Mehran University Research Journal of Engineering and Technology

https://doi.org/10.22581/muet1982.2204.15

2022, 41(4) 146-156

Evaluation of deep learning approaches for optical character recognition in Urdu language

Mehek Riaz [a], Syed Muhammad Ghazanfar Monir [b], Rija Hasan [b,*]

[a] *College of Computing and Information Sciences, Karachi Institute of Economics and Technology, Karachi Pakistan*

[b] *Department of Electrical and Computer Engineering, Mohammad Ali Jinnah University, Karachi Pakistan*

[*] Corresponding author: Rija Hasan, Email: rija.hasan@jinnah.edu

Received: 22 September 2021, Accepted: 25 February 2022, Published: 01 October 2022

KEYWORDS

Convolutional Approaches for OCR

Discriminative Analysis Densenet121

Inception V3

Optical Character Recognition

Resnet50

Support Vector Machine (SVM)

Vggnet16

ABSTRACT

With the evolving technological era, the optical character recognition systems have substantial execution, considering the widespread use of daily hand-written human transaction. Optical Character Recognition (OCR) is an implementation of Computer Vision that digitizes numerous hand dealt documents for further analysis and formatting. OCR is achieved by various ways such as discriminative analysis and deep learning. This paper focuses on evaluating deep learning models on a hand-written compiled dataset of Urdu digits. The evaluation is performed for deep convolutional learning algorithms; VGGNet16, InceptionV3, ResNet50, and DenseNet121. The convolutional models are pre-trained on the ImageNet. The model weights of fully connected layers have been evaluated, reducing the training time of the convolutional layers. The testing accuracy for the compiled dataset is observed as, ResNet50 with 96%, InceptionV3 with 95%, VGGNet16 with 95% and DenseNet121 with 94%.

1. Introduction

Optical character recognition (OCR) is a system so constructed as to evaluate a given input text into machine editable format [1]. It is an efficient way of transforming many old manuscripts, medieval texts, and old documentation as digitized files [2] and maintain daily manually recorded information in editable format [3]. This has contributed towards making information retrieval easier by efficiently accessing information that would otherwise be much time consuming, as it has removed hassle of going through manual bulk material and a simple word search has replaced this task as a result of digitization as an efficient way. The organizations such as museums and archeological researchers can greatly benefit from this approach as historically significant text captured can be converted in virtual format for analysis, the law concerning documents can be digitally preserved as well as consistency in education persistence could be easier as the information would be virtually stored [4]. There is very less work contributed to the hand-written text recognition in Urdu [5], moreover, dataset availability for Urdu digits is scarce too. Therefore, OCR for Urdu language is a huge working area of research contribution and exploration, and this research has contributed towards compilation of digitally recording Urdu handwritten digits, then building a deep learning model that recognizes and classifies the digits respectively.

OCR constitutes of a recognition and classification mechanism. The hand-written text is analyzed to extract

some key features unique to the character. These features are then classified in their respective classes. This intuition of pattern matching can be approached many ways. K-means Clustering is the algorithm that clusters the similar samples together [6]. Another method of classification is Discriminative analysis (DA), which classifies objects when groups are classified upon their discriminative features [7].

The decision making in OCR is greatly supported by an optimized method of feature selection. The features derived from the sample constitutes the actual unique characteristics that contributes to the inter class difference. Conventional ways of describing features are a derivative of diverse ways, one of which is HOG (Histogram of Oriented Gradient). The HOG is such a feature descriptor that can identify objects form images. The image under consideration for recognition is segmented, and these segments are further analyzed. The histogram of these gradients become feature descriptors. The HOG feature vector the composes the gradient description in the localized portion of the image descriptor [8]. The SURF (Speeded Up Robust Features) is a feature extraction technique that utilizes box filters for computation and is a fast and robust method for object detection or real-time tracking applications [9].

This paper focuses on evaluating deep learning model on a compiled Urdu digit dataset by evaluating different deep learning architectures. The paper is structured in the following manner. Introduction is discussed in section 1 which is followed by literature review in section 2, the methodology is described in section 3, followed by results and discussion in section 4, and conclusion is given as section 5.

2. Literature Review

Optical Character recognition (OCR) is the technique that has the ability to read scripts. It is given scanned hand-written documents or images that contain written information such as educational degrees, certificates, licenses, number plates, sign boards, etc. [5]. Many machine learning approaches have been implemented in developing OCR. Implementation of Support vector machines (SVM) for OCR in [10-11] focus on an issue of recognizing similar characters in OCR of Chinese, Japanese, and Bangla, Thai are used to classify characters. Another example of traditional OCR approach is implemented using Radial Basis Function (RBF) [12]. Random forests [13] and Naive Bayes [14] algorithm are also evaluated for optical character recognition. All the aforementioned approaches involved aspect of classification, which wouldn't have resulted in success if not for intelligent feature description. The more intelligently detailed a feature description is extracted, the easier and simplistic it is to classify based on the distinguished attributes. Most effective methods are transform based feature extraction such as DFT (Discrete Fourier Transform) and DWT (Discrete Wavelet Transform) [15], Histogram of gradients HOG [16-17] and some modified approaches [18-19] are also implemented.

Character distinguishes due to unique way of expression. A character recognition algorithm aided by any classification technique; the feature description is preferably texture based so as to distinguish the samples based on how they are written. A convolutional neural network is a composite of a feature extraction unit followed by a network of fully connected dense layers to produce binary, multiclass or regression defined outcome [20]. An approach of text recognition by segmenting lines, words and characters from Islamic manuscripts written in Arabic is proposed by Alrehali et al. [21]. An approach combining Deep Convolutional Neural Network (DCNN) and Support Vector Machine SVM is proposed by Mahmoud Shams et al. [22] for recognizing handwritten Arabic characters with 95.07% accuracy rate.

A variety of CNN architectures have been implemented in research. A research proposed by Latha et al. [23] uses a DenseNet for segmenting graphical information and text from road signs for automatic driven cars.. The model is modified to classify 12 instead of 1000. A 92.7% accuracy is observed for logo classification and for text detection the accuracy of 96.5% after finetuning parameters of DenseNet. A DenseNet implementation reported by Adriano et al. [24] caters the problem of data entry in several industries to avoid high error rate caused by manual work. The accuracy of training using DenseNet was found 98.62%.

VGGNet is a variant of CNN and is widely used for OCR. A research implemented by Cheang et al. [25] implements OCR by segmenting and recognizing vehicle license plate. By implementing this approach, the entire image is scanned, and the features are extracted. The reported error rate is 0.79%. An implementation of VGGNet by Hakim [19] utilizes the recognition of Bangla characters. The VGGNet network consists of 6 convolution layers followed by 2 fully connected layers and an output layer. The reported model classification accuracy is 92.38%. A VGGNet16 model is implemented by Murti [26] for recognizing Balinese script. The model is modified in terms of parameter description and classification accuracy of

99.74% in training data and 99.78% for test data is observed. A CNN VGGNet16 architecture is implemented by Naragudem Sarika et al. [27] for Telgu character recognition.

ResNet is a variant of CNN which is also implemented in the problem of OCR. Bartz et al. [28] proposed an approach for text detection and recognition from natural scenes. The model results a recognition accuracy of 97%. A ResNet architecture proposed by Kumar et al. [29] that deals with sign language recognition based on 3-D data. A modified ResNet with connived features is implemented, that has feedforward CNN layers of densely connected traditional ResNet. The max error rate was found 3.95%. Another example of ResNet based OCR by Yang et al. [30] which is based natural scene text detection. ResNet is effective in picking up minute details and performing online. The research yields a 0.48% error rate by implementing the original RestNet architecture with additive depth for optimization.

A CNN based architecture ideal for sequential data is Recurrent Neural Network (RNN). Long-Short-Term Memory units (LSTMs), a type of RNN that has more control and flexibility. A research proposed by Sahu et al. [31] concerns and OCR architecture for word prediction. A minimum error of 0.9% is reported. A CNN-RNN based model is proposed for cursive script recognition [32]. This approach recognizes Urdu text from printed documents. A hybrid architecture with CNN-RNN and Bi-directional LSTM are utilized to report 89.84% accuracy. An Urdu text OCR approach proposed in [33]. Features are extracted by pixel arrangement in the image. The reported accuracy is 98%. An OCR approach for Urdu text proposed for a large number of ligatures and nastaliq writing style by Rafeeq et al. [34]. The reported training accuracy is 99.6% whereas the validation accuracy is 73.13% on the convolutional layer. The deep neural network reports training and validation accuracy to be 99.6% and 95.02% respectively. An RNN architecture based OCR proposed by Achkar et al. [35] recognized hand-written medical prescriptions in English. The reported accuracy is 98%. An Urdu text recognition proposed for text written in varying font size and ligatures by Naseer et al. [36]. The model extracts Meta features with accuracy of 97.49%, character and ligature recognition with 98.05% accuracy and model accuracy of 99.05%. In a recent study by Ahmed et al. [37], an RNN based OCR system with bidirectional LSTM gates with a reported error rate between 6.04-7.93%.

Deep learning methods optimizes the traditional ML problems while producing more efficient results. Although the deep learning methods perform efficiently, still less work has been dedicated to the Urdu dataset, let alone be its evaluation on deep learning algorithms.

3. Methodology

3.1 Dataset Preparation and Preprocessing

We have created a dataset of Urdu hand-written numbers which contains 100,000 images of 0 to 9 digits. The dataset has been compiled by collecting specimen from 300 subjects. The letters were recorded on a grid of 20 rows and 10 columns as it can be seen in Fig. 1. The manually recorded images were scanned, cropped, segmented, and processed such that the data should contain skewed variations as well, as it can be seen in Fig. 2. All the images in the dataset are of fixed dimension of 126 by 120. The dataset is then composed into a directory of training with 50,000 samples, testing with 20,000 samples and validation sets with a sample count of 30,000. Evaluation of Deep Learning models.

There are four major ImageNet trained Deep Learning models evaluated upon our compiled Urdu hand-written digit datasets. The details of the evaluation are as follows.

1. VGGNet16
2. Inception
3. ResNet50
4. DenseNet121

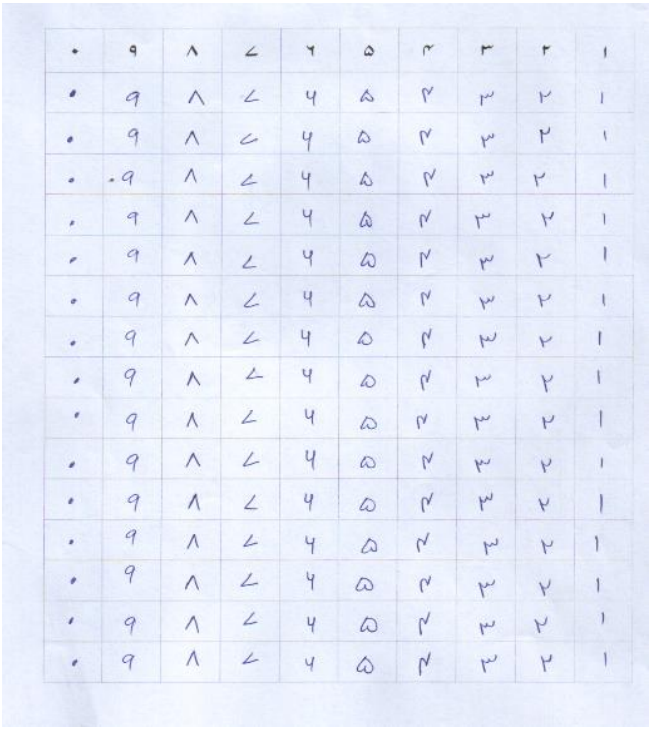

Fig. 1. Hand-written samples

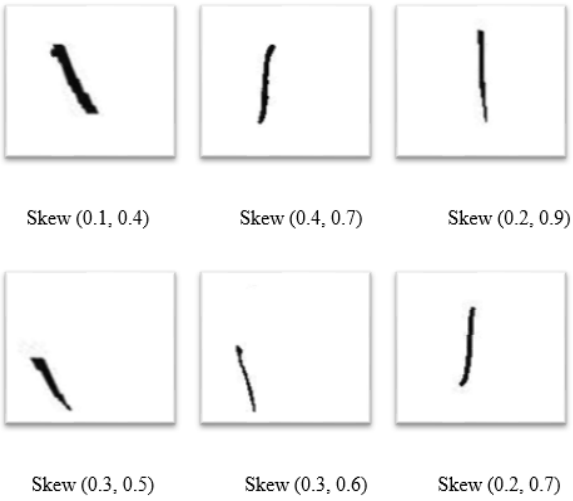

Fig. 2. Dataset augmentation

3.2 Evaluation through VGGNet16

The VGGNet16 [38] has a total of 16 weighted layers within the hidden architecture as it can be seen in Fig. 3. These 16 layers are a composition of 13 convolutional layers and 2 dense layers. Moreover, there are 5 max pooling layers in between each convolutional layer. The VGGNet16 is pre-trained on ImageNet, a vast image dataset with 1000 classes. The intuition to modify the training to our liking is to expect earlier ascend towards substantial accuracy. The pre-trained network is renovated for the ImageNet dataset, but the dense layers are replaceable for customizing the network to cater to other classes. We have coupled the pre-trained convolutional layers with our designated dense layers to respectively predict our 10 classes.

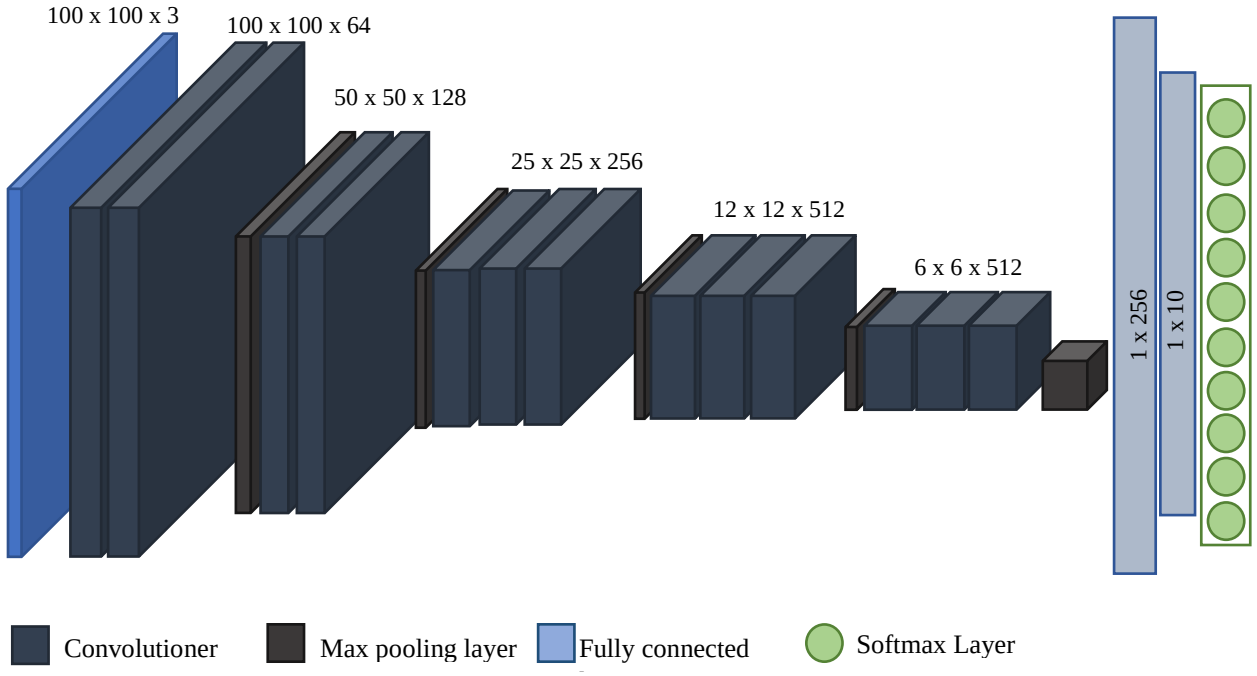

Fig. 3. VGGNet 16 network

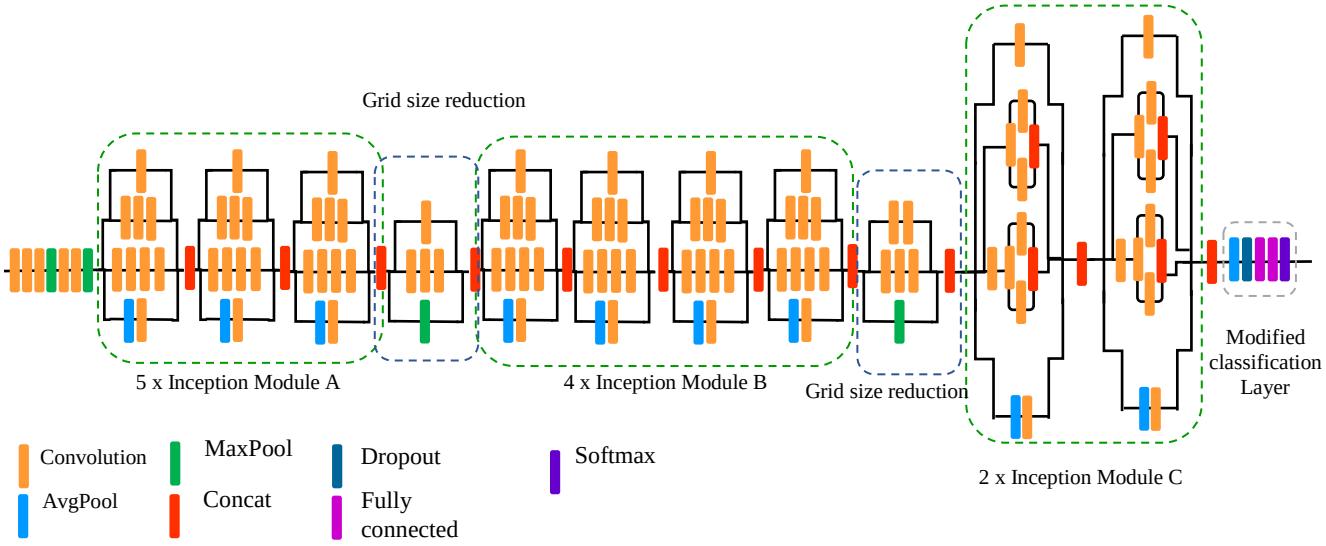

Fig. 4. InceptionV3 network

3.3 Evaluation of InceptionV3

The inception network is constructed such that the layers are stacked deeper. The InceptionV3 [39] uses inception modules and grid size reduction. There are 3 inception modules and 2 size reduction modules, all composite of 11 models in InceptionV3. The base network is already trained on ImageNet dataset and the learning has been transferred to the dense layers. At the end of the base network, a pooling layer followed by a dropout layer is present to avoid overfitting. Then the customized fully connected ReLu activated layers are present with 256 and 10 neurons respectively ending with a Softmax layer as it can be seen in Fig. 4.

3.4 Evaluation of ResNet50

The ResNet50 [40] that can be seen in Fig. 5 has 4 residual blocks. The whole network is a composite of 50 layers which is cascaded with the proposed classification dense network. Each 5 blocks are a composite of repetition of 3 group of convolution layers. First block has 3 repetitions of $\{[1 \times 1 \times 64], [3 \times 3 \times 64], [1 \times 1 \times 256]\}$, second block has 4 repetitions of $\{[1 \times 1 \times 128], [3 \times 3 \times 128], [1 \times 1 \times 512]\}$, third block has 6 repetitions of $\{[1 \times 1 \times 256], [3 \times 3 \times 512], [1 \times 1 \times 1024]\}$, and the fourth block has $\{[1 \times 1 \times 512], [3 \times 3 \times 512], [1 \times 1 \times 2048]\}$. The proposed bit is to cascade fully connected layers $[1 \times 256]$ and $[1 \times 10]$ to finally end up with Softmax layer with 10 units.

3.5 Evaluation of DenseNet121

The Densenet121 [41] is an architecture with 121 layers. There are 4 dense layers followed by 4 transition layers. Each dense layer has 2 cascaded interconnected dense blocks as it can be seen in Fig. 6. Each dense block is followed by a transition layer which has a 1×1 convolution layer and a 2×2 average pooling layer. The network average pools the output from the dense layers and forwards it to the fully connected layers $[1 \times 256]$ and $[1 \times 10]$ to finally end up with Softmax layer with 10 units. The DenseNet121 is trained upon ImageNet, the transferred learning is evaluated on the fully connected layers.

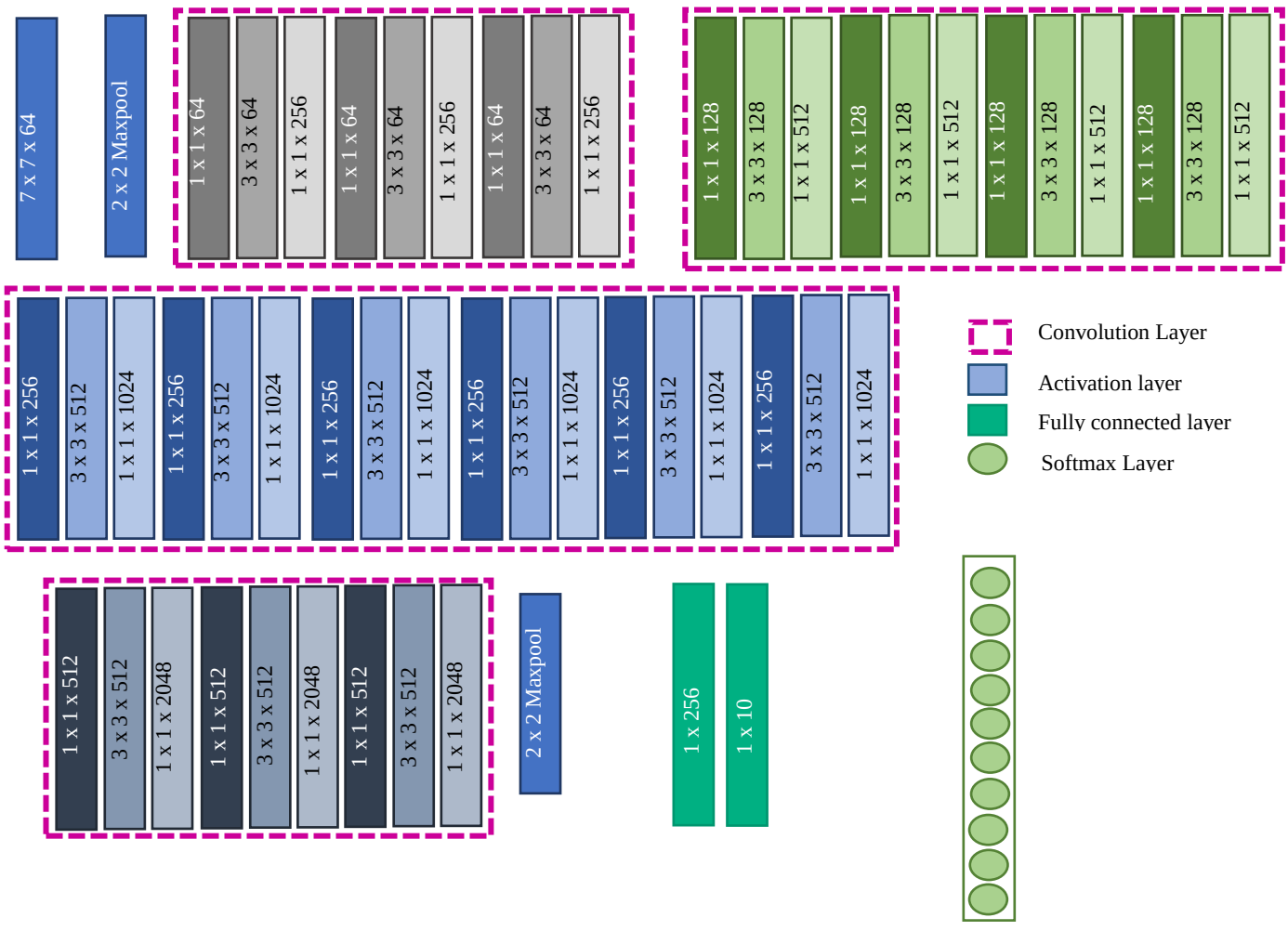

Fig. 5. ResNet50

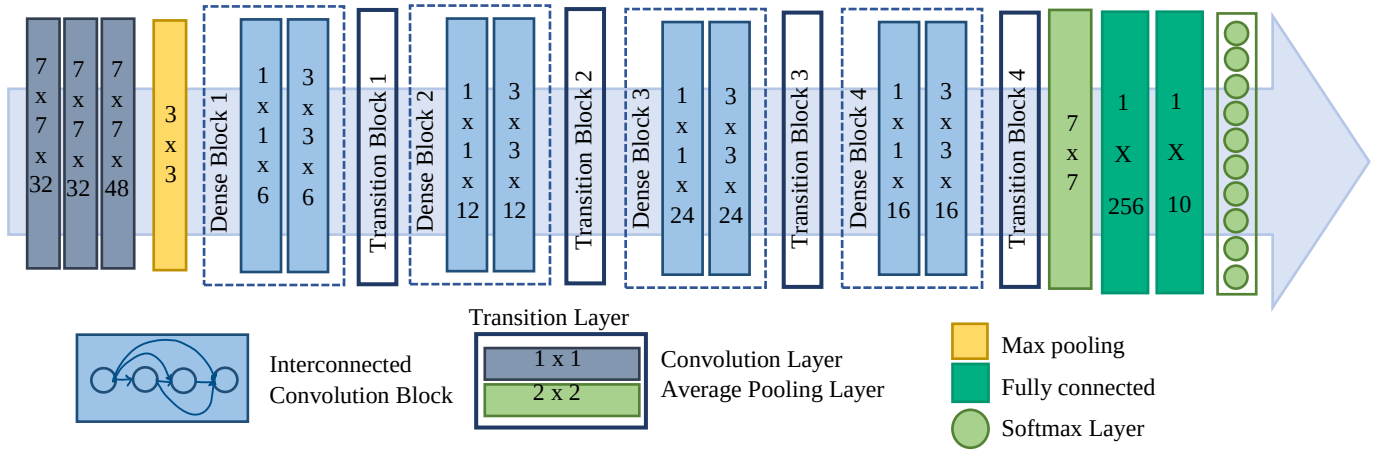

Fig. 6. DenseNet 121

4. Results and Discussion

In this research, Urdu hand-written digit classification is performed on the classical deep learning models to evaluate the performance. The evaluated deep neural network (DNN) in this research are VGGNet16, Resnet50, InceptionV3 and DenseNet121.

All the evaluated deep neural network models are pre-trained by using ImageNet weights. For further training, each model has been trained at a fixed number of 10 epochs to produce classification. As it can be seen in Table 1, the ResNet50 model comparatively gives highest accuracy of 97.05%. InceptionV3 and VGGNet16 comparably performs the same but the VGGNet16 has lesser loss percentage. The DenseNet121 performs least accurate by bearing 95.99% accuracy.

Table 1

Comparison of training accuracies of the evaluated methods

Model	Iterations	Accuracy, %	Loss, %
ResNet50	10	97.05	0.000055
InceptionV3	10	96.55	0.001563
VGGNet16	10	96.33	0.000089
DenseNet121	10	95.99	0.000055

The confusion matrices reveal more about the classification errors. For each architecture, confusion matrices have been generated to further analyze the performance that can be seen from Figs. 7 – 10. For each matrix in these figures, the horizontal and vertical green bars represent the cross comparison between each Class, whereas the blue cells in each matrix represent the true positives occurred for each Class. In Fig. 7, we can see confusion matrix of VGGNet16, we see misclassifications occur in Class 2 as Class 6, Class 4 as Class 6 and Class 6 as Class 2. For the confusion matrix of Reset50 in Fig. 8, we see misclassifications as Class 2 as Class 6 and Class 6 as Class 2. Same classification error of Class 2 as Class 6 and Class 6 as Class 2 can be observed for InceptionNetV3 in Fig. 9 and DenseNet121 in Fig. 10. The misclassification between Class 2 and Class 6 is occurred due to their appearance as mirror image of the other as it can be referred to Fig. 11. The convolutional neural networks are invariant to mirroring of the image. This can be the key aspect in misclassification because as it can be seen in Fig. 1 the Class 2 and Class 6 are written as reflection of one another, whereas the misclassification of Class 4 as Class 6 the confusion in Class 4 occurred due to manual recording discrepancies of the digit 4. In Fig. 10, we see some misclassification between Class 7 and Class 8 also. As in Fig. 11, digit 7 in Urdu if rotated a little can be confused as digit 8 in Urdu. Therefore, these minor misclassifications have occurred due to the image rotation and mirroring during the feature extractions for each convolutional network's architectures. Tables 2 – 5 show the detailed validation accuracy for each class, where precision is the ratio of truly positive against total predicted positive, recall is the ratio between true positive and actual positive, precision is the digit classification precision, F1 score measures a balanced value between precision and recall just in case if there is a class imbalance, and the support is the number of occurrences in that class.

$$precision = \frac{true\ positive}{true\ positive + false\ positive}$$

$$recall = \frac{true\ positive}{true\ positive + false\ negative}$$

$$F1 = \frac{2\ \times recall\ \times precision}{recall + precision}$$

0	2790	121	6	1	4	10	2	4	32	10
1	11	2969	0	0	0	0	0	3	1	1
2	2	13	2752	3	18	0	182	6	9	0
3	0	0	1	2982	2	0	0	0	0	0
4	1	14	14	1	2905	0	29	8	11	2
5	7	11	0	1	13	2894	0	9	46	4
6	2	22	114	8	104	0	2723	2	7	3
7	5	19	2	0	0	0	0	2985	1	0
8	20	69	7	0	3	4	10	6	2865	1
9	11	19	2	0	9	8	5	2	15	2914
	0	1	2	3	4	5	6	7	8	9

Fig. 7. Confusion matrix VGGNet16

If we consider a balanced score summary f1 score for each model performance. In Table 2, we have VGG16 scores, least score of 0.92 is observed for Class 6 and 0.94 for Class 2. If we observe the Table 3 that has DenseNet121 scores, least score of 0.89 is shared by both Class 2 and Class 6. The least accuracy is seen for Class 2 and Class 6 when the evaluations are compared.

In Table 4, the evaluation of ResNet50 reports 0.94 f1-score for both Class 2 and Class 6. Finally in Table 5, we have results of InceptionV3 that has least score of 0.93 for both Class 2 and Class 6. Once again with the reference of Fig. 1, both the digits are written as horizontal reflections of one another. Therefore, the accuracy is getting compromised as the collected features are irrespective of rotation and mirroring. As a future work, we reckon that during the feature collection, transformations that are extracted by mirroring horizontally and vertically are ruled out for any language that may contain mirrored numerals. As it can be seen in Fig. 12, the best testing performance has been observed through the InceptionNetV3 of 96%, whereas ResNet50 and VGG16 work the same with classification accuracy of 95% and DenseNet121 exhibited comparatively low accuracy of 94%.

0	2834	86	2	1	9	12	4	2	16	14
1	21	2937	1	0	2	0	1	19	4	0
2	5	5	2821	17	17	0	116	2	2	0
3	0	0	1	2973	11	0	0	0	0	0
4	2	2	7	0	2952	0	9	2	3	3
5	15	9	0	0	5	2911	0	6	32	7
6	2	10	152	3	30	0	2784	1	2	1
7	3	3	1	0	4	2	2	2968	1	1
8	41	42	9	1	10	16	12	7	2839	8
9	11	6	0	1	4	7	5	2	3	2946
	0	1	2	3	4	5	6	7	8	9

Fig. 8. Confusion matrix ResNet50

The limitation so far has been observed as the confusion between Class 2 and Class 6. The Urdu digits, 2 and 6 are written as if being mirrored along y-axis, as it can be seen in Fig. 11. The Deep Convolutional Neural Networks DCNN are designed such that they work on invariance of rotation and mirroring, which is the main reason that most misclassifications occurred for all architectures in both of these classes. The proposed approaches only work on the images of the digits. Had there been inclusion of LSTMs, there would be individual work contributed towards cognitive comprehension and remembrance of each numeral rather than just visual recognition.

0	2874	38	6	0	5	16	4	5	21	11
1	43	2930	2	0	3	0	0	3	4	0
2	5	7	2756	8	7	1	185	6	5	5
3	0	0	2	2974	2	0	7	0	0	0
4	7	5	17	3	2917	0	19	6	3	8
5	30	13	0	1	3	2876	0	12	47	11
6	6	18	150	3	17	0	2792	0	1	3
7	11	6	4	0	1	0	0	2948	1	2
8	75	42	6	0	9	6	14	6	2825	2
9	19	3	1	0	7	13	8	1	7	2926
	0	1	2	3	4	5	6	7	8	9

Fig. 9. Confusion matrix of InceptionV3

0	2835	81	2	1	5	9	16	4	10	17
1	28	2902	0	0	11	0	6	33	1	4
2	3	4	2575	2	29	0	363	1	8	0
3	0	0	1	2979	1	0	4	0	0	0
4	3	7	18	1	2942	0	8	2	1	3
5	12	8	4	2	2	2896	2	5	43	11
6	4	9	169	2	24	0	2775	1	0	1
7	3	6	1	0	6	0	6	2963	0	0
8	44	44	5	0	7	5	27	76	2837	11
9	7	7	1	0	2	5	10	2	4	2947
	0	1	2	3	4	5	6	7	8	9

Fig. 10. DenseNet121 confusion matrix

0	1	2	3	4	5	6	7	8	9
۰	۱	۲	۳	۴	۵	۶	۷	۸	۹

Fig. 11. English and Urdu numerals

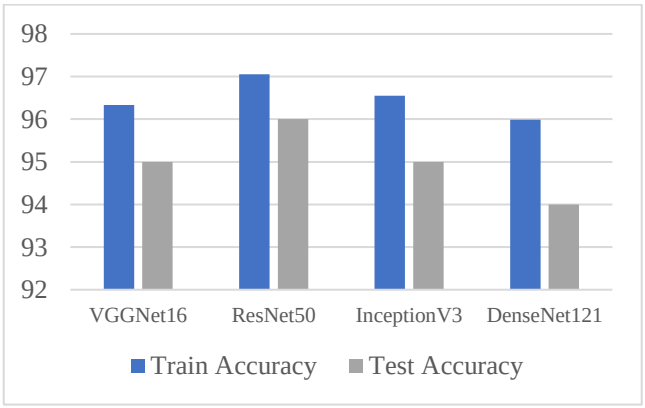

Fig. 12. Training and Testing accuracy of the evaluated methods

Table 2
VGGNet16 classification report

	precision	recall	f1-score	Support
0	0.98	0.94	0.96	2980
1	0.91	0.99	0.95	2985
2	0.95	0.92	0.94	2985
3	1.00	1.00	1.00	2985
4	0.95	0.97	0.96	2985
5	0.99	0.97	0.98	2985
6	0.92	0.91	0.92	2985
7	0.99	0.99	0.99	2985
8	0.96	0.96	0.96	2985
9	0.99	0.98	0.98	2985
Accuracy			0.96	29845
macro avg.	0.96	0.96	0.96	29845
weighted avg.	0.96	0.96	0.96	29845

Table 3
DenseNet121 classification report

	precision	recall	f1-score	support
0	0.96	0.95	0.96	2980
1	0.95	0.97	0.96	2985
2	0.93	0.86	0.89	2985
3	1.00	1.00	1.00	2985
4	0.97	0.99	0.98	2985
5	0.99	0.97	0.98	2985
6	0.86	0.93	0.89	2985
7	0.98	0.99	0.99	2985
8	0.98	0.95	0.96	2985
9	0.98	0.99	0.99	2985
accuracy			0.96	29845
macro avg.	0.96	0.96	0.96	29845
weighted avg.	0.96	0.96	0.96	29845

Table 4
ResNet50 classification report

	precision	recall	f1-score	support
0	0.97	0.95	0.96	2980
1	0.95	0.98	0.96	2985
2	0.94	0.95	0.94	2985
3	0.99	1.00	0.99	2985
4	0.97	0.99	0.98	2985
5	0.99	0.98	0.98	2985
6	0.95	0.93	0.94	2985
7	0.99	0.99	0.99	2985
8	0.98	0.95	0.96	2985
9	0.99	0.99	0.99	2985
accuracy			0.97	29845
macro avg.	0.97	0.97	0.97	29845
weighted avg.	0.97	0.97	0.97	29845

Table 5

InceptionV3 classification report

	precision	recall	f1-score	Support
0	0.94	0.96	0.95	2980
1	0.96	0.98	0.97	2985
2	0.94	0.92	0.93	2985
3	0.99	1.00	1.00	2985
4	0.98	0.98	0.98	2985
5	0.99	0.96	0.98	2985
6	0.92	0.94	0.93	2985
7	0.99	0.99	0.99	2985
8	0.97	0.95	0.96	2985
9	0.99	0.98	0.98	2985
accuracy			0.97	29845
macro avg.	0.97	0.97	0.97	29845
Weighted avg.	0.97	0.97	0.97	29845

5. Conclusion

OCR constitutes of a recognition and classification mechanism. The hand-written text is analyzed to extract some key features unique to the character. These features are then classified in their respective classes. There is scarcity of work done in optical recognition specially in the East such as for Urdu and Arabic. This work can be useful for Urdu digits detection in scanned Urdu literature eBooks or android apps, while searching for a fixed numeral in a huge content such as page numbers, historic dates, ages, years etc. Furthermore, other use of this work can benefit agricultural area, because much of the work done there is manual. The agricultural management is done in rural areas where the literacy rate is low therefore people use their domestic language in expressing costs and unit counts. An application can be made to store all the manual data of crops quantities and cost written in Urdu. This can drastically improve the commercial management and standardization values of the agricultural management system, which can contribute to elevation in the economy. The proposed system is using transfer learning from the pre-trained networks. This reduces the cost and time to train the deep convolutional network on each digit. The layers are trained upon ImageNet and is accessible for everyone. The approach of only training the model weights for classification has reduced the hassle of training the whole architecture and has worked significantly well.

This paper focuses on compiling and preparing an Urdu digit dataset and evaluating different deep learning models on it compare for accuracy. In this research we have evaluated major Deep CNN architectures that are trained upon ImageNet on Urdu hand-written numerals. We have evaluated VGGNet16, InceptionV3, DenseNet121 and ResNet50. Since these networks have 1000 classes, and our database has 10 classes, we have changed the classification end of the network by using 2 fully connected layers for each evaluation that ends up in 10 Softmax units to comply with our implementation. Since the contribution of Urdu hand-written recognition is scarce in terms of dataset compilation, we have also compiled the dataset and extended it by data augmentation. The dataset contains 0 to 9 hand-written digits in Urdu. For each digit, being represented as a class possesses 12000 samples, and after applying skewness and transformations, the dataset has been stretched to 1,00,000 samples. Training samples are 50,000, 30,000 for validation and 20,000 are for testing. Evaluation has been performed with DenseNet121, InceptionV3, ResNet50 and VGGNet16 architecture, all with original base architectures trained for ImageNet. The best testing performance has been generated by the InceptionNetV3 of 96%, whereas ResNet50 and VGG16 work the same with classification accuracy of 95% and DenseNet121 produced comparatively low accuracy of 94% as it can be seen in Fig. 12.

As a future work, we reckon that during the feature collection, transformations that are extracted by mirroring horizontally and vertically are ruled out for any language that may contain mirrored numerals. Furthermore, the existing work can be integrated with LSTMs to ensure that working is done on reembrace of each numeral as well as visual recognition.

6. References

[1] C. C. Tappert, C. Y. Suen, and T. Wakahara, "The state of the art in online handwriting recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 8, pp. 787-808, 1990.

[2] M. Kumar, S. R. Jindal, M. K. Jindal, and G. S. Lehal, "Improved recognition results of medieval handwritten Gurmukhi manuscripts using boosting and bagging methodologies", Neural Processing Letters, vol. 50, no. 1, pp. 43-56, 2019.

[3] M. A. Radwan, M. I. Khalil, and H. M. Abbas, "Neural networks pipeline for offline machine printed Arabic OCR", Neural Processing Letters, vol. 48, no. 2, pp. 769-787, 2018.

[4] J. M. M. S. R. A. K. M. Uddin, "Handwritten optical character recognition (OCR): a comprehensive systematic literature review (SLR)", IEEE Access, vol. 8, pp. 142642-142668, 2020.

[5] Noman, Zeeshan, and N. Noor, "A survey on optical character recognition system", A Survey on Optical Character Recognition System, vol. 10, no. 2, p. 4, 2016.

[6] M. C. Chang and P. Bus, "Feature extraction and k-means clustering approach to explore important features of urban identity", 16th IEEE International Conference on Machine Learning and Applications, Mexico, 2017.

[7] S. F. Rashid, F. Shafait, and a. T. M. Breuel1, "Discriminative learning for script recognition", IEEE International Conference on Image Processing, Hong Kong, p. 4, 2010.

[8] H. Ren and Z. N. Li, "Object detection using edge histogram of oriented gradient", IEEE International Conference on Image Processing, 2014.

[9] R. Verma and M. R. Kaur, "Enhanced character recognition using surf feature and neural network technique", International Journal of Computer Science and Information Technologies, vol. 5, no. 4, p. 6, 2014.

[10] A. K. Ghosh and S. Afroge, "A comparison between support vector machine (SVM) and bootstrap aggregating technique for recognizing bangla handwritten characters", 20th International Conference On Computer and Information Technology, Dhaka Bangladesh, 2017.

[11] P. Inkeaw, J. Bootkrajang, S. Marukatat, T. Gonçalves, and J. Chaijaruwanich, "Recognition of similar characters using gradient features of discriminative regions", Expert Systems with Applications, vol. 134, pp. 120-137, 2019.

[12] M. Govindarajan, "Recognition of handwritten numerals using RBF-SVM hybrid model", International Arab Journal of Information Technology, vol. 13, 2016.

[13] K. Johnson, K. Gourav, Gaurav, D. Rudrapal, and S. Debnath, "OCR for devanagari numerals using zonal histogram of angle", Journal of Statistics and Management Systems, vol. 20, no. 4, p. 17, 16 nov 2017.

[14] A. Amalia, A. Sharif, F. Haisar, D. Gunawan, and B. B. Nasution, "Meme opinion categorization by using optical character recognition (OCR) and naïve bayes algorithm", 3rd International Conference on Information and Computing, 2018.

[15] A. Lawgali, A. Bouridane, M. Angelov, and Z. Ghassemlooy, "Handwritten arabic character recognition: which feature extraction method?", International Journal of Advanced Science and Technology, vol. 34, p. 8, 2011.

[16] S. Tian et al., "Multilingual scene character recognition with co-occurrence of histogram of oriented gradients", Pattern Recognition, vol. 51, pp. 125-134, 2016.

[17] A. Boukharouba, & A. Bennia, "Novel feature extraction technique for the recognition of handwritten digits", Applied Computing and Informatics, vol. 13, no. 1, pp. 19-26, 2017.

[18] N. Bi, C. Y. Suen, N. Nobile, and J. Tan, "A multi-feature selection approach for gender identification of handwriting based on kernel mutual information", Pattern Recognition Letters, vol. 121, pp. 123-132, 2019.

[19] S. A. Hakim, "Handwritten bangla numeral and basic character recognition using deep convolutional neural network", International Conference on Electrical, Computer and Communication Engineering, Cox'sBazar, Bangladesh, 2019.

[20] S. Khan, H. Rahmani, S. A. A. Shah, and M. Bennamoun, "A Guide to Convolutional Neural Networks for Computer Vision", Morgan and Claypool, p. 209, 2018.

[21] B. Alrehali, N. Alsaedi, H. Alahmadi, and N. Abid, "Historical Arabic manuscripts text recognition using convolutional neural network", 6th IEEE Conference on Data Science and Machine Learning Applications, Riyadh Saudi Arabia, pp. 37-42, 2020.

[22] A. A. E. Mahmoud Shams, Wael. Z. ElSawy, "Arabic handwritten character recognition based on convolution neural networks and support vector machine", International Journal of Advanced Computer Science and Applications, vol. 11, no. 8, 144-149, 2020.

[23] H. N. Latha, Rudresh, S., Sampreeth, D., Otageri, S. M., & Hedge, S. S, "Image understanding: semantic segmentation of graphics and text using faster-RCNN", International Conference on Networking, Embedded and Wireless Systems, Banglore India, 2018.

[24] J. E. M. Adriano, K. A. S. Calma, N. T. Lopez, J. A. Parado, R. L. W, and J. M. & Cabardo, "Digital conversion model for hand-filled forms using optical character recognition (OCR)", IOP Conference Series: Materials Science and Engineering, Manila City, Philippines, 2019.

[25] T. K. Cheang, C. Y. S, and T. Y. H, "Segmentation-free vehicle license plate recognition using ConvNet-RNN", arXiv, vol. abs/1701.06439, pp. 1-5, 2017.

[26] I. B. T. T. Murti, "Improvement accuracy of recognition isolated Balinese characters with Deep Convolution Neural Network", Journal of Applied Intelligent System, vol. 4, pp. 22-27, 2019.

[27] N. S. Naragudem Sarika, Muni Sekhar Velpuru, "CNN based optical character recognition and applications", 6th International Conference on Inventive Computation Technologies, Coimbatore India, 2021.

[28] C. Bartz, Yang, H and C. & Meinel, "STN-OCR: A single neural network for text detection and text recognition", arXiv, vol. abs/1707.08831, pp. 1-9, 2017.

[29] E. K. Kumar, P. V. V. Kishore, M. T. K. Kumar, D. A. Kumar, and A. S. C. S. & Sastry, "Three-dimensional sign language recognition with angular velocity maps and connived feature resnet", IEEE Signal Processing Letters, vol. 25, no. 12, 2018.

[30] C. S. Yang and C. C. & Hsieh, "High accuracy text detection using resnet as feature extractor", IEEE Eurasia Conference on IOT, Communication and Engineering, Yunlin Taiwan, 2019.

[31] D. K. Sahu and C. Jawahar, "Unsupervised feature learning for optical character recognition", 13th IEEE International Conference on Document Analysis and Recognition, Tunis Tunisia, pp. 1041-1045, 2015.

[32] M. Jain, Mathew, M., and Jawahar, C. V., "Unconstrained ocr for urdu using deep cnn-rnn hybrid networks", 4th IAPR Asian Conference on Pattern Recognition, Nanjing China, 2017.

[33] S. Naz, Umar, A. I., Ahmed, S. B., Ahmad, R., Shirazi, S. H., Razzak, M. I., and Zaman, A, "Statistical features extraction for character recognition using recurrent neural network", Pakistan Journal of Statistics, vol. 34, pp. 47-53, 2018.

[34] M. J. Rafeeq, ur Rehman, Z., Khan, A., Khan, I. A., and Jadoon, W., "Ligature categorization based Nastaliq Urdu recognition using deep neural networks", Computational and Mathematical Organization Theory, vol. 25, pp. 184-195, 2019.

[35] R. Achkar, Ghayad, K., Haidar, R., Saleh, S., & Al Hajj, R, "Medical handwritten prescription recognition using CRNN", International Conference on Computer, Information and Telecommunication Systems, Beijing China, 2019.

[36] A. Naseer, & Zafar, K, "Meta features-based scale invariant OCR decision making using LSTM-RNN", Computational and Mathematical Organization Theory, vol. 25, pp. 165–183, 2019.

[37] S. B. Ahmed, Naz, S., Swati, S., & Razzak, M. I, "Handwritten Urdu character recognition using one-dimensional BLSTM classifier", Neural Computing and Applications, vol. 31, pp. 1143–1151, 2019.

[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint, no. 1409.1556, 2014.

[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision", IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas USA, pp. 2818-2826, 2016.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas USA, pp. 770-778, 2016

[41] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks", IEEE Conference on Computer Vision and Pattern Recognition, Honolulu USA, pp. 4700-4708, 2017.

