Received December 7, 2018, accepted December 18, 2018, date of publication January 21, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2890810

Automatic Visual Features for Writer Identification: A Deep Learning Approach

ARSHIA REHMAN[1], SAEEDA NAZ[1], MUHAMMAD IMRAN RAZZAK[2], AND IBRAHIM A. HAMEED[3]

[1]Higher Education Department, GGPGC No. 1, Abbottabad 22010, Pakistan
[2]The University of Sydney, Sydney, NSW 2006, Australia
[3]Department of ICT and Natural Sciences, Norwegian University of Science and Technology, 6009 Ålesund, Norway

Corresponding author: Ibrahim A. Hameed (ibib@ntnu.no)

This work was supported in part by the Higher Education Department, Khyber Pakhtun Khywa, under Grant ADP 483/170009.

ABSTRACT Identification of a person from his writing is one of the challenging problems; however, it is not new. No one can repudiate its applications in a number of domains, such as forensic analysis, historical documents, and ancient manuscripts. Deep learning-based approaches have proved as the best feature extractors from massive amounts of heterogeneous data and provide promising and surprising predictions of patterns as compared with traditional approaches. We apply a deep transfer convolutional neural network (CNN) to identify a writer using handwriting text line images in English and Arabic languages. We evaluate different freeze layers of CNN (Conv3, Conv4, Conv5, Fc6, Fc7, and fusion of Fc6 and Fc7) affecting the identification rate of the writer. In this paper, transfer learning is applied as a pioneer study using ImageNet (base data-set) and QUWI data-set (target data-set). To decrease the chance of over-fitting, data augmentation techniques are applied like contours, negatives, and sharpness using text-line images of target data-set. The sliding window approach is used to make patches as an input unit to the CNN model. The AlexNet architecture is employed to extract discriminating visual features from multiple representations of image patches generated by enhanced pre-processing techniques. The extracted features from patches are then fed to a support vector machine classifier. We realized the highest accuracy using freeze Conv5 layer up to 92.78% on English, 92.20% on Arabic, and 88.11% on the combination of Arabic and English, respectively.

INDEX TERMS Writer identification, visual features, AlexNet, multilingual, support vector machine.

I. INTRODUCTION

Handwriting plays a key role in presentation of learned behaviour of the person. It is the main identity of a person. Writer identification and authentication has got keen interest by researchers in the field of bio-metrics and forensic sciences. Automatic writer identification system helps in determining and identifying whether the given handwriting is truly matched and assigned to the claimed writer of handwriting. This system assigns the handwriting to specific and true writer out of number of writers.

The writer identification system can be classified with mode of capturing data such as online and offline. The former case deals with the spatial coordinates values while the later case deals with temporal information. Offline writer identification can also be classified with respect to textual content such as text dependent and text independent. Text-dependent methods require input image with fixed text content and measures similarity of the input with registered templates for identification. On the other hand, text-independent writer identification cope with the images of arbitrary text that do not depend on fixed text content.

For the identification of writer, one needs to extract the discriminating features from handwritten text. Features can be extracted implicitly or explicitly. Current state of the art analyze the individual on the basis of statistical, structural features or automatic model based features. Reference [1] reveals that researchers widely used statistical features like local [2]–[5] and global [6]–[8] features. Along with that structural features like graphemes [9]–[11], fragments [12]–[14] and texture of local binary patterns [15]–[17] also reported. Manual feature extraction is one

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Khurram Khan.

of the difficult task as it requires human expertise and domain knowledge to extract and select the discriminating set of features. Manual features are language dependent. Automatic features learned by deep neural networks outperformed as compare to handcrafted features [18]–[21]. The Model based automatic features are extracted by deep learning based models automatically from the raw data of images directly. These features are not dependent on language or patterns. We can easily investigate different models on the given data-set. There is no need of domain knowledge and expertise in the language or patterns.

The key contributions of the paper are indexed as follows.

- Deploy different pre-processing techniques to generate multiple effective representations of handwritten text-line images in English and Arabic languages.
- Apply sliding window technique to make patches of text-line images and use as an input to a model to learn discriminating patterns.
- Investigating optimal visual features from different freeze layers of AlexNet architecture.
- Classifying and identifying the writer by deploying support vector machine using QUWI data-set.

The rest of the paper is organized as follows: Section II presents the background of deep learning and transfer learning. Section III reports the closely related work for writer identification using deep learning. Section IV presents the proposed pipeline along with experimental setups followed by the analysis of results in section V. Finally in section VI, we draw the conclusion of the study.

II. DEEP LEARNING AND TRANSFER LEARNING

Machine learning is the study of algorithms and mathematical models that give the ability to the computer system to make predictions or decisions on given data. Deep learning is a field of machine learning. The term *deep learning* was first introduced to the machine learning community by *Rina Dechter* in 1986 [22]. The deep learning based algorithms and methods inspired by biological neural network for learning of data representation. Deep learning models such as deep neural networks (DNN), recurrent neural networks (RNN), convolutional neural networks (CNN), and different architectures of CNN like AlexNet, GoogleNet, ResNet, VGG etc have widely been used in different fields like computer vision, artificial intelligence, machine translation, pattern recognition, speech recognition, natural language processing etc. The deep learning based models need large number of samples of images for learning features. In some areas, availability or development of large data-sets are not possible, very expensive or time-consuming then transfer learning mechanism helps in identification of pattern in such cases.

Transfer learning or inductive learning is the technique in which the knowledge acquired through one problem is stored and later applied to the other problem [23]. In transfer learning, network is trained on a large data-set (base data-set) and then transfer the learned features to the small data-set (target data-set). This process will tend to work if the features are general, that is, suitable to both base and target data-sets, instead of being specific to the base data-set. ConvNet features are more inclusive in the early layers and more specific to the base data-set in the later layers.

Depending on the size of target data-set and the similarity of target data-set to the base data-set, the approach for using transfer learning will be different. Therefore the most common ways for features extraction using transfer learning are:

- Freeze or fixed features
- Fine-tune or reuse features

In first scenario, different layers of the pre-trained model on base task are fixed so that the ConvNet act as a fixed feature extractor for the target data-set. Finally the target data-set can be classified using linear classifiers (SVM or softmax) on the fixed or freeze features in the literature. In the second scenario, replace the ConvNet on the target data-set and fine-tune the weights of pre-trained ConvNet for continuing back propagation. It is depending on the problem whether to fine-tune all layers or keep some layers. The size of the target data-set and parameters in earlier layer are under consideration while fine tune the features to the new task or freezing the transferred layers. When the size of the target data-set is small and the features are large, fine-tuning cause over-fitting. Thus freezing the feature layers is appropriate in this case. However, when the data-set is large and the features are small then fine tuning improve the performance of target data-set and may not suffer from over-fitting.

III. RELATED WORK

In this section, we will elaborate the related work in the field of automated writer identification using deep learning techniques. Although enough effort using traditional techniques found in this domain since last few decades but deep learning approaches were rarely published in the literature. The first writer recognition system using deep learning technique was presented in 2015 by Fiel and Sablatnig [20]. They conducted experiments on IAM, ICDAR 2011, ICDAR 2013 and CVL databases. They deployed eight layer CNN and extract the features from fully connected or penultimate layer. These CNN based activation features served as feature vector. χ^2 distance used to measure the distance between samples. They achieved highest recognition rate of 98.6% on ICDAR-2011 97.6% on CVL. But they also show the worse accuracy of 40.5% on ICDAR-2013 in hard criteria. A reason might be missing Greek training data and wrong segmentation.

Christlein *et al.* [19] computed local descriptors using activation features from CNN. Exemplar SVM was employed for classification. They evaluated 0.989 and 0.994 mAP in Top 1 using ICDAR 2013 and CVL database. Another effort by the same author found in [2]. They used CNN activation features from image patches of 32×32. They computed LeNet and ResNet based features and reported 99.5% on CVL, 99.6% on ICDAR13, KHATT in Top-1 respectively. In [18], LeNet and ResNet CNN models were used to learn the activation features. The activation features were utilized as local features by encoding with VLAD. They conducted experiments on CVL, ICDAR13 and KHATT data-set using Exemplar SVM and VLAD encoding schemes. They achieved the accuracy of 99.5% on CVL, 99.6% on ICDAR13, 99.6% on KHATT in Top-1.

In online domain, Yang *et al.* conducted experiment on CASIA-OLHWDB1.0 data-set to employ deep convolutional network [24]. They had substantially improved the performance with data augmentation techniques and Drop-Stroke. They extracted automated features from CNN and named as path-signature features. They attained 99.52% accuracy. Xing and Qiao presented deep writer multi-stream CNN based system to learn features. The architecture consists of two branches with convolutional layers [25]. They conducted experiments on IAM and HWDB data-sets by setting the input line in the form of patches of size 113×113 using patch scanning strategies. They deployed data augmentation learning for the significant performance. They pre-trained their model on HWDB data-set and then fine tune on IAM and reported best accuracy of 98.01% on 301 writers with 4 English alphabets as input. While direct training on IAM gave 98.80% accuracy. Similarly, they also used IAM to pre-train the model and fine-tuning on HWDB produced 93.85% accuracy and on direct training of HWDB gave 93.45% accuracy. They used end to end training of writers, thus comparison with other systems becomes impossible. Tang and Wu experimented on ICDAR-2013 and CVL data-sets. They deployed CNN for feature extraction globally rather than local image patches [26]. They generated the feature vector by randomly segmenting words from original writings. They had generated 500 training and 20 samples per writer for testing. Baysian Network and *Fc7*log likelihood was used for the computation of similarity. They retrieved the accuracy of 99.0% on ICDAR 2013 and 99.7% on CVL, respectively.

Nasuno and Arai conducted experiments on Japanese data-set with 100 kind of words from each 100 writers. They utilized 90 words for training and 10 for testing. Alexnet CNN was deployed for feature extraction. They calculated approximately 90% accuracy [27]. The assessment of all the methods discussed in the literature (Table.1) had been accomplished on different databases other than QUWI. Literature also divulges that proposed deep learning techniques had not investigated the different layers of ConvNet of base task to the target task. These reasons are main motivation of this study to investigate the visual patterns form different layers of base task for identification of writer of target task. Therefore, our study is presented as a pioneer study using QUWI Dataset (English and Arabic languages) in the field of writer identification. Due to this reason the achieved results in this paper cannot be directly compared with others.

TABLE 1. The summarized related work using deep learning approaches for writer identification.

Reference	Features	Model	data-set	Accuracy (%)
Fiel and Sablatnig [20]	Raw pixels by CNN	χ^2 distance	ICDAR 2011 ICDAR2013 CVL	98.6 97.6 40.5
Christlein et al. [19]	Raw pixels by CNN	Exemplar SVM	ICDAR 2013 CVL	0.989 0.994 mAP
Christlein et al. [2]	SIFT, Raw pixels by ResNet	Exemplar SVM	ICDAR 2017-Historical-WI	88.9
Christlein et al. [18]	Raw pixels by LeNet and ResNet	Exemplar SVM	ICDAR13 CVL KHATT	99.6 99.5 99.6
Yang et al. [24]	Raw pixels by CNN	CNN	CASIA-OLHWDB1.0	99.52
Xing and Qiao in [25]	Raw pixels by CNN	CNN	IAM HWDB	98.80 93.45
Tang and Wu [26]	Raw pixels by CNN	Baysian Network and CNN	ICDAR2013 CVL	99.0 99.7
Nasuno and Arai [27]	Raw pixels by Alenxnet	Alenxnet	Japnease (100 words)	90

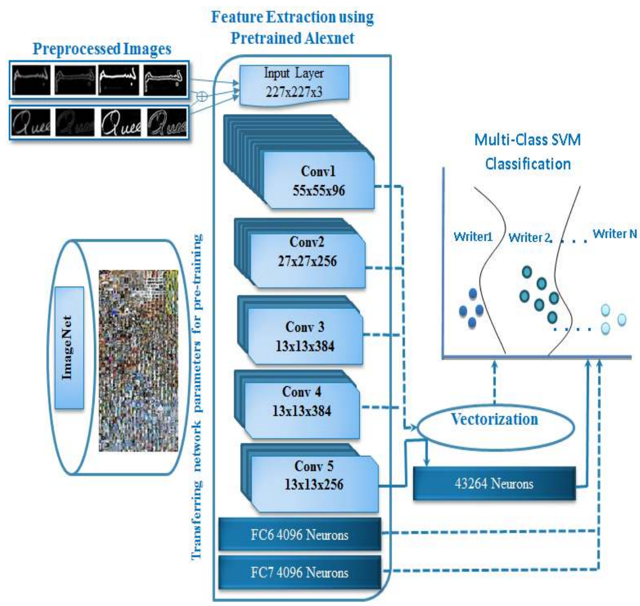

FIGURE 1. AlexNet based automatic deep writer identification system.

IV. METHODOLOGY

The proposed identification technique deploy the pre-trained AlexNet architecture of CNN on multiple representations of patches of text-lines images. The network is trained using the features extracted from freeze/fixed layers (Conv1, Conv2, Conv3, Conv4, Conv5, Fc6, Fc7 and fusion of Fc6 and Fc7) of Alexnet architecture. The proposed identification technique deploy the pre-trained AlexNet architecture of CNN on multiple representations of patches of text-lines images. The network is trained using the features extracted from freeze/fixed layers (Conv1, Conv2, Conv3, Conv4, Conv5, Fc6, Fc7 and fusion of Fc6 and Fc7) of AlexNet architecture. The overall pipeline encompasses of three main steps: pre-processing, feature extraction and classification as depicted in Figure 1. The pre-processing step involved in this study are skew detection, skew correction, normalization and segmentation.

Overlapped sliding window strategy is then employed to extract patches from text line images of English and Arabic languages. Data augmentation proved itself as a successful mechanism in the field of deep machine learning approaches for expanding training data which result in promising classification results. The traditional data-augmentation techniques are cropping, translation, rotation etc. To produce the enormous amount of data for the deep model, unlike traditional techniques of data augmentation are applied like contours, negatives and sharpness in our study. The proposed data augmentation methods allow to producing effective representations of text-lines data that has expanded our data-set. The pre-trained AlexNet architecture of CNN is employed to extract discriminating visual features from multiple representations of image patches. The different freeze layers (Conv3, Conv4, Conv5, Fc6, Fc7 and fusion of Fc6 and Fc7) of AlexNet are investigating for better performance of the model. Finally, the extracted automatic features are fed to the SVM for classification. Before presenting the detail on the stages of proposed pipeline, we would like to give an overview of the data-set that is used in our proposed study.

TABLE 2. Statistical analysis of QUWI data-set.

QUWI Details	Statistics
Number of Writers	1017
Total Documents	5085
Total Number of Digitized Pages	4068
Pages per Writer	4
Approx. Number of Arabic words for text-independent analysis	60,000
Approx. Number of English words for text-independent analysis	100,000
Approx. Number of Words By a Writer	60
Approx. total text lines	9
Approx. total text lines in first page for Arabic	6
Approx. total text lines in first page for Arabic	11
Approx. total text lines in first page for English	6
Approx. total text lines in first page for English	14

A. DATA-SET

We mainly conducted our experimental evaluations on the QUWI data-set [28]. The data-set consist of handwritten text assembled from 1017 writers in both Arabic and English scripts. Each individual were asked to write 4 pages. Furthermore, the data-set encompasses 4068 digitized pages, approximately 60,000 Arabic words for text-independent analysis and more than 100,000 Arabic words for text-dependent analysis and same statistical analysis for English script. The first page comprises approximately 6 handwritten lines in the Arabic language. The third page contains about 6 handwritten lines in English. The first and the third pages are utilize for text-independent writer identification tasks. Similarly, the second page encompasses an Arabic text of 3 paragraphs with the average of 11 lines while the fourth page comprises 14 English text lines approximately. The second and fourth pages are to be used for text-dependent writer identification tasks. Succeeding table 2 summarizes the statistical details of QUWI data-set.

After having described the data-set, we will now present our proposed pipeline for characterizing the writer of a handwritten text in the forthcoming sections.

B. PRE-PROCESSING

Pre-processing is the first step in pattern recognition and digital image processing. In this step, irrelevant information is removed from the data. The input images are transformed to a form which is appropriate for further processing. In this paper, pre-processing steps involve skew detection and correction, normalization, segmentation, sliding window strategy for patches, contours extraction, computation of sharp images and negatives.These steps are elaborated in the forth-coming sub-sections with examples

1) SKEW DETECTION AND CORRECTION

Skew is introduced in images due to improper scanning. Skew detection and correction is the process of identifying the skew angle and then correcting that angle. Skew in the document images introduce difficulties in segmentation process so it needs to be corrected in pre-processing step. The algorithm proposed in [29] based on Probablistic Hough Transform [30] is used in our study to detect the lines, skew and its correctness.

2) SEGMENTATION

Segmentation divides an image into disjoint regions such that pixels within a region share some common attributes. In the segmentation step, the text of a paragraph is segmented into lines. There are numerous methods which work robustly in order to classify the information as text or graphics in an image of a document. However, there is a difference between English and Arabic text line segmentation. Perfect text line extraction of Arabic script is one of the complex and hot issue because of the cursive writing of Arabic script and other associated challenges like overlaps of ligatures or sub-words, graphism, positions and number of diacritics and dots etc.

As a function of the segmentation, all handwritten images may require a re-sampling to a predefined size. This normalization is carried out in a way as to preserve the aspect ratio of the image. For this purpose the image width is adjust to a default value-adjusted and the height will change without any change on height-to-width ratio. Handwritten images of QUWI are normalized to the width of 1250 pixels and the aspect ratio is kept maintained with respect to height. We convolved the images with hamming window of size 80 and compute horizontal projection profile. To crop the images into lines up Minima and down Minima value is calculated [31]. The resultant segmented lines after skew detection and correction are depicted as in Figure 2 and Figure 3.

3) SLIDING PATCHES

In order to extract the patterns that a writer employs frequently as he writes, we first need to carry out segmentation of writing into small sub-images (patches/fragments). Before proceeding to the analysis of small writing fragments,

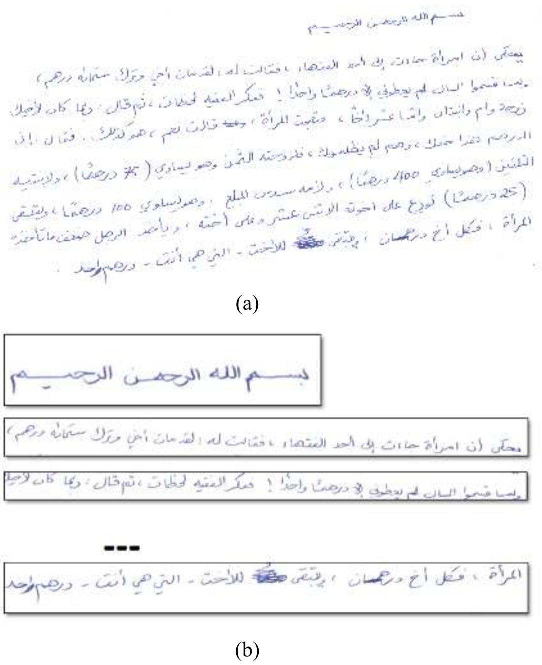

FIGURE 2. Results of Arabic line segmentation. (a) Original image of writer Id 0001_01. (b) Segmented lines after skew detection and correction.

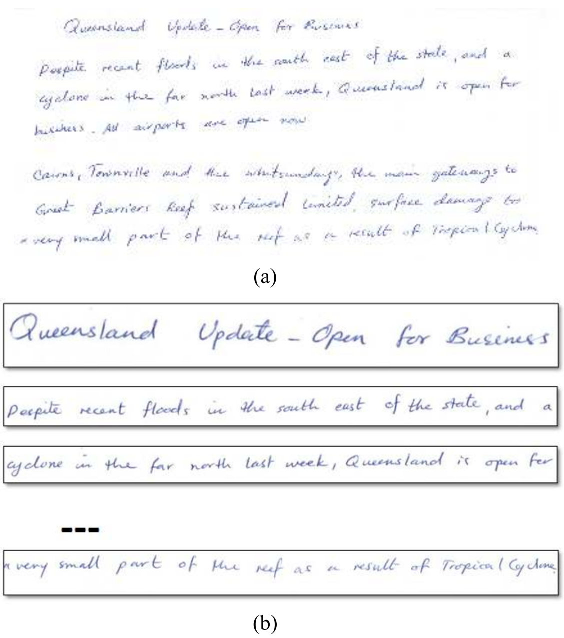

FIGURE 3. Results of English line segmentation. (a) Original image of writer Id 0001_03. (b) Segmented lines after skew detection and correction.

the document images are converted into grey scale resolution. Although the images of QUWI might carry additional information e.g., pen pressure, different inks etc. We choose to work on the grey scale images which simplify the representation and comparison of two handwriting forms.

Patches extraction is needed because we deployed CNN AlexNet that is a deep learner and require a large number of annotated data for the best results. It is to be noted that

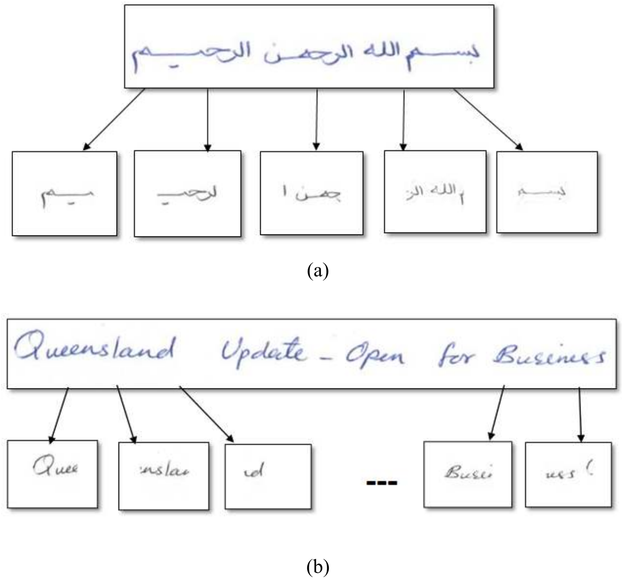

FIGURE 4. Results of patches extraction. (a) Patches of line 1 of writer Id 0001_01. (b) Patches of line 1 of writer Id 0001_03.

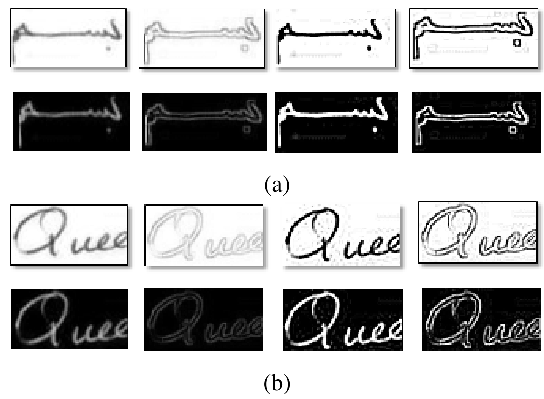

FIGURE 5. patches of original, contoured, sharped, sharped contours and their Negatives [left to right]. (a) Arabic samples. (b) English samples.

paragraphs or lines is not a sufficient data for the AlexNet. Thus we divide the lines into the patches of size 227 × 227. This division is carried out by positioning small windows over the writing from right to left and top to bottom while considering Arabic script that is 1 and 2 page of QUWI dataset. Similarly, 227 × 227 windows is slide on text line from left to right and top to bottom while considering English text. The resultant patches representation is shown in Figure 4.

4) DATA-AUGMENTATION: CONTOUR, SHARP AND NEGATIVE IMAGES GENERATION

In order to increase the performance of CNN AlexNet, we have further work for expanding the samples of patches of text-lines of QUWI dataset in pre-processing step. Unlike data-augmentation techniques, first we take the contours of the images by eroding the image with structuring element. Then perform the set differences of image and its resultant erosion followed by image un-sharping masking to sharp all patches at final step. The resultant images of contours, negatives and sharpness are represented in Figure 5.

TABLE 3. Features description.

No.	Features	Description	Dimensions
1	Conv1	96 filters of 11×11, stride=4, size=55×55×96 = 290,400 neurons	290,400
2	Conv2	256 filters of 5×5, stride=2, size=27×27×256 = 186,624 neurons	186,624
3	Conv3	384 filters of 3×3, stride=1, size = 13×13×384 = 64,896 neurons	64,896
4	Conv4	384 filters of 3×3, stride=1, size = 13×13×384 = 64,896 neurons	64,896
5	Conv5	256 filters of 3×3, stride=1, size=13×13×256 = 43,264 neurons	43,264
6	Fc6	4,096 neurons	4,096
7	Fc7	4,096 neurons	4,096
8	Fc6+Fc7	4,096+4,096 = 8,192 neurons	8,192

C. AUTOMATIC VISUAL FEATURES EXTRACTION

After pre-processing, the next step is feature extraction. It is one of the most significant step of any research domain in machine learning. In our approach, we deploy AlexNet architecture of CNN to extract the learned freeze activation features. It is the easiest and fastest way to use the representational power of pre-trained deep network i.e. Alexnet architecture of CNN. AlexNet CNN [32] is a deep learning model that is pre-trained on ImageNet [33]. AlexNet architecture encompasses five convolutional layers and three fully connected layers. Deeper layers contain higher-level features, constructed using the lower-level features of earlier layers. Number of experiments conducted to extract features using the convolutional and down sampling fully connected layers in our proposed system. In the features extraction step, we have selected freeze automated visual features computed on each CovNet layer to train a network in experiment-1 – experiment-8. These visual features are independent of domain, pattern, language or script. The feature detail is presented in Table 3. The first convolutional layer apply 96 filter of 11 × 11 with 4 strides. This Conv layer produces 290,400 neurons, 363 weights, 1 bias and the total parameters of 105,705,600. *tanh* function is applied in Relu layer to make improvement six times faster. After the Relu layer, local response normalization is applied from the relation:

$$b^i_{x,y} = a^i_{x,y} \Big/ \left(k + \alpha \sum_{j=max(0,i-n/2)}^{min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^{\beta} \tag{1}$$

where n is 5 channels per element, k is the stride having value 2, $\alpha = 10^{-4}$ and $\beta = 0.75$. The max pooling layer applies maximum operation with 3 × 3 filters and a stride of 2. Similarly second convolutional apply same operations but with the 256 filters of 5 × 5. The third, fourth and fifth Convolution layers apply 384, 384 and 256 filters with the filter size of 3 × 3 each followed by Relu layer respectively. Two fully connected layers of 4096 neurons are used followed with dropout layer of 50% and Relu layers.

So as to investigate which feature of the AlexNet has greater expressive capability, two fully connected layers Fc6, Fc7 and all convolution layers Cov1-Cov5 are extracted and then fed to SVM. The fixed features from each convolutional layer is unified by vectorization. Vectorization is the process of unifying the different dimensions into a single feature vector. Each Convolutional layer is vectorize to single feature vector and then passed to the classifier. Despite the fact that internal representations of CNN layers are difficult to decode, an intuitive guess may be made by way of visualizing the output of various layers. Figure 6 visualize the strongest activation from each convolutional layer and fully connected layer. It can be seen that Conv5 layer is appropriately suitable for learning discriminated features in this case. As fully connected layer produces distortion activation, this lead us to choose the features from Conv5.

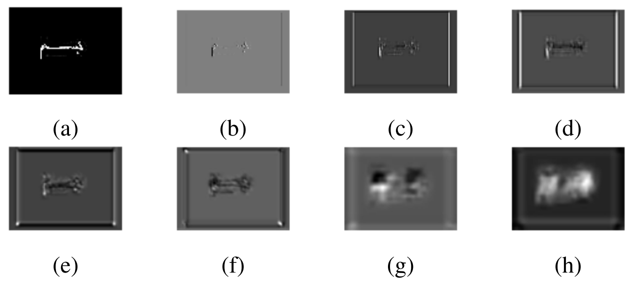

FIGURE 6. Feature visualization of strong activation channels. (a) Orignal image. (b) Conv1. (c) Conv2. (d) Conv3. (e) Conv4. (f) Conv5. (g) Fc6. (h) Fc7.

D. CLASSIFICATION AND IDENTIFICATION

The classification accuracy increase with the increase of the depth of the number of layers in deep learning model. The various feature vectors from different layers are extracted and then pass to the SVM classifier for classification.

In our study, Conv5 features are abstract, effective and more expressive in our research problem. Thus 13 × 13 × 256 dimensions are unified to a single feature vector of 43,264 through vectorization. Then feature vector is fed to the multi-class SVM for classification and identification of writer(s).

V. EVALUATION METRICS

The effectiveness of proposed method is evaluated by computing evaluation matrices like accuracy, sensitivity, specificity and precision for each representation. We also combine multiple representations and calculate evaluation matrices.

- **Accuracy**: Accuracy is the fraction of labels that the network predicts correctly computed by the following relations:

$$Accuracy = \frac{tp + fn}{tp + tn + fp + fn} \tag{2}$$

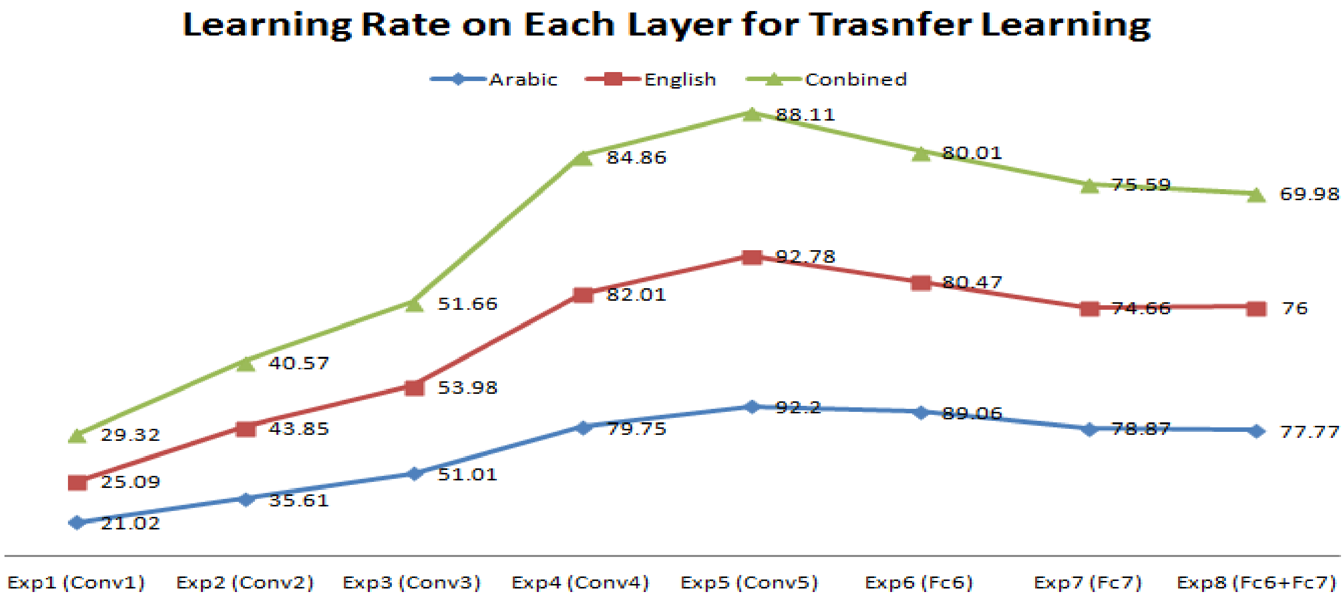

FIGURE 7. Reported accuracy against convolutional and fully connected layers fed to SVM for identification of writer using QUWI data-set's samples.

- **Sensitivity**: Sensitivity determines the ability of system to correctly classify the writers and is defined as the proportion of true positives and calculated by:

$$Sensitivity = \frac{tp}{tp + fn} \tag{3}$$

- **Specificity**: Specificity is the test ability of the model to correctly classify the actual writers form the data corpus and is calculated by:

$$Specificity = \frac{tn}{tn + fp} \tag{4}$$

- **Precision**: Precision is the true positive relevant measure and is calculated by:

$$Precision = \frac{tp}{tp + fp} \tag{5}$$

where tp is the true positive rates, tn is the true negative rates, fp is the false positive rates and fn is the false negative rates.

VI. RESULTS AND ANALYSIS

As we discussed in features extraction section that number of experimental studies for extraction of features was carried out on the QUWI database comprising data from 1017 writers. The 80% of database was used in training while 20% data for testing. We have generated several representations of patches of text-line images including images of original text, contours, sharp and negatives in pre-processing step. In this paper, two studies are carried out to evaluate the performance of AlexNet architecture of CNN. In first study, eight series of experiments were conducted using ConvNet of conv1, conv2, conv3, conv, conv5, fc6, fc7 and fc6 + fc7 (as shown in Fig. 7). For conv5 feature vector, network-5 achieved highest identification rate up to 92.20%, 92.78% and 88.11% on Arabic, English and Combined languages, respectively.

TABLE 4. Summary of identification rate using multiple representations for features vector from Conv5 layer.

Exp-No	Combination of Different Presentations of Data	Arabic	English	Combined
Exp-1	Original Patches	77.74	71.44	70.65
Exp-2	Original and Contoured Patches	85.62	78.30	80.50
Exp-3	Sharped and Contoured Patches	82.00	76.67	78.34
Exp-4	Original, Contoured and Sharped Patches	76.85	82.18	76.98
Exp-5	Original, Contoured, Sharped and Negative Patches	92.20	92.78	88.11

The network-5 in experiment5 shows promising performance as compare to other networks trained on other convolutional layers and fully connected layers in Experiment1-experiment4 and experiment6-experiment8. The Fig. 7 shows that identification rate increased from conv1 to con5 and achieved highest accuracy then it drops out. The reason of dropping the identification rate of writer is due to different samples data in target data-set (QUWI) as compare to samples data of base data-set (ImageNET). The target task has handwritten image patches of text-lines in Arabic and English while the base task has natural images like cat, dog, pencils, cap etc. The extracted features on fc6 and fc7 layers are more specific to the base data-set (ImageNet).

In second study, the different representations are investigated using AlexNet architecture. We carried out five series of experiments on features extracted by conv5 using different representations of data of QUWI data-set as depicted in table 4. The experiment-1 was carried on conv5 based features using original patches and showed accuracy up to 77.74%, 71.44% and 70.65% on Arabic, English and combined languages. The original and contoured patches are used in experiment-2, sharped and contoured patches in experiment-3, original, contoured, sharped patches in experiment-4 and finally using fusion of image patches of original, contoured, sharped and negatives in experiment-5. Investigating and comparing the different combination of representations of images, it is concluded that combination of all representations leads to the highest accuracy in experiment-5 as compare to other representations in from experiment-1 to experiment-4. There are two main reasons, one is due the large number of samples of data-set and second reason is negative images which increased frequency of black pixel information in the images. As AlexNet is pre-trained on the ImageNet that contains natural images, the black information leads to relate with natural images rather than other representations. Furthermore it is also seen in experiment-5 that 92.78% accuracy achieved on English language. The 92.20% accuracy is achieved on Arabic language.

We also computed different evaluation matrix for fusion of all representations of data in experiment-5 (showed in table 5). However, fusing both languages also yields to promising and satisfactory results of 88.11%. This shows that handwriting of the writer in English is also affected and helpful for identification handwriting of the writer in Arabic and vice versa. This study is also pioneer study in the field of writer identification that how the two languages (Arabic and English) are helping for each other in identification of writer.

TABLE 5. **Different types of evaluation matrix for combination of all representations using Conv5 based features.**

Language	Accuracy	Sensitivity	Specificity	Precision
Arabic	92.20	92.24	98.60	92.16
English	92.78	92.62	98.35	93.02
Combine	88.11	88.07	97.89	89.01

For the generalization of results, we also carried out two types of experiments concerning cross validation. The reason of cross validation is to analyze the effect of universal information samples on the learning of our proposed system. First, we conducted k-fold cross validation since it guarantees that each sample eventually become the part of training as well as testing sets. The QUWI data-set consists of 1017 writers. In the first case, we divide the number of writer into 9 folds.Thus we have placed the data of 113 writers in each set. We have evaluated experiment in such a way that 113 writers is used for testing and the remaining 904 writers in 8 folders are used for training purpose. The overall split of the data-set for 9-fold cross validation and the results obtained by 9-fold cross validation are shown in Figure 8. Consistently higher accuracy on all partitions of the data indicate that our network indeed achieves good generalization and no training/test bias was present in the initial experiments.

In the second type of cross-validation, repeated random sub-sampling approach is applied. The data-set is shuffled and randomly divided into 80% training set and 20% testing set then experiment carried out. The same process is repeated ten times to carried out 10 experiments. Finally identification

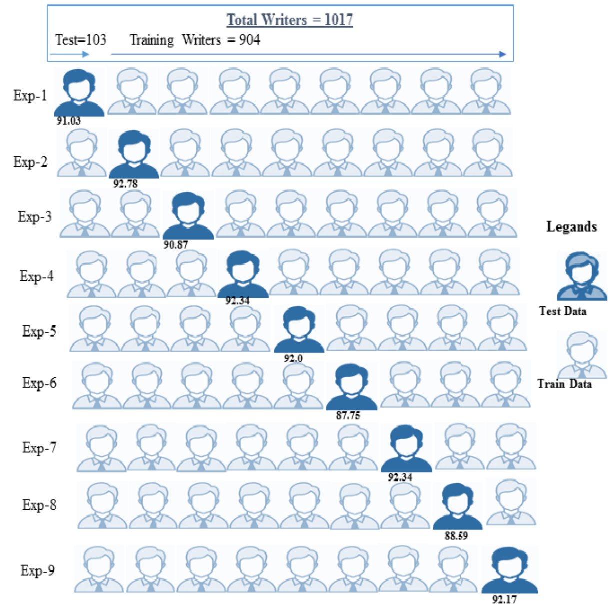

FIGURE 8. **The overall split of the QUWI data-set for 9-fold cross validation and accuracy of each experiment.**

results are then averaged of all experiments. In these experiments an average accuracy of $92 \pm 0.78\%$ is achieved.

No direct comparison of performance of our proposed writer identification system using QUWI data-set is possible with reported other systems in the literature (summarized in Table.1) for writer identification due to use of different data-sets like ICDAR 2011, ICDAR 2013, ICDAR 2017, CVL, IAM, KHATT etc. Our proposed approach is a pioneer study using QUWI data-set which investigated the different layers for features extraction from base task and transferred to the target task for writer identification. In this study, it is also investigated that sharing of the visual patterns of handwriting in different languages or scripts helped across these languages and provided the marginal performance in combined languages for writer identification.

VII. CONCLUSION

Automatic writer identification is very intriguing research problem in the field of document analysis and handwriting recognition. The effective implementation of writer identification systems can be applicable in forensic and historical analysis, banks, check processing, signature analysis, graphology, legal documents, ancient manuscripts, digital rights administration, and document analysis methods. The objective of our study is to explore the visual patterns for automatic writer identification from unconstrained offline scanned text-lines images of handwriting.

We have presented a pioneer study for writer identification and authentication in handwritten documents using QUWI data-set. The proposed approach used pre-trained CNN model named as AlexNet architecture using freeze layers. The features are extracted from base data-set named as ImageNet and then transferred the learned freeze features for classification using target data-set named as QUWI data-set. We conducted number of experiments for identification of writer using different freeze or fixed layers like conv3, conv4, conv5, fc6, fc7 and fusing of fc6 and fc7. The freeze cov5 layer outperform as compare to other layers and fusion of fc6 and fc7 and achieved the highest accuracy 92.78% on English, 92.20% on Arabic and 88.11% on combining both languages, respectively. The deep transfer learning shows significant results on QUWI data-set as a pioneer study. It is empirically shown that our proposed method is suitable for Latin as well as Arabic like scripts.

In future, we will employ various deep learning models like GoogleNet, VGG, ResNet and its variants for writer identification. We also explore the fusion of different architectures of CNN. We also aim to evaluate the deep transfer learning techniques for multi-script writer identification.

REFERENCES

[1] A. Rehman, S. Naz, and M. I. Razzak, "Writer identification using machine learning approaches: A comprehensive review," *Multimedia Tools Appl.*, pp. 1–43, Sep. 2018. doi: 10.1007/s11042-018-6577-1.

[2] V. Christlein, M. Gropp, S. Fiel, and A. Maier. (2017). "Unsupervised feature learning for writer identification and writer retrieval." [Online]. Available: https://arxiv.org/abs/1705.09369

[3] S. Naz *et al.*, "Offline cursive Urdu-Nastaliq script recognition using multidimensional recurrent neural networks," *Neurocomputing*, vol. 177, pp. 228–241, Feb. 2016.

[4] N. K. Garg *et al.*, "Writer identification system for handwritten gurmukhi characters: Study of different feature-classifier combinations," in *Proc. Int. Conf. Comput. Intell. Data Eng.* Singapore: Springer, 2018, pp. 125–131.

[5] S. Naz, A. I. Umar, R. Ahmad, S. B. Ahmed, S. H. Shirazi, and M. I. Razzak, "Urdu Nasta'liq text recognition system based on multi-dimensional recurrent neural network and statistical features," *Neural Comput. Appl.*, vol. 28, no. 2, pp. 219–231, 2017.

[6] C. Adak, B. B. Chaudhuri, and M. Blumenstein. (2017). "An empirical study on writer identification & verification from intra-variable individual handwriting." [Online]. Available: https://arxiv.org/abs/1708.03361

[7] I. Siddiqi and N. Vincent, "A set of chain code based features for writer recognition," in *Proc. 10th Int. Conf. Document Anal. Recognit. (ICDAR)*, Jul. 2009, pp. 981–985.

[8] R. Kumar and M. Kaur, "A character based handwritten identification using neural network and SVM," *Int. J. Sci. Res. Sci., Eng. Technol.*, vol. 3, no. 1, pp. 120–124, 2017.

[9] P. Pandey and K. R. Seeja, "Forensic writer identification with projection profile representation of graphemes," in *Proc. 1st Int. Conf. Smart Syst., Innov. Comput.* Singapore: Springer, 2018, pp. 129–136.

[10] J. J. Miller, R. B. Patterson, D. T. Gantz, C. P. Saunders, M. A. Walch, and J. Buscaglia, "A set of handwriting features for use in automated writer identification," *J. Forensic Sci.*, vol. 62, no. 3, pp. 722–734, 2017.

[11] A. Durou, I. Aref, S. Al-Maadeed, A. Bouridane, and E. Benkhelifa, "Writer identification approach based on bag of words with OBI features," *Inf. Process. Manage.*, vol. 56, no. 2, pp. 354–366, 2019.

[12] I. Siddiqi and N. Vincent, "Combining contour based orientation and curvature features for writer recognition," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Berlin, Germany: Springer, 2009, pp. 245–252.

[13] G. Ghiasi and R. Safabakhsh, "Offline text-independent writer identification using codebook and efficient code extraction methods," *Image Vis. Comput.*, vol. 31, no. 5, pp. 379–391, 2013.

[14] A. A. Ahmed, H. R. Hasan, F. A. Hameed, and O. I. Al-Sanjary, "Writer identification on multi-script handwritten using optimum features," *Kurdistan J. Appl. Res.*, vol. 2, no. 3, pp. 178–185, 2017.

[15] S. He and L. Schomaker, "Writer identification using curvature-free features," *Pattern Recognit.*, vol. 63, pp. 451–464, Mar. 2017.

[16] D. Bertolini, L. S. Oliveira, and R. Sabourin, "Multi-script writer identification using dissimilarity," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 3025–3030.

[17] Y. Hannad, I. Siddiqi, and M. El Youssfi El Kettani, "Writer identification using texture descriptors of handwritten fragments," *Expert Syst. Appl.*, vol. 47, pp. 14–22, Apr. 2016.

[18] V. Christlein and A. Maier, "Encoding CNN activations for writer recognition," in *Proc. 13th IAPR Int. Workshop Document Anal. Syst. (DAS)*, Apr. 2018, pp. 169–174.

[19] V. Christlein, D. Bernecker, A. Maier, and E. Angelopoulou, "Offline writer identification using convolutional neural network activation features," in *Proc. German Conf. Pattern Recognit.* Cham, Switzerland: Springer, 2015, pp. 540–552.

[20] S. Fiel and R. Sablatnig, "Writer identification and retrieval using a convolutional neural network," in *Proc. Int. Conf. Comput. Anal. Images Patterns*. Cham, Switzerland: Springer, 2015, pp. 26–37.

[21] S. Naz *et al.*, "Urdu Nastaliq recognition using convolutional–recursive deep learning," *Neurocomputing*, vol. 243, pp. 80–87, Jun. 2017.

[22] I. Goodfellow, Y. Bengio, A. Courville, and F. Bach, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[23] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Classification in BioApps*. Cham, Switzerland: Springer, 2018, pp. 323–350.

[24] W. Yang, L. Jin, and M. Liu, "Chinese character-level writer identification using path signature feature, dropstroke and deep CNN," in *Proc. 13th Int. Conf. Document Anal. Recognit. (ICDAR)*, Aug. 2015, pp. 546–550.

[25] L. Xing and Y. Qiao, "Deepwriter: A multi-stream deep CNN for text-independent writer identification," in *Proc. 15th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Oct. 2016, pp. 584–589.

[26] Y. Tang and X. Wu, "Text-independent writer identification via CNN features and joint Bayesian," in *Proc. 15th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Oct. 2016, pp. 566–571.

[27] R. Nasuno and S. Arai, "Writer identification for offline Japanese handwritten character using convolutional neural networl," in *Proc. 5th IIAE Int. Conf. Intell. Syst. Image Process. (ICISIP)*, Honolulu, HI, USA, 2017, pp. 94–97.

[28] S. Al Maadeed, W. Ayouby, A. Hassaíne, and J. M. Aljaam, "QUWI: An Arabic and English handwriting dataset for offline writer identification," in *Proc. Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Sep. 2012, pp. 746–751.

[29] R. Ahmad, S. F. Rashid, M. Z. Afzal, M. Liwicki, A. Dengel, and T. Breuel, "A novel skew detection and correction approach for scanned documents," in *Proc. 12th Int. IAPR Workshop Document Anal. Syst. (DAS)*, 2016, p. 2.

[30] R. S. Stephens, "Probabilistic approach to the hough transform," *Image Vis. Comput.*, vol. 9, no. 1, pp. 66–71, 1991.

[31] R. Ahmad, M. Z. Afzal, S. F. Rashid, M. Liwicki, and A. Dengel, "Text-line segmentation of large titles and headings in arabic like script," in *Proc. 1st Int. Workshop Arabic Script Anal. Recognit. (ASAR)*, Apr. 2017, pp. 168–172.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.

Authors' photographs and biographies not available at the time of publication.