Exploiting Multi-scale Fusion, Spatial Attention and Patch Interaction Techniques for Text-Independent Writer Identification

Abhishek Srivastava[1(✉)], Sukalpa Chanda[2], and Umapada Pal[1]

[1] Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India
abhisheksrivastava2397@gmail.com, umapada@isical.ac.in

[2] Department of Computer Science and Communication, Østfold University College, Halden, Norway

Abstract. Text independent writer identification is a challenging problem that differentiates between different handwriting styles to decide the author of the handwritten text. Earlier writer identification relied on handcrafted features to reveal pieces of differences between writers. Recent work with the advent of convolutional neural network, deep learning-based methods have evolved. In this paper, three different deep learning techniques - spatial attention mechanism, multi-scale feature fusion and patch-based CNN were proposed to effectively capture the difference between each writer's handwriting. Our methods are based on the hypothesis that handwritten text images have specific spatial regions which are more unique to a writer's style, multi-scale features propagate characteristic features with respect to individual writers and patch-based features give more general and robust representations that helps to discriminate handwriting from different writers. The proposed methods outperforms various state-of-the-art methodologies on word-level and page-level writer identification methods on three publicly available datasets - CVL, Firemaker, CERUG-EN datasets and give comparable performance on the IAM dataset.

Keywords: Convolutional neural network · Writer identification · MSRF-Net

1 Introduction

Handwriting of an individual is unique and this particular phenomenon has been utilized by forensic handwriting experts for many decades. Handwriting experts today are aided by computer programs which actually can identify an individual on the basis of his handwriting, this technique of identifying a writer from a document image using a software is termed as "Writer Identification". Over the last

C. Wallraven et al. (Eds.): ACPR 2021, LNCS 13189, pp. 203–217, 2022.
https://doi.org/10.1007/978-3-031-02444-3_15

two decades many work has been published on "Writer Identification". But text independent Writer Identification in a limited data scenario is still a challenging task. It has found various applications in forensic [17] and historical [1] document analysis. Before the advent of deep-learning techniques, handcrafted features like gradient, chain-code, allograph, texture etc., were mostly used for writer identification. These feature extraction techniques render discriminating features in predicting the identity of the writer. Deep-learned features have shown impressive performance in various types of image classification problem and "Writer Identification" is also not an exception. Deep learning-based methods in general demand a huge amount of annotated text for proper training. For an application like "Writer Identification" it might not be possible always to procure enough annotated data. Over the last few years, some deep-learning based methods [16], [11] have explored writer identification. To tackle these issues methods which require limited data for identification of the authors are required. Word level writer identification is challenging since very limited information about writer's pattern and technique is available to make a decision. Few deep learning based methodologies are available, for example, He et al. [6] proposed fragment based deep neural network to use convolution neural networks (CNN) for writer identification. CNN were able to learn high level features of the text block and recognize various discriminative features in the word image. CNN's have been previously used to capture local features at the sub-region and character level and combining them for writer identification. Attention based mechanisms are well suited to identify characteristic and discriminative region in an image and enhance the performance of visual recognition based systems. In case of text independent writer identification, the word image is constituted of various segments which capture the unique style of the person's handwriting. Previous deep learning methodologies fail to exploit the contribution of more informative regions of the text image. Recent advances in computer vision has generated interest in fusion of multi-scale features to obtain diverse and rich feature representations [26]. Various resolution scales in handwritten text capture different aspects of a writer's style and structure of his/her handwriting, exploiting multi scale features and their fusion for eventual classification that obtains higher accuracy.

We devise three deep learning techniques to address and exploit those above mentioned facts and compare them to study the impact of different deep learning techniques for writer identification at word and page level. The contributions of our work are as follows:

- We propose a Spatial Attention network (SA-Net) which incorporates spatial attention to enhance relevant and informative feature maps and suppress irrelevant features for effective writer identification performance. Another potential discriminative features in text images are multi-scale features.
- To achieve efficient multi-scale fusion, we customized the MSRF-Net [22] to a classification network suitable for writer identification.
- Inspired by He and Schomaker [6] we propose another patch based CNN named PatchNet which has separate pathways for each patch and uses a

Dual Patch Dense Feature Exchange (DPDFE) block to exchange information across various patches, and making separate writer identity prediction for each patch.
– We attained new benchmarks on CVL, Firemaker and CERUG-EN datasets on word-level and page-level writer identification tasks.

The structure of this paper is as follows. Section 2 provides an overview of related methods and strategies introduced over previous years. Section 3 introduces our proposed methodologies for text independent writer identification. The details of our experiment settings and datasets used are presented in Sect. 4. In Sect. 5 we report the results attained by our methods and their comparison with other state-of-the-art methods on word-level and page-level writer identification tasks, we conclude our paper in Sect. 6.

2 Related Work

The initial works in the field of writer identification were guided by handcrafted feature generation and later with the advent of deep-learning, deep-learning based writer identification methods were proposed. Before the deep-learning methods a wide variety of classifiers like SVM, K-NN, Neural Network were used along with different tools like PCA and LDA to magnify the discriminativeness of various hand crafted features. In the following two subsections we will have a brief discussion on handcrafted features for writer identification followed by deep-learning based approaches.

2.1 Hand Crafted Feature Based Writer Identification

Difference in visual shapes in handwritten characters has been exploited by considering Connected component contour shapes, textural and allograph level features in [2], Schomaker and Bulacu [19] proposed connected-component contours and its probability density function for writer identification. Bulacu et al. [2] exploited to identify the writer. He et al. [8] used Hidden Markov Tree (HMT) in wavelet domain for writer identification. Tan et al. [24] developed a Continuous Character Prototype Distribution feature extraction technique and made classification using Minimum Distance method. Jain and Doermann [9] used K adjacent segments (KAS) to model character contours. The KAS features were clustered using a technique called affinity propagation to build a codebook for the bag of features model. Jain and Doerman [10] captured shape and curvature using contour gradients and used psuedo alphabets as features. Then writer identification was performed using K-Nearest Neighbour classifier. He et al. [7] extracted features such as junction detection, final junction refinement quill and hinge and linked it with a learned codebook to increase performance. Chahi et al. [3] used connected components of the sub-images to extract features referred to as Cross multi-scale Locally encoded Gradient Patterns (CLGP). These CLGP histogram feature vectors were fed into a Nearest Neighbor classifier for writer identification.

2.2 Deep Learning Feature Based Writer Identification

Recently deep learning has drawn attention as convolutional neural networks(CNN) have proven effective in extracting discriminative features from handwritten texts. Initially, Fiel and Sablatnig [4] trained a CNN classifier and used the output of second last fully connected layer as features to perform nearest neighbour classification. Tang and Wu [25] performed data augmentation on handwritten documents to allow training of a deep CNN. The CNN is then used for feature extraction and Joint Bayesian technique is used for writer identification. DeepWriter [27] used multi-stream CNNs to learn diverse representation of text images. Rehman et al. [18] augmented text images using various techniques like contour, negatives and sharpness using text line images. Multiple patches were generated from the text images and fed into an architecture similar to AlexNet pretrained on Imagenet to generate features. These features were classified using a support vector machine classifier. Keglevic et al. [12] designed a triplet network to calculate similarity measure between different patches, and trained it by maximizing inter-class distance and minimizing intra-class distance. Global features of document is then calculated by aggregating vector of local image patch descriptors. Nguyen et al. [16] generated tuples of text images by randomly sampling characters as input for their CNNs. They trained CNNs to extract sub-region, character and global level features and effectively aggregated them to predict the identity of writer. He et al. [6] designed FragNet which first builds a global feature pyramid and then a local fragment pathway which leverages fragments of global feature pyramids to make separate writer identity prediction for each writer. Javidi and Jampour [11] quantified the thickness of handwritten documents using handwriting thickness descriptors(HTD). Resnet-18 was used to extract features from the text images and they were combined with HTDs for classification. In this work, we propose three different deep learning models which uses different architecture based components suitable for identifying and capturing various aspects of a writer's technique and style.

3 Methodology

In this section we discuss about our proposed approaches. We have developed the following methods.

1. We develop a spatial attention based mechanism for identifying various author specific features of the word image. The characteristic style and features of the word occupy a very limited region in the word image. Generating a spatial attention map can help enhancing the features exploited from such regions. This serves as the basis of our spatial attention network(SA-Net) for writer identification.
2. Multi-Scale features can capture information of varying spatial and receptive field sizes. The word images can have key discriminative features of diverse scale sizes which convey various characteristic features of writer. Thus, it is advantageous to design a writer identification system which effectively

leverage multi-scale features while predicting the identity of our writer. We convert our MSRF-Net [22] to MSRF(Multi-Scale Residual Fusion) Classification network to effectively fuse multi-scale features and leverage them into predicting the identity of the writer more accurately.

3. Inspired by FragNet [6] we develop a patch based convolutional neural network called PatchNet. We use a different stream for each patch used and densely exchange various patch features using our Dual Patch Dense Feature Exchange (DPDFE) blocks. Each local patch predictions are then averaged over to make our final writer identity prediction.

This section is structured as follows. In Sect. 3.1 we describe our spatial attention network(SA-Net), Sect. 3.2 describes how we amend our MSRF network to a classification network and exploit multi-scale features of word images to develop a more accurate system for writer identification. Finally, in Sect. 3.3 we describe our proposed Patch-Net.

3.1 Spatial Attention Network

In this section we introduce our spatial network for writer identification. Specific regions of word images have characteristic textural and shape information which is unique to a specific writer. Characters in the word images also have a unique style in the manner they are written. To allow the identification and recognition of these regions we develop a spatial attention mechanism. Let I_w denote word images where ($I_w \in R^{W \times H}$). The framework resembles a VGG-style network [21] where each I_w is initially processed by a convolutional block. Each convolutional block has 2 consecutive convolutional layers with 3×3 kernel size followed by batch-normalization and ReLU activation. This is described in Eq. 1 where X denotes the input tensor.

$$X_{conv} = ReLU(BN(Conv(Conv(X)))) \tag{1}$$

The convolutional blocks are followed by a spatial attention unit (see Fig. 1). This block comprises of two convolutional layers followed by a sigmoid activation function which calculates attention coefficient for each spatial location in the feature maps (see Eq. 2). These attention maps are denoted as A_{att}. We multiply these attention maps described in Eq. 3 to suppress regions which are non relevant and enhance the spatial location of relevant and important feature maps.

$$A_{att} = \sigma(Conv(X_{conv})) \tag{2}$$

$$X_{spa} = X_{conv} \otimes A_{att} \tag{3}$$

X_{spa} denotes the spatial attention enhanced feature maps which are then halved using max pooling. The number of feature maps in a convolutional and spatial attention unit are set to [64, 128, 256, 512] respectively. We use adaptive average pooling at the last layer and a fully connected layer to make the final prediction or writer identity. For page level prediction we make predictions for all word

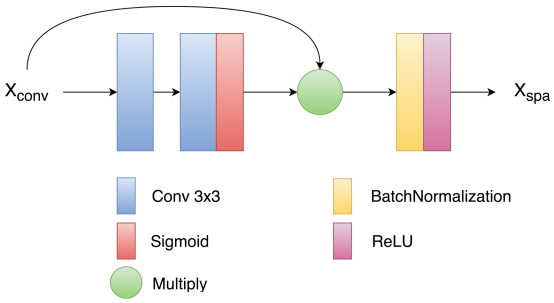

Fig. 1. Architecture of spatial attention unit employed in SA-Net

images in the page and average over them as described in Eq. 4, where N are the total word images in a page and ID_{page} represents the identity of the writer.

$$ID_{page} = \frac{1}{N} \sum_{n=1}^{n=N} P(I_w^n) \tag{4}$$

3.2 MSRF Classification Network

Multi-scale feature exchange has been studied in past years in the field of computer vision. Fusion of multi-scale features result in diverse representations consequently generating richer and accurate feature maps. The word images are also structured such that different scale features capture varying writer characteristics. We use this motivation to convert our MSRF-Net [22] into a classification network (see Fig. 2). Dual scale dense fusion (DSDF) blocks used in MSRF-Net serves the purpose of fusion of two different scaled features. The dense nature of the blocks allows features of various receptive fields to be generated and the residual connections allow relevant high-level and low-level features to be maintained while making final predictions. We modify the MSRF-Sub-network to translate it into a classification head. Contrary to the MSRF sub-network which aimed to fuse and exchange multi-scale features across all scales, we ensure that all different scaled representations are able to flow in the last scale level of the classification network (see Fig. 3.2). To improve gradient flow, we allow last scale level of the MSRF classification network to make prediction before and after each DSDF block in the last scale level as shown in Fig. 2. We use an adaptive pooling module and a fully connected layer in succession to make predict writer of the word image. Finally we average over all the predictions of to make our final predictions as shown in Eq. 5, where C represents the number of classification layers in the MSRF classification network and ID_{word} represents the identity of the writer for the word image.

$$ID_{word} = \frac{1}{C} \sum_{k=1}^{k=C} P(I_w^k) \tag{5}$$

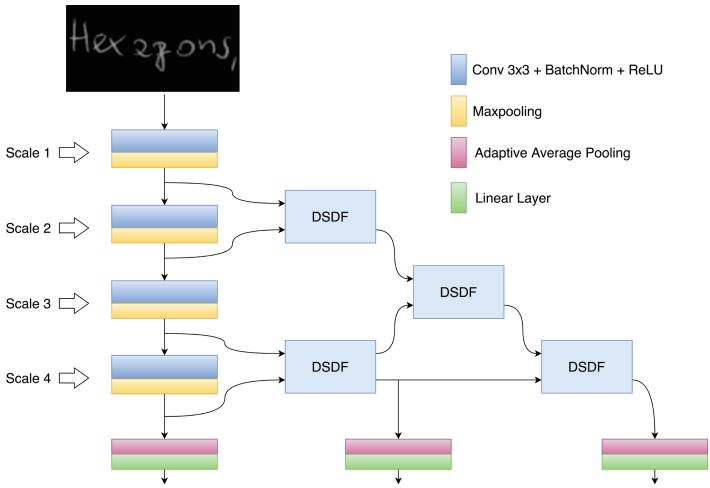

Fig. 2. Architecture of multi-scale residual fusion classification network

In order to make page level prediction, we again average over all word level predictions contained in the page as shown in Eq. 6.

$$ID_{page} = \frac{1}{N} \sum_{n=1}^{n=N} P(I_w^n) \tag{6}$$

3.3 PatchNet

Inspired by FragNet [6] we develop a patch based classification network (see Fig. 4). The $\mathcal{I}_w$ is divided into patches of size 64×64. We generate 5 patches from the original 64×128 $\mathcal{I}_w$ and make different pathways for each patch. Each path has a initial convolutional unit of two successive convolutional layers, batch-normalization and ReLU activation. Which is followed by a maxpooling layer to reduce the spatial dimension by a factor of 2. To exchange information between two patches we design dual patch feature exchange (DPDFE) block. The entire convolutional unit, DPDFE blocks and max-pooling sequence is repeated 4 times to make patch level predictions. We also use a global prediction pathway which has a similar architecture as SA-Net without the spatial attention unit. Each patch level predictions and global prediction are averaged to make the final prediction. Page level predictions are made according to Eq. 7.

$$ID_{page} = \frac{1}{N} \sum_{n=1}^{n=N} P(I_w^n) \tag{7}$$

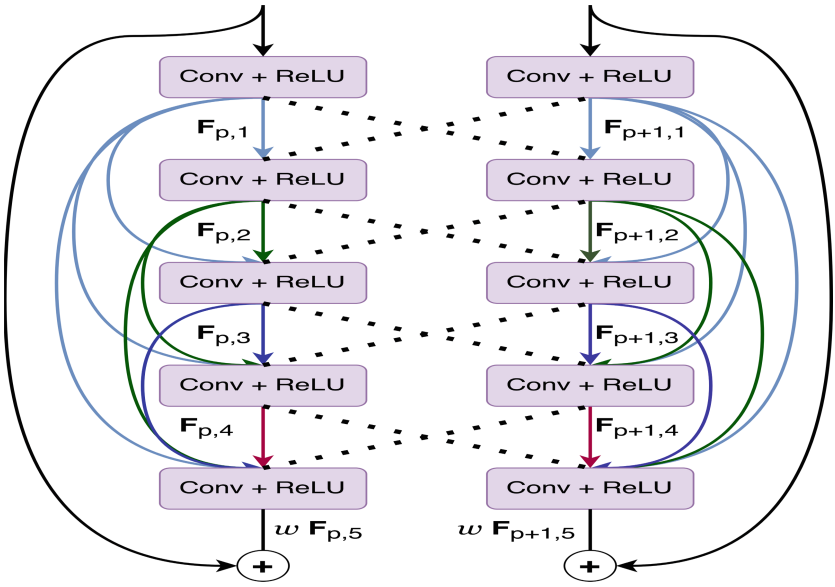

Fig. 3. Architecture of dual patch dense feature exchange block (dotted lines represent features incoming from parallel patch stream)

Dual Patch Dense Feature Exchange Blocks. In this section, we describe the structure of our dual patch dense feature exchange blocks. Let two successive patches be denoted by $\mathcal{I}_p$ and $\mathcal{I}_{p+1}$. The feature maps generated by convolutional unit of each patch stream be denoted by $M_{p,l}$, where l denotes how many layers of DPDFE blocks the feature maps have been processed by and initially $l = 0$. The DPDFE blocks are residual dense blocks which takes feature maps of two different patches and process each of them using two different densely connected streams (see Fig. 3). Each stream has 5 densely connected convolutional layers, Let the output of each such layer be $F_{p,c}$ where p denotes which patch is being processed and c denotes which convolutional layer has processed the feature maps in the dense stream. After each convolutional layer in the dense stream, the two different patch streams exchanges features as described in Eq. 8 ($M_{p,l}$ and $F_{p,0}$ are the same).

$$M_{p,l+1} = F_{p,c} \oplus F_{p,c-1} \oplus F_{p,c-2} \oplus \cdots \oplus F_{p+1,0} \tag{8}$$

We again scale the output features of DPDFE blocks by a factor of $w = 0.4$ to avoid instability [14,23] and add it back to the input of the respective DPDFE block as shown in Eq. 9.

$$M_{p,l+1} = w \times M_{p,l+1} + M_{p,l} \tag{9}$$

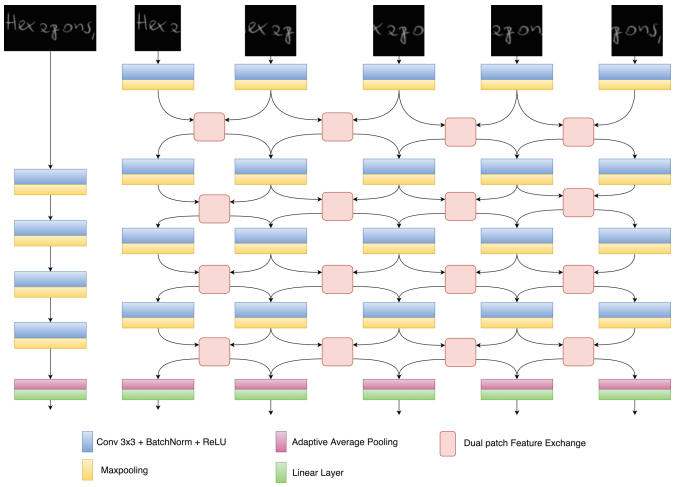

Fig. 4. Architecture of our proposed PatchNet

4 Experiments

In this section we describe the writer identification datasets used for our experiments. We also describe the implementation details of our three deep learning-based writer identification methods. We use the training and testing split used by He and Schomaker in FragNet [6]. It is ensured that each word image from each page either occurs in the training split or in the testing split, which makes the methods suitable for both word-level and page-level writer identification tasks.

4.1 Datasets

We benchmark our methods on four publicly available datasets namely: CERUG-EN [7], Firemaker [20], CVL [13] and IAM [15]

1. CERUG-EN [7] has 105 documents, predominantly from Chinese students. There are two paragraphs in English where one paragraph is used for training and another is used for testing. Since word images are not provided separately we use the roughly segmented word images provided by He and Schomaker in the publicly released code of FragNet.
2. Firemaker [20] has 250 different writers where each writer writes four pages. First page is used for training and the fourth page is used for testing.
3. CVL [13] has 310 writers. Each person has written five pages of text with 27 writers contributing seven pages. First three pages are used for training and the rest are used for testing.
4. IAM [15] has 610 different writers contributing varying amount of text. When more than one page is available for a writer, we choose one page for training and rest for testing. When only one page is available the lines

are divided into training and testing subsets. The word images are publicly available.

4.2 Implementation Details

We pre-process the images to 64×128 while maintaining aspect ratio. To avoid distortion white pixel padding is done. We use a batch size of 16 and train all methods for 50 epochs. We follow the training setting of FragNet to ensure fair and complete comparison. Adam optimizer is used with initial learning rate 0.0001 and a weight decay of $1e{-}4$. We decay the learning rate by a factor of 0.5 after every 10 epochs. The feature maps in each level of classification networks are [64, 128, 256, 512] for all 3 methods. The FLOPs of WordImgNet are 1.05G, whereas FLOPs of FragNet-64, FragNet-32, Frag-Net-16 are approximately 7.14G, 7.41G and 3.90G, respectively. The FLOPs of MSRF-Classification network, SA-Net and PatchNet, and ResNet18+HTDs are around 5.5G, 4.10G, and 7.65G respectively. The proposed models are available at https://github.com/NoviceMAn-prog/SA-Net-MSRF-CNet-and-PatchNet-for-Writer-Identification.

5 Results and Discussion

In this section we will compare our MSRF Classification Network, SA-Net and PatchNet with other published state-of-the-art methods on word-level and page-level writer identification task. It is worth mentioning here that there exists many deep-learning methods for writer identification, and even though those experiments were conducted on public datasets, lack of publicly available source code of those published methods creates hindrance towards a fair comparison. Additionally we chose methods that were designed for word level writer identification. Keeping those factors in mind, we could compare our methods with [6,11] as those methods have released their code. We train and test those methods using the same set of training and test images as we did for our proposed methods for an unbiased comparison. We establish new state-of-the-art writer identification results on three benchmark datasets - CVL, Firemaker and CERUG-EN datasets. Section 5.1 describes various other state-of-the-art deep learning methods we select for comparison with our methods on word-level and page-level writer identification tasks. In Sect. 5.1 we provide quantitative comparison of our methods with other baselines on word level writer identification tasks. Section 5.1 provides writer identification results on page level writer identification task.

5.1 Comparison with Other Published Methods

To provide exhaustive comparison of our methods with other baselines we select ResNet18 [5], ResNet18 conjugated with handwriting thickness descriptors(HTD) [11], WordImgNet [6] and FragNet-q [6] where q represents the q $\times$ q fragment size.

1. ResNet18 is a standard computer vision classification baseline.
2. FragNet is fragmentation based CNN with two streams. First global feature pyramid used for extraction of features. Second stream is a fragment pathway to process fragments of the original image and receive fragments from the global feature pyramid to make prediction. Each fragment has its own prediction and the final prediction is made by averaging over all local fragment predictions. We use FragNet-64, FragNet-32 and FragNet-16 for our experiments.
3. WordImgNet is designed such that the entire image is fed into a CNN framework identical to the fragment pathway to make a single global prediction.
4. ResNet18 + HTDs, ResNet18 captures high level features of the input text image. HTDs are spatial descriptor that analyze a writer's handwriting thickness depending upon factors like pressure of pen, unique style. The features extracted by ResNet18 are concatenated with HTDs which serves as additional discriminative features.

Result on Word Level Writer Identification. In this section we compare the Top-1 and Top-5 writer identification accuracy of our proposed methods with other state-of-the-art methods at word level. In Table 1 we present the detailed comparison of all methods on all four datasets. In CERUG-EN [7] we observe that SA-Net gives the best performance outperforming the previous state-of-the-art performer FragNet-64 by 4.7% accuracy in Top-1 and by 1.5% in Top-5. We can notice that along with SA-Net, MSRF Classification also beats FragNet-64 in performance in both Top-1 and Top-5 accuracy, PatchNet is comparable to it in performance. For Firemaker dataset, our SA-Net again gives the best performance gaining 3.2% and 0.8% in Top-1 and Top-5 accuracy over FragNet-64. MSRF-Net gives the second best performance gaining 2.2% and 0.8% in Top-1 and Top-5 accuracy over FragNet. In CVL dataset writer identification problem, MSRF Classification obtains the best performance achieving 91.4% Top-1 and 97.6% Top-5 accuracy outperforming FragNet-64 by 1.2% and 0.1% in Top-1 and Top-5 accuracy. SA-Net also performs better than FragNet-64 achieving 90.7% and 97.4% Top-1 and Top-5 accuracy. PatchNet gives a comparable 86.1% Top-1 and 96.3% Top-1 and Top-5 accuracy respectively. On the writer identification task on IAM dataset, FragNet-64 reports the best Top-1 accuracy of 85.1% while the best Top-5 accuracy is shared between FragNet-64 and MSRF Classification network, both achieving 95%. The superior performance of SA-Net on two datasets i.e. Firemaker and CERUG-EN shows the potential of spatial attention mechanism's ability to extract relevant differentiating elements of a writer's handwriting. The multi scale features obtained and fused in MSRF classification network obtains the highest Top-1 and Top-5 accuracy on CVL dataset. This displays the capacity of multi-scale features to identify the characteristics of writer's style in his handwriting. Although PatchNet outperforms previous state-of-the-art methods on only one dataset, it shows the potential of patch or fragment based networks for writer identification.

Table 1. Result comparison on word level writer identification

Method	IAM		CVL		Firemaker		CERUG-EN	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
ResNet18 [5]	83.2	94.3	88.5	96.7	63.9	86.4	70.6	94.0
ResNet18+HTD [11]	76.9	91.6	85.1	95.6	60.7	82.6	70.1	91.8
WordImgNet [6]	81.8	94.1	88.6	96.8	67.9	88.1	77.3	96.4
FragNet-16 [6]	79.8	93.3	89.0	97.2	59.6	83.2	60.6	90.3
FragNet-32 [6]	83.6	94.8	89.0	97.3	65.0	86.8	62.3	90.1
FragNet-64 [6]	**85.1**	**95.0**	90.2	97.5	69.0	88.5	77.5	95.6
Patch (proposed)	80.2	93.5	86.1	96.2	62.4	84.9	77.1	96.5
SA-Net (proposed)	83.4	94.6	90.7	97.4	**72.2**	**89.3**	**82.2**	**97.1**
MSRF-Net (proposed)	84.6	**95.0**	**91.4**	**97.6**	71.2	**89.3**	79.6	96.8

Table 2. Result comparison on page level writer identification

Method	IAM		CVL		Firemaker		CERUG-EN	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
ResnNet18+HTD [11]	95.2	98.0	98.3	98.3	98.0	99.2	98.0	**100**
WordImgNet [6]	95.8	98.0	98.8	99.4	97.6	98.8	97.1	100
FragNet-16 [6]	94.2	97.4	98.5	99.4	92.8	98.0	79.0	97.1
FragNet-32 [6]	95.3	98.0	98.6	99.4	96.0	99.2	84.7	97.1
FragNet-64 [6]	**96.3**	98.0	99.1	99.4	97.6	**99.6**	98.1	**100**
Patch (proposed)	93.6	96.9	99.0	99.3	95.6	98.4	98.1	**100**
SA-Net (proposed)	94.7	**98.2**	**99.4**	99.4	**98.0**	**99.6**	**99.1**	**100**
MSRF-Net (proposed)	94.8	98.1	**99.4**	**99.6**	97.2	99.2	98.1	**100**

Result on Page Level Writer Identification. In this section we provide the quantitative analysis of the comparison between our proposed methods and other state-of-the-art methods on page level writer identification. In the Firemaker dataset writer identification task, the proposed SA-Net outperforms FragNet-64 by 0.4% in Top-1 accuracy. SA-Net reports a Top-5 accuracy of 99.6% which is equal to the Top-5 accuracy to FragNet-64. SA-Net reports the highest Top-1 page level accuracy on CERUG-EN of 99.1%. Additionally, FragNet-64, PatchNet, SA-Net, MSRF classification network and ResNet18+HTDs all tie for the best Top-5 performance of 100% on CERUG-EN. MSRF-Net and SA-Net both outperforms FragNet-64 by 0.3% on Top-1 page level accuracy on the CVL dataset. MSRF-Net report the highest 99.6% Top-5 page level accuracy while FragNet-64 and SA-Net gives 99.4% Top-5 page level accuracy. For IAM dataset, FragNet-64 reports the highest 96.3% Top-1 page level accuracy. SA-Net and MSRF-Net reports the first and second best Top-5 accuracy of 98.2% and 98.1%, respectively. We notice that again SA-Net and MSRF-Net attains new

benchmarks on IAM, CVL, Firemaker and CERUG-EN datasets, exhibiting the potential of amplified features on the basis of spatial attention and multi-scale features (Table 2).

6 Conclusion

In this paper we proposed three deep learning based solutions for text-independent writer identification. Our proposed SA-Net was able to identify and enhance the feature flow from spatial regions more relevant and significant in determining the identity of the writer. MSRF Classification network performed multi-scale feature fusion to gather more diverse representations consisting of features having varying receptive fields. The residual nature of the dual scale dense fusion (DSDF) blocks allow an effective combination of high- and low-level feature representations to be available at the disposal of final classification layer to make more accurate predictions. On the other-hand, PatchNet allows effective feature exchange between different patch streams to make more robust predictions. Our methods were able to outperform previous state-of-the-art methods for word-level and page-level writer identification on CVL, Firemaker and CERUG-EN datasets, while giving comparable performance on the IAM dataset. We show that developing deep learning based mechanisms exploiting spatially relevant regions and multi scale features is also a viable option to increase performance of writer identification systems.

Acknowledgement. This is a collaborative research work between Indian Statistical Institute, Kolkata, India and Østfold University College, Halden, Norway. The experiments in this paper were performed on a high performance computing platform "Experimental Infrastructure for Exploration of Exascale Computing" (eX3), which is funded by the Research Council of Norway.

References

1. Brink, A., Smit, J., Bulacu, M., Schomaker, L.: Writer identification using directional ink-trace width measurements. Pattern Recogn. **45**(1), 162–171 (2012)
2. Bulacu, M., Schomaker, L.: Text-independent writer identification and verification using textural and allographic features. IEEE Trans. Pattern Anal. Mach. Intell. **29**(4), 701–717 (2007)
3. Chahi, A., Ruichek, Y., Touahni, R., et al.: Cross multi-scale locally encoded gradient patterns for off-line text-independent writer identification. Eng. Appl. Artif. Intell. **89**, 103459 (2020)
4. Fiel, S., Sablatnig, R.: Writer identification and retrieval using a convolutional neural network. In: Azzopardi, G., Petkov, N. (eds.) CAIP 2015. LNCS, vol. 9257, pp. 26–37. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23117-4_3
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
6. He, S., Schomaker, L.: FragNet: writer identification using deep fragment networks. IEEE Trans. Inf. Forensics Secur. **15**, 3013–3022 (2020)

7. He, S., Wiering, M., Schomaker, L.: Junction detection in handwritten documents and its application to writer identification. Pattern Recogn. **48**(12), 4036–4048 (2015)
8. He, Z., You, X., Tang, Y.Y.: Writer identification of Chinese handwriting documents using hidden Markov tree model. Pattern Recogn. **41**(4), 1295–1307 (2008)
9. Jain, R., Doermann, D.: Offline writer identification using K-adjacent segments. In: 2011 International Conference on Document Analysis and Recognition, pp. 769–773. IEEE (2011)
10. Jain, R., Doermann, D.: Writer identification using an alphabet of contour gradient descriptors. In: 2013 12th International Conference on Document Analysis and Recognition, pp. 550–554. IEEE (2013)
11. Javidi, M., Jampour, M.: A deep learning framework for text-independent writer identification. Eng. Appl. Artif. Intell. **95**, 103912 (2020)
12. Keglevic, M., Fiel, S., Sablatnig, R.: Learning features for writer retrieval and identification using triplet CNNs. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 211–216. IEEE (2018)
13. Kleber, F., Fiel, S., Diem, M., Sablatnig, R.: CVL-database: an off-line database for writer retrieval, writer identification and word spotting. In: 2013 12th International Conference on Document Analysis and Recognition, pp. 560–564. IEEE (2013)
14. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 136–144 (2017)
15. Marti, U.V., Bunke, H.: The IAM-database: an English sentence database for offline handwriting recognition. Int. J. Doc. Anal. Recogn. **5**(1), 39–46 (2002)
16. Nguyen, H.T., Nguyen, C.T., Ino, T., Indurkhya, B., Nakagawa, M.: Text-independent writer identification using convolutional neural network. Pattern Recogn. Lett. **121**, 104–112 (2019)
17. Pervouchine, V., Leedham, G.: Extraction and analysis of forensic document examiner features used for writer identification. Pattern Recogn. **40**(3), 1004–1013 (2007)
18. Rehman, A., Naz, S., Razzak, M.I., Hameed, I.A.: Automatic visual features for writer identification: a deep learning approach. IEEE Access **7**, 17149–17157 (2019)
19. Schomaker, L., Bulacu, M.: Automatic writer identification using connected-component contours and edge-based features of uppercase western script. IEEE Trans. Pattern Anal. Mach. Intell. **26**(6), 787–798 (2004)
20. Schomaker, L., Vuurpijl, L., Schomaker, L.: Forensic writer identification: a benchmark data set and a comparison of two systems (2000)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
22. Srivastava, A., et al.: MSRF-Net: a multi-scale residual fusion network for biomedical image segmentation. arXiv preprint arXiv:2105.07451 (2021)
23. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. In: Proceedings of AAAI Conference on Artificial Intelligence, vol. 31 (2017)
24. Tan, G.X., Viard-Gaudin, C., Kot, A.C.: Automatic writer identification framework for online handwritten documents using character prototypes. Pattern Recognit. **42**(12), 3313–3323 (2009)
25. Tang, Y., Wu, X.: Text-independent writer identification via CNN features and joint Bayesian. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 566–571. IEEE (2016)

26. Wang, J., et al.: Deep high-resolution representation learning for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **43**, 3349–3364 (2020)
27. Xing, L., Qiao, Y.: DeepWriter: a multi-stream deep CNN for text-independent writer identification. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 584–589. IEEE (2016)