Journal of Information Assurance and Security.
ISSN 1554-1010 Volume 8 (2021) pp. 085-093

The Effectiveness of Transfer Learning for Arabic Handwriting Recognition using Deep CNN

Mohamed Elleuch[1], Safa Jraba[2] and Monji Kherallah[3]

[1] National School of Computer Science (ENSI), Manouba University,
Campus Universitaire de la Manouba, 2010 Manouba, Tunisia
mohamed.elleuch.2015@ieee.org

[2] Faculty of Sciences, Gafsa University,
Campus Universitaire Sidi Ahmed Zarroug, 2112 Gafsa, Tunisia
jrabasafa@gmail.com

[3] Faculty of Sciences, Sfax University,
Route de la Soukra km 4 - BP 1171, 3000 Sfax, Tunisia
monji.kherallah@fss.usf.tn

***Abstract*: Recognition of handwritten words is a widely used system in the modern world that remains a significant challenge. Traditional machine learning techniques require creative engineering and considerable domain expertise to transform the raw data into a feature vector from which the classifier could rank the input model. To deal with this problem, the popular convolutional neural networks (CNNs), and Its Derivatives introduced recently, have effectively replaced handcrafted descriptors with network features and have been found to provide significantly better results than traditional methods. It is one of the fastest growing fields in machine learning, promising to reshape the future of artificial intelligence. However, the problem with deep learning is that it requires large data sets for training due to the large number of parameters needed to be tuned by a learning algorithm. In our proposed approach, Deep Convolutional Neural Networks (DCNN) is adapted to perform the classification phase. Notably the architectures Inception-v3, ResNet, and VGG16, using an enriched dataset containing images extracted from the IFN/ENIT database. To refine pre-trained models for deep characteristics extraction, we utilize Transfer Learning technique. This technique has shown very good performance for the frequent recognition problems. According to the results obtained, the developed system gives very interesting recognition rates.**

***Keywords*:** Convolutional Neural Networks, Deep Learning, ResNet, VGG16, Inception-v3, Transfer learning, handwritten**.**

I. Introduction

Pattern recognition by classification of offline and online data has been a major interest that has been practiced in several research fields such as computer vision [1, 2], automatic speech recognition (ASR) [3] and EEG signal classification [4]. As a result, handwriting recognition remains a very popular research field due to technological developments such as handwriting capture tools.

Because of the difficulty of this subject, the recognition of Arabic script has been studied in depth for about twenty years by researchers who have used different approaches, such as, the Hidden Markov Model (HMM), deep neural networks (DNN) and recurrent neural networks (RNN), support vector machine (SVM), multilayer perceptron (MLP), and convolutional neural networks (CNN), etc. The results have been diverse and satisfactory: these learning systems have proven their reliability and power in a wide range of applications, and as a result, they have been able to perform optical character recognition (OCR) in Asian and Latin languages [5, 6]. The large number of parameters remains the major drawback of these architectures, which can lead to over-adjustment. Regarding the recognition of offline Arabic script, our research has highlighted the recognition aspects. Handwritten characters and superimposed characters are subject to great variation, due to the diversity of shapes, strokes, curves and concavities, so special importance has been given to the recognition of handwritten Arabic words [7], convolutional neural network classification and support vector machine is adopted in the field of handwritten Arabic [8]. On the other hand, to improve the performance of our architecture, we are interested in CNN, which is designed as a hierarchical neural network followed by a good-capacity representation that is used to learn the right features for the various levels of the hierarchy. In addition, our model has been applied by performance to various computer vision problems such as handwriting and visual object recognition.

CNN is composed of a number of convolutional layers as well as fully connected layers. The fully connected layers are the same as for a multilayer perceptron. However, there are two major limitations of the multilayer perceptron that are encountered in the classification phase: first, there is no theoretical relationship between the classification task and the structure of the multilayer perceptron. Secondly, the MLP multilayer perceptron extracts hyperplanar separation surfaces,

in the feature representation space, talking about the margin between two different classes, the results obtained are not optimal.

Compared to other classical machine learning methods, depth learning has proven its performance for image classification. The extraction of features is done automatically, without human intervention.

The choice of architecture is very complex, so it is important to study and explain architectures that are effective in inspiring our research. The VGGNet [9] has been widely used because of its architectural simplicity. On the other hand, it suffers from the need for a lot of calculations.

The VGGNet [9] has a simple architecture, which is why it is widely used. The architecture of GoogLeNet [10] has the power to solve the problems of memory and computing costs. For GoogLeNet [11], the emphasis is on the number of parameters used, which is reduced, compared to AlexNet and VGGNet [11] [12]. Thanks to the low computational costs of Inception, researchers have used Inception networks for image classification based on a powerful data set [13].

Inception-v3 architecture consist of three inception module (A, B and C) punctuated with grid size reduction step. At the end of the training operation, when accuracies were approaching saturation, the auxil-iary classifiers participate as regularizer and specially when they had Dropout or BatchNorm techniques.

Data augmentation has an important role in increasing the number of training images, thus improving the performance of deep learning techniques in computer vision domains [14][12]. The results show that the CNN methods with particular data augmented datasets yield the highest accuracies. The main objective of this work is to identify, from an image, the healthy leaves and the sick leaves of a given data set.

However, the ResNet architecture, reaches an error rate of 3.57% (error rate of the top 5) [15], where the recipe for the success of this architecture to form such a deep network is that it has residual connections and get the accuracy of a much deeper network.

CNN-based deep learning approaches have successful in image processing; however, due to the small data size, CNN models can overfit. The problem of training based on a small data set has been solved by Transfer Learning (TL).

Transfer learning is an interesting technique in machine learning that has been used well in image classification. The CNN model can be used in three different ways: (1) train the CNN from scratch; (2) use the transfer learning strategy to take advantage of the characteristics of a pre-trained model on a larger data set; and (3) keep the transfer learning strategy and refine the weight of the CNN architecture.

In this work, we study the applicability of CNNs using transfer learning strategies on two datasets.

The consecutive parts of this document are organized as follows: for the next part, the basic concepts will be described through the Deep CNN architectures. Part 3 details the proposed methodology and describes the transfer learning technique. In part 4, we have reviewed the experimental results. Finally, we close the document with observations and some further drafts.

II. Deep Convolutional Neural Networks

In this section I will discuss CNN, ResNet, VGG16 Model, and transfer learning strategy. The transfer learning technique shows promising performance in classifying the dataset used.

A. Convolutional Neural Network

CNN is preferred as a deep learning method in our work. It is easily used to identify and classify objects with minimal preprocessing, succeeds to analyze visual images and can easily separate required functionality with its multi-layered structure. It consists of four main layers: layer convolutional 'CONV', activation function layer 'ReLu', grouping layer 'Pooling', and Fully Connected Layer 'FCL'.

The layers are organized in such a way that the data flows through each layer in the systems, converts the data and authorizes the data on the next layer. The system will explicitly benefit from the data and develop the multiples the quality and detail of everything it collects from the layering sequence. The execution of the structure can be enhanced by managing various layers of the system and using pre-trained CNNs in light of how the frame can slow down in an optimum neighborhood specific to the starting space of the training data.

The architecture of CNN is illustrated in Figure 1 and allows efficient processing of image data and avoids manual selection of non-linear features. In the convolution layer, first of all a feature extraction is performed and then the result is passed to the activation function. At the grouping layer, the feature map is reduced in size to ensure robust learning results for the incoming data. When running through the individual steps of the convolution and grouping layers and starting from the incoming data, global characteristics are obtained. At the end, the extracted characteristics are transferred to the fully connected layer to realize the classification.

Various architectures that are based on CNN have been developed and are designed for the 1000-class image classification such as AlexNet [12], VGGNet [9], ResNet [15] and GoogLeNet [10].

The convolution operation is defined as:

$$x_j^l = f\left(\sum_i y_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (1)$$

Where

- x_j^l represents the *j-th* convolution result of the *l-th* layer,
- y_i^{l-1} is the *i-th* output map of its previous layer,
- '*' denoted the convolution operation,
- k_{ij}^l represents the convolution kernel and b_j^l is the bias of the *j-th* output layer

Non linearity has been performed in CNN by activation functions and pooling layers. We exploited the activation function ReLU (Rectified Linear Units):

$f(x) = \max(0, x)$ (2)

The FCL takes the function written as follows:

$$y_j^l = \max\left(0, \sum_i y_i^{l-1}.w_{i,j}^l + b_j^l\right) \quad (3)$$

Where y_j^l represents the *j-th* node in the *l-th* layer, $w_{i,j}^l$ represents the weights between y_j^l and y_i^{l-1}, and b_j^l is the bias.

The last step of foretelling a distribution $P(y_i)$ is to handle a softmax over the outputs z_i (shapes of characters):

$$y_i = \frac{\exp(z_i)}{\sum_k \exp(z_k)} \quad (4)$$

$$z_i = \sum_j y_j^{l-1}.w_{i,j} + b_i \quad (5)$$

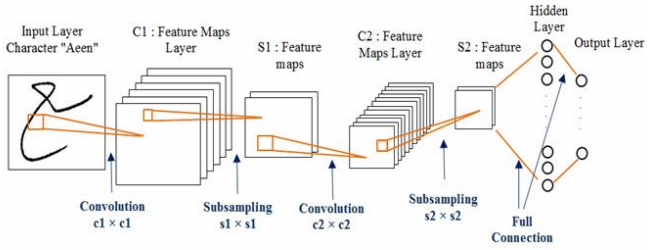


Figure 1. Convolution Neural Network (CNN) Architecture

B. *Inception-v3*

Inception-v3 is a CNN architecture from the Inception family (v1, v2/v3 and v4), presented as GoogLeNet with 22 layers, that makes several improvements including using Label Smoothing, Factorized 7 x 7 convolutions, and the use of an auxiliary classifier to propagate label information lower down the network [13]. It's composed of parallel connections (See Fig. 2).

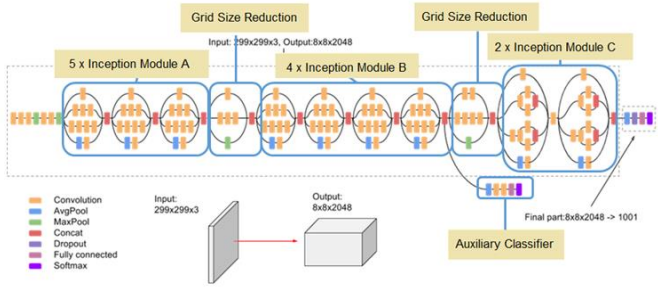


Figure 2. Inception-v3 Architecture

Inception-v3 breaks down the convolutions by employing smaller 1-D filters as indicated in Figure 3 to minimize number of Multiply-and-Accumulates (MACs) and weights, as well as benefit from the factorizing Convolutions, in order to go deeper to 42 layers.

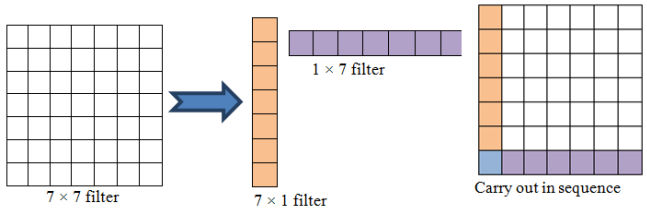


Figure 3. Decomposing greater filters into reduced filters: Building a 7×7 support from 1×7 and 7×1 filter

C. *VGG16*

Simonyan and Zisserman are the developers of VGGNet. It consists of 16 convolutional layers. It is characterized by a very uniform architecture, which makes it very attractive. It is now the most powerful choice for extracting features in an image.

VGG16 contains mainly three different parts: convolution, pooling and fully connected layers, it starts with two convolution layers followed by pooling, then two more convolutions followed by pooling, after this repetition three convolutions followed by pooling, then finally three fully connected layers. The following figure shows the architecture of the VGG-16 model. VGG model weights are available on different platforms and can be used for further analysis - model and application development. The idea of using model weights for different tasks points to the birth of transfer learning.

The public availability of the VGGNet weight configuration served as the main tool for feature extraction. However, VGGNet falls into a management problem, because of the 138 million parameters it contains [9].

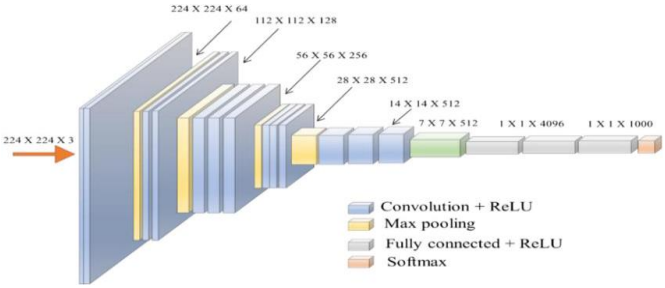


Figure 4. VGG16 Architecture

VGG was introduced in 2014 and still produces better precision for some specific tasks compared to other classifiers thanks to many advanced image classifiers which arrived after 2014. Although the model with transfer learning does not provide satisfactory results in this experiment, using other layers of the VGG for the feature extraction process or fine-tuning the parameters may produce better accuracy. If the given number of training samples (images) is not sufficient to build a classifier, in this case, VGG can be easily exploited for feature extraction as it is trained on millions of images.

We applied in our work the VGG16 architecture with TL, which showed its strength in terms of high recognition rate compared to other architectures.

D. *Residual Network (ResNet)*

ResNet is a continuation of the deep networks, it is the layer that revolutionized CNN's architectural race by introducing the concept of residual learning in CNN (See Figure 5), and it developed an effective methodology for the formation of deep networks [16].

ResNet nominated CNN's 152-deep layers, which won the 2015- ILSVRC Competition. It is also placed in CNN networks based on multiple paths. ResNet proposed a CNN of 152 layers in depth, which won the 2015 ILSVRC competition award.

It has a lower complexity than VGGNet, It achieves an error rate of 3.57% in the top 5, which is higher than the human performance of this dataset.

The good performance of ResNet on image recognition and localization tasks has shown that the depth is one critical importance for many visual recognition tasks.

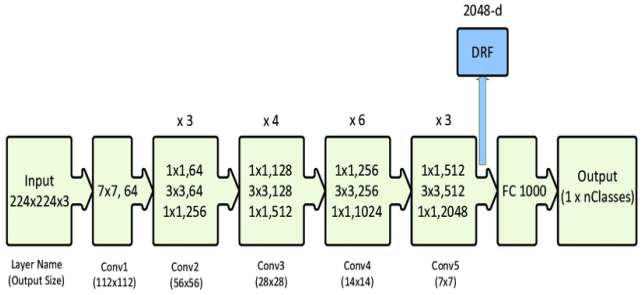


Fig. 5. ResNet Architecture

We applied in our work the pre-trained ResNet architecture, which has shown its strength in terms of high recognition rates compared to other architecture.

E. Transfer Learning Technique

Deep neural networks may not be able to learn properly when training data is not sufficient. Transfer learning can play an important role in solving this problem and adjusting the model according to the new task. Transfer learning is an advanced machine learning technique that can improve the performance of models. However, the intelligent use of the transfer learning technique is a necessary condition to obtain the appropriate advantages.

Transfer learning is a method of machine learning in which a model developed for one task is reused as the starting point for a model on a second task. Transfer learning generally used to speed up training time and possibly improve model performance". While taking advantage of transfer learning, we must consider that the benefits of using transfer learning are not obvious.

There are a few factors that we can look for when applying transfer learning.

- ✓ Higher start: The initial phase of the model with transfer learning should outperform the model without transfer learning.
- ✓ Higher slope: The performance slope or rate of improvement during the learning phase of the source model is steeper than it would otherwise be.
- ✓ Higher asymptote: Pre-trained should converge more easily.

Three of the most famous models that can be used for transfer learning which are Inception-v3, VGG16 and ResNet.

III. Proposed Method

Our approach is to recognize the Arabic handwritten words extracted from the IFN/ENIT database, for which we used extensive learning. This is why the following figure shows our model to follow. In the present work, to recognize Arabic handwritten words extracted from the IFN/ENIT database, the model adopted is the deep convolutional neural network (DCNN), as well as the Inception-v3, ResNet and the VGG16. The system is composed of three main steps: data pre-processing, learning and recognition.

In the first step, data pre-processing is performed, followed by a conversion of the raw data into a normalized two-dimensional image in the range of [0,1]. In addition, the data set is divided into a training set and a test set. In the second step, the validation set is used to select an appropriate model structure. In the third step, the test set is selected to validate the performance of the model.

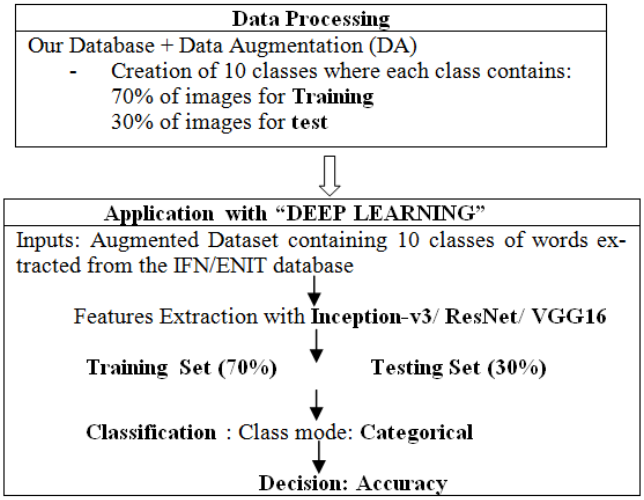


Figure 6. Overview of proposed Framework

A. CNN Model

The convolutional neural network uses a special mathematical operation called "convolution". The CNN has four layers of fundamental components: convolution, RELU, pooling, and fully connected layers (See Figure 7). Recently, a fully connected network uses more parameters than a convolutional neural network with calculation of the convolution over small regions of the input space and some sharing of parameters between regions. This has made it possible to train the models on larger sequence windows, thus improving recognition in our work.

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	(None, 64, 64, 3)	0	
conv2d_22 (Conv2D)	(None, 64, 64, 32)	896	input_2[0][0]
batch_normalization_20	(BatchNo(None, 64, 64, 32)	128	Conv2d_22[0][0]
activation_20 (Activation)	(None, 64, 64, 32)	0	batch_normalization_20[0][0]
conv2d_23 (Conv2D)	(None, 64, 64, 32)	9248	activation_20 [0][0]
batch_normalization_21	(BatchNo(None, 64, 64, 32)	128	conv2d_23[0][0]
activation_21 (Activation)	(None, 64, 64, 32)	0	batch_normalization_21[0][0]
conv2d_24 (Conv2D)	(None, 64, 64, 32)	9248	activation_21[0][0]
batch_normalization_22	(BatchNo(None, 64, 64, 32)	128	conv2d_24[0][0]
Add_10 (Add)	(None, 64, 64, 32)	0	activation_20[0][0] batch_normalization_22[0][0]
activation_22 (Activation)	(None, 64, 64, 32)	0	Add_10[0][0]
conv2d_25 (Conv2D)	(None, 64, 64, 32)	9248	activation_22 [0][0]

Figure 7. CNN Model

B. ResNet Model

When the input is on the network, the convolution process precedes the ResNet block, because it is the only one that has been done. In the convolution and grouping layers, the steps, rows and columns will be set to 2, so as to reduce the height and width of the feature map. The ResNet blocks are represented by configuration as the sequence {BN-LReLU-Conv-BN-LReLU-Conv} [17]. In the ResNet block, the size of the feature map does not change. The number of parameters is reduced when the network ends with the average global pool instead of flattening. Then, a fully connected layer is used for recognition. The process of the ResNet model is shown in Figure 8.

batch_normalization_18	(BatchNo(None, 16, 16, 128)	512	Conv2d_20[0][0]
activation_18 (Activation)	(None, 16, 16, 128)	0	batch_normalization_18[0][0]
conv2d_21(Conv2D)	(None, 16, 16, 128)	147584	activation_18 [0][0]
batch_normalization_19	(BatchNo(None, 16, 16, 128)	512	conv2d_21[0][0]
add_9(Add)	(None, 16, 16, 128)	0	activation_17 [0][0] batch_normalization_19[0][0]
activation_19 (Activation)	(None, 16, 16, 128)	0	add_9[0][0]
average_pooling2d_1	(None, 2, 2, 128)	0	activation_19[0][0]
flatten_1 (Flatten)	(None, 512)	0	average_pooling2d_1[0][0]
Dense1 (Dense)	(None, 10)	5130	flatten_1 [0][0]

Figure 8. ResNet Model

C. *VGG16 Model*

The VGG16 model is classified among the 5 best models, it represents the highest accuracy of image classification. The VGG16 network is designed to expand the amount of data and to classify the input image into several categories. The network is considered as a "deep network", it contains 16 weight layers. The weight layers are only the convolutional and fully connected layers because they contain the parameters that can be learned (See Figure 9). Deep neural networks also require a large amount of computing power.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 224, 224, 3)	0
Block1_conv1 (conv2D)	(None, 224, 224, 64)	1792
Block1_conv2 (conv2D)	(None, 224, 224, 64)	36928
Block1_pool (Maxpooling2D)	(None, 112, 112, 64)	0
Block2_conv1 (conv2D)	(None, 112, 112, 128)	73856
Block2_conv2 (conv2D)	(None, 112, 112, 128)	147584
Block2_pool (Maxpooling2D)	(None, 56, 56, 128)	0
Block3_conv1 (conv2D)	(None, 56, 56, 256)	295168
Block3_conv2 (conv2D)	(None, 56, 56, 256)	590080
Block3_pool (Maxpooling2D)	(None, 28, 28, 256)	0
Block4_conv1 (conv2D)	(None, 28, 28, 512)	2359808
Block4_conv2(conv2D)	(None, 28, 28, 512)	2359808
Block4_conv3 (conv2D)	(None, 28, 28, 512)	2359808
Block4_pool (Maxpooling2D)	(None, 14, 14, 512)	0
Block5_conv1 (conv2D)	(None, 14, 14, 512)	2359808
Block5_conv2 (conv2D)	(None, 14, 14, 512)	2359808
Block5_conv3 (conv2D)	(None, 14, 14, 512)	2359808
Block5_pool (Maxpooling2D)	(None, 7, 7, 512)	0
flatten_1 (Flatten)	(None, 25088)	0
Dense_1 (Dense)	(None, 10)	250890
Total params 14 965 578		

Figure 9. VGG16 Model

D. *Inception-v3 Model*

The Inceptionv3 is the third iteration of the inception architecture, first developed for the GoogLeNet model (See Figure 10). Inception v3 was trained on ImageNet.

In the Inception v3 model, divers techniques have been suggested to optimize the grid to loosen up constraints to facilitate model conditioning. The techniques include factorized convolutions, regularization, dimension reduction, and parallel computations.

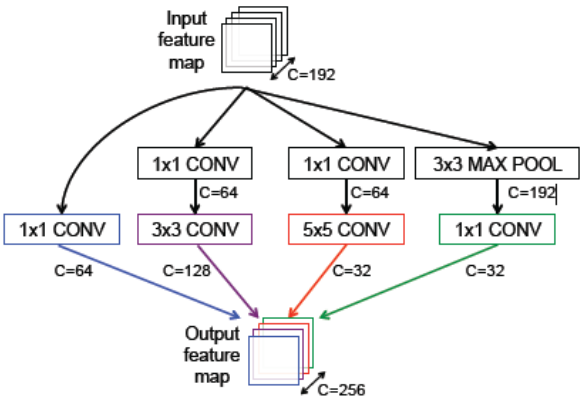


Figure 10. Inception module from GoogleNet [21]

IV. Experiments and Results

In this part, we present our dataset for images of Arabic handwritten words / characters for experiments. Next, we detail and discuss the frameworks for the experiments. The results obtained with the proposed systems are presented and compared. For each method, different sets were used for training, validation and testing.

In addition, we use the training dataset for training, the test set one to assess recognition. We used Python 3.7, the programming language and OpenCV computer library of all deep learning models.

A. *Dataset*

Our Database is composed of pictures of handwritten Arabic words (names of Tunisian villages), it is distributed over 10 classes. The samples of our base are classified in two groups: Training Base with 450 images and Test Base with 150 images. The number of samples in each class is equal to 60. Thereafter, we apply the technique of data augmentation. In the adopted split, some samples images of our dataset are shown in Table1.

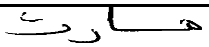
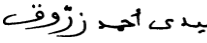
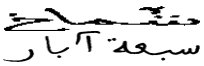
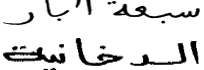
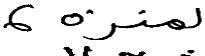
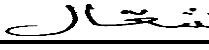
N°	Image of classes	Arabic pronunciation	French Pronunciation
1		مارث	Mareth
2		أكودة	Akoudah
3		دوار اللواته	Douwaralouwatah
4		نقّة	Nagah
5		سيدي أحمد زروق	Sidi ahmedzarouk
6		شمّاخ	Chamakh
7		سبعة آبار	Sab'ataAbar
8		الدخانية	Aldakhaniya
9		المنزه6	Almanzah 6
10		شعّال	Cha'al

Table1. Classes of each form of handwritten Arabic words

On the other hand, to measure the effectiveness of our system proposed for high-low level dimension of data, words are segmented into letters (See Figure 11) from set 'a' to 'e' (See Table 2). We have kept 3.360 images as train data and 1.120 images as test data. These images include 56 shapes of characters. Details of the class for each shape are presented in Table 3.

Set	Number of words	Number of characters	Number of PAWs
a	6537	51984	28298
b	6710	53862	29220
c	6477	52155	28391
d	6735	54166	29511
e	6033	45169	22640
Total	*32492*	*257366*	*138060*

Table 2. Different IFN/ENIT Datasets

Arabic Script	Shape	class	Arabic Script	Shape	class
Aeen (ع)	ع	1	Laam (ل)	ل	29
	عـ	2		لـ	30
	ـع	3		ـل	31
	ـعـ	4		ـلـ	32
Alif (ا)	ا	5	Lam_Alif (لا)	لا	33
	ـا	6		ـلا	34
Baa (ب)	ب	7	Meem (م)	م	35
	بـ	8		مـ	36
	ـب	9		ـم	37
	ـبـ	10		ـمـ	38
Daal (د)	د	11	Noon (ن)	ن	39
	ـد	12		ـن	40
Faa (ف)	ف	13	Raa (ر)	ر	41
	فـ	14		ـر	42
	ـف	15	Saad (ص)	ص	43
	ـفـ	16		صـ	44
Haa (ه)	ه	17		ـص	45
	هـ	18		ـصـ	46
	ـه	19	Seen (س)	س	47
	ـهـ	20		سـ	48
	ھ	21		ـس	49
Hamza (ء)	ء	22		ـسـ	50
Jeem (ج)	ج	23	Taa (ط)	ط	51
	جـ	24		طـ	52
	ـج	25	Waao (و)	و	53
	ـجـ	26		ـو	54
Kaaf (ك)	ك	27	Yaa (ى)	ى	55
	كـ	28		ـى	56

Table 3. Class for each shape of an Arabic script

Figure 11. Samples of Arabic letters generated from the IFN/ENIT database.

B. Data Augmentation

Our Database illustrates a medium number of images of Arabic handwritten words (Names of Tunisian villages), our goal to obtain a deep convolutional neural network accurate more than a million parameters. Therefore, we use the data augmentation technique; we apply it to amplify the number of learning images and thus improve the results of our model. Increasing the learning set and gives solution to reduce the problem of over-fitting.

The Keras image data generator class was used to increase the data and the best results were obtained by setting the parameters. The following figure shows an example of the original photos and their pre-processed and enhanced versions are shown:

Figure 12. Samples of word "أكودة" after the application of Data Augmentation

C. Experimental Settings

In order to evaluate the effectiveness of the proposed system based on the CNN classifier, we studied its performance in terms of training and word (and character) recognition of the names of Tunisian villages extracted from the IFN/ENIT database. In order to effectively train the model on a larger number of data to perfectly manage the variability of the writing, the execution of our model is divided as follows:

- **Pre-processing:** in this phase, we have to use a non-standardized data set. It is interesting to confirm that the basic pre-processing activities take place during the development of the database.
- **Parameterization:** Transfer learning has turned to be part of many applications, where the current standard is to take an existing architecture conceived for natural image datasets like "ImageNet", with the corresponding pre-trained weights, then fine-tunes the pattern to the intended imaging data (IFN/ENIT). Training is provided for 10 epochs, hence convergence is ensured.
- **Extraction of features:** they represent the information extracted from the image of a word or a character, in order to build the classifiers of the classification phase, these characteristics are utilized.

The challenge is to determine which features are most appropriate for classification. In this paper, the CNN and Its Derivatives were used for this experiment.
The system is implemented with the Python TensorFlow deep learning library, being an open-source machine learning algorithm generated by Google. For our proposed system (based on Inception-v3/ResNet/VGG16) to be valid, it is useful to use our databases containing handwritten Arabic words (and characters) of Tunisian village names, which is extracted from the IFN/ENIT database. Both databases are divided as follows: a training set and a test set.

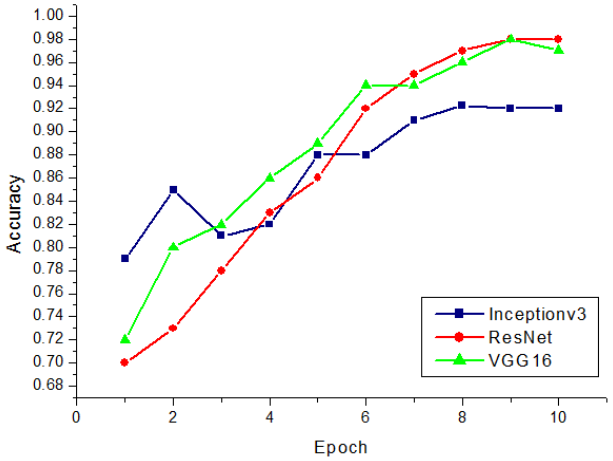
D. Results and Discussion

In our experimental study, we evaluated the performance of the CNN, as well as the Inception-v3, ResNet and VGG16 architectures with TL technique derived from it.
Firstly, the three proposed models are used on the database consisting only of characters. Then the generated models (pre-trained) which have refined parameters, will be applied to the database of words IFN/ENIT.
The overall recognition rates provided by these models are presented in the following table:

Methods	Character database *(56 classes)*	Word database IFN/ENIT (*10 classes)*	
	Learning with TL	*Learning from scratch*	*Learning with TL*
Inception-v3	97.16 %	92.27 %	95.70 %
ResNet	**99.73 %**	98 %	**98.99 %**
VGG16	98.51 %	98 %	98.10 %

Table 4. Accuracy rate of our proposed methods using characters/words Databases

It was obviously noted that our models based on transfer learning outperforms any other strategy learning from scratch [18]. In table 4, it is shown that ResNet model with TL outperforms moderately Inception-v3 classifier and lightly VGG16 model once tested on handwritten words database with IFN/ENIT or characters database.
Consequently, these results indicate that transfer learning technique helped improving the classification performance of convolutional neural network (CNN) in all cases (inception v3, ResNet and VGG).



(a)

(b)

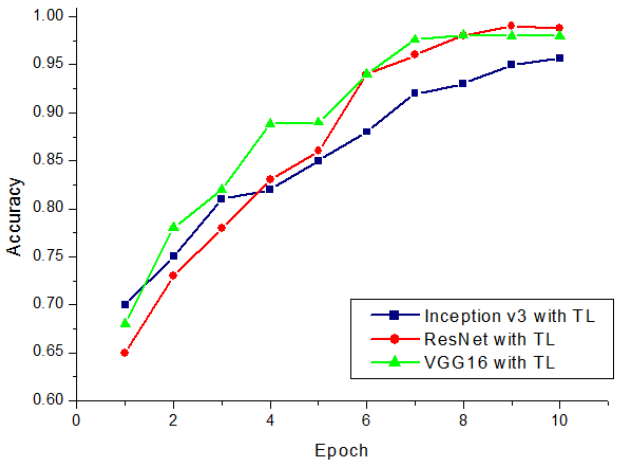


Figure 13. Performance evaluation of 3 models (a) learning from scratch (b) with transfer learning

From the results obtained in figure 13, we noticed that both architectures ResNet and VGG16 with TL, recorded very high recognition rates with a percentage of 98.99% and 98.10% respectively, which demonstrates the effectiveness of features of the neural convolution approach for the recognition of offline Arabic handwritten script.
The comparison with the other results is presented in Table 5. We observed that the performance of our offline Arabic handwriting recognition system based on transfer learning strategy achieved a promising result compared to other research works in the same field founded on traditional methods and deep learning approaches.
From the above, we can conclude that the VGG16 / ResNet classifier provides better recognition rates compared to other classifiers [18-23].

Authors	**Architectures**	**Recognition Rate (RR)**	
		of Characters	**of Words**
Our Models	Inception-v3 with TL	97.16%	95.70%
	ResNet with TL	**99.73%**	**98.99%**
	VGG16 with TL	98.51%	**98.10%**
[19]	CNN based SVM with Dropout	92.95%	-
[20]	DBN	-	82%
[21]	CDBN	95.54%	91.55%
[22]	CNN-HMM	-	89.23%
[18]	CNN	-	93%

Table 5. Performance Comparison using handwritten Arabic words/characters IFN/ENIT Database

Thanks to these experimental results which allowed us to achieve higher recognition rates, it is demonstrated that the transfer learning technique has a major effect on huge convolution architecture by utilizing a small data set.
In fact, the goal of pre-training on large benchmarks is to take advantage of the re-use of scaling pre-trained weights, which leads to better speed of convergence.

V. Conclusion

In this work, we have explored the applicability of transfer learning technique in our proposed models (Inception-v3, ResNet and VGG16) on Arabic handwritten recognition and demonstrated the efficiency of the system for Arabic handwritten Script recognition applied on IFN/ENIT database. As we can see, we compared Deep CNN based models with learning from scratch versus transfer learning strategy. So ResNet and VGG models with TL achieve results with an accuracy of 98.99% and 98.10% respectively, applied to a set of images from the Arabic handwritten word IFN/ENIT, which represent promising and encouraging results.

In general, we conclude that the proposed system based on the TL technique provided promising recognition results on the Arabic handwritten image dataset.

As future work, we plan to utilize more advanced methods to enhance the current study, such as Deep LSTM network, recurrent nets with attention modeling [24].

References

[1] H. Lee. R. Grosse, R. Ranganath, A Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Communications of the ACM*, 54(10) (2011) 95-103.

[2] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 42(2), pp. 513-529, 2012.

[3] H. Lee, PT. Pham, Y. Largman, AY. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," *Advances in Neural Information Processing Systems* (NIPS), pp. 1096-1104, 2009.

[4] Y. Ren, Y. Wu, "Convolutional Deep Belief Networks for Feature Extraction of EEG Signal," *International Joint Conference on Neural Networks (IJCNN)*, pp. 2850-2853, 2014.

[5] D.C. Ciresan, U. Meier, J. Schmidhuber, "Transfer Learning for Latin and Chinese Characters with Deep Neural Networks," *In Proceedings of International Joint Conference on Neural Networks*, 2012.

[6] D.C. Ciresan, J. Schmidhuber, "Multi-Column Deep Neural Networks for Offline Handwritten Chinese Character Classification," *In Proceedings of CoRR*. 2013.

[7] R. AI-Hajj, L. Likforman-Sulem, C. Mokbel, "Combining slanted frame classifiers for improved HMM-based Arabic handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, voI 31(7), pp. 1I65-1177, 2009.

[8] J.H. AIKhateeb, J. Ren, J. Jiang, H. AI-Muhtaseb, "Offline handwritten Arabic cursive text recognition using hidden Markov models and re-ranking," *Pattern Recognit. Lett.*, vol. 32(8), 2011.

[9] Simonyan K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv* preprint arXiv:1409.1556 (2014).

[10] Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[11] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture.European Conference on Computer Vision*, 2018.

[12] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012): 1097-1105.

[13] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[14] A. Sutskever , G.E and I. Hinton, "Data Augmentation for Plant Classification" . *European Conference on Computer Vision*, September, 2017.

[15] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[16] T.-Y. Lin et al., "Microsoft coco: Common objects in context," *in European conference on computer vision*, pp. 740–755, 2014.

[17] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Proc. of the 32nd International Conference on Machine Learning*, vol. 37, pp. 448-456, 2015.

[18] Jraba, S., Elleuch, M., & Kherallah, M. Arabic Handwritten Recognition System using Deep Convolutional Neural Networks. *In Intelligent Systems Design and Applications: 15th International Conference on Intelligent Systems Design and Applications (ISDA 2020)*, December 12–15, 2020. Springer Nature.

[19] M. Elleuch, R. Maalej and M. Kherallah, "A new design based-SVM of the CNN classifier architecture with dropout for offline Arabic handwritten recognition". *Procedia Computer Science*, Vol. 80, pp. 1712-1723, 2016.

[20] K. Jayech, M. A. Mahjoub and N. E. B. Amara, "Arabic handwritten word recognition based on dynamic bayesian network". *Int. Arab J. Inf. Technol.*, Vol. 13(6B), pp. 1024-1031, 2016.

[21] M. Elleuch and M. Kherallah, "Off-line Handwritten Arabic Text Recognition using Convolutional DL Networks". *International Journal of Computer Information Systems and Industrial Management Applications*, Vol. 12, pp. 104-112, 2020.

[22] M. Amrouch, M. Rabi and Y. Es-Saady, "Convolutional feature learning and CNN based HMM for Arabic handwriting recognition". *In International conference on image and signal processing*, pp. 265-274, Springer, Cham, 2018.

[23] M. Elleuch and M. Kherallah, "Boosting of Deep Convolutional Architectures for Arabic Handwriting Recognition". *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, Vol. 10(4), pp. 26-45, 2019.

[24] Lee, C. Y., & Osindero, S. (2016). Recursive recurrent nets with attention modeling for ocr in the wild. In

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2231-2239).

Author Biographies

Mohamed Elleuch received the B.S. degree in computer science from The Faculty of Economics and Management of Sfax-Tunisia (FSEGS) and completed his master's degree in New Technologies of Dedicated Computer Systems from the National School of Engineering of Sfax (ENIS), Tunisia. He obtained the Ph.D. degree in Computer Science at the National School of Computer Science (ENSI), University of Manouba-Tunisia, in 2016. His main research concerns pattern recognition, handwriting recognition, deep neural networks, computer vision, image and signal processing.

Safa Jraba She was born in Gafsa (Tunisia). She received the master's degree in computer Science and completed his Master Degree in 2020 from The Faculty of Sciences of Gafsa-Tunisia (FSG). She currently continues his research on pattern recognition.

Monji Kherallah graduated in Electrical Engineering 1989, obtained a Ph.D. in Electrical Engineering in 2008. He is now a professor in Electrical & Computer Engineering at the University of Sfax. His research interest includes applications of intelligent methods to pattern recognition and industrial processes. He focuses his research on handwritten documents analysis and recognition, handwritten Arabic recognition, biometrics, pattern recognition and image processing. He is member of the editorial board of "Pattern Recognition Letters." He was a member of the organization committee of the International Conference on Machine Intelligence ACIDCA-ICMI'2005.