A high-performance word recognition system for the biological fieldnotes of the *Natuurkundige Commissie*[⋆]

Mahya Ameryan[1] and Lambert Schomaker[1]

[1]Artificial Intelligence, Faculty of Science and Engineering, University of Groningen,
Groningen, The Netherlands
mahya.ameryan@gmail.com
l.r.b.schomaker@rug.nl

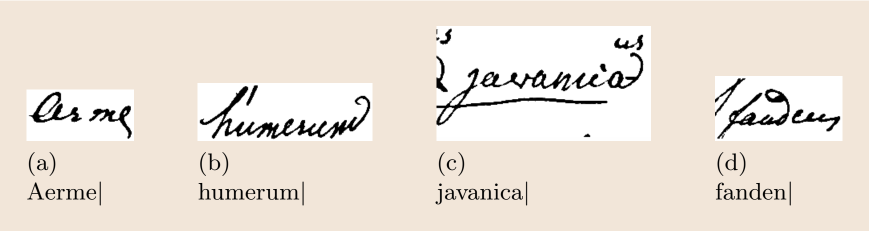
Abstract. In this research, a high word-recognition accuracy was achieved using an e-Science friendly deep learning method on a highly multilingual data set. Deep learning requires large training sets. Therefore, we use an auxiliary data set in addition to the target data set which is derived from the collection *Natuurkundige Commissie*, years 1820-1850. The auxiliary historical data set is from another writer (van Oort). The method concerns a compact ensemble of Convolutional Bidirectional Long Short-Term Memory neural networks. A dual-state word-beam search combined with an adequate label-coding scheme is used for decoding the connectionist temporal classification layer. Our approach increased the recognition accuracy of the words that a recognizer has never seen, i.e., out-of-vocabulary (OOV) words with 3.5 percentage points. The use of extraneous training data increased the performance on in-vocabulary words by 1 pp. The network architectures in an ensemble are generated randomly and autonomously such that our system can be deployed in an e-Science server. The OOV capability allows scholars to search for words that did not exist in the original training set.

Keywords: Historical handwriting recognition · Convolutional Bidirectional Long Short-Term Memory (CNN-BiLSTM) · E-Science server.

1 Introduction

Historical manuscripts are an important aspect of cultural heritage [22]. Extracting information from them by e-Science servers would be helpful for scholars and the general public. An e-Science server is the application of computationally intensive modern methods for data collection, preparation, experimentation, result

[⋆] This work is part of the research program Making Sense of Illustrated Handwritten Archives with project number 652-001-001, which is financed by the Netherlands Organization for Scientific Research (NWO). We would like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.



(a) Aerme| (b) humerum| (c) javanica| (d) fanden|

Fig. 1: Allographic variability and oscillatory slur in handwritten words. The German word, 2 *Aerme*, contains two different shapes of 'e'; (b) The Latin word, *humerum*, has two allographs for the letter 'm'; (c) The Latin word, *javanica*, has two allographs for the letter 'a'; (d) The German word, *fanden*, has two allographs for the letter 'n'. The gray-scale image samples are taken from the MkS data set and labeled using an extra-separator coding scheme [2]. Also note the post-hoc superscript note suggesting the spelling *javanicus.*

dissemination, and long-term maintenance, e.g., *Monk* [19, 20, 6, 15, 16]. The *Monk* e-Science server is a live, web-driven search engine for character and word recognition, annotation, and retrieval. It contains various handwritten historical and contemporary manuscripts in numerous languages: Chinese, Dutch, Thai, Arabic, English, German, and Persian. Additionally, intricate machine-printed documents such as Egyptian hieroglyphs and Fraktur are available in the *Monk* e-Science server.

Ideally, the use of deep learning methods should be beneficial in an e-Science server context. Deep learning paradigms, especially Long short-term memory (LSTM [7]), have shown superior performance in solving handwriting-recognition problems. However, these methods have two drawbacks. First, they demand large training sets. Secondly, the design of optimal neural architectures requires human supervision [21], which is in contradiction with the principle of autonomy in artificial intelligence. Therefore, the plain application of such a fine-tuned LSTM-based recognizer is not feasible in e-Science servers. In this research, we use a very recent proven homogeneous ensemble of end-to-end trainable convolutional Bidirectional -LSTMs, conveniently applicable on e-Science servers. The target data set is derived from the 17,000-page manuscript of biological field notes of the Dutch *Natuurkundige Commissie* (NC) in the Indonesian Archipelago between the years 1820 to 1850. Although authors attempt to track their knowledge and observations in a disciplined way along with fascinating drawings, this systematic does not occur in the hand-written script, e.g., allographs differ depending on the serial position in a word. Namely, a letter at the beginning and middle of a word is mostly well-formed. However, it can have an oscillatory slur as an ending character. Fig. 1 shows different allographs for the letters 'e', 'm', 'a', and 'n'. The samples are derived from the MkS data set. Furthermore, the ground-truth label in each data set is presented in a particular way, due to the designers handcrafting. As an example, in the ground-truth labels of a standard benchmark Arabic data set, characters or ligatures are separated by a token (|)

[13]. Moreover, there are specific requirements for each recognizer method. The introduction of new form-based alphabets or feature extraction approaches has addressed this significance [10, 4]. As a consequence, we considered it necessary to see whether the labeling systematics of Latin-based handwritten scripts has an effect on recognition performance. A lot of effort has been put into labeling the massive target collection, which has resulted in near 10k labeled word images. In deep learning, this is not enough data to obtain optimal recognition accuracy. Therefore, we are faced with the question of whether another data set that has textual and stylistic similarities to the target data set can be useful. An auxiliary data set is extracted from hand-written diaries of Pieter van Oort [23], written in a very neat way contrary to the NC collection. Still, this text comes from the same historical period and the same geographical context. It is a limitation if a recognizer can only recognize the lexical words from the training set. We invistigate what is the performance on lexical test words that were never seen by the recognizer. Another question is whether it is possible to design an optimized recognizer less dependent on the presence of a machine-learning expert using generating neural networks autonomously within an e-Science server.

Background

In current literature, good results are obtained using LSTM networks, however, at the expense of a tremendous amount of training, using large ensembles of up to a thousand neural networks [17]. In our work, we aim at reaching similar performances using a more compact approach, using a small ensemble. In [11], an ensemble is composed of eight recognizers: four architectures of a recurrent neural network (RNN); a grapheme-based MLP-HMM; two different variants of a context-dependent sliding window based on GMM-HMM. Such heterogeneous architecture demands a lot of engineering efforts. Furthermore, in this study, we will use the dual-state word-beam search (DSWBS) CTC decoder [14]. This method concerns two states in the word-search process: A word state and a non-word state. Each character can be either a *word-character* or *non-word-character*. A search beam is an evolving list of most likely word candidates. A search beam's temporal evolution is based on its state. A beam in the word-state can be extended by entering a *word-character*. Entering a *non-word character* brings the system to the non-word state. A beam in the non-word state is extended by entering a *non-word-character* while entering a *word character* ends the beam and brings the system to the word state. This *word character* is the beginning of a new beam. This state information can be used in conjunction with a prefix table to trace the best possible (partial) word hypotheses, from an uncertain begin state until an uncertain end-of sequence state.

The label-coding scheme may have a significant effect on performance due to the varying demands of different methods [10, 4]. In [2], it is shown that the combination of DSWBS and a proper label-coding scheme is effective compared to using DSWBS combined with a Plain label-coding scheme. In the Plain coding scheme only the characters in the word image appear in the corresponding ground-truth label. It is reported that stressing the word-ending shape with an extra token is beneficial when DSWBS is used as the CTC decoding method (this is the 'extra-label coding scheme' [2]). However, stressing the start-of-word shapes with a token is detrimental for recognition accuracy.

Given the lack of labeled data in handwriting recognition, as compared to, e.g., speech recognition, it is important to be able to use existing labeled corpora. Even more than in the case of speech, the use style variation in handwriting often inhibits the effective use of pre-existing data sets. To address this issue, transfer learning has been applied. However, this may only have a clearly improved performance in case of style similarity, e.g., in case of a well-written two-author single-language historical script [3]. Transfer over different historical script style periods remains notoriously difficult. The new element in this study is that we will use additional labeled data from a different writer and different document type, but from the same historical period and a comparable colonial context. The research question is whether it is possible to increase the performance on a target manuscript, even if style and content are different?

2 Method

2.1 Pre-processing and augmentation

The input consists of human-labeled, isolated-word grey-scale images from the the *Monk* e-Science server [15, 16]. The pre-processing and data augmentation are performed in three steps: random stretching/squeezing of a gray-scale image in the direction of width; re-sizing an image into 128 × 32 pixels; and finally, an intensity normalization. These three steps are conducted in each epoch of the training process, yielding different augmentations per epoch. Baseline alignment or deslanting methods are not applied in this procedure.

2.2 Label-coding scheme

We used a novel extra-separator label-coding scheme for words. This method delivered an improvement of 2 to 3 percentage points in accuracy [2]. In this scheme, an additional unique character (e.g. '|') is concatenated to the end of the ground-truth label. This character gives an extra hint relating the ending-of-word shape condition to the recognizer, which we have shown to be effective in several data sets [2].

2.3 Neural Network

We used a proven ensemble of five end-to-end training Convolutional Bidirectional Long Short-Term Memory neural networks (CNN-BiLSTMs), Fig. 2 [2, 1]. Each of the five CNN-BiLSTMs consists of five front-end convolutional layers, three BiLSTM layers, and a connectionist temporal classification (CTC [5]) layer. The architecture of the CNN-BiLSTMs in the ensemble only differ in the number of hidden units of three of the five convolutional layers, layer 2, 3, and 4. Table 1 shows the number of hidden units for the five CNN-BiLSTMs, A_i, $i = 1..5$.

The convolutional layers in a CNN-BiLSTM are used for feature extraction. The first convolutional layer is fed by the pixel intensity of the input word-image. Each convolutional layer has four stages: (1) convolution operation; (2)

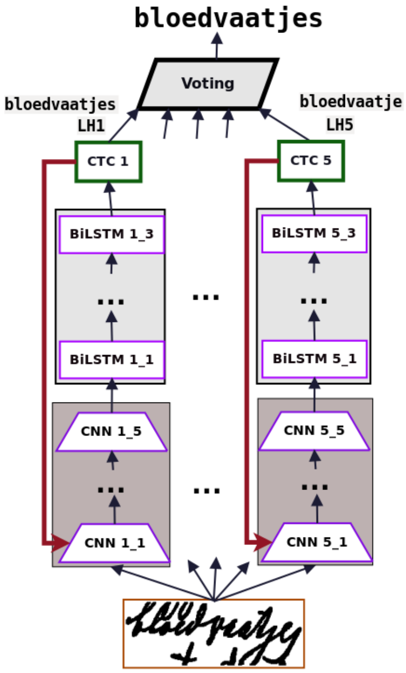


Fig. 2: The ensemble of five end-to-end training convolutional BiLSTM neural networks. Each network provides a label hypothesis with its relative likelihood (LH_i, $i = 1..5$.)

Table 1: Number of hidden units in the convolutional layers of five CNN-BiLSTMS architectures (A_i, $i = 1..5$). l_j indicates jth layer of a CNN front-end, where $j = 1..5$.

Layer / Arch.	Hidden unit size				
	l_1	l_2	l_3	l_4	l_5
A_1	128	256	256	256	512
A_2	128	256	512	512	512
A_3	128	128	256	256	512
A_4	128	128	512	512	512
A_5	128	128	128	256	512

intensity normalization; (3) the ReLU activation function [12]; (4) a max-pooling layer. The networks use RMSProp [18] as the gradient descent function. No drop out is applied. The size of the batches is 50 word images. The output of the fifth convolutional layer is fed to the three-layer BiLSTM. Each BiLSTM layer contains 512 hidden units.

2.4 A connectionist temporal classification (CTC)

There are $|A + 2|$ output units in the CTC layer, where $|A|$ is the size of the alphabet of the training lexicon labeled using the plain label-coding scheme [2]. The other two residual units are for the extra separator (e.g., '|'), and a special blank [5] for CTC, which presents observing 'no label' and differ the space character. Dual-State Word-Beam Search (DSWBS) [14] is used for CTC decoding. In our research, DSWBS uses a prefix tree formed a given lexicon without using any statistical language model.

2.5 A voting module

For an input image, each of five CNN-BiLSTMs produces a word hypothesis with its relative likelihood value. The five hypotheses and likelihood values are the input of the voting module. The hypotheses are divided into subsets. Afterward, the final label of the input image is determined using three rules:

1. Plurality → choose it.
2. Tie → choose the subset with the maximum average likelihood value.
3. Only singleton sets → choose the subset with maximum likelihood value.

3 Data sets

The MkS data set is derived from 17,000 pages of biological field-notes of the *Natuurkundige Commissie* in the Indonesian Archipelago between the years 1820 to 1850 [8, 9]. The manuscript is stored in the Naturalis Biodiversity Center in Leiden, the Netherlands. Sparsely, 950 pages are labeled in this project (NNM001001033-7). The manuscript has a wide variety of languages, including German, Latin, Dutch, Malay, Greek, and French. After the random data split for 5-fold cross-validation, the resulting proportion of out-of-vocabulary (OOV) words, i.e., words in the test set not appearing in the training set, is 31.9% (case-sensitive) and 29.5% (case-insensitive count). An in-vocabulary (INV) word exists both in the test set and in the training set. The daily routine of the committee is described in the field diary of Pieter van Oort, one of the major draftsmen and collectors [23]. The data sets used in this paper are summarized in Table 2.

The implementation of the network is based on the Tensorflow framework. The bar character (|) is selected as the stressed-ending sign because it is not present in the ground-truth labels of the MkS and Van Oort data sets. The evaluations were performed in a case-insensitive manner. We conducted 5-fold cross validation experiments. There are two training sets in our experiments: T_1 and T_2. The T_1 training set exclusively contains images of three folds of from MkS. T_2 has all of word images of the van Oort data set plus the images from three folds of the MkS data set. The proportion of OOV words in the test set equals 29.4% when T_1 is the training set [2] and 26.9% when T_2 is used as the training set.

Table 3 shows a comparison of word-recognition accuracy (%) of single recognizers (Table 1), and of the ensemble, separate for the training sets T_1 and T_2. Results are presented of average word accuracy and its standard deviation (av ±sd). Two CTC decoder variants are compared: lexicon-free best-path (BP [5]) and dual-state word-beam search (DSWBS [14]). A complete lexicon for this task is fed to DSWBS consisting of a list of all words occurring in the data sets (Table2). Two label-coding schemes (Plain vs Extra-separator) are compared when the training set is T_2. **Best path vs Dual-state word-beam search**:

Table 2: The data sets used in the experiments. Subscript *cs* denotes 'case sensitive', *ci* denotes 'case insensitive' lexicon size counting.

Set		Image samples#	Lexicon$\#_{cs}$	Lexicon$\#_{ci}$
MkS	Train	5,778	2648	2,497
	Validation	1,926	1,188	1,142
	Test	1,926	1,188	1,142
MkS	Full	9,630	3,730	3,494
Van Oort	Full	15,295	5,066	4,813

Table 3: A comparison of word-recognition accuracy of this work (T_2) with [2] (T_1) using Best-path (BP) and the dual-state word-beam search (DSWBS) CTC decoders in terms of average ±standard deviation (avg ±sd) and The ensemble. The Concise dictionary is applied for DSWBS. The Extra separator and Plain label-coding schemes are compared in this work.

	Architecture	[2](T_1)	This work (T_2)	This work (T_2)
CTC decoder	Label-coding scheme	Extra-coding scheme	Plain	Extra-coding scheme
DSWBS	a1	82.65	81.90	85.8
	a2	83.29	82.98	86.9
	a3	81.86	82.53	85.3
	a4	82.22	81.92	86.6
	a5	82.58	81.97	84.84
	avg ±sd	**82.52 ±1.1**	**82.26 ±1.0**	**85.88 ±1.0**
	Ensemble	**85.72 ±0.7**	**86.30 ±0.4**	**89.22 ±0.2**
BP	a1	53.57	54.75	55.48
	a2	54.30	56.49	57.15
	a3	54.27	55.32	55.12
	a4	53.26	55.04	57.53
	a5	54.80	54.65	54.77
	avg ±sd	**54.04 ±1.1**	**55.25 ±1.5**	**56.01 ±1.8**
	Ensemble	**61.57 ±1.5**	**63.76 ±0.5**	**64.20 ±1.0**

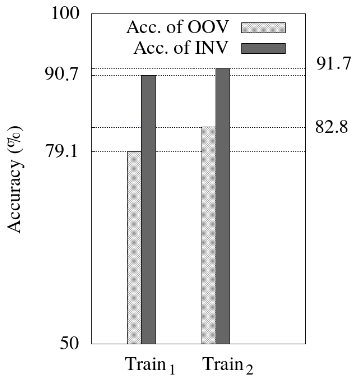
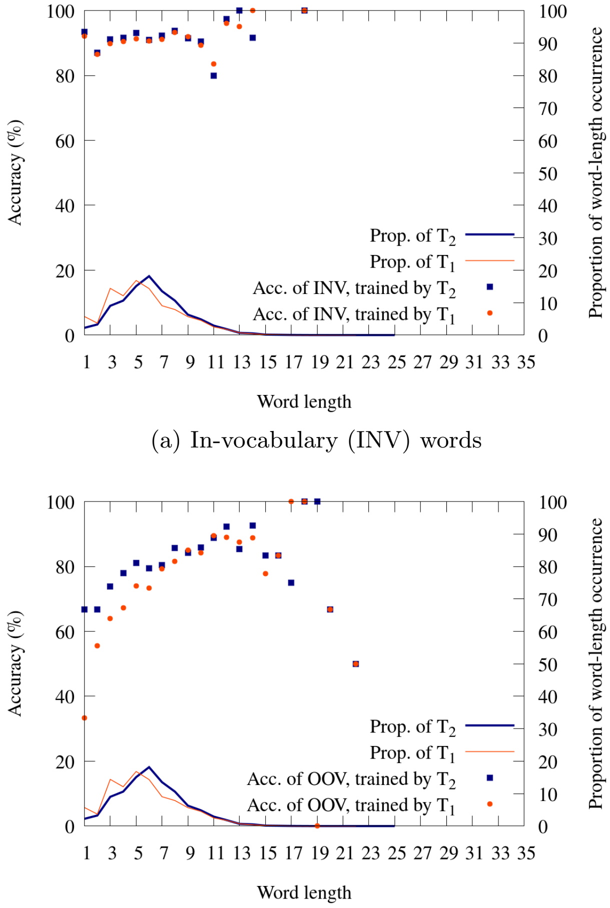


Fig. 3: A comparison of word-recognition accuracy of using the MkS data set as the training set (T_1 [2]) and using the MkS and van Oort data sets as the training set (T_2) for in-vocabulary (INV) and out-of-vocabulary (OOV) words. Using the additional training data yields to an improved accuracy ($p < 0.05$).

Results show that using a lexicon-based search method in the CTC layer significantly increases the word-recognition accuracy (more than 28 pp), as expected. **Single network vs Ensemble**: The ensemble voting improves the performance while its effect is stronger on a weaker method: 8 pp in the case of BP, and 3 pp in the case of DSWBS. The solution of ties increases the performance 1.3 pp when the DSWBS CTC decoder is used and 3.3 pp in the case of best-path CTC decoder, when the training set is the combined set T_2. **The five folds vs the five architectures**: There is no particular information for an architecture in one of the folds (Chi-squared test, N.S. $p >> 0.05$). **The five architectures**: The architectures within the ensemble do not differ (Chi-squared test, N.S. $p >> 0.05$).



(a) In-vocabulary (INV) words

(b) Out-of-vocabulary (OOV) words

Fig. 4: Comparison of word-recognition accuracy (%) of the ensembles when using T_1 and T_2 as training sets in term of word length of In-vocabulary (INV) and out-of-vocabulary (OOV) words per word length. The continuous lines indicate the proportion of words with a particular word length of the training sets.

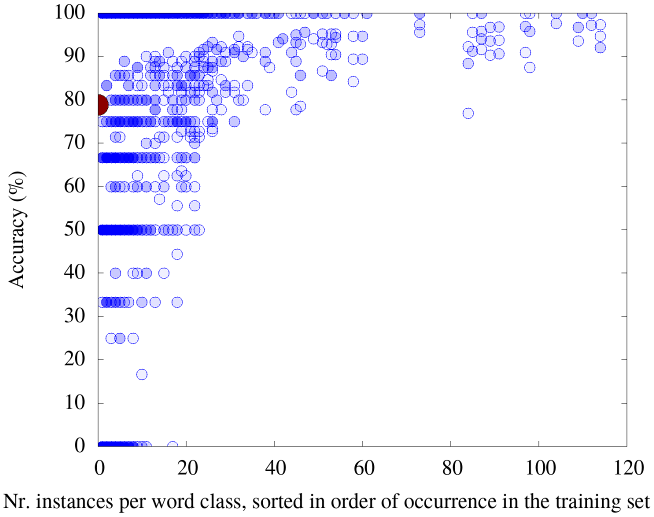


Fig. 5: The word-recognition accuracy of test words achieved by the five networks (Table 1) using 5-fold cross-validation and the T_2 training set. The *X-axis* present the number of instances per word class, sorted in the order of occurrence in the training set. The blue circles represent the in vocabulary (INV) word-test classes. The dark red circle shows the out-of-vocabulary (OOV) words. The horizontal stripes are due to the quantization of the test-set results, e.g., the horizontal stripe at an accuracy of 50% is the performance of 1 in 2, 2 in 4, 3 in 6 etc. instances per test-set word class.

The T_1 vs T_2 training sets: Using the van Oort data set along MkS data set (T_2) increases the word-recognition performance 2 to 3 pp (Chi-squared test, $p < 0.05$, significant). **The extra-separator label-coding scheme vs the plain label-coding scheme**: Using the extra-separator label-coding scheme increased the performance when DSWBS is used (Chi-squared test, $p << 0.05$, significant).

Since the OOV rate differs when T_1 and T_2 are training sets, we consider the words that do not appear in the lexicon of both training sets as the 'real' OOV words. Subsequently, the words that exist in the lexicon of both training sets are INV words. Therefore, 26.9% of the test set is counted as the OOV words and 70.6% of the test set concerns the INV words. Fig. 3 shows a comparison of word-recognition accuracy (%) of the ensemble when T_1 with T_2 are used as the training set, for OOV and INV words, separately. The figure illustrates that using T_2 increases the performance 1 pp on the INV words to 91.7% and 3.7 pp on OOV words to 82.8%. The reliable recognition of OOV words is important because it allows the scholars in the humanities to search for a word that does not exist in the training set. We scrutinized the performance of the ensembles in recognizing the INV and OOV words per word length, and we conduct a comparison when T_1 and T_2 are used as the training sets in Fig. 4. These figures also show the proportion of words with a particular word-length of both training sets. Fig. 4(a) shows that using the T_2 data set is slightly beneficial for the INV words which their word lengths are up to 8 characters. Fig. 4 (b) shows that using the van Oort data set along the MkS data set (T_2) increases the performance of the ensemble from 1 to 8 characters word-length in out-of-vocabulary words. For longer words its effect is not clear.

Fig. 5 shows the word-recognition accuracy of test words achieved by the five networks (Table 1) using 5-fold cross-validation and the T_2 training set. The *X-axis* shows the number of instances per word class, sorted in the order of occurrence in the training set. In an e-Science server, the X-axis roughly corresponds to time, starting with just a few instances per word class, increasing as more labels enter the learning system. The blue circles represent the in-vocabulary (INV) word-test classes. The dark red circle shows the average of word-recognition accuracy of the out-of-vocabulary (OOV) words. The average of recognition of OOV words is near 80% for a single recognizer when there is not any example of them in the training set. There are several *'threads'* in the figure. This is due to the quantized number of samples per word in a test set. For instance, for a word in the lexicon with three instances in the test set, the accuracies can only be 0%, 33%, 66% or 100%. The easy words are in the stripe at an accuracy of 100% and can be recognized well with just a few numbers of instances in the training set. On the other end, there are difficult words at the stripe of 0%, which are still not recognized after a dozen of training examples. This density plot gives a more realistic view into the origins of the performance, compared to a single average word accuracy that is computed over all words and all instances.

4 Discussion and conclusion

In this study, high word-recognition accuracy was achieved using a compact ensemble of five end-to-end convolutional neural networks. The method is applicable to e-Science servers where human intervention for hand-crafting the hyperparameters needs to be minimal. The ensemble uses Plurality voting, with special provision for ties. A proper label-coding scheme is applied. At the decoding stage, i.e., the CTC layer, a search method is applied which uses a prefix tree for a given lexicon, without using any statistical language model. The given intrinsic lexicon consists of all the words which appear in the target and auxiliary data sets. DSWBS shows only a slight drop of 0.4 pp in performance when using a large external lexicon, e.g., 30 times larger [1] and including the intrinsic word list. The results should give an indication for practical applications where the intrinsic lexicon is not known and a large external (public) lexicon is used. The target data set (MkS) is extracted from a historical collection that belonged to *Natuurkundige Commissie* in the Indonesian Archipelago between the years 1820 to 1850. MkS (T_1) is highly multilingual, in multiple styles, often in sloppy handwriting. We used an auxiliary data set, the van Oort handwritten diary, to boost the performance. This data set, unlike MkS, is written very neatly but belongs to the same historical period and geographical context.

The combination of MkS and van Oort (T_2) in training increased the performance of the ensemble 3.5 pp to 89%, where 1.3 pp is the effect of the solution of ties. The effect is more clear (+4pp) on the test words which the recognizer has never seen, i.e., the out-of-vocabulary (OOV) words. The comparison of two CTC decoder methods (with and without lexicon) confirms the significant effect of lexicon application (+25pp). The obtained high OOV word-recognition accuracy gives the user ability to search for terms that are not in the training set. The use of T_2 notably increased the performance for OOV words with a length of 1 to 8 characters. However, this effect is not apparent for longer OOV words. Reporting average accuracy hides the differences over word classes. We provided a detailed view of the word accuracy per class as a function of the number of examples. There may exist INV words which demand only a few numbers of instances in a training set to yield 100% accuracy. These are the easy words. On the contrary, for difficult words, the performance does not clearly increase when the number of instances for such a word-class increases in the training set. Some words do not reach 100% accuracy despite having more than 100 examples in the training data. The small increase of 1 percentage-point(pp) in the performance of the system on the in-vocabulary words when using auxiliary data may indicate that the number of word samples was probably already large enough. However, the 3.7 pp improvement of the performance on out-of-vocabulary words due to auxiliary training data reveals that for generalization, the system benefits from such data, even if it is partially different in style and content.

The proposed system can be deployed in an e-Science server such as the *Monk* e-Science server [15, 16]: The network architectures in an ensemble are generated randomly and autonomously. The approach also allows scholars to search for words that did not exist in the training set that they used.

The good results of this study are due to (a) a limited-size ensemble of LSTM networks using effective plurality voting; (b) an adapted label-coding scheme stressing on word ending shapes; (c) the use of dual-state word-beam search using a prefix lexicon and (d) the use of related but dissimilar handwriting in the training process. For future work, we intend to develop a handwritten line recognizer by extending our word-recognition approach.

References

1. Ameryan, M., Schomaker, L.: A limited-size ensemble of homoge-neous CNN/LSTMs for high-performance word classification (2019), https://arxiv.org/abs/1912.03223.
2. Ameryan, M., Schomaker, L.: Improving the robustness of LSTMs for word classification using stressed word endings in dual-state word-beam search. In: 17th Int. Conf. Frontiers in Handwriting Recognition (2020).
3. Dutta, K., Krishnan, P., Mathew, M., Jawahar, C.V.: Improving CNN-RNN hybrid networks for handwriting recognition. In: 2018 16th Int. Conf. Frontiers in Handwriting Recognition (ICFHR). pp. 80–85 (2018).
4. Fischer, A., Riesen, K., Bunke, H.: Graph similarity features for HMM-based handwriting recognition in historical documents. In: 12th Int. Conf. on Frontiers in Handwriting Recognit. pp. 253–258 (2010).
5. Graves, A., Fern´andez, S., Gomez, F., Schmidhuber, J.: Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In: Proc. 23rd Int. Conf. Mach. Learn. pp. 369–376 (2006).

6. He, S., Samara, P., Burgers, J., Schomaker, L.: Image-based historical manuscript dating using contour and stroke fragments. Pattern Recognition 58, 159–171 (2016).
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 9, 1735–1780 (1997).
8. Klaver, C.: Inseparable friends in life and death: the life and work of Heinrich Kuhl (1797-1821) and Johan Conrad van Hasselt (1797-1823). Barkhuis, Groningen (2007).
9. Mees, G., Achterberg, C.: Vogelkundig onderzoek op nieuw guinea in 1828. Zoologische Bijdragen 40, 3–64 (1994).
10. Menasri, F., Vincent, N, Cheriet, M., Augustin, E: Shape-based alphabet f or off-line arabic handwriting r ecognition. Nnth I nt. Conf. Document Analysis and Recognition (ICDAR 2007) 2, 969–973 (2007).
11. Menasri, F., Louradour, J., Bianne-Bernard, A.L., Kermorvant, C.: The A2iA French handwriting recognition system at the Rimes-ICDAR2011 competition. In: Viard-Gaudin, C., Zanibbi, R. (eds.) Document Recognition and Retrieval XIX. vol. 8297, pp. 263–270. International Society for Optics and Photonics, SPIE (2012).
12. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: Proc. the 27th Int. Conf. Machine Learning. pp. 807–814 (2010).
13. Pechwitz, M., Maddouri, S.S., Mrgner, V., Ellouze, N., Amiri, H.: IFN/ENIT database of handwritten Arabic word. In: 7th Colloque Int. Francophone sur l'Ecrit et le Document (2002).
14. Scheidl, H., Fiel, S., Sablatnig, R.: Word beam search: A connectionist temporal classification decoding algorithm. In: The Int. Conf. Frontiers of Handwriting Recognition (ICFHR). pp. 253–258. IEEE Computer Society (2018).
15. Schomaker, L.: A large-scale field test on word-image classification in large historical document collections using a traditional and two deep-learning methods. ArXiv (2019).
16. Schomaker, L.: Handwritten Historical Document Analysis, Recognition, and Retrieval State of the Art and Future Trends (chap.), in: Lifelong Learning for Text Retrieval and Recognition in Historical Handwritten Document Collections. World Scientific (November 2020).
17. Stuner, B., Chatelain, C., Paquet, T.: Handwriting recognition using Cohort of LSTM and lexicon verification with extremely large lexicon (2016).
18. Tieleman, T., Hinton, G.: Lect. 6.5-RMSProp: Divide the gradient by a running average of its recent magnitude. COURSERA:Neural Netw. for Mach learn. 4, 26–31 (2012).
19. Van der Zant, T., Schomaker, L., Haak, K.: Handwritten-word spotting using biologically inspired features. IEEE Trans. Pattern Anal. Mach. Intell. 30, 1945–1957 (2008).
20. Van der Zant, T., Schomaker, L., Zinger, S., Van Schie, H.: Where are the search engines for handwritten documents? Interdisciplinary Science Reviews 34(2-3), 224–235 (2009).
21. Voigtlaender, P., Doetsch, P., Ney, H.: Handwriting recognition with large multidimensional long short-term memory recurrent neural networks. In: 15th Int. Conf. Frontiers in Handwriting Recognition. pp. 228–233 (2016).
22. Weber, A., Ameryan, M., Wolstencroft, K., Stork, L., Heerlien, M., Schomaker, L.: Towards a digital infrastructure for illustrated handwritten archives. In: Ioannides, M. (ed.) Final Conf. of the Marie Sklodowska-Curie Initial Training Network for Digital Cultural Heritage, Olimje, Slovenia, May 23-25, 2017, LNCS, vol. 10605, 155-166. Springer International Publishing (2018).
23. Weber, A.: Collecting colonial nature: European naturalists and the Netherlands Indies in the early nineteenth century. Low Countries Historical Review 134-3, 72–95 (2019).