Towards an IMU-based Pen Online Handwriting Recognizer

Mohamad Wehbi[1](✉), Tim Hamann[2], Jens Barth[2], Peter Kaempf[2],
Dario Zanca[1], and Bjoern Eskofier[1]

[1] Machine Learning and Data Analytics Lab, Friedrich-Alexander-Universität
Erlangen-Nürnberg, Erlangen, Germany
{mohamad.wehbi,dario.zanca,bjoern.eskofier}@fau.de
[2] STABILO International GmbH, Heroldsberg, Germany
{tim.hamann,jens.barth,peter.kaempf}@stabilo.com

Abstract. Most online handwriting recognition systems require the use of specific writing surfaces to extract positional data. In this paper we present a online handwriting recognition system for word recognition which is based on inertial measurement units (IMUs) for digitizing text written on paper. This is obtained by means of a sensor-equipped pen that provides acceleration, angular velocity, and magnetic forces streamed via Bluetooth. Our model combines convolutional and bidirectional LSTM networks, and is trained with the Connectionist Temporal Classification loss that allows the interpretation of raw sensor data into words without the need of sequence segmentation. We use a dataset of words collected using multiple sensor-enhanced pens and evaluate our model on distinct test sets of seen and unseen words achieving a character error rate of 17.97% and 17.08%, respectively, without the use of a dictionary or language model.

Keywords: Online handwriting recognition · Digital pen · Inertial measurement unit · Time-series data

1 Introduction

The field of handwriting recognition has been studied for decades, increasing in popularity with the advancements of technology. This increase in popularity is due to the substantial number of people using handheld digital devices that provide access to such technologies, and the desire of people to save and share digital copies of written documents. The aim of a handwriting recognition system is to allow users to write without constraints, then digitize what was written for a multitude of uses.

Handwriting recognition (HWR) is widely known to be separated into two distinct types, offline and online recognition [37,38]. For offline recognition, a static scanned image of the written text is given as input to the system. Offline recognition, also known as optical character recognition (OCR), is the more common recognition technique used in a wide range of applications for reading

J. Lladós et al. (Eds.): ICDAR 2021, LNCS 12823, pp. 289–303, 2021.
https://doi.org/10.1007/978-3-030-86334-0_19

specific details on documents, such as in healthcare and legal industry [42], banking [36], and postal services [43]. Online handwriting recognition (OHWR), alternatively, requires input data in the form of time series and includes the use of an additional time dimension within the data to be digitized. This results in a dynamic spatio-temporal signal that characterizes the shape and speed of writing [25]. OHWR systems are deployed in applications on tablets and mobile phones for users to digitize text using stylus pens or finger inputs on touch screens. Such systems use precise positions of the writing tip. However, one drawback that can be perceived is the need for a positional tracking system, whether it be a mobile touch screen, or any other pen tracking application. The need for such a tracking system restrains the user from writing on any surface and limits the usability of the system as well as the capability of the writer [13].

Another approach to applying OHWR is the use of inertial measurement units (IMUs) as, or integrated within, a writing tool. These tools provide movement data, such as accelerometer or gyroscope signals, which can be used for classification and recognition tasks. The major disadvantage of IMU sensors is that they are prone to error accumulation over time which, if not corrected, can lead to significant errors in the data recordings. Furthermore, IMU sensors generate noisy output signals, which is further intensified when touching a surface due to surface friction, which successively leads to lower performance at the task required due to deficient input data quality. However, when coupled with the correct models, IMUs produce beneficial data from which precise information can be extracted, such as the specific movements of a pen during handwriting. Moreover, IMUs are sourceless, self-contained and require no additional tools for data collection and extraction, and hence, a major advantage of IMU-based recognition systems is that no specific writing surface is required and systems rely only on the signals collected from the sensors.

In this paper, we discuss further the latter approach and introduce an OHWR system that uses sensor data recordings from pen movements to recognize writing on regular paper. We present an end-to-end system that processes sensor recordings in the form of time series data, and outputs the interpreted digital text on a tablet. Our system surpasses previous sensor-based pen systems in recognition rates, and is the first IMU-based pen recognizer that recognizes complete words and is not restricted to single character or digit recognition on paper. We use, as a digitizer, a regular ballpoint pen integrated with multiple sensors, and designed with a soft grip, that allows the user to write on a plain paper surface without constraints.

The rest of the paper is structured as follows: Sect. 2 summarizes available OHWR systems, distinguishing between positional-based and IMU-based systems. Section 3 presents the digitizer used in our system and explains the data acquisition process. Section 4 describes our end-to-end neural network architecture and describes the model training process, the hyperparameters used, and the data splitting. Section 5 reports the results and discusses the results obtained on distinct test sets. Section 6 outlines the future work to be implemented in our system.

2 Related Work

HWR has been a topic of interest in research for many years. Reviews about recognition systems [25,37] present pre-processing techniques, extracted features, in addition to different recognition models such as segment-and-decode methods and end-to-end recognition systems. In our work, we focus on the difference in the type of data used rather than methods of recognition. We briefly describe some previously developed recognition systems while distinguishing between systems using positional data and ones using IMU data.

2.1 Positional-Data Based Systems

Basic OHWR systems were presented by [6,48] in the late 1990s using Hidden Markov Model (HMMs) and Artificial Neural Networks (ANNs) that model the spatial structure of handwriting. A system developed using a multi-state time delay neural network was presented in [21] using a dictionary of 5000 words with pen position and pen-up/pen-down data. Models in which both an image of the text along with pen tracking data were used to develop a Japanese handwriting recognition system [22]. Different language specific systems were implemented to apply recognition systems for different writing styles, such as Arabic [45] and Chinese [28].

The availability of public datasets considerably increased research in this field. The UNIPEN dataset [18] is a collection of characters with recorded pen trajectory information including coordinate data with pen-up/down features, which was used to implement a character recognition model using time delay [19] and Convolutional [33] neural networks. An Arabic recognition system [4] applied HMMs using the SUSTOLAH dataset [34].

The IAM-ONDB dataset [30] is considered the most popular dataset in the OHWR domain. It includes pen trajectories of sentences written on a smart whiteboard with an infrared device mounted on the corner of the board to track the position of writers' pens, in addition to image data of written text from a collection of 86,272 word instances. [30] also introduced a HMM-based model with segmented data reaching up to about 66% recognition rate. This rate increased to 74% and consecutively to 79% when recurrent neural networks (RNNs) and bi-directional Long Short-Term Memory-Networks (BLSTMs) were implemented with non-segmented data [16,32]. An unconstrained recognition system was introduced in [15] with the integration of External Grammar models, achieving a word error rate (WER) of 35.5% using HMMs and 20.4% using BLSTMs. The combination of diverse classifiers led to a word level accuracy of 86.16% [31].

A multi-language system [24], supporting up to 22 different scripts for touch enabled devices, was based on several components: character model, segmentation model, and feature weights. The model was trained and evaluated on different public and internal datasets, leading with error rates as low as 0.8% to 5.1% on different UNIPEN test sets, also achieving a character error rate (CER) and WER of 4.3% and 10.4%, respectively, on the IAM-ONDB dataset.

The model results improved with the introduction of Bézier Curves, an end-to-end BLSTMs architecture, and language specific models with CER and WER of 2.5% and 6.5%, respectively [8].

The above metioned systems describe different methods for data preprocessing, feature extraction, and classification models, achieving different results for OHWR, while using input of either raw coordinate strokes or features extracted from these strokes. These systems were designed for position-based data that was extracted using specially designed hardware writing surfaces or touch screens and thus still pose a limitation if a system for digitzing paper-writing is required.

2.2 IMU-data Based Systems

The use of IMU data for HWR has been presented in different forms throughout the past years. Accelerometer-based digital pens for handwritten digit and gesture trajectory were developed in [23,46] with an accuracy of 98% & 84.8% over the ten digits, respectively. A 26 uppercase alphabet recognition system was developed using an inertial pen with a KNN classifier with 82% accuracy [41]. Pentelligence [40] combined the use of writing sounds with pen-tip motion from a digital pen equipped with microphones and IMU sensors for digit recognition reaching an accuracy of 98.33% for a single writer.

More recent studies used the Digipen [27] for the recognition of lowercase Latin alphabet characters. A recognition rate of 52% was achieved using LSTMs. The Digipen [47] was also used for the classification of uppercase and lowercase Latin alphabet characters separately, with a different dataset than what was used in [27]. A 1-Dimensional Convolutional Neural Network (1D-CNN) model achieved an accuracy of 86.97%.

At the time of the development of these systems, no public dataset for this task was available for a concrete evaluation of different systems. The On-HW dataset [35] was the first published IMU-based dataset and consisted of recordings of the complete Latin alphabet characters. It was released with baseline methods having an accuracy of 64.13% for the classification of 52 classes. These results were based on the writer-independent scenarios described in the papers (when available), since a writer-dependent model is not a feasible model when developing an OHWR system for general use.

For word level recognition, wearable technologies were implemented as approaches for HWR using IMUs. Airwriting is a tracked motion of continuous sensor stream, in which writing is of a single continuous stroke. It suffers from no surface friction and allows writing in free space. A digital glove equipped with accelerometers and gyroscopes for airwriting was designed in [5], achiveing a WER of 11% using an HMM model following the segment-and-decode approach, evaluated on nine users writing 366 words using a language model consisting of 60000 words. A CNN-RNN approach for in-air HWR [12] achieved a word recognition rate (WRR) 97.88% using BLSTMs. Similar work was presented in [11] with a recognition rate of 97.74% using an encoder-decoder model. More recent

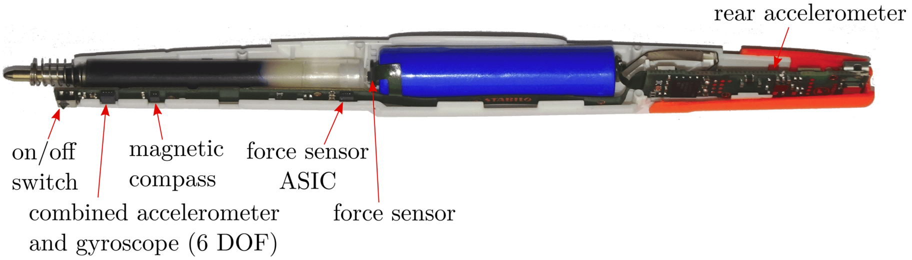

Fig. 1. The Digipen sensor placement

work presented a wearable ring for on-surface HWR [29] which provided acceleration and the angular velocity data from the finger resulting in 1.05% CER and 7.28% WER on a dataset of 643 words collected by a single writer.

Differently from positional-based systems or wearable systems, IMU-based digital pen systems are still limited to single digit or character recognition, with a solution for word recognition not demonstrated yet. This is due to the fact that writing on a paper surface introduces a considerable amount of noise in the data which makes the learning of a recognition model challenging. Furthermore the evaluation of such systems has been conducted on very different setups and on limited data. Here we propose a system that aims at filling the gap between pen-based systems and other approaches. We show that our IMU-based pen recognizer is practical for word recognition, and achieves significant improved results in comparison to previous pen devices.

3 Data Acquisition and Description

In this section, we introduce the pen used as a digitzer and the data collection process, and describe the data that was used to train our model. We base our system on a set of digital pens of the same model to ensure that our work is not biased towards one single instance of the device. The pen model used in our system is the STABILO Digipen which was used in [27,35,47]. The selection of this digitizer was based on the two main factors:

- **Suitability**: The Digipen is a ballpoint pen that can be used to write on paper like any regular pen. It is equipped with five different sensors, a combined accelerometer and gyroscope module at one side of the pen close to the pen tip, another single accelerometer module close to the other end of the pen, a magnetometer module, in addition to a force sensor at the tip of the pen that provides data about when the tip touches a surface, displayed in Fig. 1. This tool also includes a Bluetooth module that allows the transmission of collected sensor data to other devices in real-time. Accordingly, a trained recognition model can be integrated within a mobile app and can be used without the need for any further equipment but a mobile device. The Digipen streams sensor data via Bluetooth Low Energy with a sampling frequency 100 Hz.

Detailed information about the pen dimensions, sensor modules and ranges can be found in [2,35].

- **Availability**: The Digipen was not specifically designed for our work, and is available in a line of products. This implies that the system we developed is not for a single use case study, but can be further extended for different use case scenarios when required. Additionally, the availability of the pen allows the collection of data in a parallel manner which can accelerate a study that uses this tool.

3.1 Data Collection Application

To collect the ground truth labels of the data, the Digipen is provided with a Devkit [1] guide for the development of a mobile application for interaction with the pen. The application provides two files with similar timestamps of the recording session that can be used to extract data samples in the form of training data with the relative labels. A sample is defined as a complete word recording that consists of a series of timesteps.

3.2 Data Recording

Recording sessions were conducted in parallel using 16 different Digipens, with each session taking up to 45 min of recording time. A set of 500 words was used to collect the main set of data used for the system. Single words were displayed on the screen of a tablet, the users were asked to write the word on paper, in their own handwriting style, using the Digipen.

The dataset included recordings from 61 participants who volunteered to contribute to our study, with some participants contributing less than the 500 required words due to time constraints. The number of samples collected was 27961 word samples.

In addition to the main dataset, a separate dataset (unseen words set) was recorded from two other individuals. This recording consisted of random words selected from a set of 98463 words, different from the main set, serving as a second test set, with the purpose of testing the results of the system on unseen words. The final count of this set was 1006 sample recordings. Figure 2 shows histograms of the data count of both datasets with respect to the lengths of the samples and labels separately.

3.3 Data Preparation

Data recording is subject to faults during the process. To ensure that our model was trained on valid data, all hovering data before and at the end of each single sample recording was trimmed out. This was achieved by removing the data associated with force sensor readings below a pre-specified threshold at the beginning and the end of a recording. From the first time this threshold was exceeded within that recording, all data was kept even when the force reading temporarily

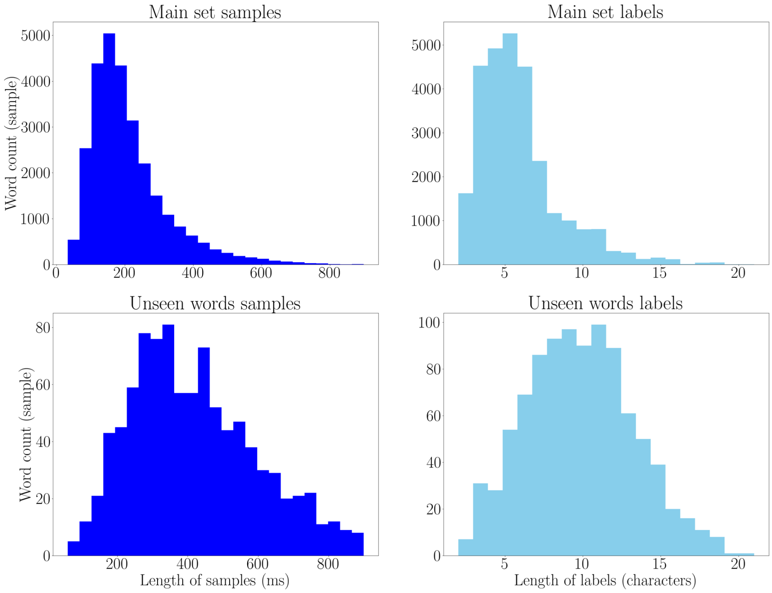

Fig. 2. A histogram displaying the number of samples in relation with (left) length of samples and (right) length of labels, in both (above) the main set and (below) the second test set.

fell below that threshold. The threshold was determined experimentally through monitoring the highest force sensor values while hovering with the pen. Additionally, samples which appeared too short or long to be correct recordings were considered as faulty recordings and removed from the dataset. The data was then normalized per sample using the z-score normalization in order to input data features of a similar scale into the model. No further preprocessing or feature extraction was applied.

4 End-to-End Models

The architectures described in this section were inspired by research aimed at developing end-to-end recognition models in handwriting [8,16] and speech recognition [10]. The use of Recurrent Neural Networks (RNNs), distinctively, Long Short-Term Memory networks (LSTMs), is common in the applications of handwriting and speech recognition due to the ability to transcribe data into sequences of characters or words while preserving sequential information. Bidirectional RNNs (BRNNs) make use of both past and future contextual information at every position of the input sequence in order to calculate the output sequences, and Bidirectional LSTMs (BLSTMs) have shown to achieve the best recognition results in the context of phoneme recognition when compared to other neural networks [17].

The models presented in this paper take input multivariate time series data samples of different lengths, comprised of 13 channels, representing the tri-axial measurements of the three IMU sensors and the magnetometer, in addition to the force sensor. In this section we evaluate different model architectures, describe the data splitting and the training process using raw sensor data.

4.1 Model Architectures

In the context of handwriting recognition using positional data, a model consisting of BLSTM layers proved sufficient to achieve the best recognition rates with the use of extracted feature vectors [16], or resampled raw stroke data [8], while CNN models obtained the best character classification accuracy in systems using raw sensor data with the Digipen [35,47].

Following recent studies, we included the CNN model in our study. The model included four 1D-Convolutional layers, consisting of $1024, 512, 256, 128$ feature maps, consecutively, with kernel sizes of $5, 3, 3, 3$, respectively, and a fully connected layer of 100 units.

In contrast to positional data, in our case the input sequences are long due to a high sampling rate. Downsampling is not a viable option with IMU data because it leads to the loss of critical information [7]. Therefore, in addition to the CNN model, we implemented a CLDNN model (including Convolutional, LSTMs, and fully connected layers), which is typically used in speech recognition [10], where data samples are of high sampling rates and BLSTM models lead to latency constraints. The Convolutional layers reduce the dimensionality of the input features, which reduces the temporal variations within the LSTMs, which are then fed into the Dense layers where the features are tranformed into a space that makes that output easier to classify [39]. Hence, a CLDNN model allows to avoid latency constraints and slow training and prediction times which occur with BLSTM models. The model consisted of three Convolutional layers, followed by two BLSTM layers and a single fully connected layer. The Convolutional layers comprised of $512, 256, 128$ feature maps with kernel sizes of $5, 3, 3$, respectively. The BLSTM layers were of 64 units each, and the fully connected layer included 100 units. A grid search was implemented to determine the optimal hyperparameters setup.

In both described models, Batch Normalization [20] and Max Pooling (of size 2) were applied after each Convolutional layer. The Relu activation was used in the Convolutional and the fully connected layers, while Tanh was used in the BLSTM layers. Random dropout [44] with a dropout rate of 0.3 was applied after each layer to prevent overfitting and improve robustness of the system.

Similarly to the current developed systems in the field, we relied in our model on the Connectionist Temporal Classification (CTC) loss [14] with a Softmax output layer which provides an implicit segmentation of the data. The CTC is an RNN loss function that enables labeling whole sequences at once. It uses the

Table 1. Number of samples in each set per fold (word samples).

Sets	Folds				
	1	2	3	4	5
Training	18452	18655	17588	17579	17598
Validation	4614	4664	4397	4395	4400
Test	4895	4642	5976	5987	5963

network to provide direct mapping from an input sequence to an output label without the need of segmenting the data. It introduces a 'blank' character that is used to find the best alignment of characters that best interprets the input.

4.2 Model Training

We split the main dataset into five folds, distributed into 49 users in the training set and 12 users in the test set, and train our model on the different folds separately. No writer appears in both sets to consider a writer-independent recognition task. The training data for each fold was divided into an 80/20 (training/validation) split. The unseen words dataset was used to test the effectiveness of the models for unseen word data. Table 1 shows the different training, validation, and seen test sets, per each fold, not including any unseen words data.

For the implementation of our models, we used Keras/Tensorflow(v1) python libraries [3,9], which include standard functions required for our work. The models were trained using a batch size of 64 samples and optimized with the Adam Optimizer [26] with a starting learning rate of 10^{-2}. A learning rate scheduler was implemented to monitor the validation loss and decrease the learning rate with a patience of 10 epochs and a factor of 0.8. We trained the models until the validation loss showed no decrease for 20 iterations after the minimum learning rate of 10^{-4} was reached, and saved the best model determined by the lowest validation loss during training. Finally, the evaluation of our model required the decoding of the CTC output into a word interpretation for which we used the Tensorflow standard CTC decoder function with a greedy search that returns the most likely output token sequence without the use of a dictionary.

5 Evaluation and Discussion

Table 2 presents the average results obtained. In terms of word recognition, the CNN model achieved the higher error rate of 35.9% and 31.65% average CER for seen and unseen words, respectively, which implies that the even though a CNN model achieved good results in character recognition [35,47], it was not sufficient for the CTC to find the best character alignment within a word sample. The higher recognition rates were achieved by the CLDNN model, with an average of 17.97% and 17.10% CER. The models recognized unseen words

Table 2. Average error rates with respect to model properties.

Models	Seen words Avg. % CER	UnSeen Words Avg. % CER	Avg. epochs	Trainable parameters	Seconds per epoch
CNN	35.90 (± 2.01)	31.65 (± 1.07)	153	2,154,957	53
CLDNN	17.97 (± 1.98)	17.10 (± 1.68)	236	743,373	102

without distinction from seen words, since the CTC learns to identify individual characters within the data. Additionally, having users in two distinct test sets different from users in the training sets provided a user-independent recognition model.

Considering the different models in regard with the model complexity and time performance, Table 2 shows the training time with respect to the trainable parameters of each model, in addition to the training iterations required to converge to the best performance. The CNN model consisted of a larger number of training parameters, however required lower training and prediction times. The CLDNN model achieved the better complexity to performance ratio with a significantly better recognition rate yet a longer training time relative to the CNN.

In addition to the average CER, Fig. 3 reports the average Levenshtein Distance per label length for both test sets using the CLDNN model. This shows the minimum number of character edits, including insertions, deletions and substitutions, required to change a predicted word into the ground truth label. This means that the prediction of our model was on average divergent by 0.98 and 1.66 character edits for the average length 5.59 and 9.72 characters for the seen and unseen test sets, respectively. A detailed analysis of the errors showed that an average of 68% of the predicted words were missing characters, which is due to cursive writing. 26% of the prediction were of a substitution nature, which occurs between characters that look similar in both uppercase and lowercase, such as ‘P-p’, ‘K-k’, and ‘S-s’, while 6% only included more characters than the relative ground truth, which occurs with multiple stroke characters.

The model used in our system followed the common used model in HWR systems, both offline and online, which is a stack of Convolutional or Recurrent layers trained with the CTC loss, and achieved an overall recognition rate similar to previous position-based models that did not make use of languages models [16]. However, this result is not directly comparable with previous systems, since these systems were trained on different data types, with sentence data, while our dataset consists of word data. Additionally, the state-of-the-art models in positional-based systems make use of complex language models. Moreover, the public IAM-OnDB dataset includes a higher number of classes in comparison to our dataset. Nonetheless, the presented results suggest that our system is on an adjacent level in terms of recognition rates without the use of a dictionary.

Our system did not show the same level of recognition rates in comparison with the wearable systems described, which were trained on different datasets using distinct hardware. These systems followed the segment-and-decode approach with separate system-specific extracted features from uni-stroke data. Such

Table 3. IMU-based pen recognizers with respective data type and performance, including the results of the presented CLDNN model. Digipen represents the pen used in this work.

	Data type	Recognition type	Accuracy (%)
[23]	Acc	10 Digits	84.8
[46]	Acc	10 Digits	98
[40]	Acc, Gyr, Sound	10 Digit	98.33
[27]	**Digipen**	26 Characters	52
[41]	Acc, Gyr, Mag	26 Characters	82
[47]	**Digipen**	26 Characters	86.97
[35]	**Digipen**	52 Characters	64.13
Ours	**Digipen**	**Words**	**82.92 (CRR)**

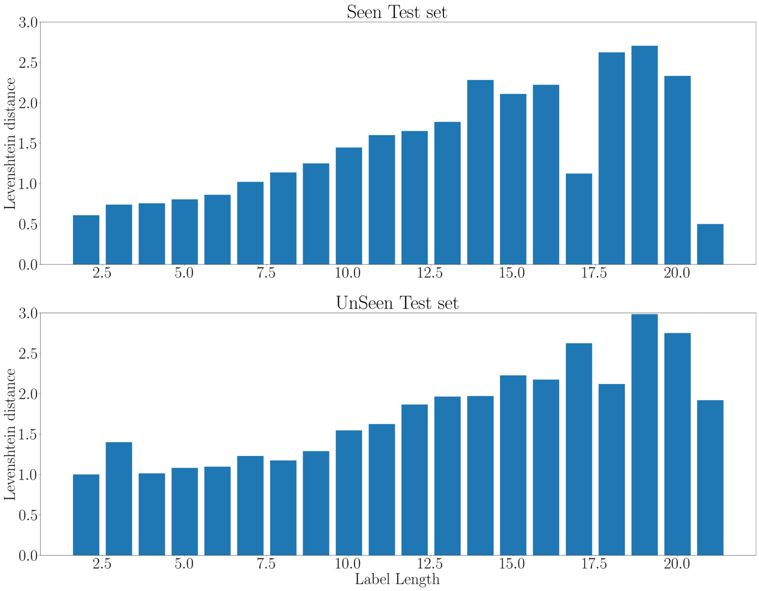

Fig. 3. A bar graph displaying the average Levenshtein Distance per label length for the seen and unseen test sets evaluated using the CLDNN model.

systems provided air-writing capability, which does not fit for our paper-writing recognizer. The wearable ring presented in [29] was designed for on-surface writing, however, the system was developed and evaluated for a single specific writer. Also, writing with finger does not present the same efficiency in comparison with pen writing.

Considering paper-writing recognition using sensor-equipped pens, our system achieved significant results in comparison to previously developed systems. Even though some previous systems used different hardware, our system, to the best knowledge of the authors, is the first IMU-based pen system that enables word recognition. Table 3 shows a summary of the described sensor-equipped pens in Sect. 2. Moreover, our model achieved an improved character recognition rate (CRR) by 18.79% for the 52 Latin alphabet characters relatively to previous systems using the same hardware.

6 Conclusion and Future Work

In this paper, we presented a system that applies OHWR by writing on normal paper using an IMU-enhanced ballpoint pen. We described the data collection tools and process in detail, and provided a complete system setup. We trained CNN and CLDNN end-to-end models that take normalized raw sensor data as input, and output word interpretations using the CTC loss with a greedy search decoder. The models were trained and evaluated using a five-fold cross-validation method, with test users being different from the users in the training set. We also evaluated the models on a separate test set to evaluate the efficiency of our system for unseen words. The presented CLDNN model showed the best performance without distinction between seen and unseen words.

Our system showed significant improvements in comparison with previously presented character recognition systems using digital pens. With the results presented in this work, we showed that sensor-enhanced pens are efficient and yield promising results in the OHWR field in which digitizing writing on paper is required. Accordingly, to further improve the applications of OHWR using digital pens, the dataset used in this work is planned to be published for use in the scientific community. Future work following this will include complete sentence recognition, in addition to including digits and punctuation marks. Finally, the end-to-end model we presented requires minimal preprocessing, and mainly depends on the data, and thus to increase the robustness of a language recognizer, we plan to pair our model with a distinct dictionary or language model specific to the language to be recognized.

Acknowledgments. This work was supported by the Bayerisches Staatsministerium für Wirtschaft, Landesentwicklung und Energie as part of the EINNS project (Entwicklung Intelligenter Neuronaler Netze zur Schrifterkennung) (grant number IUK-1902-0005 // IUK606/002). Bjoern Eskofier gratefully acknowledges the support of the German Research Foundation (DFG) within the framework of the Heisenberg professorship program (grant number ES 434/8-1).

References

1. The digipen devkit. https://stabilodigital.com/devkit-demoapp-introduction/. Accessed 30 Jan 2021
2. The digipen hardware. https://stabilodigital.com/sensors-2021/. Accessed 30 Jan 2021

3. Abadi, M., et al.: Tensorflow: a system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), pp. 265–283 (2016)
4. Abd Alshafy, H.A., Mustafa, M.E.: Hmm based approach for online Arabic handwriting recognition. In: 2014 14th International Conference on Intelligent Systems Design and Applications, pp. 211–215. IEEE (2014)
5. Amma, C., Georgi, M., Schultz, T.: Airwriting: hands-free mobile text input by spotting and continuous recognition of 3d-space handwriting with inertial sensors. In: 2012 16th International Symposium on Wearable Computers, pp. 52–59. IEEE (2012)
6. Bengio, Y., LeCun, Y., Nohl, C., Burges, C.: LEREC: a NN/HMM hybrid for on-line handwriting recognition. Neural Comput. **7**(6), 1289–1303 (1995)
7. Bersch, S.D., Azzi, D., Khusainov, R., Achumba, I.E., Ries, J.: Sensor data acquisition and processing parameters for human activity classification. Sensors **14**(3), 4239–4270 (2014)
8. Carbune, V., et al.: Fast multi-language LSTM-based online handwriting recognition. In: International Journal on Document Analysis and Recognition (IJDAR), pp. 1–14 (2020)
9. Chollet, F., et al.: Keras (2015). https://keras.io
10. Feng, Y., Zhang, Y., Xu, X.: End-to-end speech recognition system based on improved CLDNN structure. In: 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), pp. 538–542. IEEE (2019)
11. Gan, J., Wang, W.: In-air handwritten English word recognition using attention recurrent translator. Neural Comput. Appl. **31**(7), 3155–3172 (2019)
12. Gan, J., Wang, W., Lu, K.: A unified CNN-RNN approach for in-air handwritten English word recognition. In: 2018 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2018)
13. Gerth, S., et al.: Is handwriting performance affected by the writing surface? comparing tablet vs. paper. Frontiers in psychology 7 (2016)
14. Graves, A., Fernández, S., Gomez, F., Schmidhuber, J.: Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 369–376 (2006)
15. Graves, A., Fernández, S., Liwicki, M., Bunke, H., Schmidhuber, J.: Unconstrained online handwriting recognition with recurrent neural networks. In: Advances in Neural Information Processing Systems 20, NIPS 2008 (2008)
16. Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J.: A novel connectionist system for unconstrained handwriting recognition. IEEE Trans. Pattern Anal. Mach. Intell. **31**(5), 855–868 (2008)
17. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. **18**(5–6), 602–610 (2005)
18. Guyon, I., Schomaker, L., Plamondon, R., Liberman, M., Janet, S.: Unipen project of on-line data exchange and recognizer benchmarks. In: Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5), vol. 2, pp. 29–33. IEEE (1994)
19. Halder, A., Ramakrishnan, A.: Time delay neural networks for online handwriting recognition, June 2007. https://doi.org/10.13140/RG.2.2.25975.52641

20. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456. PMLR (2015)
21. Jaeger, S., Manke, S., Reichert, J., Waibel, A.: Online handwriting recognition: the NPEN++ recognizer. Int. J. Doc. Anal. Recogn. **3**(3), 169–180 (2001)
22. Jäger, S., Liu, C.L., Nakagawa, M.: The state of the art in Japanese online handwriting recognition compared to techniques in western handwriting recognition. Doc. Anal. Recogn. **6**(2), 75–88 (2003)
23. Jeen-Shing, W., Yu-Liang, H., Cheng-Ling, C.: Online handwriting recognition using an accelerometer-based pen device. In: 2nd International Conference on Advances in Computer Science and Engineering (CSE 2013), pp. 231–234. Atlantis Press (2013)
24. Keysers, D., Deselaers, T., Rowley, H.A., Wang, L.L., Carbune, V.: Multi-language online handwriting recognition. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1180–1194 (2016)
25. Kim, J.H., Sin, B.-K.: Online handwriting recognition. In: Doermann, D., Tombre, K. (eds.) Handbook of Document Image Processing and Recognition, pp. 887–915. Springer, London (2014). https://doi.org/10.1007/978-0-85729-859-1_29
26. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
27. Koellner, C., Kurz, M., Sonnleitner, E.: What did you mean? an evaluation of online character recognition approaches. In: 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 1–6. IEEE (2019)
28. Liu, C.L., Jaeger, S., Nakagawa, M.: 'online recognition of Chinese characters: the state-of-the-art. IEEE Trans. Pattern Anal. Mach. Intell. **26**(2), 198–213 (2004)
29. Liu, Z.T., Wong, D.P., Chou, P.H.: An IMU-based wearable ring for on-surface handwriting recognition. In: 2020 International Symposium on VLSI Design, Automation and Test (VLSI-DAT), pp. 1–4. IEEE (2020)
30. Liwicki, M., Bunke, H.: IAM-ONDB-an on-line English sentence database acquired from handwritten text on a whiteboard. In: Eighth International Conference on Document Analysis and Recognition (ICDAR 2005), pp. 956–961. IEEE (2005)
31. Liwicki, M., Bunke, H., Pittman, J.A., Knerr, S.: Combining diverse systems for handwritten text line recognition. Mach. Vis. Appl. **22**(1), 39–51 (2011)
32. Liwicki, M., Graves, A., Fernàndez, S., Bunke, H., Schmidhuber, J.: A novel approach to on-line handwriting recognition based on bidirectional long short-term memory networks. In: Proceedings of the 9th International Conference on Document Analysis and Recognition, ICDAR 2007 (2007)
33. Mandal, S., Prasanna, S.M., Sundaram, S.: Exploration of CNN features for online handwriting recognition. In: 2019 International Conference on Document Analysis and Recognition (ICDAR), pp. 831–836. IEEE (2019)
34. Musa, M.E.: Towards building standard datasets for Arabic recognition. Int. J. Eng. Adv. Res. Technol. (IJEART) **2**(2), 16–19 (2016)
35. Ott, F., Wehbi, M., Hamann, T., Barth, J., Eskofier, B., Mutschler, C.: The onhw dataset: Online handwriting recognition from imu-enhanced ballpoint pens with machine learning. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 3, pp. 1–20 (2020)
36. Palacios, R., Gupta, A., Wang, P.S.: Handwritten bank check recognition of courtesy amounts. Int. J. Image Graphics **4**(02), 203–222 (2004)
37. Plamondon, R., Srihari, S.N.: Online and off-line handwriting recognition: a comprehensive survey. IEEE Trans. Pattern Anal. Mach. Intell. **22**(1), 63–84 (2000)

38. Priya, A., Mishra, S., Raj, S., Mandal, S., Datta, S.: Online and offline character recognition: a survey. In: 2016 International Conference on Communication and Signal Processing (ICCSP), pp. 0967–0970. IEEE (2016)
39. Sainath, T.N., Vinyals, O., Senior, A., Sak, H.: Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4580–4584. IEEE (2015)
40. Schrapel, M., Stadler, M.L., Rohs, M.: Pentelligence: Combining pen tip motion and writing sounds for handwritten digit recognition. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–11 (2018)
41. Shaikh Jahidabegum, K.: Character recognition system for text entry using inertial pen. Int. J. Sci. Eng. Technol. Res. (IJSETR) 4 (2015)
42. Singh, A., Bacchuwar, K., Bhasin, A.: A survey of OCR applications. Int. J. Mach. Learn. Comput. **2**(3), 314 (2012)
43. Srihari, S.N.: Recognition of handwritten and machine-printed text for postal address interpretation. Pattern Recogn. Lett. **14**(4), 291–302 (1993)
44. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
45. Tlemsani, R., Belbachir, K.: An improved Arabic on-line characters recognition system. In: 2018 International Arab Conference on Information Technology (ACIT), pp. 1–10. IEEE (2018)
46. Wang, J.S., Chuang, F.C.: An accelerometer-based digital pen with a trajectory recognition algorithm for handwritten digit and gesture recognition. IEEE Trans. Industr. Electron. **59**(7), 2998–3007 (2011)
47. Wehbi, M., Hamann, T., Barth, J., Eskofier, B.: Digitizing handwriting with a sensor pen: a writer-independent recognizer. In: 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR),pp. 295–300. IEEE (2020)
48. Yaeger, L.S., Webb, B.J., Lyon, R.F.: Combining neural networks and context-driven search for online, printed handwriting recognition in the newton. AI Mag. **19**(1), 73–73 (1998)