Writer Identification: The effect of image resizing on CNN performance

Abdelillah Semma [1,*], Said Lazrak [2], Yaâcoub Hannad [3], Mohamed Boukhani [4], Youssfi El Kettani [5]

[1] Ibn Tofail University, Kenitra, Morocco – semma_abdelillah@yahoo.fr
[2] Ibn Tofail University, Kenitra, Morocco – s.lazrak@mmsp.gov.ma
[3] Faculty of Educational Sciences, Mohammed V University, Rabat, Morocco – y.hannad@gmail.com
[4] National Institute of Statistics and Applied Economics, Rabat, Morocco – m.boukhani@mmsp.gov.ma
[5] Ibn Tofail University, Kenitra, Morocco – elkettani@univ-ibntofail.ac.ma

KEY WORDS: Writer Identification, Deep Learning, Convolutional Neural Networks, Image resizing.

ABSTRACT:

Introducing Deep Learning has been successful in improving the performance of automated writer identification systems. However, using very large patch sizes as input to CNN consumes a lot of machine resources and requires a lot of training time. To overcome this problem, many researchers use resized images.
In this paper, we will try to make a comparative study between several patches sizes which were then resized to a normalized size of 32x32. Our aim is to elaborate the best recommendations for choosing the image resizing in order to increase the CNN performance. Thus, we will carry our tests on three databases. The first is CVL, a Latin dataset with 310 writers, the second is CERUG-CH a Chinese dataset with 105 writers and the last is KHATT that contains the Arabic writings of 1000 writers. To see if the type of CNN model impacts the results conducted on resized images, we deploy two models: ResNet-18 and MobileNet. The main finding is that the best results correspond to the resizing values of the images which makes it possible to have the average line height of the writings closer to the height of the CNN patches.

1. INTRODUCTION

Even if both handwriting recognition and writer identification of a script seem to fall under the same field relating to handwritten texts, but they do not pose the same types of difficulties, as handwriting recognition seeks to eliminate variability between various writers and to recognize the real text drawn in a handwritten image. While in the writer identification approach we seek to know the different commonalities or the invariants in the writings of each writer in order to attribute a handwritten document to its true author. Two writings from the same author are not 100% alike but they vary according to several factors related to the psychic state of the writer (Nosary et al., 1999) and to the tools used such as the pen or the writing support.

There are two approaches that help recognize the writer of a handwritten document, the first is writer identification and the second is writer verification. In the writer verification approach, we seek to verify whether two or more handwritten documents were produced by the same writer. This approach is document-oriented. In the writer identification approach, we have a document that we seek to recognize the corresponding writer among a given set of writers using a database of handwritten training documents. This approach is writer-oriented.

Introducing of deep learning (Christlein and Maier, 2018; Fiel and Sablatnig, 2015; Xing and Qiao, 2016; Semma et al.,2022; Rehman et al., 2019; Javidi and Jampour, 2020; Chahi et al., 2020; Semma et al., 2021) has contributed to the improvement of automated systems for identifying writers. This enabled, for example, to achieve an identification rate of 99.6% (Christlein and Maier, 2018) on the Arabic dataset KHATT instead of 84.4% (Hannad et al., 2019) and 84.5% (Asi et al., 2017) obtained by former systems. However, as we will see in the related work section, researchers do not use a well-defined standard for choosing input patches for their CNN. This is clearly seen in the entertaining of the patch sizes used and in the resizing rates of the images in the various research papers.

In our study, we try to verify the impact of resizing handwritten images by testing 12 values varying from 8 to 100 percent of the original size of the images. Thus, we took original patches of sizes ranging from 400x400 to 32x32 which were resized to 32x32 and introduced them into two Convolutional Neural Networks ResNet-18 and MobileNet. The tests were carried out on three datasets: Arabic (KHATT) with 1000 writers, Latin (CVL) with 310 writers and Chinese (CERUG-CN) with 105 writers. The key highlights of our study include:

- Reviewing CNN performance using multiple image resizing rates.
- Checking the relationship between script line heights and good CNN performance.
- Conducting tests on three datasets of several languages and sizes

This paper contains 4 sections. After a brief introduction, we will refer to some research works related to the field of handwriting identification and we focus on works that have used convolutional neural networks. In the datasets section, we present the three datasets used in our study. In section 4 we detail the methodology adopted as well as the datasets used and the neural networks deployed. Before concluding we present in section 5 the different results obtained.

2. RELATED WORK

The first work that dealt with writer identification using deep learning is (Fiel and Sablatnig (2015)) which used a CNN composed of three convolutional layers and two others fully connected with 56x56 input patches. The proposed descriptors are those relating to the second fully connected layer of dimension 4096 and the deployed classifier is a nearest neighbor. To identify the writer of a manuscript the author compares the average vector of all the descriptors of the patches in the manuscript with the vectors of the training images using the distance $\chi 2$.

---

* Corresponding author

In Xing and Qiao (2016), the authors proposed a deepWriter composed of two half deepWriter, each of which has the structure of an alexNet with 5 convolutional layers and two dense. The sum of the ouputs of the two half deepWriter is introduced in a Softmax classification function. The authors tested three square patch sizes 113x113, 131x131, and 227x227 in two IAM and HWDB datasets.

Christlein et al. (2015) proposed a system based on the GMM encoding vectors of the penultimate layer of CNN with an SVM classifier. He conducted his study on two datasets ICDAR13 and CVL. The same author proposed another unsupervised classification system (Christlein et al., 2017) based on a Resnet residual networks trained on SIFT patches of size 32x32 and the indices of clusters of SIFT descriptors as targets. To classify the test images, the author used an exemplary SVM as the classifier and the VLAD encoding vectors of the Resnet activations as descriptors. The system achieved an accuracy of 84.1% on CLaMM16 and 88.9% on Historical-WI. A third study, conducted by the same author (Christlein and Maier, 2018), in which he used the VLAD and Triangulation Embedding encoding of the activations of the penultimate layer on three databases CVL, KHATT and ICDAR13. The author tested 4 types of CNN (LeNet-A, LeNet-B, Resnet-8 and Resnet-20) with 32x32 patches.

In (Rehman et al., 2019) the author proposed a system based on CNN AlexNet with input patches of size 227x227. To avoid overfitting, the author employed the data-augmentation technique using the contour, negative and sharpness patches of the original patches. All training and test images were normalized to a fixed width of 1250 pixels with respect for the ratio between height and width of each image. The evaluation of the proposed system on the bilingual QUWI dataset allowed the achievement of 92.20 % in Arabic, 92.78 % in English and 88.11 % in the combination of the two languages.

Among the most recent studies relating to writer identification, we find (Javidi and Jampour, 2020) which introduced an extended version of Resnet with as input fixed rectangular patches of size 64x180. Their approach consists in adding a vector of dimension 45 relating to a descriptor of the thickness of the handwriting (HTD) to the penultimate layer which has a size 512 and which will have after the addition of the vector HTD a size of 512 + 45. Their method allowed to surpass the performance of the original version of Resnet and to achieve important results in four datasets IAM, Firemaker, CVL, and CERUG-EN.

In (Chahi et al., 2020) the authors deployed a new Local gradient full-Scale Transform Patterns (LSTP) method that exploits local regions in handwritten images to extract connected components. They used a KNN classifier to identify the writers of seven datasets.

In (Semma et al., 2021), the authors proposed a system based on extracting 32x32 size patches around FAST and Harris key-points. To identify the writers of each handwritten image, the proposed system uses the VLAD encoding technique of the 30th convolutional layer of ResNet-34 and a KNN classifier. The tests conducted allowed to reach 99.8%, 99.7%, 99.7%, 99.3% and 99.5% in five datasets QUWI-Ar, QUWI-En, QUWI-Ar + QUWI-En, IFN/ENIT and IAM respectively.
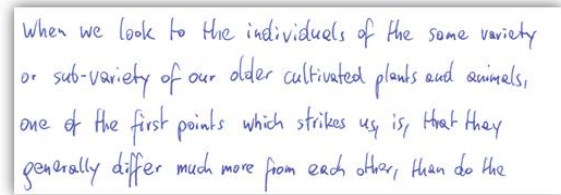
3. DATASETS

3.1 CVL

The CVL database (Kleber et al., 2013) comprises the handwritten images belonging to 311 writers in German and English languages. Each participant produced 7 texts (6 in English and 1 in German). The database is available to the public in an original version including printed and handwritten texts and in a cropped version with only handwritten texts. The CVL dataset is accessible to the public via the following link: https://zenodo.org/record/1492267.
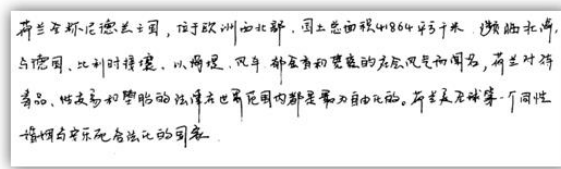
In our study, we use the cropped version with the handwritten of 310 writers with only four images for each writer. We took the first image as a test image and three other images for the CNN training and validation phase.
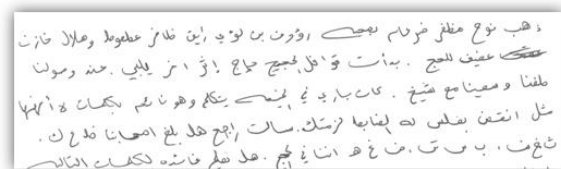
(a)

(b)

(c)

Figure 1. Handwriting samples from (a) KHATT, (b) CVL and (c) CERUG-CN dataset

3.2 KHATT

KHATT (Mahmoud et al., 2014) is a database of Arabic manuscript documents containing the writings of 1000 people. Each writer filled out four pages with fixed and random text. The participants are from different countries such as Saudi Arabia or Morocco, and belong to different age group and sex with 67% male and 33% female.

For each of the 1000 writers, we consider the first page as the test image (25%) and the other 3 pages (75%) as training and validation images.

3.3 CERUG-CN

CERUG (He et al., 2015) is a bilingual dataset containing the handwritten texts of 105 writers. Each participant produces 4 pages. On the first page the writer copied two paragraphs in Chinese language, on the second page he wrote free text in Chinese, on the third page he copied two paragraphs in English. The third page has been divided into two pages, each of which contains a single paragraph. The participant also produced a 4th page containing two paragraphs containing the names of some cities and countries in English and Chinese. This last page was divided into two pages.

In our study we used only the two first pages in Chinese, one for the test and the other for the CNN training.

Figure 1 shows some examples of handwritten documents from KHATT, CVL and CERUG-CN datasets
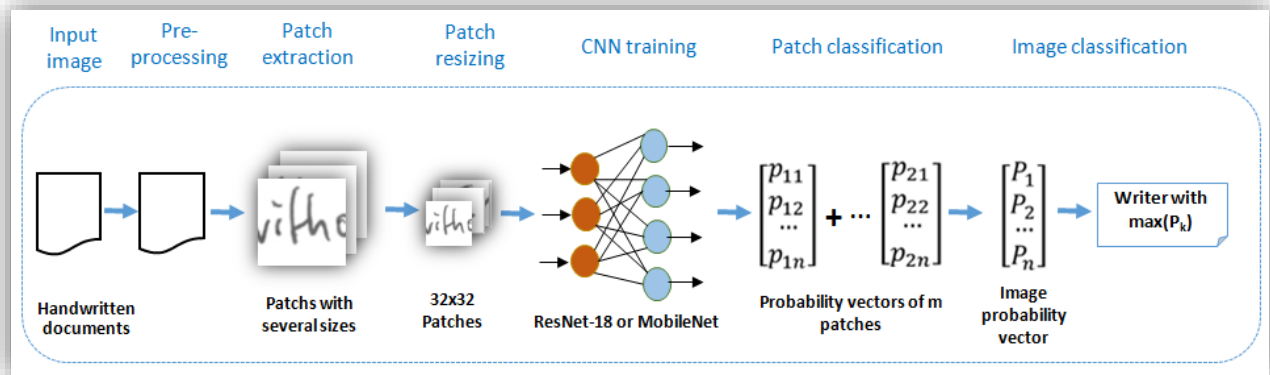
4. METHODOLOGY

Figure 2. CNN classification

We present, via this section, our methodology that aims assess the impact of resizing images on the performance of convolutional neural networks in research work relating to writer identification.

4.1 Pre-processing

Before introducing the handwritten images into the convolutional neural network for the training phase, the images are first Pre-processed. This operation involves:

- Converting the images to gray mode.
- Image skew adjustment using the method developed by Abecassis (2011). This process begins by detecting the block containing the handwritten text in the requested image. Then we calculate the rotation angle of the text block. Finally, we correct the skew of the text by applying a reverse rotation angle to the image.

4.2 Extraction of Patches

In our study we use a Resnet-18 and MobileNet with 32x32 input patches.

To check the effect of resizing images, we take 12 relative resizing values. Each value corresponds to a transformation of patches from an original size to a size of 32x32. Thus, the value of 16% corresponds to resizing from patche size 200x200 to patche size 32x32.

Patches are extracted randomly around pixels containing text (black pixels of the binary image). For a fair evaluation of the different values of resizing, we try to save for each image the coordinates of the first random extraction to take them in consideration during the other tests. Thus, the 200x200 and 64x64 and the other patch sizes are centralized around the same points of each image of the same dataset. An illustration of sample patches of different resize values are depicted in Figure 3.

For CVL dataset, we extract 50 patches from each image, and thus 200 patches for each writer. These patches are distributed as follows: 50 test patches, 135 training patches, and 15 validation patches. For CERUG-CN, we extracted 400 patches from each image to train ResNet-18 and MobileNet.
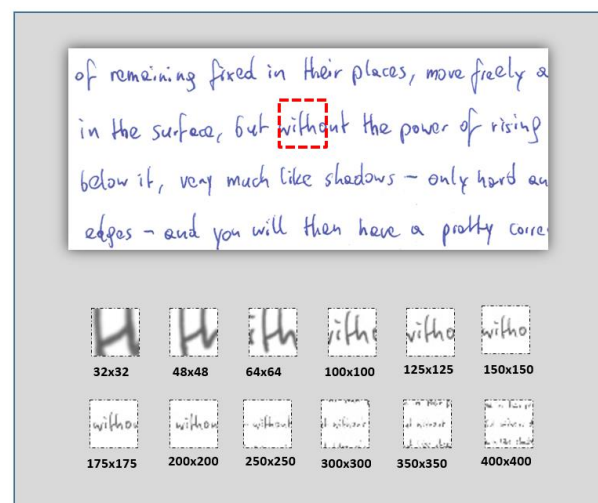

Figure 3. Samples of original patches of various sizes taken from the CVL dataset. These patches are subsequently resized to size 32x32 patches.

The KHATT dataset contains 4 pages for each writer. We take the first image as a test image and the other three images as training and validation images. The number of patches extracted is 50 for each image to train ResNet-18 and 200 for MobileNet.

To train and test the two networks ResNet-18 and mobileNet, we consider the same patches for the two databases CVL and CERUG-CN. But for KHATT database, we have changed the extraction points and the number of patches per image.
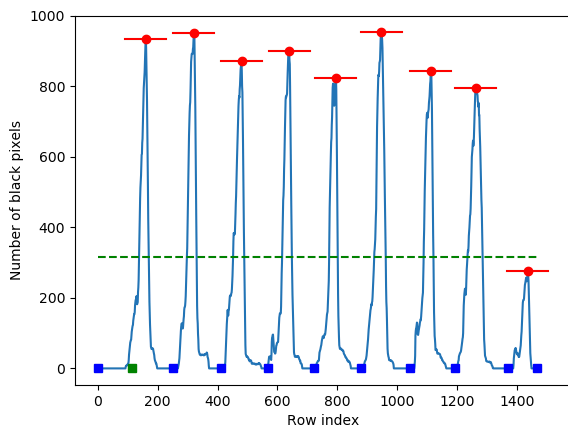
4.3 CNN

Figure 4. Horizontal histogram of a handwritten image from the CVL dataset

ResNet-18 : After the success of Alexnet in the ImageNet competition in 2012, the Resnet residual networks won the one in 2015. To handle the overfitting problem the residual networks proposed a system based on residual blocks which allow, thanks to their techniques of skip connection to save the identity of the block input after two or three layers of the block. In this work We used a modified version of Resnet-18 (He et al. 2016a) which allows the addition of the identity after applying the activation function of the block, as opposed to the original version (He et al. 2016b) where the addition of the identity is done before the activation function.

MobileNet: It is a convolutional network designed to meet the constraints of insufficient resources in environments such as mobile devices.

For the CNN training we have adopted a batch size of 500 with a learning rate initialized at 0.001 and multiplied by 0.1 after the 9th, 13th and 15th epochs.

4.4 Average line height

To see if there is a relationship between the average line height and the size of the original patches for which we have good CNN performance, we have chosen a line segmentation method for the different datasets.

The chosen line segmentation technique (Omar et al. 2018) is based on peaks and valleys. It succeeded in giving 98% of the well-segmented lines.
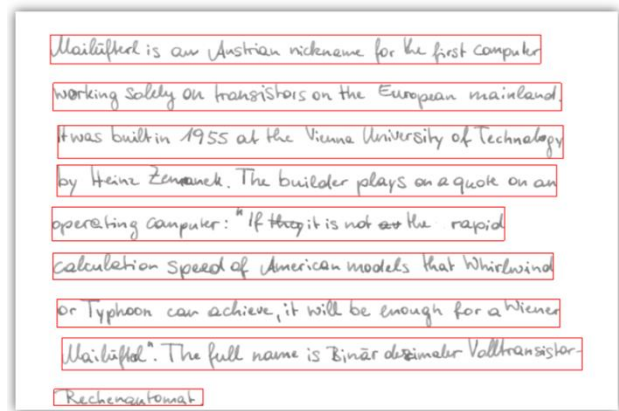

Figure 5. Line segmentation of a handwritten image of the CVL dataset

The segmentation algorithm begins with the horizontal projection of the binary version of the handwritten image (as can be seen in Figure 4.

The second step consists in calculating the threshold through the following equation:

$$threshold = max(h_i)/3$$

Where h_i is the intensity of the row i.
Therefore, any row having an intensity (h_i greater than threshold) and greater than the intensity of the preceding and the following rows will be considered as a peak. Then we detect the valley between each two peaks relative to the line whose intensity is minimal.

This technique can detect all the main lines, but small lines may be missed. To solve this problem we go to the third step which consists in calculating the difference between two successive valleys, and if this value is greater than the threshold2 then there would probably be another peak-valley between the two.

$$threshold2 = 1.5 * (average\ distance\ between\ peaks)$$

To determine the edges of the rectangular containing each line, we take each two consecutive valleys and we try to eliminate the space closest to the two valleys, as can be seen in the Figure 5

After having calculated the boundaries of the lines, we save the distance between each two successive boundaries, which considered as the value of the line spacing.

The average values of line height and line spacing of the three datasets are presented in Table 1.

Dataset	Average line height (px)	Average of line spacing (px)
KHATT	126	13
CERUG-CN	122	46
CVL	105	38

Table 1. Average values of line height and line spacing of the different datasets

5. EXPERIMENTS & RESULTS

As mentioned in the methodology section, the input patches of the two models ResNet-18 and MobileNet are 32x32 size. To resize the original images, we extracted patches of different sizes varying from 32x32 to 400x400 from the same extraction points. So in the remainder of the paper, the original patches of size NxN refers to resizing value of the original images with a rate of (32/N) and not the input patches of the two CNNs, because CNN input is the result of resizing said patches to 32x32.

To check the impact of resizing the original images on the performance of convolutional networks in the writer identification domain, we will compare the following parameters for each resize value:

- CNN convergence speed.
- CNN loss value.
- Classification score of test images.
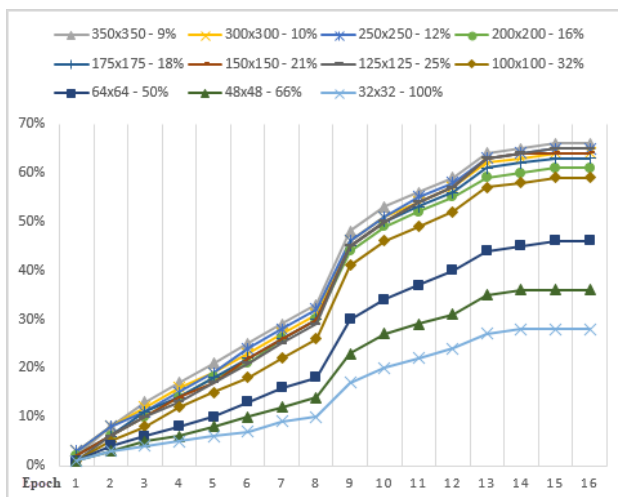
5.1 Training Phase

Figure 6. Evolution of the accuracy of the Resnet-18 training phase using various patche sizes of the KHATT dataset resized to 32x32

As depicted in Figure 6, the score of good classified training patches becomes important as resizing value of the original images decreases. This can be explained by the fact that large original patch sizes contain more important information than small ones.

We also note that the original patch sizes greater or close to the average value of the line height (126 pixels for KHATT dataset) converge faster than smaller sizes.

The same observation for the classification rate of the training patches can be applied for the Loss value where we see that it is greater for small original patch size which implies a higher error rate for patches containing little information.

These remarks are valid for the KHATT dataset with the Resnet-18 model as for the other datasets and the MobileNet model.
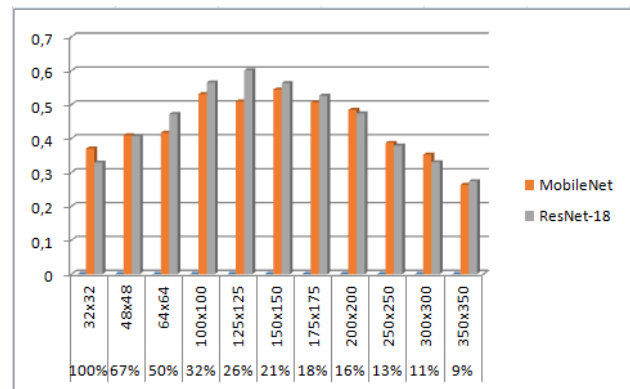
5.2 Test Phase

Figure 7. Classification of test images for the KHATT dataset using both Resnet-18 and MobileNet models with several original patch sizes

KHATT: For this dataset, we changed the extraction points, and the number of extracted patches per image between the two models. The chosen number of training patches is 50 patches per image for the ResNet-18 model and it is 200 for the MobileNet model. As depicted in Figure 7 the evolution is almost similar for the two models even if we use different patches.

The main finding is that the best score for the identification rate corresponds to the original 125x125 patch size for ResNet-18 which is close to the average line height for the KHATT dataset (126 pixels). We also note that the other important scores correspond to original patches of size 100x100 and 150x150. For MobileNet the two best classification rates are recorded for the original patches of size 150x150 and 100x100 with 54% and 53% respectively against 51% for 125x125.

The figure also shows that for very small original patch sizes like 32x32 and 48x48, the classification rate is very low: 33% and 41% using Resnet-18 and 37% and 41% using MobileNet. The same thing applies to the original patches of very large sizes such as 300x300 and 350x350 whose identification rate is 33% and 27% using Resnet-18 and 35% and 26% using mobileNet.
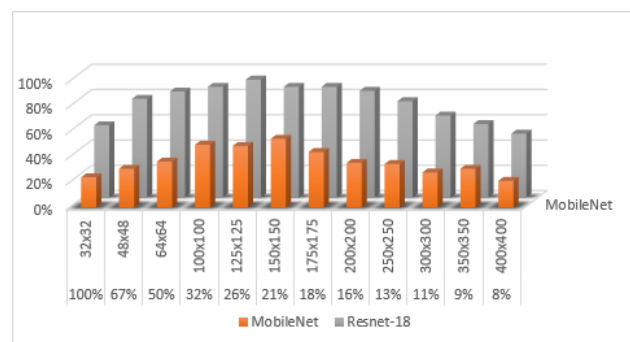

Figure 8. The Top-1 classification rate of test images from the CERUG-CN dataset for both Resnet-18 and MobileNet models with several sizes of the original patches

CERUG-CN: For the CERUG-CN datasets, we used the same extraction point and the same number of extracted patches to train both Resnet-18 and MobileNet. The high score recorded for the classification of test images with the Resnet-18 model is 93% for

125x125 original patche size. The second best is 88% corresponding to 100x100, 150x150 and 175x175. For the MobileNet model, the best score is recorded for original 150x150 patches with a rate of 54%. The other scores that are closer correspond to original patches of size 100x100, 125x125 and 175x175 with successive rates of 50%, 49% and 44%.
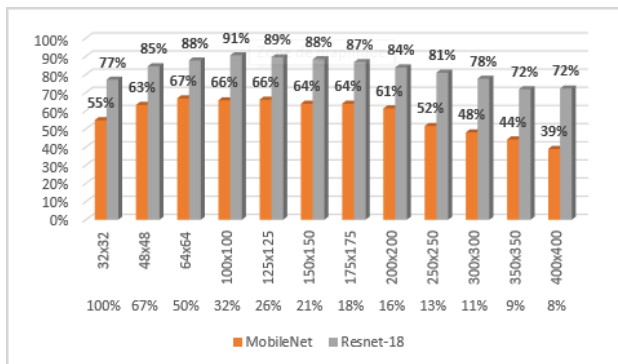

Figure 9. The Top-1 classification rate of test images from the CVL dataset for both Resnet-18 and MobileNet models with several sizes of the original patches

CVL dataset: To train Resnet-18 and MobileNet on the Latin CVL database, we took three training images and one test image and extracted 50 patches per image. As for CERUG-CN, we have trained the two models on the same patches.

From the illustrative Figure 9 we can see that the best scores relative to the classification of the test images are recorded for the original patches of sizes between 64x64 and 175x175. The maximum of the Resnet-18 model is reached for patches size 100x100 with a classification rate of 91%. Whereas for MobileNet the maximum is reached for the original patches of size 64x64 with a rate of 67%. The second observation is that the scores of the MobileNet model in the range of 64x64 to 175x175 remain comparable with values varying between 64% and 67%. Likewise for resnet-18 the scores in the same interval vary between 85% and 91%.
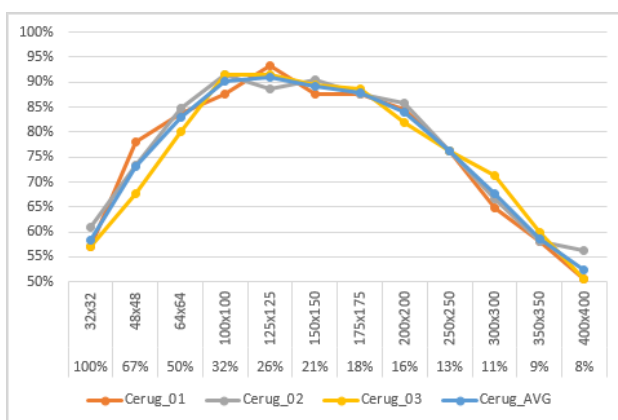

Figure 10. Results of three tests per patch size using the ResNet-18 model and the CERUG-CN dataset

To verify whether the evolution of the classification rates of the test images compared to the resizing values of the original images remains the same for several tests, we performed for each resizing value in CERUG-CN dataset three tests with Resnet-18

model. In each test we have modified the order of the input patches of the training phase.

As we can see in Figure 10, the numbers relating to the classification rates are not stable and for two patch sizes 100x100 and 125x125, we have in the first test a classification rate of 125x125 greater than that of 100x100, while for the second test the opposite is true, while for the third test they are equal. The same remark can be concluded for several other patches.

Even sometimes the classification rates of test images vary between two successive original patch sizes, but the general evolution remains the same with low classification rates for very small and very large sizes, and high classification rates for original patch sizes close to the average line height.

According to the analysis of all the tests carried out on the three datasets KHATT, CERUG-CN and CVL and using the two networks Resnet-18 and MobileNet, we can deduce that there is not a standard rule which proves that a given resizing value allows to have the maximum performance of CNN. Because, as we can see in Figure 10 even if we take the same CNN model and the same number and extraction points of patches, the change of the order of the input patches of the training phase clearly affects the classification results.

However and according to the results obtained, we can say that:

- The original patches of size close to the average line height of the dataset give classification rates which are the best or among the best.
- Original patches very small or very large in size than the average line height probably give lower classification rates.
- If the average interline is very large then the interval of the original patch sizes which allow good results will be very wide (this is the case of the two CVL and CERUG-CN datasets which have an average interline of 38 and 46 pixels respectively).

6. CONCLUSION

Many researchers are building writer identification systems using Convolutional Neural Networks with patches extracted from resized images.

To verify the impact of image resizing values on classification rates, we performed several tests with two different convolutional neural networks ResNet-18 and MobileNet and three datasets containing handwritten documents in Arabic (KHATT), Latin (CVL) and Chinese (CERUG-CN).

Tests have shown that the choice of the resize rate of the original images greatly affects the performance of CNN in the field of writer identification. In addition, the best performance was observed for original patches of size close to the average line height.

We performed the tests by resizing the original patches of different sizes to 32x32 size patches that we put as our CNN input patches. However, the study can be generalized to sizes other than 32x32.

REFERENCES

Abecassis, F., 2011. Opencv-morphological skeleton. Retrieved from Felix Abecassis Projects and Experiments: http://felix. abecassis.me/2011/09/opencv-morphological-keleton/geological mapping at Cuprite Nevada: a rule-based system. International Journal of Remote Sensing 31, 7.

Asi, A., Abdalhaleem, A., Fecker, D., M¨argner, V., El-Sana, J., 2017. On writer identification for arabic historical manuscripts. International Journal on Document Analysis and Recognition (IJDAR) 20, 173–187.

Chahi, A., Ruichek, Y., Touahni, R., et al., 2020. Local gradient full-scale trans-form patterns based off-line text-independent writer identification. Applied Soft Computing , 106277.

Christlein, V., Bernecker, D., Maier, A., Angelopoulou, E., 2015. Offline writer identification using convolutional neural network activation features, in: German Conference on Pattern Recognition, Springer. pp. 540–552.

Christlein, V., Gropp, M., Fiel, S., Maier, A., 2017. Unsupervised feature learning for writer identification and writer retrieval, in: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), IEEE. pp. 991–997.

Christlein, V., Maier, A., 2018. Encoding cnn activations for writer recognition, in: 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), IEEE. pp. 169–174.

Fiel, S., Sablatnig, R., 2015. Writer identification and retrieval using a convolutional neural network, in: International Conference on Computer Analysis of Images and Patterns, Springer. pp. 26–37.

Hannad, Y., Siddiqi, I., Djeddi, C., El-Kettani, M.E.Y., 2019. Improving Arabic writer identification using score-level fusion of textural descriptors. IET Biometrics 8, 221–229.

He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

He, K., Zhang, X., Ren, S., Sun, J., 2016b. Identity mappings in deep residual networks, in: European conference on computer vision, Springer. pp. 630–645.

He, S., Wiering, M., Schomaker, L., 2015. Junction detection in handwritten documents and its application to writer identification. Pattern Recognition 48, 4036–4048.

Javidi, M., Jampour, M., 2020. A deep learning framework for text-independent writer identification. Engineering Applications of Artificial Intelligence 95, 103912.

Kleber, F., Fiel, S., Diem, M., Sablatnig, R., 2013. Cvl-database: An off-line database for writer retrieval, writer identification and word spotting, in: 2013 12th international conference on document analysis and recognition, IEEE. pp. 560–564.

Mahmoud, S.A., Ahmad, I., Al-Khatib, W.G., Alshayeb, M., Parvez, M.T., M¨argner, V., Fink, G.A., 2014. Khatt: An open arabic offline handwritten text database. Pattern Recognition 47, 1096–1112.

Nosary, A., Heutte, L., Paquet, T., Lecourtier, Y., 1999. Defining writer's invariants to adapt the recognition task, in: Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318), IEEE. pp. 765–768.

Omar, W., Ahmed, s.H., Ibrahim, R., Abdelrahman, E., 2018. Writer Identification : Pattern recognition Course. URL: https://github.com/OmarBazaraa/WriterIdentifier/blob/master/docs/description.pdf.

Rehman, A., Naz, S., Razzak, M.I., Hameed, I.A., 2019. Automatic visual features for writer identification: A deep learning approach. IEEE access 7, 17149–17157.

Semma, A., Hannad, Y., El Kettani, M.E.Y., 2022. Impact of the cnn patch size in the writer identification, in: Networking, Intelligent Systems and Security.Springer, pp. 103–114.

Semma, A., Hannad, Y., Siddiqi, I., Djeddi, C., El Kettani, M.E.Y., 2021. Writer identification using deep learning with fast keypoints and harris corner detector. Expert Systems with Applications 184, 115473.

Xing, L., Qiao, Y., 2016. Deepwriter: A multi-stream deep cnn for text-independent writer identification, in: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), IEEE. pp. 584–589.