IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 21, 2020, accepted November 13, 2020, date of publication November 18, 2020,
date of current version December 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039003

HDSR-Flor: A Robust End-to-End System to Solve the Handwritten Digit String Recognition Problem in Real Complex Scenarios

ARTHUR FLOR DE SOUSA NETO[1], BYRON LEITE DANTAS BEZERRA[1], (Member, IEEE), ESTANISLAU BAPTISTA LIMA[1], AND ALEJANDRO HÉCTOR TOSELLI[2]

[1]Polytechnic School, Universidade de Pernambuco, Recife 50100-010, Brazil
[2]Pattern Recognition and Human Language Research Center (PRHLT), Universitat Politècnica de València, 46022 València, Spain

Corresponding author: Byron Leite Dantas Bezerra (byron.leite@upe.br)

This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES)—Finance Code 001, and in part by the National Council of Scientific and Technological Development [Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)] under Grant 315251/2018-2.

ABSTRACT Automatic handwriting recognition systems are of interest for academic research fields and for commercial applications. Recent advances in deep learning techniques have shown dramatic improvement in relation to classic computer vision problems, especially in Handwritten Text Recognition (HTR). However, several approaches try to solve the problem of deep learning applied to Handwritten Digit String Recognition (HDSR), where it has to deal with the low number of trainable data, while learning to ignore any writing symbol around the digits (noise). In this context, we present a new optical model architecture (Gated-CNN-BGRU), based on HTR workflow, applied to HDSR. The International Conference on Frontiers of Handwriting Recognition (ICFHR) 2014 competition on HDSR were used as baselines to evaluate the effectiveness of our proposal, whose metrics, datasets and recognition methods were adopted for fair comparison. Furthermore, we also use a private dataset (Brazilian Bank Check - Courtesy Amount Recognition), and 11 different approaches from the state-of-the-art in HDSR, as well as 2 optical models from the state-of-the-art in HTR. Finally, the proposed optical model demonstrated robustness even with low data volume (126 trainable data, for example), surpassing the results of existing methods with an average precision of 96.50%, which is equivalent to an average percentage of improvement of 3.74 points compared to the state-of-the-art in HDSR. In addition, the result stands out in the competition's CVL HDS set, where the proposed optical model achieved a precision of 93.54%, while the best result so far had been from Beijing group (from the competition itself), with 85.29%.

INDEX TERMS Deep learning, handwritten digit string recognition, handwritten text recognition, neural network, string classification.

I. INTRODUCTION

Handwritten Digit String Recognition (HDSR) has attracted intense attention in recent years as an academic research topic, due to its vast applications in industry [1]. Moreover, automatic HDSR systems are very helpful in offline handwriting recognition applications, where the recognition is performed through images. Therefore, developing effective algorithms to HDSR has become the target in several application domains, such as form processing, bank check processing and postal code identification [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Shorif Uddin.

The HDSR task has been a subject of study for many years, and a variety of methods and algorithms have been proposed to improve recognition precision [3]. Traditional approaches are based on over-segmentation strategy, to then classify the detected digits. However, this kind of methods faced many problems in practice, such as various styles of handwriting, connected characters and background noise [4]. Recently, end-to-end systems are used to deal with these problems, where neural network models extract features from images and recognize the digits, without prior detection [5]–[8].

In the last years, advances in deep learning have revolutionized the field of computer vision, and efficient techniques are proving to be very impactful for many tasks, such as image classification, object detection and pattern recognition [2]. Handwritten Text Recognition (HTR) research field, for example, has demonstrated a dramatic advance through Convolutional Recurrent Neural Networks (CRNN) [9]–[12] as optical models, in which the convolutional layers are responsible for extracting the features from text images, while the recurrent layers propagate and decode the extracted features through a Connectionist Temporal Classification (CTC), resulting in the final text.

On the other hand, the HDSR research field, even with a less complex objective compared to HTR systems (restricted to digits only), still has to deal with an unlimited number of handwriting styles, sizes, variations of digit patterns, and noises [13]. In addition, the very low volume datasets for training optical models represent a great challenge for deep learning techniques, in which the vanishing gradient problem increases according to the complexity of the model (how deep) [14]. Thus, several works were developed to solve this problem and achieve better results, such as models composed by K-sparse Auto Encoders [5], CRNN [15], ResNet with RNN-CTC [6], [7], CNN [16], ResNet-41 [17], and YoLo [8].

In this context, we propose a new Gated Convolutional Recurrent Neural Network using CTC (Gated-CRNN-CTC) as optical model for HDSR. The proposed architecture is motivated by the promising results achieved by Puigcerver [10] and Bluche *et al.* [9] in handwritten text lines recognition, and its main objectives are:

- Improve state-of-the-art recognition rate in the HDSR research field;
- Able to handle very low volumes of data in various HDSR scenarios;
- Offer a low complexity model through the low number of trainable parameters (thousands).

A variety of experiments were carried out to validate the proposed optical model, as well as the Bluche and Puigcerver models, in which we used the results from International Conference on Frontiers of Handwriting Recognition (ICFHR) 2014 competition on HDSR [13] as baseline.

We used the metrics and 3 datasets proposed by the competition, and in addition, a private dataset for analysis. Within each dataset, we also still performed different arrangements in training and validation sets for the optical model, in which it varied from 90%/10% to 10%/90%. Thus, we tested the models in several volumes of data and used the best result for the final analysis, which showed that the proposed optical model outperformed the previous ones. Finally, an open source implementation is also provided[1].

The remaining of this paper is structured as follows. In section II, reference approaches in the literature are described. Then, in section III, the proposed optical model is presented. In section IV, the experimental methodology are detailed. In section V, the experimental results obtained in each dataset are discussed. Finally, section VI presents the conclusions.

[1] https://github.com/arthurflor23/handwritten-text-recognition

II. RELATED WORKS

There are several approaches to deal with digit recognition through Handwritten Digit String Recognition (HDSR) systems. Thus, in the following subsections we present the traditional methods from ICFHR 2014 competition on HDSR [13], the current state-of-the-art in HDSR and also the current state-of-the-art in Handwritten Text Recognition (HTR).

A. METHODS FROM ICFHR 2014 COMPETITION

Initially, five research groups participated in the competition, in which one group presented two different approaches, totaling six methods for recognition of handwritten digit string.

The Beijing method [13] first used a neural network to classify the images as simple or complex. To clean and improve the image, Otsu's binarization [18] was used for simple images, while two Multilayer Perceptron (MLP) [19] followed by Sauvola's binarization [20] were applied for complex images. The next step is responsible for correcting the writing angles and normalizing the image, where it generates the candidate character patterns. The candidate patterns are classified using polynomial classifier with gradient direction histogram feature, and beam search algorithm [21] is used to find the path with minimum cost or maximum score.

The Pernambuco [13] method preprocessed the images to detect non-digit structures [22], and in the case of gray level images, the algorithm chooses the best threshold level using an MLP neural network [19]. Finally, a classifier composed of Multi-Dimensional Recurrent Neural Network (MDRNN) [23] and Support Vector Machine (SVM) [24] is used.

In the Shanghai [13] method, digits are segmented based on adaptive binarization and analysis of connected components. Finally, directional characteristics are used in classifiers, such as SVM [25] and Self-Organizing Map (SOM) [26], and their recognition results for each digit image are ordered on confidences.

The Singapore method [13], preprocesses the images to extract vectors of sequential features using Histogram of Oriented Gradients (HOG) [27]. Finally, a Recurrent Neural Network (RNN) [28] is applied for recognition.

Finally, Tébessa I [13] method uses a direct segmentation strategy based on the oriented sliding window [29]. Recognition is carried out by projecting the vertical histogram (single digit image) or by using the Radon transform [30] to obtain the orientation angle of the sliding window, combined with SVM [25]. The second method proposed by the same group, Tébessa II [13], uses the same description as above, but with multiple orientation angles of the sliding window [29].

B. STATE-OF-THE-ART IN HDSR

In recent years, considerable efforts have been made to employ deep learning techniques within HDSR, especially the Convolutional Neural Networks (CNNs) and its advanced automatic feature extractor [31]. In this way, Saabni [5] proposed an algorithm that trains k-Sparse Auto Encoders and stacks its hidden layers as pre-trained hidden layers within a deep neural network, in which he used sliding window technique to achieve high recognition rates.

Later, Zhan *et al.* [6] presented the combination of the ResNet [32] model with Bidirectional Long Short-Term Memory (BLSTM) layers and CTC (RNN-CTC architecture). Indeed, the high performance of ResNet, to extract features from images, resulted in a major improvement for HDSR, especially when dealing with long string labeling.

Besides, Wang *et al.* [33] developed a sequence labeling convolutional network for handwritten string recognition, in which Spatial Pyramid Pooling (SPP) is utilized to handle arbitrary string length. Moreover, a flexible pooling strategy, called FSPP, is used to adapt the network to the recognition of strings, avoiding the digit segmentation in the process.

On the other hand, Hochuli *et al.* [34] verified that sequences of handwritten digits, isolated or in contact, can be recognized only through CNNs and also without a segmentation step. Even so, the approach used four CNN classifiers trained on synthetic data for two specific tasks, string length classification and digit classification. In addition, the CNN system presented is still limited to dealing with background noises.

Furthermore, a different model approach to string recognition was presented by Zhan *et al.* [16], where it uses an entirely CNN architecture and avoids the disadvantages brought by RNNs. Where next, Zhan *et al.* [35] also presented a CRNN architecture combined with residual connections (called Residual CRNN, or Res-CRNN), bringing a universal model for sequence labeling tasks. At the same time as Xu *et al.* [7], slightly improved Zhan's previous work [6], which uses Bidirectional Gated Recurrent Unit (BGRU) layers combined with ResNet.

More recently, Lupinski *et al.* [36] built a new approach using Attention Mechanism in a Seq2Seq based model for HDSR, through one CNN and two RNNs (Encoder-Decoder). Beyond that, Ma *et al.* [37] developed an end-to-end system as a simplified target detector (character level) for recognition of digits in low resolution images through bounding boxes.

Finally, Xu *et al.* [17] used ResNet-41 to resolve the HDSR problem in low resolution based on object detection, but performed badly in scenarios of lack of data diversity. On the other hand, Hochuli *et al.* [8] brought an end-to-end approach using YoLo-based model [38], [39] as an automatic detection and recognition process. In this system, the preprocessing and segmentation of the images are eliminated, and in addition, it handles well in the recognition of long sequences.

C. STATE-OF-THE-ART IN HTR

The Handwritten Text Recognition (HTR) research field has the challenge of offering robust systems to deal with different scenarios in production environments, which usually have the handwritten sequence surrounded by background noise, or even combined with other irrelevant information [2], [40]. Therefore, state-of-the-art optical models in HTR fit with the objective of the HDSR research field presented above.

Historically, HTR systems have been formulated as a sequence matching problem: a sequence of features extracted from the input data is matched to an output sequence composed of text characters. The transcription task was initially performed by segmentation and graph search algorithm [41], followed by Hidden Markov Models (HMMs) [42], [43]. However, HMMs fail to make use of context information in a text sequence, due to the Markovian assumption that each observation depends only on the current state.

In the last few years, CNNs have been used to overcome the limitations of HMMs [44], [45], as well as the BLSTM layers has been widely used for the propagation of features, due to its low complexity and high performance achieved [9], [10]. Finally, the output of the recurrent layers is passed to CTC [46], which it uses to calculate the loss value (training mode) and decode it into final text (inference mode).

In this context, simpler optical models were developed in order to improve performance compared to traditional ones. Thus, Puigcerver [10] presented the CNN-BLSTM architecture as a way to minimize the computational cost, and also to achieve better results over state-of-the-art models that used Multidimensional LSTM (MDLSTM) layers.

Likewise, Bluche *et al.* [9] introduced an Gated-CNN-BLSTM architecture, where it uses the Gated mechanism in the convolutional layers as a way to extract more relevant features, allowing a parameter reduction of the optical model significantly and reaching impressive results.

Finally, Neto *et al.* [11], [12] demonstrated improvement in the HTR area through the Gated-CNN-BGRU architecture, where it uses the Gated mechanism in the convolutional layers as a way to extract more relevant features, and BGRU in the recurrent layers. Thus, introduced a new optical model of low computational cost and high performance in recognition, even in scenarios of low data volume. In addition, the optical model proved to be robust in different datasets, surpassing the results of previous models.

III. PROPOSED OPTICAL MODEL

The proposed optical model is a three-step pipeline based on HTR systems [9]–[12], applied to the HDSR context. Thus, the three steps are composed by capture, propagation, and decoding of features, so: *(i)* the CNNs layers extract the most relevant features from the input images through the Gated mechanism proposed by Dauphin *et al.* [47]; *(ii)* the BGRU layers propagate the extracted resources along the sequence, instead of the traditional BLSTMs; and *(iii)* CTC calculates the loss value for the training process and decodes the model output in the final text using Beam Search algorithm, instead of the decoding process through HMM and language model.

In CRNN architecture (steps *i* and *ii*), the optical model was designed to aim at high precision and low number of trainable parameters, consequently, low computational cost. Thus, we use the Gated mechanism [47] to compose the convolutional layers. This approach aims to extract more relevant features from the images, through the pointwise multiplication operation between the features of a layer (h_1), in which half have sigmoid activation (σ) and the other does not (h_2):

$$y = \sigma(h_1) \odot h_2 \tag{1}$$

In this way, the convolutional block (shown in Figure 1) consists of 5 mini-blocks with traditional and gated convolutions, ending with a convolutional layer. The extracted feature maps follow the values 16, 32, 40, 48, 56, and 64. The 3×3 kernel is applied along the layers, except in the third and fifth, where it changes to 2×4. In addition, the stride values corresponds to 1×1 for 3×3 kernels, with the exception of the first layer with 2×2, and 2×4 value for 2×4 kernels. We also used He uniform and Parametric Rectified Linear Unit (PReLU) [48] as initializer and activator, respectively. Finally, batch renormalization [49] is applied after all convolutional layers, followed by dropout (rate of 0.2) [50] in the last three gated.

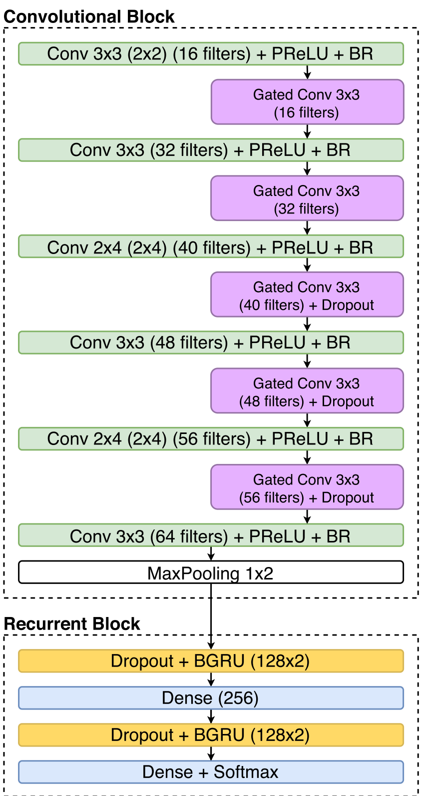

FIGURE 1. The proposed optical model architecture.

The recurrent block (shown in Figure 1) contains 2 BGRUs with 128 GRU cell units in each and a dropout system (rate of 0.5) [51]. In addition, a dense layer with 256 hidden units is applied between the two bidirectional layers. Finally, the model has a dense layer of equal size to the character set + 1 (CTC blank symbol) to be decoded.

IV. MATERIALS AND METHODS

In order to demonstrate the effectiveness of the proposed optical model, we conducted experiments on three datasets from the ICFHR 2014 competition [13] and a private one, according to the competition assessment metrics.

A. DATASETS

The ORAND-CAR dataset [13] consists of 11,719 images obtained from the Courtesy Amount Recognition (CAR) field of real bank checks, and is composed by two sources, which give the images different characteristics, such as type of noise and handwriting style. Therefore, considering the two different sources, this dataset is divided into two subsets: *(i)* CAR-A, that comes from an Uruguayan bank; and *(ii)* CAR-B, that comes from a Chilean bank.

The CAR-A dataset consists of 5,793 images, where 2,009 are for training and 3,784 for testing. Through the ground truth labels, this set has only digits (0-9) for recognition, ignoring other writing symbols present in the images, and has minimum, maximum, and average string length of 2, 8, and 4.5, respectively. In addition, the images have a variety of noises, such as background patterns, and user strokes. Figure 2 shows a sample of images from the CAR-A dataset.

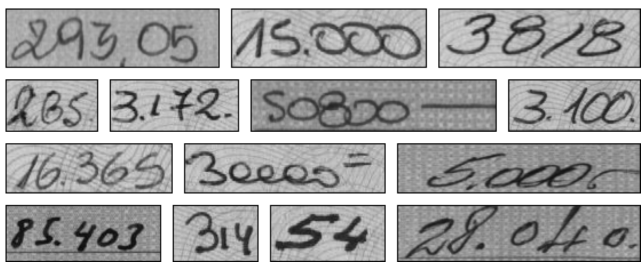

FIGURE 2. Sample images from the CAR-A dataset.

The CAR-B dataset consists of 5,936 images, where 3,000 are for training and 2,936 for testing. Through the ground truth labels, this set has only digits (0-9) for recognition, ignoring other writing symbols present in the images, and has minimum, maximum, and average string length of 3, 8, and 5.6, respectively. The images also have the variety of noises, such as background patterns, and user strokes. Figure 3 shows a sample of images from the CAR-B dataset.

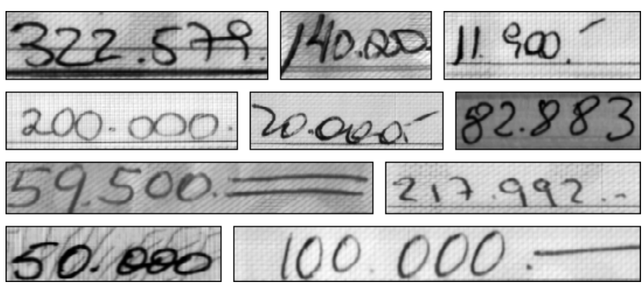

FIGURE 3. Sample images from the CAR-B dataset.

The Computer Vision Lab (CVL) Handwritten Digit String (HDS) consists of 7,960 images from 300 writers, where only 1,262 are for training and 6,698 for testing. Through the ground truth labels, this set has only digits (0-9)

TABLE 1. Volume of data from different training and validation partition arrangements.

Arrangement	CAR-A			CAR-B			CVL HDS			BBC-CAR		
Train/Valid. (%)	Train	Valid.	Test	Train	Valid.	Test	Train	Valid.	Test	Train	Valid.	Test
90/10	1,808	201		2,700	300		1,135	127		31,762	3,530	
80/20	1,607	402		2,400	600		1,009	253		28,233	7,059	
70/30	1,406	603		2,100	900		883	379		24,704	10,588	
60/40	1,205	804		1,800	1,200		757	505		21,175	14,117	
50/50	1,004	1,005	3,784	1,500	1,500	2,936	631	631	6,698	17,646	17,646	7,151
40/60	804	1,205		1,200	1,800		504	758		14,116	21,176	
30/70	603	1,406		900	2,100		378	884		10,587	24,705	
20/80	402	1,607		600	2,400		252	1010		7,058	28,234	
10/90	201	1,808		300	2,700		126	1136		3,529	31,763	

for recognition, ignoring other writing symbols present in the images, and has minimum, maximum, and average string length of 5, 7, and 6.1, respectively. Unlike the previous sets, the CVL HDS has no background noises, and on the other hand, is the set with smaller volume of data for training. Figure 4 shows a sample of images from the CVL HDS dataset.

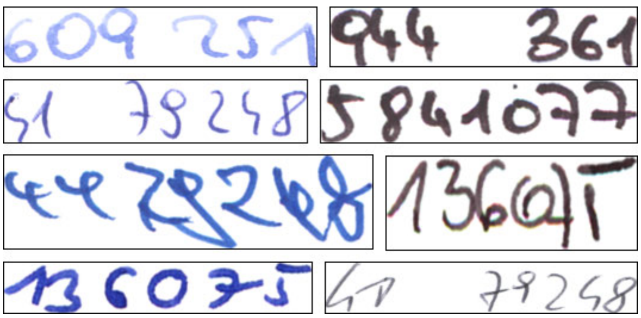

FIGURE 4. Sample images from the CVL HDS dataset.

The private dataset, named Brazilian Bank Check - Courtesy Amount Recognition (BBC-CAR), has the largest volume of data, composed of 42,443 images in total, where 35,292 are for training and 7,151 for testing. Through the ground truth labels, this set considers the punctuation marks in the values, so it has digits (0-9) and comma“,” (Brazilian delimiter for decimals) for recognition, ignoring other writing symbols present in the images. In addition, the minimum, maximum, and average string length is 4, 9, and 5.5, respectively. The images are in black and white mode with 200 DPI, and have a great variety of noise, cursive styles, printed values and unimportant information, such as additional punctuation marks, or part of values from other fields of bank checks. Figure 4 shows a sample of images from the BBC-CAR dataset.

Finally, the datasets presented do not have a validation partition, which would normally use 10% random data from the training partition to compose the validation data. However, to complement the experiments, and to have a better analysis of the proposed optical model, we trained the models for different scenarios of data volume, varying in steps of 10%,

FIGURE 5. Sample images from the BBC-CAR dataset.

and dividing the training and validation partitions, respectively, in 90%/10% up to 10%/90%. In addition, we used the value of *seed* $= 42$ for reproducibility of the experiments. Table 1 details the data volume scenarios with different arrangements in the training and validation partitions.

B. EXPERIMENTAL EVALUATION

The two metrics (hard and soft), proposed by the ICFHR 2014 competition on HDSR [13], were used in this work. In the hard metric, we measured the precision (recognition rate), that is, the number of digit strings recognized correctly divided by the total number of strings. The three best guesses were evaluated (TOP-3), in which we used the three best paths found by CTC’s beam search algorithm.

The soft metric evaluates how close a resulting answer is to a target. Thus, the Levenshtein Distance (LD), also known as edit distance (insertion, deletion, and substitution), was used. In this way, we used the Normalized Levenshtein Distance (NLD) to avoid any bias with respect to the string length, which can be computed by:

$$NLD(a_T, a_R) = \frac{LD(a_T, a_R)}{\mid a_T \mid}, \tag{2}$$

where, a_T is the target string (ground truth), a_R is the recognized string, $\mid a_T \mid$ is the length of the string a_T and LD is the Levenshtein Distance. We also used the Average Normalized Levenshtein Distance (ANLD) defined as:

$$ANLD = \frac{\sum_{i=1}^{T} NLD(a_T^i, a_R^i)}{T}, \tag{3}$$

where, T is the number of test digit strings. Then, the ANLD will provide the final score of the optical models. In addition, ANLD is an inverse performance metric, in other words, low value indicates high performance, while a high value indicates low performance.

Finally, for statistical testing, we performed ten training runs of the optical models in each different arrangement of the datasets, and used Wilcoxon signed-rank test [52] with 5% significance. Thus, we considered the null hypothesis $H_0 : \mu_1 \geq \mu_2$, and alternative hypothesis $H_1 : \mu_1 < \mu_2$, where μ_1 is the NLD of the proposed optical model and μ_2 is the NLD of the other optical model in comparison. This means that the p-value must be lower than $\alpha = 0.05$ to assume that the proposed optical model offer significantly lower error rate.

C. EXPERIMENTAL SETUP

To compose the methods and models of the experiment, we used the same protocol presented at the ICFHR 2014 competition on HDSR [13]. Besides, the models most recently presented in the state-of-the-art in HDSR, and the state-of-the-art optical models in HTR (Puigcerver [10] and Bluche *et al.* [9]) were compared with our proposed optical model.

It is worth mentioning that the works presented by Puigcerver [10] and Bluche *et al.* [9] were evaluated following different experimental methodologies according to the original proposals. Thus, in order to make a fair comparison between the optical models and to statistically validate the results from the same perspective, we used the same workflow and hyperparameters for the optical models in each dataset.

Therefore, we trained optical models in order to minimize the validation loss value of the CTC function. Then, we used RMSprop optimizer with the learning rate of 0.001 and mini-batches of 8 images per step. Reduce Learning Rate on Plateau (factor 0.1) and Early Stopping mechanisms were also applied after 10 and 20 epochs, respectively, without improving the value of the loss of validation. Furthermore, a common charset was used for encoding and decoding (composed of digits and comma), and *beam_width* = 25 was used to setup the beam search decoding algorithm.

No treatment preprocessing was applied to the images, but only the adjustment for the optical model input. Thus, standardization (Z-score normalization) and resizing to $1024 \times 128 \times 1$ (Height $\times$ Width $\times$ Channel) with padding was applied. In addition, data augmentation, to increase the variety of images, was applied to the training partition during the training process of the optical models. Morphological and displacement functions with random parameters were used in each step, such as erosion (up to 5×5 kernel), dilation (up to 3×3 kernel), rotation (up to 3 degrees), resizing (up to 5%), and displacement of height and width (up to 5%).

Finally, all training was conducted on the Google Colab platform, which offers Linux operating system with 12GB memory and GPU NVIDIA Tesla P100 16GB.

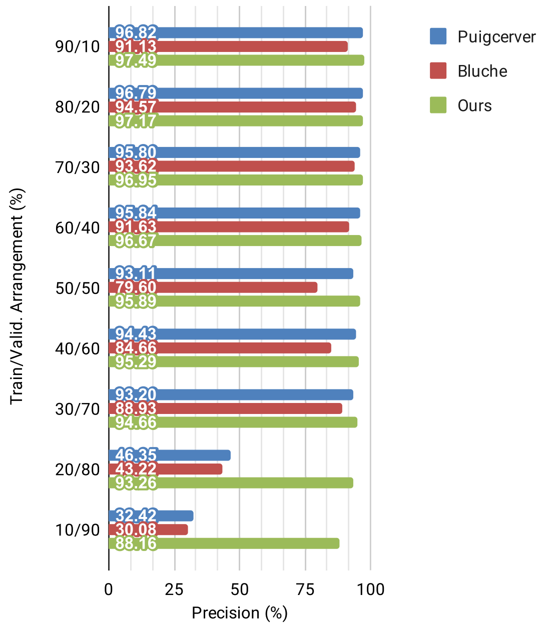

FIGURE 6. Results in different arrangements of training and validation partitions in the CAR-A dataset.

V. RESULTS AND DISCUSSION

First, we used the different arrangements in training and validation partitions for an initial analysis of the optical models of Puigcerver, Bluche, and the one proposed. Then, we selected the best result obtained by each optical model to compare with those presented by the competition methods and state-of-the-art approaches. It is worth mentioning that the results of the Shanghai method [13], is the only one that has TOP-3 smaller than its own TOP-2.

In this context, the arrangement of the CAR-A dataset that provided the best precision results were 90%/10%, for the proposed optical model and Puigcerver model, while 80%/20% for Bluche model. Thus, the optical models reached 97.49%, 96.82% and 94.57%, respectively. It is worth mentioning that the proposed optical model maintained high precision even with low data, as in the case of arrangements 20%/80% and 10%/90%, with only 402 and 201 data for training, respectively. Figure 6 shows the results obtained throughout the different arrangements for each optical model.

The proposed optical model improved the best results from the current state-of-the-art (Hochuli *et al.* [8]) by about 1.29 percentage points in the TOP-1, with 97.49%. In addition, we reached 98.52% and 98.96% in TOP-2 and TOP-3, respectively. Furthermore, the ANLD value was much lower (better) than that presented in the ICFHR 2014 competition on HDSR [13]. Table 2 details the results obtained, as well as the previous works ("-" represents not informed).

In the CAR-B dataset, 80%/20% arrangement had the best precision results for the three optical models. In this way, the proposed optical model reached 98.48%, Puigcerver 98.34% and Bluche 97.41%. The CAR-B set was slightly

TABLE 2. ANLD and Precision results in the CAR-A test set.

Method	ANLD	Precision TOP-1	TOP-2	TOP-3
Beijing [13]	0.05	80.73%	86.34%	86.97%
Pernambuco [13]	0.07	78.30%	89.16%	91.99%
Shanghai [13]	0.22	49.50%	53.78%	49.50%
Singapore [13]	0.17	52.30%	61.80%	65.40%
Tébessa I [13]	0.27	37.05%	45.59%	47.20%
Tébessa II [13]	0.25	39.72%	44.77%	48.18%
Saabni [5]	-	85.80%	88.90%	-
Zhan et al. [6]	-	89.75%	-	-
Wang et al. [33]	-	82.61%	-	-
Hochuli et al. [34]	-	50.10%	-	-
Zhan et al. [16]	-	92.20%	-	-
Xu et al. [7]	-	91.89%	-	-
Zhan et al. [35]	-	91.97%	-	-
Ma et al. [37]	-	72.50%	-	-
Lupinski et al. [36]	-	84.01%	-	-
Xu et al. [17]	-	83.17%	-	-
Hochuli et al. [8]	-	96.20%	-	-
Puigcerver	0.02	96.82%	98.04%	98.42%
Bluche	0.03	94.57%	96.70%	97.46%
Ours	**0.01**	**97.49%**	**98.52%**	**98.96%**

easier than CAR-A, considering the amount of data greater than CAR-A, in which it provided results very close between the optical models. Finally, Puigcerver and Bluche models did not maintain high precision in the last arrangement with only 300 trainable data (90%/10%), unlike the proposed one. Figure 7 shows the results obtained throughout the different arrangements for each optical model.

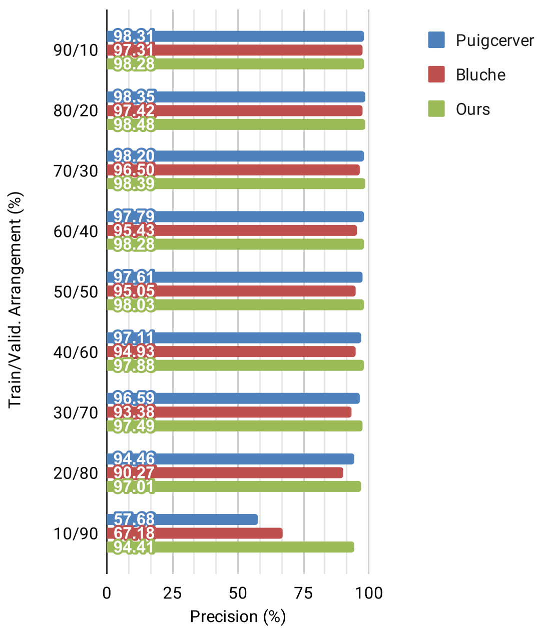

FIGURE 7. Results in different arrangements of training and validation partitions in the CAR-B dataset.

The proposed optical model also improved the best results from the current state-of-the-art (Hochuli *et al.* [8]) by about 1.68 percentage points in the TOP-1, with 98.48%. In addition, we reached 99.12% and 99.34% in TOP-2 and TOP-3, respectively. The ANLD value was also lower (better) than that presented in the ICFHR 2014 competition on HDSR [13]. Table 3 details the results obtained, as well as the previous works ("-" represents not informed).

TABLE 3. ANLD and Precision results in the CAR-B test set.

Method	ANLD	Precision TOP-1	TOP-2	TOP-3
Beijing [13]	0.08	70.13%	76.38%	77.79%
Pernambuco [13]	0.06	75.43%	87.46%	90.09%
Shanghai [13]	0.33	28.09%	31.20%	28.09%
Singapore [13]	0.13	59.60%	67.70%	71.30%
Tébessa I [13]	0.31	26.62%	34.01%	35.68%
Tébessa II [13]	0.30	27.72%	31.37%	34.11%
Saabni [5]	-	85.80%	88.90%	-
Zhan et al. [6]	-	91.14%	-	-
Wang et al. [33]	-	83.32%	-	-
Hochuli et al. [34]	-	40.20%	-	-
Zhan et al. [16]	-	94.02%	-	-
Xu et al. [7]	-	93.79%	-	-
Zhan et al. [35]	-	93.87%	-	-
Ma et al. [37]	-	76.10%	-	-
Lupinski et al. [36]	-	82.97%	-	-
Xu et al. [17]	-	84.79%	-	-
Hochuli et al. [8]	-	96.80%	-	-
Puigcerver	**0.01**	98.35%	99.03%	99.24%
Bluche	0.02	97.42%	98.33%	98.68%
Ours	**0.01**	**98.48%**	**99.12%**	**99.34%**

The CVL HDS dataset was the most challenging among the sets, especially for deep learning models. The volume of data is the smallest among the four sets, reaching up to 126 trainable data in the last arrangement (10%/90%), which provided results different from any other dataset.

Therefore, 90%/10% arrangement provided the best result for the proposed optical model and for Bluche, with 93.53% and 50.69%, respectively. While 70%/30% was the one for the Puigcerver model, with 54.41%. In this set, only the proposed optical model achieved high precision in all arrangements, in which the lowest result was 89.67% using only 126 trainable data from the last arrangement. Figure 8 shows the results obtained throughout the different arrangements for each optical model.

Until our proposed system, the state-of-the-art result for the CVL HDS set was still in ICFHR 2014 competition on HDSR [13], by the Beijing group with 85.29% (TOP-1). Furthermore, the most recent works, which already use deep learning techniques, in general did not deal well with this set. It is worth mentioning that the works of Saabni [5] and Hochuli *et al.* [8], [34] did not work with this dataset.

However, our proposed optical model proved to be quite competent when dealing with a minimum volume of data, reaching 93.54% as the best result on TOP-1 and 95.65%, 96.51% on TOP-2 and TOP-3, respectively. On the other hand, the Puigcerver and Bluche models did not achieve significant results, when compared to other state-of-the-art

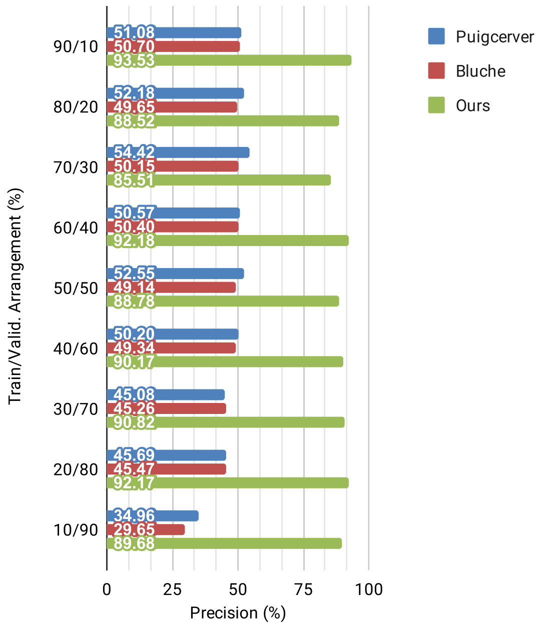

FIGURE 8. Results in different arrangements of training and validation partitions in the CVL HDS dataset.

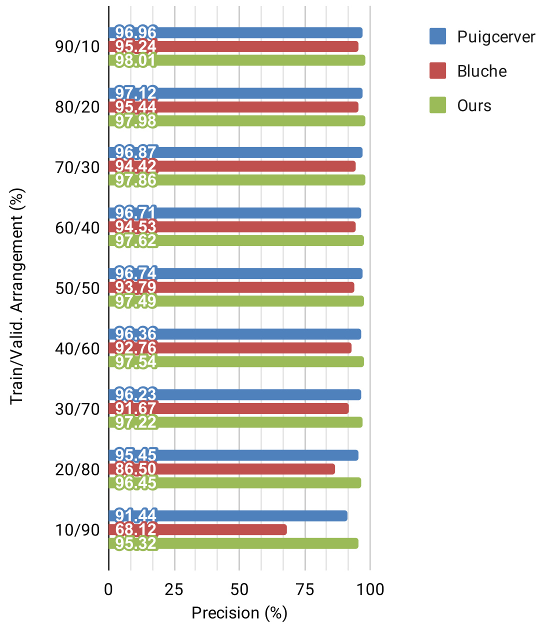

FIGURE 9. Results in different arrangements of training and validation partitions in the BBC-CAR dataset.

TABLE 4. ANLD and Precision results in the CVL HDS test set.

Method	ANLD	Precision TOP-1	Precision TOP-2	Precision TOP-3
Beijing [13]	0.03	85.29%	91.28%	91.89%
Pernambuco [13]	0.12	58.60%	68.50%	72.34%
Shanghai [13]	0.16	48.93%	54.00%	48.93%
Singapore [13]	0.12	50.40%	60.60%	65.40%
Tébessa I [13]	0.12	59.30%	65.75%	66.90%
Tébessa II [13]	0.11	61.23%	65.27%	68.24%
Zhan et al. [6]	-	27.07%	-	-
Wang et al. [33]	-	79.23%	-	-
Zhan et al. [16]	-	42.69%	-	-
Xu et al. [7]	-	63.03%	-	-
Zhan et al. [35]	-	28.23%	-	-
Ma et al. [37]	-	65.90%	-	-
Lupinski et al. [36]	-	22.45%	-	-
Xu et al. [17]	-	78.67%	-	-
Puigcerver	0.30	54.42%	56.64%	58.18%
Bluche	0.32	50.70%	53.04%	54.45%
Ours	**0.04**	**93.54%**	**95.65%**	**96.51%**

approaches. In addition, the ANLD value was lower (better) than that presented in the ICFHR 2014 competition on HDSR [13]. Table 4 details the results obtained, as well as the previous works ("-" represents not informed).

The BBC-CAR dataset has the largest amount of data in all arrangements among the other sets. This shows that optical models are also capable of handling large sets. In fact, all the results of the optical models were good throughout the arrangements, except for the last one for the Bluche model, which even having 3,529 trainable data, the model did not maintain the high precision rate. Figure 9 shows the results obtained throughout the different arrangements for each optical model.

As the BBC-CAR dataset is private, the analysis is only between the optical models of Puigcerver, Bluche and proposed one. In this context, our proposed optical model reached the best result in 90%/10% arrangement, with 98.01% (TOP-1), while Puigcerver and Bluche reached their best results in 80%/20% arrangement, with 97.12% and 95.44%, respectively. In addition, the ANLD value of the proposed model was slightly lower than the others. Table 5 details the results obtained in the BBC-CAR test set.

TABLE 5. ANLD and Precision results in the BBC-CAR test set.

Method	ANLD	Precision TOP-1	Precision TOP-2	Precision TOP-3
Puigcerver	0.02	97.12%	97.85%	98.18%
Bluche	0.02	95.44%	96.85%	97.47%
Ours	**0.01**	**98.01%**	**98.68%**	**98.91%**

In general, the optical models missed certain characters in the recognition, for example, when part of the digit is missing, or the digit has a writing pattern very similar to some other, or depending on the image noise is very high. However, even in very noisy scenarios and with irrelevant information around the digit, mainly from the BBC-CAR dataset, the proposed optical model handled it very well. Figure 10 shows some examples of the recognition results obtained from optical models (TOP-1) in each dataset presented.

Finally, for the statistical analysis, we used the all results obtained in datasets, considering the TOP-1 value of the optical models of Puigcerver, Bluche and the proposed one. Then, we computed the NLD p-value lower than 0.01 in all datasets (different arrangements included). This is below the standard $\alpha = 0.05$, meaning that we can assume that the

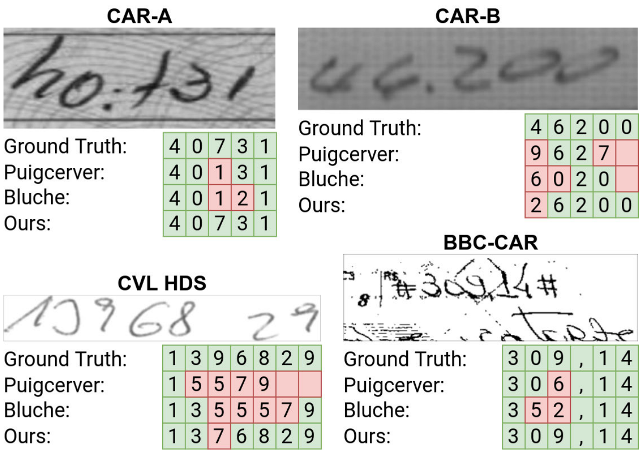

FIGURE 10. Examples of the recognition results obtained from optical models in each dataset presented.

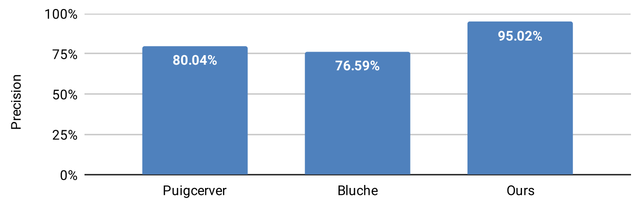

FIGURE 11. Consolidated of the precision results obtained by the optical models in all datasets and their arrangements.

proposed optical model, based on Gated-CNN-BGRU, has a significantly lower NLD in the test partitions of each tested dataset. Moreover, we also consolidated the precision values (shown in Figure 11) obtained by the optical models in all datasets (also considering their arrangements). In this context, we highlight the difference of the proposed model, which surpasses Puigcerver and Bluche models around 14.98 and 18.43 percentage points, respectively.

In addition, another important evaluative requirement for deep neural networks is the complexity of its architecture, which impacts the size and decoding time of the optical model. In this context, the proposed optical model is close to the Bluche model, in number of trainable parameters (thousands), which is already well below that of Puigcerver (millions). For the decoding time, we calculated the average of each optical model in all runs of the experiment. Thus, the proposed model was the intermediate one, with Bluche model being the fastest and Puigcerver the slowest. Table 6 shows the number of parameters and average decoding time using a standard notebook with dual core CPU (Intel i7-7500U).

Therefore, the improvements observed in the recognition rates of the proposed optical model are explained by the combination of the applied Gated mechanism [47], and non-traditional deep learning techniques in the context of HTR, such as BGRU layers, PReLU activators, and batch renormalization [48], [49]. In this way, we can extract and propagate the characteristics of the images efficiently,

TABLE 6. Number of trainable parameters and average decoding time of the optical models.

Optical Model	# of params	Decoding Time
Puigcerver	9.4 M	70 ms/image
Bluche	0.7 M	59 ms/image
Ours	0.8 M	65 ms/image

maintaining the balance between the low number of parameters and high performance. This application is highlighted in the CVL HDS dataset, whose proposed optical model achieved about 8.25 percentage points higher than the best previous result known in this dataset, since the ICFHR 2014 competition on HDSR [13].

VI. CONCLUSION

In this paper, we have presented a new Gated-CNN-BGRU architecture, based on Handwritten Text Recognition (HTR) workflow, for Handwritten Digit String Recognition (HDSR) systems, with focus on improving the state-of-the-art recognition rate of this research field.

In this context, we evaluated the optical model proposed with the current state-of-the-art in HDSR in four different challenging datasets, where three are from the ICFHR 2014 competition on HDSR [13], and another private one. For a better analysis, mainly in low data volume scenarios, we also made different arrangements between the partitioning of data for training and validation, varying from 90%/10% to 10%/90%.

The proposed optical model proved to be efficient in recognizing sequences in images with high noise, discarding irrelevant information around the target text. It demonstrated robustness even under low data volume, reaching an average precision of 96.50% in the competition datasets, which is equivalent to an average of 3.74 percentage points in improvement to the current state-of-the-art. In addition, we highlight the results obtained in the CVL HDS set, where the proposed optical model reached precision of 93.54%, while the best result so far had been from Beijing group (from the competition itself), with 85.29%.

In the future, we want to investigate the quality of the proposed optical model to recognize other numeric fields, which present even more image noise and longer digit strings. Furthermore, we intend to explore low volume datasets for the recognition of fields with digits and letters, since the proposed optical model also fits the Handwritten Text Recognition scenario.

REFERENCES

[1] H. Sajedi, "Handwriting recognition of digits, signs, and numerical strings in persian," *Comput. Electr. Eng.*, vol. 49, pp. 52–65, Jan. 2016.
[2] B. L. D. Bezerra, C. Zanchettin, A. H. Toselli, and G. Pirlo, *Handwriting: Recognition, Development and Analysis*. Hauppauge, NY, USA: Nova, Jul. 2017.
[3] M. Diem, S. Fiel, and F. Kleber, *Handwritten Digit and Digit String Recognition: Benchmarking State-of-the-Art Systems*. Singapore: World Scientific, Apr. 2018, pp. 67–88.

[4] A. Gattal, Y. Chibani, and B. Hadjadji, "Segmentation and recognition system for unknown-length handwritten digit strings," *Pattern Anal. Appl.*, vol. 20, no. 2, pp. 307–323, May 2017.
[5] R. Saabni, "Recognizing handwritten single digits and digit strings using deep architecture of neural networks," in *Proc. 3rd Int. Conf. Artif. Intell. Pattern Recognit. (AIPR)*, Sep. 2016, pp. 1–6.
[6] H. Zhan, Q. Wang, and Y. Lu, "Handwritten digit string recognition by combination of residual network and RNN-CTC," in *Proc. Int. Conf. Neural Inf. Process. (ICONIP)*. Cham, Switzerland: Springer, 2017, pp. 583–591.
[7] X. Xu, J. Zhou, and H. Zhang, "Screen-rendered text images recognition using a deep residual network based segmentation-free method," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Beijing, China, Aug. 2018, pp. 2741–2746.
[8] A. G. Hochuli, A. S. Britto, J. P. Barddal, R. Sabourin, and L. E. S. Oliveira, "An End-to-End approach for recognition of modern and historical handwritten numeral strings," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.
[9] T. Bluche and R. Messina, "Gated convolutional recurrent neural networks for multilingual handwriting recognition," in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Nov. 2017, pp. 646–651.
[10] J. Puigcerver, "Are multidimensional recurrent layers really necessary for handwritten text recognition?" in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, Nov. 2017, pp. 67–72.
[11] A. F. S. Neto, B. L. D. Bezerra, A. H. Toselli, and E. B. Lima, "HTR-Flor++: A handwritten text recognition system based on a pipeline of optical and language models," in *Proc. ACM Symp. Document Eng.*, New York, NY, USA, 2020, pp. 1–4.
[12] A. F. de Sousa Neto, B. L. D. Bezerra, A. H. Toselli, and E. B. Lima, "HTR-Flor: A deep learning system for offline handwritten text recognition," in *Proc. 33rd SIBGRAPI Conf. Graph., Patterns Images Tuts. (SIBGRAPI)*, 2020, pp. 1–8.
[13] M. Diem, S. Fiel, F. Kleber, R. Sablatnig, J. M. Saavedra, D. Contreras, J. M. Barrios, and L. S. Oliveira, "ICFHR 2014 competition on handwritten digit string recognition in challenging datasets (HDSRC 2014)," in *Proc. 14th Int. Conf. Frontiers Handwriting Recognit.*, Sep. 2014, pp. 779–784.
[14] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*. New Jersey, NJ, USA, 2010, pp. 249–256.
[15] B. Shi, X. Bai, and C. Yao, "An End-to-End trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017.
[16] H. Zhan, S. Lyu, and Y. Lu, "Handwritten digit string recognition using convolutional neural network," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 3729–3734.
[17] Y. Xu and J. Guo, "Low resolution handwritten digit string recognition based on object detection network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–7.
[18] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[19] J. C. B. Rabelo, C. Zanchettin, C. A. B. Mello, and B. L. D. Bezerra, "A multi-layer perceptron approach to threshold documents with complex background," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2011, pp. 2523–2530.
[20] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognit.*, vol. 33, no. 2, pp. 225–236, Feb. 2000.
[21] Q.-F. Wang, F. Yin, and C.-L. Liu, "Integrating language model in handwritten Chinese text recognition," in *Proc. 10th Int. Conf. Document Anal. Recognit.*, Jul. 2009, pp. 1036–1040.
[22] B. L. D. Bezerra, G. D. C. Cavalcanti, and J. Rabelo, "Detecting and treating invasion in the courtesy amount field on bank checks," in *2008 11th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, 2008, pp. 1–6.
[23] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 545–552.
[24] B. L. D. Bezerra, C. Zanchettin, and V. B. de Andrade, "A MDRNN-SVM hybrid model for cursive offline handwriting recognition," in *Artificial Neural Networks and Machine Learning–(ICANN)*. Berlin, Germany: Springer, 2012, pp. 246–254.
[25] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[26] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.
[27] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.
[28] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Nat. Acad. Sci. USA*, vol. 79, no. 8, pp. 2554–2558, 1982.
[29] A. Gattal and Y. Chibani, "Segmentation strategy of handwritten connected digits (SSHCD)," in *Image Analysis and Processing–(ICIAP)*. Berlin, Germany: Springer, 2011, pp. 248–254.
[30] M. Hejazi, G. Shevlyakov, and Y.-S. Ho, "Modified discrete radon transforms and their application to rotation-invariant image analysis," in *Proc. IEEE Workshop Multimedia Signal Process.*, Oct. 2006, pp. 429–434.
[31] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst.*, vol. 25, Jan. 2012, pp. 84–90.
[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[33] Q. Wang and Y. Lu, "A sequence labeling convolutional network and its application to handwritten string recognition," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2950–2956.
[34] A. G. Hochuli, L. S. Oliveira, A. S. Britto, Jr., and R. Sabourin, "Handwritten digit segmentation: Is it still necessary?" *Pattern Recognit.*, vol. 78, pp. 1–11, Jun. 2018.
[35] H. Zhan, S. Lyu, X. Tu, and Y. Lu, "Residual CRNN and its application to handwritten digit string recognition," in *Neural Information Processing*, T. Gedeon, K. W. Wong, and M. Lee, Eds. Cham, Switzerland: Springer, 2019, pp. 49–56.
[36] T. Lupinski, A. Belaid, and A. K. Echi, "On the use of attention mechanism in a Seq2Seq based approach for off-line handwritten digit string recognition," in *Proc. Int. Conf. Document Anal. Recognit. (ICDAR)*, Sep. 2019, pp. 502–507.
[37] Y. Ma, J. Guo, and W. Wei, "An exceedingly fast model for low resolution handwritten digit string recognition," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Oct. 2019, pp. 282–288.
[38] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[39] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.
[40] J. A. Sánchez, V. Romero, A. H. Toselli, M. Villegas, and E. Vidal, "A set of benchmarks for handwritten text recognition on historical documents," *Pattern Recognit.*, vol. 94, pp. 122–134, Oct. 2019.
[41] C. D. Stefano, A. Marcelli, A. Parziale, and R. Senatore, "Reading cursive handwriting," in *Proc. 12th Int. Conf. Frontiers Handwriting Recognit.*, Nov. 2010, pp. 95–100.
[42] H. Bunke, M. Roth, and E. G. Schukat-Talamazzini, "Off-line cursive handwriting recognition using hidden Markov models," *Pattern Recognit.*, vol. 28, no. 9, pp. 1399–1413, Sep. 1995.
[43] A. H. Toselli and E. Vidal, "Handwritten text recognition results on the bentham collection with improved classical N-Gram-HMM methods," in *Proc. 3rd Int. Workshop Historical Document Imag. Process. (HIP)*, 2015, pp. 15–22.
[44] A. Graves, S. Fernández, M. Liwicki, H. Bunke, and J. Schmidhuber, "Unconstrained on-line handwriting recognition with recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 20, Jan. 2007, pp. 577–584.
[45] P. Voigtlaender, P. Doetsch, and H. Ney, "Handwriting recognition with large multidimensional long short-term memory recurrent neural networks," in *Proc. 15th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Oct. 2016, pp. 228–233.
[46] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in *Proc. 23rd Int. Conf. Mach. Learn.*, Jan. 2006, pp. 369–376.
[47] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, 2017, pp. 933–941.
[48] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[49] S. Ioffe, "Batch renormalization: Towards reducing minibatch dependence in batch-normalized models," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 1942–1950.
[50] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[51] V. Pham, T. Bluche, C. Kermorvant, and J. Louradour, "Dropout improves recurrent neural networks for handwriting recognition," in *Proc. 14th Int. Conf. Frontiers Handwriting Recognit.*, Sep. 2014, pp. 285–290.
[52] F. Wilcoxon, *Individual Comparisons by Ranking Methods*. New York, NY, USA: Springer, 1992, pp. 196–202.

ARTHUR FLOR DE SOUSA NETO received the B.S. degree in information systems from the University of Pernambuco, Brazil, in 2016, and the M.S. degree in computer engineering with emphasis on artificial intelligence from the Polytechnic School, University of Pernambuco Brazil, in 2020. His current research interests include artificial intelligence, machine learning, deep learning, computer vision, handwritten text recognition, and natural language processing.

BYRON LEITE DANTAS BEZERRA (Member, IEEE) received the Ph.D. degree with an emphasis on artificial intelligence from the Federal University of Pernambuco, Brazil, in 2008. He began his activities as a Professor with the Polytechnic School, University of Pernambuco, in 2009, where he is currently an Associate Professor, a Researcher, and the Post-Graduation Director. At the leadership of the Pattern Recognition Research Group, he has developed dozens of research and technological innovation projects. He has experience in computer science, working mainly on the following topics: document processing, handwriting recognition, gesture recognition, recommendation systems, information filtering, and retrieval. Through fruitful partnerships with companies, he has contributed to the design and improvement of document capture and imaging systems, forms processing, handwriting recognition, and signature verification software. After ten years of use, his algorithms have already processed more than one billion documents.

ESTANISLAU BAPTISTA LIMA received the B.Sc. degree in computer science from the State University of Paraíba, Brazil, in 2010, the M.Sc. degree in software engineering from Chongqing University, China, in 2012, and the Ph.D. degree in computer application technology from the Ocean University of China, in June 2018. He currently a Postdoctoral Fellow with the Polytechnic School, University of Pernambuco, Brazil. His current research interests include computer vision, AI, machine learning, deep learning, object recognition/detection, and scene text recognition.

ALEJANDRO HÉCTOR TOSELLI received the M.S. degree in electrical engineering from the Universidad Nacional de Tucuman, Argentina, in 1997, and the Ph.D. degree in computer science from the Universitat Politècnica de València, Spain, in 2004. From 2005 to 2008, he worked as a Research Member of the Instituto Tecnológico de Informática, Universitat Politècnica de València, participating in different research projects with technological transfer to industry. From 2008 to 2009, he was a Postdoctoral Fellow with the Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA), Rennes, France, in the Recognition and interpretation of Images and Documents Research Group (IMADOC). Since 2010, he has been a full-time Researcher with the Pattern Recognition and Human Language Research Center (PRHLT), Universitat Politècnica de València, being active in different European Research Projects, like EU READ and HIMANIS, focused mainly on the topic of handwritten text recognition. His current research interests include the general subject of document analysis and processing under traditional and user-interactive pattern recognition view. Particularly, he focuses on handwritten text recognition and keyword spotting on historical documents by employing recognition methodologies, such as neural networks and hidden Markov models. He has teaching experience in imparting different tutorial courses about the handwritten text recognition and other related topics.