



Estimating the Optimal Training Set Size of Keyword Spotting for Historical Handwritten Document Transcription

Giuseppe De Gregorio^(✉)  and Angelo Marcelli 

DIEM, University of Salerno, Via Giovanni Paolo II 132, 84084 Fisciano, SA, Italy
{gdegregorio, amarcelli}@unisa.it

Abstract. We address the problem of estimating the tradeoff between the size of the training set and the performance of a KWS when used to assist the transcription of small collections of historical handwritten documents. As this application domain is characterized by a lack of data, and techniques such as transfer learning and data augmentation require more resources than those that are commonly available in the organizations holding the collections, we address the problem of getting the best out of the available data. For this purpose, we reformulate the problem as that of finding the size of the training set leading to a KWS whose performance, when used to support the transcription, allows to obtain the largest reduction of the human efforts to achieve the complete transcription of the collection. The results of a large set of experiments on three publicly available datasets largely adopted as a benchmark for performance evaluation show that a training set made of 5 to 8 pages is enough for achieving the largest reduction, independently of the actual pages included in the training set and the corresponding keyword lists. They also show that the actual time reduction depends much more on the keyword list than on the KWS performance.

Keywords: Historical document processing · Keyword spotting · Performance evaluation

1 Introduction

The quick advances that the field of Artificial Intelligence (AI) and particularly Machine Learning (ML) has made in recent years are leading to the development and use of a wide variety of tools that enable people to rethink the way they approach problems in different domains. This new approach enables more efficient interaction with information, deeper and faster data analysis, and leads to improvements in decision-making and workflows that are undergoing effective transformation [2, 5].

ML algorithms learn from training data to define a model capable of transforming the input information with the goal of generating an output capable

of solving a given problem. The learned models, thus, strictly depend on training data and the success of a model depends on the ability to train with large amounts of data [19]. However, in some application domains, it is hardly possible to have large amounts of training data and this can be a strong obstacle to the profitable use of ML techniques. Techniques such as transfer learning [25], data augmentation and synthetic data generation [19] try to propose a solution to the lack of training data by attempting to integrate the original training set with other datasets that are, in some way, related to the referring domains. However, the application of these techniques is sometimes not easy to make. In fact, it is required for a skilled user to be able to adapt and choose the pre-trained models for the desired application use. In other words, the user who wants to utilise the system must have the technical skills to integrate transfer learning or data augmentation solutions to his own process, and he must also be able to choose the pre-training data domain that most suit the desired solution, provided that this domain exists and is available.

In this paper, we would like to propose a different view, that in our opinion fit with the overall context sketched above. We assume that the profitability of the KWS in supporting the transcription can be evaluated by the ratio between the amount of human effort required to achieve the complete and error-free transcription of the collection with and without the support of the system, and then asked ourselves whether it is better to spend the human efforts mostly for producing the training set leading to the best performance of the KWS, or rather to train the KWS on a smaller (than in the previous case) dataset and spending most of the human efforts required to validate the outputs of the KWS on a larger (than in the previous case) number of pages. As a case study, we considered the transcription process of collections of handwritten documents of historical interest using a KeyWord Spotting system (KWS) as an ML tool to help the transcription [1]. This application area is of particular interest because the lack of data is a distinctive feature of this domain; some collections of historical interest are inherently made up of smallish data, and the stylistic and graphical features may be specific to the collection and therefore unique. Consequently, documents drawn up at different times or in different geographical areas can have extremely different characteristics, even when the content is expressed in the same language. This implies that the data sets built on particular collections are poorly able to adapt to the characteristics of collections produced at different times and locations. Therefore, and particularly for collections of limited size, the only way left to have data available to build the training set for the KWS is to manually transcribe part of the collection itself. Also, small collections are often held at small organizations, such as small museums, local archives or libraries. While the hardware and human resources required to use modern ML techniques may not be a problem for large organizations, this may not be the case for small organizations and archives. Small organizations' hardware resources are often limited, and processing large amounts of data can be difficult. Furthermore, modern image processing techniques based on artificial intelligence or deep learning technologies sometimes require not only particular and specific

hardware, but also adequate technical skills, and therefore the presence of highly qualified and trained personnel to fully exploit the potential of the technologies used. Therefore, solutions that are simple to apply to small data collections and that limit their use to only the information available from the collection itself can be useful, which however allows to simplify and reduce the human efforts required for achieving the transcription of the entire collection.

The process of transcribing such collections usually involves the manual transcription part of the collection by the user to be used to train the KWS system. Once such a system is trained, it can be used to support the transcription of the remaining pages of the collection. In this perspective, and considering that the output of the KWS must be validated for an error-free transcription, the performance of the supporting ML system takes a back seat. While it is easy to imagine that the larger the training set, the better the KWS system performs, building a large training set comes at a cost, which in our case is represented by the time the user has to spend transcribing the pages (of the training set) without a support system. Moreover, expanding the training set, i.e. transcribing a greater portion of the collection, leaves fewer documents to be transcribed with the help of the KWS to complete the entire collection, and the fewer the pages left to be transcribed, the lower the benefits introduced by the KWS on the time required for the complete transcription.

The results of a large set of experiments, performed on three public historical documents datasets largely used as benchmarks for performance evaluation and aimed at evaluating how long it takes to get a correct transcription of the whole collection as the size of the training set for the KWS system increases, show that training sets made up of 8 to 10 pages allow achieving the greatest gain in terms of human efforts, in all the cases and independently of the actual pages composing the training set. They also show that higher recall rates of the KWS lead to higher gains in the transcription time, mostly independently of the precision rate.

The remaining of the paper is organized as follows: in Sect. 2 we review the work proposing either TL or DA to deal with data scarcity in the case of historical documents, to highlight the reason why they might not be viable in the case of small collections of documents, while in Sect. 3 we describe the implementation of the transcription process that has been used for the experiments. The experimental setting is described in Sect. 4, and the results of the experiments we have designed and performed are reported in Sect. 5. Eventually, in Sect. 6, we discuss the experimental findings, draw some preliminary conclusions and outline future investigations.

2 Related Works

When it comes to machine learning, one common issue is the lack of available data to train models. However, there are two potential solutions - transfer learning (TL) and data augmentation (DA). The approach of TL involves first training a model on a more general task that contains a vast amount of data. This initial model can then serve as a starting point for training a second model that aims

to solve a different task [24]. On the other hand, the DA technique allows for generating new training data by manipulating the original data through transformations. The goal of DA is to expand and enhance a small set of training data [19].

These techniques have also been explored in the field of Historical Document Analysis, which is a difficult domain since historical documents are collections with specific and particular characteristics and generally can be of small size [12].

Transfer learning is commonly employed in computer vision to take advantage of the availability of public image datasets. However, applying this approach to historical records can be challenging due to the distinct nature of such data. Studer et al. [20] demonstrate that leveraging pre-trained ImageNet networks can enhance the accuracy of certain historical data analysis tasks. Nevertheless, this technique might lower the performance of other tasks, such as semantic segmentation. Despite the diversity of domains, this technique can generally improve performance [8, 9, 22] but the need to add a small amount of target data in learning to obtain a minimum rate of performance is always evident.

One common strategy for augmenting training data involves applying various transformations to the original images, such as flipping, rotation, or scaling. Noise can also be added or data can be purposefully degraded. [6, 11, 14] Recently, more advanced techniques have emerged, such as generative methods that generate entirely new training elements or combine different components (e.g., backgrounds, text, and images) to produce new documents [4, 10, 16]. Lately, some generative networks of the GAN type have been used to generate documents of a historical type with the aim of obtaining documents reporting a reference style [13, 23].

Both methods have the ability to enhance the efficiency of pre-existing models, but they require labelled starting datasets to work, even if they are not extensive. Additionally, the efficiency of transfer learning is affected not only by the starting pre-training dataset but also by the specific task it is attempting to address. For example, layout analysis displays more significant performance enhancements when compared to the gains achievable with handwritten text recognition. In regards to data augmentation, it is essential to use this technique with caution since going overboard can lead to the introduction of unwanted noise and artefacts during training, resulting in a decline in model performance. This is especially important to keep in mind when the initial dataset size is small because the small dimension can also limit the effectiveness of augmentation techniques.

However, both techniques require a minimum amount of labelled real data; DA needs real instances to apply transformations to or as reference instances for generation, while TL needs a fine-tuning phase on real data. When, as in the case of transcription of small handwritten collections using a KWS system, these data are not available the methods do not avoid the need to prepare such datasets manually.

3 The Transcription Process

The human efforts required by the transcription process of a small collection of handwritten documents of cultural and historical interest can be reduced by adopting ML tools and technologies, and among them, Keyword Spotting system (KWS) has shown better performance than handwritten text recognition to deal with the writing style variations occurring in documents produced at different times and places. A KWS has the task of finding instances of words of which it knows a representation, in the pages of the collection to be transcribed. In the preparation phase (*training*) of the KWS system, the knowledge base of the system is built up, which consists of *keywords*, i.e. words of whom the system knows both the representation and the correct transcription. The running system thus aims to retrieve the words whose representations are most similar to those of the keywords in the entire collection and link the transcription of the keywords to them. In this way, the system attempts to retrieve words without having to explicitly recognise the text contained in an image, and this property allows such systems to adapt to situations with limited data [1].

A user who wishes to use a KWS for transcription intents must create a list of keywords to be used to support the process. For this purpose, in the absence of preliminary data, the user must transcribe a part of the collection, which we call *TS* (Training Set), and use this as training information to prepare the KWS system. Once the KWS has been trained, the system can be used to support the transcription of the remaining part of the collection, which we will call *DS* (Data Set). The system's task is to recover the transcription of the words in the keyword list that are present in the *DS* so that the user no longer has to enter the transcription of these words manually.

Since the aim of the process is to obtain an error-free transcription, a validation phase of the output of the KWS system on the *DS* set is required. In other words, the user must check the system's output, validate the words correctly recognised by the KWS, correct the errors made by the system and, finally, produce a transcription for the words outside the vocabulary (OOV - Out Of Vocabulary), i.e. for the words that appear only in *DS* and for which the KWS system cannot provide a transcription. The validation process of a correct KWS output must be done by an extremely simple and fast procedure, e.g. a simple click of the mouse while scrolling through the list of options provided by the system. It is important that this procedure is faster than the time needed to transcribe a word manually because in this way the KWS can bring an effective improvement of the time needed to transcribe the whole collection. Once all the correct responses have been validated, the user has to provide the correct transcription for the words that the system did not recognize and for the words that the system is unable to recognize, i.e. the OOV words. The transcription of these words must be provided manually.

The process then expects the user to spend a time T_{TS} to transcribe the words in *TS* and create the keyword list, and then a time T_{DS} to validate and correct the system's output on *DS*. The use of the KWS system is beneficial for the transcription process if the sum of the times T_{TS} and T_{TD} is less than the

time T_m needed for the same user to transcribe the whole collection manually without the help of a KWS:

$$T_{TS} + T_{DS} < T_m \quad (1)$$

At this point, it becomes clear how important the size of the training set TS is. The larger it is, the more training data is available to prepare the KWS system. Moreover, by increasing TS the number of OOV words in the DS set decreases, simply because the cardinality of the keyword list increases. This leads to the assumption that large TS sets enable the KWS system to perform better and thus reduce the time T_{DS} . On the other hand, to get a large training set, the user has to manually transcribe more words, which increases the time T_{TS} . Since it is the sum of the two times that determines the usefulness of the system, the size of the set TS turns out to be a parameter with crucial importance.

4 Experimentation Details

4.1 Datasets

Two small datasets composed of handwritten cursive script dating back to the 18th century were considered for the experimentation, namely the George Washington dataset [17] and the Bentham Collection [18]. Both datasets collect 20 pages of handwritten documents written by a single writer. A third dataset is considered, the Parzival dataset [7] which is a record consisting of 47 pages by three writers. These pages were taken from a 13th-century medieval German manuscript containing the epic poem Parzival by Wolfram von Eschenbach. The Fig. 1 shows three excerpts from the various datasets and highlights the differences in the visual characteristics and writing style of the three collections. Table 1 reports the size of the three datasets in terms of words contained. The table shows both the total number of words contained in the collections and the number of unique words, i.e. the number of different words present. Looking at the relationship between the number of words in the collection and the number of pages, we find that the pages of the Bentham collection contain a smaller number of words, while the Parzival collection is the one with the most words per page, having almost three times as many words per page as the Bentham collection.

Table 1. Dataset details.

Dataset	Num Pages	Num Words	Num Unique Words
<i>Washington</i>	20	4819	1187
<i>Bentham</i>	20	3478	1091
<i>Parzival</i>	47	23412	4616

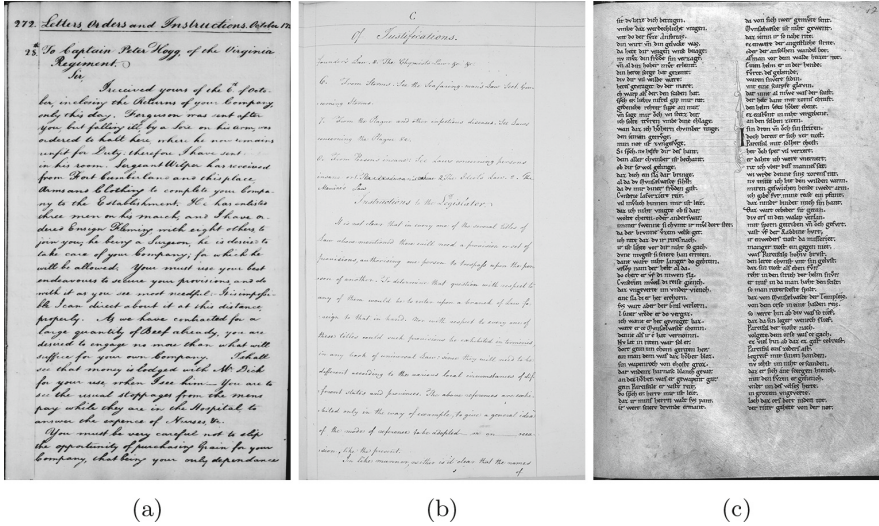


Fig. 1. Examples of documents from the three collections analysed: (a) Washington; (b) Bentham; (c) Parzival.

4.2 KWS System

The KWS used during the experiments is based on the PHOCNet [21], which has been configured to be used in a segmentation-based QbS scenario. First, the words contained in TS are transcribed manually and the labelled data is used to train the PHOCNet. During the query time, we extract all unique transcriptions in TS and use their PHOC representation as a query list. Then, the similarity between the images from DS and the words in the keyword list is calculated using the Bray-Curtis dissimilarity [3]. As a performance measure, recall and precision are calculated on DS by varying the distance acceptance threshold.

4.3 Temporal Gain

Having established the performance indices for the accuracy and recall of the KWS system, it is possible, given the size in words of the sets TS and DS , to estimate the time saving that can be achieved in transcribing the entire collection by using the performance estimation model presented in [15]. The model provides the percentage time gain G obtainable by using a KWS to transcribe documents after the validation and correction process that the user has to go through in order to obtain an error-proof transcription of the entire collection. The temporal gain can be calculated as:

$$G = 1 - T_u/T_m \quad (2)$$

where T_m is the manual transcription time, while T_u is the time taken to complete the transcription using the assistance system. While the time T_m depends only on the capabilities of the user who is transcribing, T_u also depends on the performance of the KWS system and therefore on the size of the keyword list.

In order to assess how the size of TS affects the time needed for transcription, we calculated the time gain obtainable by letting the number of pages used to build TS vary. This was done by starting with a single page TS and adding a new page to it until the entire collection was used as the training set. To generalise the results obtained, three randomly defined page orders were considered for each dataset and the results for each of the trials were recorded. Finally, the results are given by averaging the results of each trial.

5 Results

It is interesting to see how the number of OOV words and in-vocabulary words varies in the different collections as the pages of TS vary. Figure 2 shows how the distribution of words changes as the number of training pages increases. It is interesting to note that the trend of the curves is similar in all cases and that the number of OOV words tends to be relatively low for a TS which consists of a page count between 5 and 10. A difference can be seen in the Bentham dataset, as in this case, the ratio between OOV and in-vocabulary words tends to decrease less slowly than in the other two datasets. This could mean that the transcription of the Bentham dataset is more complex due to the larger number of OOV words.

Figure 3 shows the Precision/Recall curves of the KWS system recorded on DS when the pages used to define the TS of the different datasets vary. Looking at the curve plot, it is immediately noticeable that the KWS system, as easily expected, shows increasingly better performance as the training set dimension is increased. In fact, the KWS system continues to learn over the entire collection. However, it should be noted that in all cases, the performance of the network with very few training pages (less than 5) is always unsatisfactory. However, when the training set consists of more than 5 pages, the KWS performance seems to improve as the size of TS increases, but the performance gain is limited. A slightly different case is that of the Bentham dataset, where the network has more difficulty learning and more pages are needed in TS to achieve satisfactory performance. As can be seen from the Table 1, the Bentham is the dataset where the pages have the least written words, and therefore with the least useful information per page. It is therefore not surprising that it turns out to be the dataset on which the KWS has the most difficulty learning.

Finally, Fig. 4 shows the gain in transcription time obtained by varying the pages of the TS set. Interestingly, all systems achieve the maximum gain with a TS set consisting of 5 to 8 pages, regardless of the total size of the collection. It is also interesting that the maximum gain is related to the performance of the KWS system. The highest gain among the three cases is obtained with the Parzival dataset, the same dataset where the KWS system could achieve the best performance. In contrast, the lowest gain was obtained in the Bentham dataset, where the KWS system performed the worst.

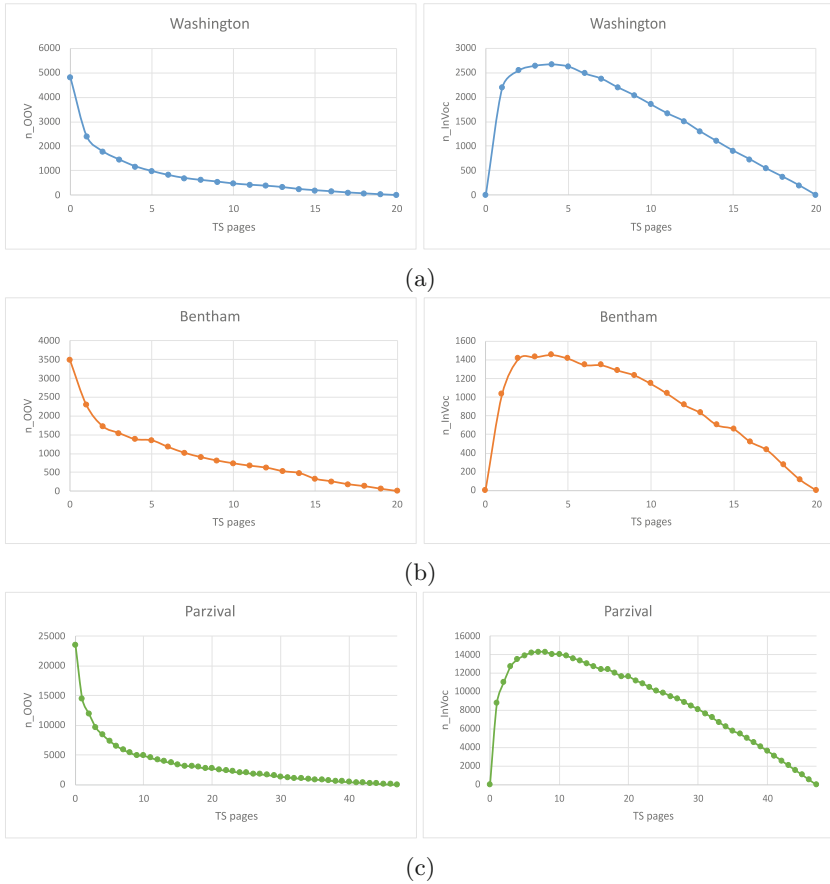


Fig. 2. Trend of OOV words (on the left) and in-vocabulary words (on the right) as the number of *TS* pages vary for the different datasets: (a) Washington; (b) Bentham; (c) Parzival.

6 Discussion and Conclusion

With this work, we have investigated how the time required to obtain a complete and error-free transcription of a small collection of handwritten documents using a KWS system to support the process varies depending on the size of the training set provided to the KWS.

Taking into account the distinctive features of the collections we are interested in and the cultural institutions that hold them, we assume that only information obtained from the collection itself can be used for training the KWS. In the absence of data from other datasets, then, training the KWS system requires manually transcribing a portion of the collection to create the training set. This process must be done manually by a user and takes some time. Once the training set is built and the KWS is trained, the user must validate and correct the

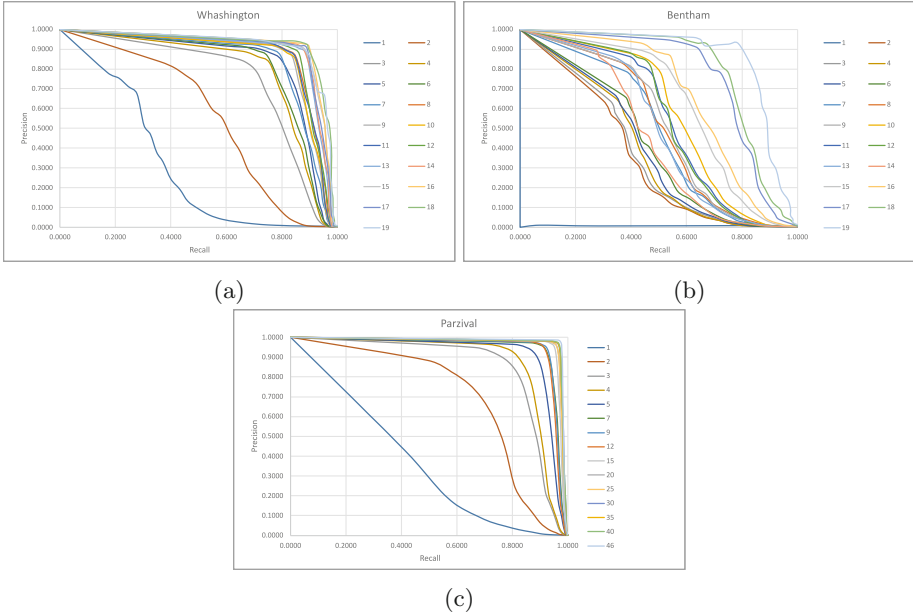


Fig. 3. Precision/Recall curve of the KWS system as TS pages vary for the different datasets: (a) Washington; (b) Bentham; (c) Parzival.

solutions proposed by the system to obtain an error-free transcription of the entire collection. It follows that the use of the KWS system becomes profitable when the sum of these times is less than the time required for the same user to transcribe the same collection, as described in Eq. 1. So the question arises on whether spending most of the human efforts to provide the KWS with the largest affordable training set so as to achieve top performance is the best strategy to achieve the largest reduction of the human efforts required for the complete transcription.

The experiments performed on the three datasets of different sizes showed that focusing on the performance of the KWS and trying to maximize it does not allow the user to achieve the best reduction of the time required for transcription. From the curves in Fig. 3, it can be seen that the KWS continues to improve its performance as the amount of training set *TS* increases. On the other hand, observing Fig. 4, it can be seen that a *TS* made up of a few pages is already enough to obtain the largest user time gain. It is interesting to note that the maximum time gain was achieved with a *TS* consisting of a number of pages between 5 and 8 in all three datasets, regardless of actual pages, the list of keywords of the training set, the distribution between in-vocabulary and OOV words, and size of the collection.

It is also clear from the curves of Fig. 4 that the nature of the data set plays an important role in the achievable gain. The lowest gain was recorded for the Bentham dataset, which is the smallest collection in terms of the number of

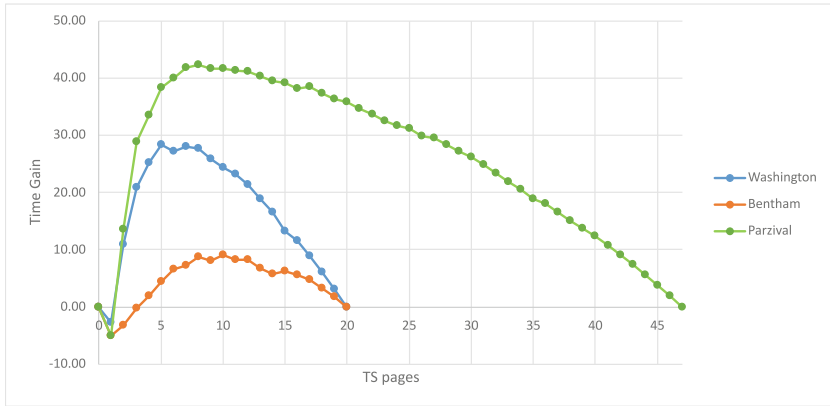


Fig. 4. The curves show the trend of the time gain obtainable by varying the pages in TS for the three different datasets considered.

words and has the largest ratio of OOV words to in-vocabulary words. This collection is the one that would take the least time of the three in the case of manual transcription, but it is also the collection that requires the user to consume the most resources in the validation and correction phase of the DS set due to the low power of the KWS and the high OOV word rate. The other extreme is the behaviour of the Parzival collection. This, in contrast, is the largest collection with a low ratio of OOV to in-vocabulary. However, it is interesting to point out that in both cases the best temporal gain was recorded with a TS consisting of 8 pages. We can therefore conclude that, although the ability to train a well-performing KWS is important, it is the nature of the dataset, its size, the length of the keyword list, and the distribution of the OOV words that affect performance in terms of transcription time gain. Eventually, the precision-recall curves in Fig. 3 indicate that the recall rate of the KWS plays a much relevant role than precision on the actual gain, and therefore KWS is capable of spotting OOV words may allow for a big leap in performance when used to assist the transcription of small collections of handwritten historical documents.

References

1. Ahmed, R., Al-Khatib, W.G., Mahmoud, S.: A survey on handwritten documents word spotting. *Int. J. Multimed. Inf. Retr.* **6**, 31–47 (2017)
2. Ashok, M., Madan, R., Joha, A., Sivarajah, U.: Ethical framework for artificial intelligence and digital technologies. *Int. J. Inf. Manage.* **62**, 102433 (2022)
3. Bray, J.R., Curtis, J.: An ordination of the upland forest communities of southern wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957)
4. Capobianco, S., Marinai, S.: Docemul: a toolkit to generate structured historical documents. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), vol. 1, pp. 1186–1191. IEEE (2017)

5. Cockburn, I.M., Henderson, R., Stern, S.: The impact of artificial intelligence on innovation: an exploratory analysis. In: *The Economics of Artificial Intelligence: An Agenda*, pp. 115–146. University of Chicago Press (2018)
6. Fischer, A., Visani, M., Kieu, V.C., Suen, C.Y.: Generation of learning samples for historical handwriting recognition using image degradation. In: *Proceedings of the 2nd International Workshop on Historical Document Imaging and Processing*, pp. 73–79 (2013)
7. Fischer, A., et al.: Automatic transcription of handwritten medieval documents. In: *2009 15th International Conference on Virtual Systems and Multimedia*, pp. 137–142. IEEE (2009)
8. Granet, A., Morin, E., Mouchère, H., Quiniou, S., Viard-Gaudin, C.: Transfer learning for a letter-ngrams to word decoder in the context of historical handwriting recognition with scarce resources. In: *27th International Conference on Computational Linguistics (COLING)*, pp. 1474–1484 (2018)
9. Granet, A., Morin, E., Mouchère, H., Quiniou, S., Viard-Gaudin, C.: Transfer learning for handwriting recognition on historical documents. In: *7th International Conference on Pattern Recognition Applications and Methods (ICPRAM)* (2018)
10. Journet, N., Visani, M., Mansencal, B., Van-Cuong, K., Billy, A.: Doccreator: a new software for creating synthetic ground-truthed document images. *J. Imaging* **3**(4), 62 (2017)
11. Kieu, V., Visani, M., Journet, N., Domenger, J.P., Mullet, R.: A character degradation model for grayscale ancient document images. In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR 2012)*, pp. 685–688. IEEE (2012)
12. Lombardi, F., Marinai, S.: Deep learning for historical document analysis and recognition—a survey. *J. Imaging* **6**(10), 110 (2020)
13. Madi, B., Alaasam, R., Droby, A., El-Sana, J.: HST-GAN: historical style transfer GAN for generating historical text images. In: Uchida, S., Barney, E., Eglin, V. (eds.) *DAS 2022*. LNCS, vol. 13237, pp. 523–537. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-06555-2_35
14. Maini, S., Groleau, A., Chee, K.W., Larson, S., Boarman, J.: Augraphy: a data augmentation library for document images. arXiv preprint [arXiv:2208.14558](https://arxiv.org/abs/2208.14558) (2022)
15. Marcelli, A., De Gregorio, G., Santoro, A.: A model for evaluating the performance of a multiple keywords spotting system for the transcription of historical handwritten documents. *J. Imaging* **6**(11) (2020). <https://doi.org/10.3390/jimaging6110117>. <https://www.mdpi.com/2313-433X/6/11/117>
16. Monnier, T., Aubry, M.: Docextractor: an off-the-shelf historical document element extraction. In: *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 91–96. IEEE (2020)
17. Rath, T.M., Manmatha, R.: Word spotting for historical documents. *Int. J. Doc. Anal. Recogn.* **9**(2–4), 139 (2007)
18. Sánchez, J.A., Romero, V., Toselli, A.H., Vidal, E.: ICFHR 2014 competition on handwritten text recognition on transcriptorium datasets (HTRTS). In: *2014 14th International Conference on Frontiers in Handwriting Recognition*, pp. 785–790. IEEE (2014)
19. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. *J. Big Data* **6**(1), 1–48 (2019)
20. Studer, L., et al.: A comprehensive study of imagenet pre-training for historical document image analysis. In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 720–725. IEEE (2019)

21. Sudholt, S., Fink, G.A.: Phocnet: a deep convolutional neural network for word spotting in handwritten documents. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), pp. 277–282. IEEE (2016)
22. Todorov, K., Colavizza, G., et al.: Transfer learning for historical corpora: an assessment on post-OCR correction and named entity recognition. In: CHR, pp. 310–339 (2020)
23. Vögtlin, L., Drazyk, M., Pondenkandath, V., Alberti, M., Ingold, R.: Generating synthetic handwritten historical documents with OCR constrained GANs. In: Lladós, J., Lopresti, D., Uchida, S. (eds.) ICDAR 2021. LNCS, vol. 12823, pp. 610–625. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86334-0_40
24. Weiss, K., Khoshgoftaar, T.M., Wang, D.: A survey of transfer learning. *J. Big Data* **3**(1), 1–40 (2016)
25. Zhuang, F., et al.: A comprehensive survey on transfer learning. *Proc. IEEE* **109**(1), 43–76 (2020)