compared to previously utilized methods. A comprehensive evaluation was conducted to thoroughly assess the effectiveness of various conventional approaches, such as the Support Vector Machine, Multilayer Perceptron, Random Forest Algorithm, Random Tree, Naïve Bayes, Bayes Net, and the j48 Decision Tree. Table V systematically compares the outcomes obtained from these methods in previous research endeavors, highlighting the exceptional performance of the proposed methodology in this study [22].

TABLE V
PERFORMANCE COMPARISON

Classifiers Name	Accuracy
Multilayer Perceptron	90.37%
Support Vector Machine	87.97%
Random Forest	85.75%
Bayes Net	84.35%
Naïve Bayes	81.85%
J48	79.51%
Random Tree	85.6%
Proposed Model	98.71%

This comparison in Table V reflects a performance improvement and justifies using a three-layer convolutional model in MNIST handwritten digit data set. Three-layer CNN gives precise results where else more layer addition don't increase the efficiency and the addition of more layer also increases the complexity of the model.

B. Fashion MNIST Dataset

The proposed model for the Fashion MNIST dataset leads to 91.39% accuracy. Figure 5 is a picture representation of training and validation accuracy vs. epoch that describes the accuracy of this model for this dataset.

Fig. 5. Training and validation accuracy vs epoch

Figure 6 describes some of the outputs from 9134 correctly predicted labels among 10000 testing labels. Displayed in the image are a series of grayscale samples extracted from a dataset. It is worth noting that the machine learning model has correctly classified 9,134 instances, as indicated by the identical predicted and actual class labels positioned above each image. These findings demonstrate the model's exceptional accuracy in deciphering and classifying handwritten numerical data.

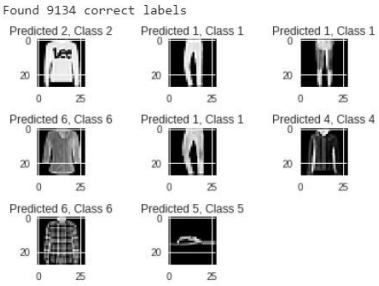

Fig. 6. Sample of some correctly predicted labels

This model performed better than the output accuracy of this dataset's previous learning algorithm. To gauge and investigate the performance of the selected methods or algorithms, namely Decision Tree Classifier, Extra Tree Classifier, K Neighbors Classifier, Linear SVC Classifier, Random Forest, SVC, and Logistic Regression, with the proposed method to compare the data set results. The following Table VI compares the results of various algorithms implemented in previous experiments and compares them with the submitted work [22].

TABLE VI
PERFORMANCE COMPARISON

Classifiers Name	Accuracy
Decision Tree Classifier	79.9%
Extra Tree Classifier	77.5%
K Neighbors Classifier	85.4%
Linear SVC Classifier	83.6%
Random Forest Classifier	77.6%
SVC	89.7%
Logistic Regression	84.2%
Proposed Model	91.39%

The comparative analysis underscores a notable performance enhancement, advocating for applying the three-layer convolutional model to the Fashion MNIST dataset. The rationale behind employing a three-layer CNN architecture is its ability to process and analyze complex datasets effectively.

VI. CONCLUSION

In this study, we developed distinct models for two datasets: one consisting of grayscale images of handwritten digits and the other featuring images of fashion items. Our deployment of these models led to highly efficient and accurate outcomes without any bias. Specifically, the three-layer convolutional model applied to the MNIST handwritten digit dataset achieved an impressive accuracy of 98.71%. In comparison, the three-layer convolutional model achieved a commendable accuracy of 91.39% for the Fashion MNIST dataset without any bias. Improvements in data sorting and validation processes will further enhance the accuracy of these models. Moving forward, we aim to extend the application of our model to more realistic and higher-resolution images and explore its potential to recognize various languages and alphabets. This will broaden the scope of its applicability and lead to exciting new directions for our research.

REFERENCES

[1] W. Sultani, C. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun. 2018. doi: 10.1109/cvpr.2018.00678.

[2] A. Boyle, G. B. Ross, and R. B. Graham, "Machine Learning and Deep Neural Network Architectures for 3D Motion Capture Datasets," 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Jul. 2020. doi: 10.1109/embc44109.2020.9176426.

[3] Md. N. Absur, "Anomaly detection in biomedical data and image using various shallow and deep learning algorithms," in Algorithms for intelligent systems, 2022, pp. 45–58. doi: 10.1007/978-981-16-6460-1_3. Available: https://doi.org/10.1007/978-981-16-6460-1_3

[4] X.-L. Ren and A.-X. Chen, "Solving the VRP Using Transformer-Based Deep Reinforcement Learning," 2023 International Conference on Machine Learning and Cybernetics (ICMLC), Jul. 09, 2023. doi: 10.1109/icmlc58545.2023.10327956.

[5] K. F. Ahmed Nasif, Md. Nurul Absur, and Md. Al Mamun, "Order Dependency in Sequential Correlation," 2019 3rd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Dec. 26, 2019. doi: 10.1109/icecte48615.2019.9303557.

[6] G. Schwartz and K. Nishino, "Recognizing Material Properties from Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 8, pp. 1981–1995, Aug. 2020, doi: 10.1109/tpami.2019.2907850.

[7] S. P and R. R, "A Review of Convolutional Neural Networks, its Variants and Applications," in Proc. of the 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Feb. 2023, doi: 10.1109/iciscois56541.2023.10100412.

[8] X. Zhang et al., "End-to-End Latency Optimization of Multi-view 3D Reconstruction for Disaster Response," in Proc. of the 2022 10th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud), Aug. 2022, doi: 10.1109/mobilecloud55333.2022.00010.

[9] M. Sheng et al., "Facial Expression Recognition Based on Sparse Autoencoder and Shallow Convolutional Neural Network," in Proc. of the 2020 15th International Conference on Computer Science & Education (ICCSE), Aug. 2020, doi: 10.1109/iccse49874.2020.9201819.

[10] C. Farabet et al., "Learning Hierarchical Features for Scene Labeling," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1915–1929, Aug. 2013, doi: 10.1109/tpami.2012.231.

[11] A. H. Abdulnabi et al., "Multimodal Recurrent Neural Networks With Information Transfer Layers for Indoor Scene Labeling," IEEE Transactions on Multimedia, vol. 20, no. 7, pp. 1656–1671, Jul. 2018, doi: 10.1109/tmm.2017.2774007.

[12] H. Le et al., "Guided Anchoring Cascade R-CNN: An Intensive Improvement of R-CNN in Vietnamese Document Detection," in Proc. of the 2021 8th NAFOSTED Conference on Information and Computer Science (NICS), Dec. 2021, doi: 10.1109/nics54270.2021.9701510.

[13] L.-C. Chen et al., "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs," arXiv:1412.7062 [cs], Jun. 7, 2016, available at: https://arxiv.org/abs/1412.7062v4.

[14] D. K. Yadav, N. Kumari, and S. Harron, "Advances in Convolutional Neural Networks for Object Detection and Recognition," in Proc. of the 2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC), Jan. 2024, doi: 10.1109/icocwc60930.2024.10470695.

[15] Z. Yan et al., "HD-CNN: Hierarchical Deep Convolutional Neural Network for Large Scale Visual Recognition," arXiv:1410.0736 [cs, stat], May 15, 2015, available at: https://arxiv.org/abs/1410.0736.

[16] T. Xiao et al., "The Application of Two-Level Attention Models in Deep Convolutional Neural Network for Fine-Grained Image Classification," arXiv:1411.6447 [cs], Nov. 24, 2014, available at: https://arxiv.org/abs/1411.6447.

[17] S. Ji, W. Xu, M. Yang, and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 221–231, Jan. 2013, doi: 10.1109/tpami.2012.59.

[18] E. Xhaferra, E. Cina, and L. Toti, "Classification of Standard FASHION MNIST Dataset Using Deep Learning Based CNN Algorithms," in Proc. of the 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Oct. 2022, doi: 10.1109/ismsit56059.2022.9932737.

[19] E. Rani et al., "MNIST Handwritten Digit Recognition Using Machine Learning," in Proc. of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Apr. 2022, doi: 10.1109/icacite53722.2022.9823806.

[20] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms," [Online]. Available: https://github.com/zalandoresearch/fashion-mnist, 2017.

[21] Mybridge, "Learn Machine Learning from Top 50 Articles for the Past Year (V.2019)," Medium, 29 Dec. 2018. [Online]. Available: https://medium.mybridge.co/learn-machine-learning-from-top-50-articles-for-the-past-year-v-2019-15842d0b82f6. [Accessed: Mar. 31, 2024].

[22] S. M. Shamim et al., "Handwritten Digit Recognition Using Machine Learning Algorithms," *Indonesian Journal of Science and Technology*, vol. 3, no. 1, p. 29, Apr. 10, 2018. doi: 10.17509/ijost.v3i1.10795.

[23] Deng, L. (2012). The mnist database of handwritten digit images for machine learning research. IEEE Signal Processing Magazine, 29(6), 141–142.

[24] Xiao, H., Rasul, K., Vollgraf, R. (2017, August 25). Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. https://arxiv.org/abs/1708.07747