ICDAR 2024 Competition on Handwriting Recognition of Historical Ciphers

Alicia Fornés[1,2][0000−0002−9692−5336], Jialuo Chen[0000−0002−7808−6567]1, Pau Torras[1][0000−0003−0327−9046] Carles Badal[1,2][0000−0001−8948−9946], Beáta Megyesi[0000−0002−4838−6518]3, Michelle Waldispühl[0000−0002−5603−6083]4, Nils Kopal[5], and George Lasry[6]

[1] Computer Vision Center, Universitat Autònoma de Barcelona, Spain
[2] Computer Science Department, Universitat Autònoma de Barcelona, Spain
{afornes, jchen, ptorras}@cvc.uab.cat, carles.badal@uab.cat
[3] Stockholm University, Sweden
beata.megyesi@ling.su.se
[4] University of Oslo, Norway
michelle.waldispuhl@ilos.uio.no
[5] University of Siegen, Germany
nils@kopaldev.de
[6] Independent researcher, Israel
george.lasry@gmail.com

Abstract. Handwritten Text Recognition (HTR) in low-resource scenarios (i.e. when the amount of labeled data is scarce) is a challenging problem. This is particularly true for historical encrypted manuscripts, commonly known as ciphers, which contain secret messages and were typically used in military or diplomatic correspondence, records of secret societies, or private letters. To hide their contents, the sender and receiver created their own secret method of writing. The cipher alphabets often include digits, Latin or Greek letters, Zodiac and alchemical signs, combined with various diacritics, as well as invented ones. The first step in the decryption process is the transcription of these manuscripts, which is difficult due to the great variation in handwriting styles and cipher alphabets with a limited number of pages. Although different strategies can be considered to deal with the insufficient amount of training data (e.g., few-shot learning, self-supervised learning), the performance of available HTR models is not yet satisfactory. Thus, the proposed competition, which includes ciphers with a large number of symbol sets and scribes, aims to boost research in HTR in low-resource scenarios.

Keywords: Handwritten Text Recognition · Historical Documents · Cipher Recognition · Competition

1 Introduction

Handwritten Text Recognition (HTR) has significantly advanced in the deep learning era. However, training deep learning-based HTR models is challenging

in low-resource scenarios, commonly defined as scenarios in which the amount of labeled data is scarce or minimal. This is particularly the case of historical manuscripts with rare scripts (e.g., cuneiform, Egyptian hieroglyphs) or unknown alphabets (e.g., historical encrypted documents, so-called ciphers). Encrypted manuscripts are a specific type of historical documents that contain secret messages, typically used in military reports, diplomatic letters, records of secret societies, and in private correspondence. With the aim to hide the content of the messages in ciphers, the sender and receiver created their own secret method of writing by transposing or substituting characters, using special symbols, or inventing a completely new alphabet of symbols. For example, the cipher alphabets often include digits, Latin or Greek letters, Zodiac and alchemical signs combined with various diacritics, as well as invented symbols.

Given the particularities of encrypted documents, it is necessary to join the expertise in computer vision, computational linguistics, philology, cryptanalysis, and history for a successful decryption of such manuscripts [9]. The first step in the decryption process is the transcription of these historical handwritten documents. Not surprisingly, deep learning-based HTR systems are not satisfactory when there is little available training data, especially when manuscripts contain high variations in handwriting styles. Moreover, the recognition of ciphers becomes even harder when the alphabet is invented because no dictionaries or language models are available to help in the training process.

Lately, several strategies have been considered to deal with the lack of training data in handwriting recognition, such as one-shot, zero-shot, and few-shot learning [1, 15, 5], unsupervised and self-supervised learning [13, 14, 11, 3], as well as data generation and augmentation [7, 4, 12, 17, 10]. However, the performance of HTR models in historical handwritten ciphers is still far from satisfactory [16]. For this reason, we believe that a competition on the recognition of encrypted documents, as an example of low-resource scenario, can boost the research in this direction. In addition, the recognition of ciphers is an example of low-resource scenario with a high historical interest. Thousands of enciphered historical manuscripts are buried in libraries and archives. Therefore, transcribing and decrypting the information contained in these special sources is important for understanding our cultural heritage, as it helps to shed new light on and even to (re-)interpret our history.

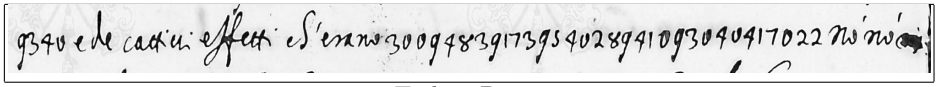
The rest of the paper is organized as follows. Section 2 describes the competition framework, including datasets, tasks and metrics. Section 3 is devoted to describing the participant's methods and baselines. The results are presented in Section 4. Finally, Section 5 presents the conclusions of this competition.

2 Competition Framework

In this section, we describe the datasets, tasks and metrics used.

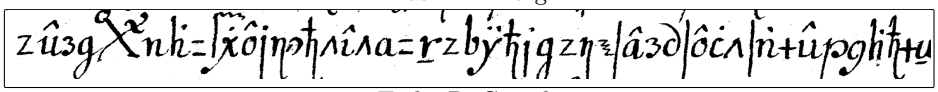
2.1 Tasks and Datasets

For this competition, we have selected a variety of ciphers, covering a large number of symbol sets and scribes. Specifically, we have provided 691 pages of encrypted manuscripts. Given the variability of cipher alphabets, we have divided them into ciphers with digits (task 1) and ciphers with symbol alphabets (task 2A, 2B, 3A, and 3B), which are written using an alphabet of symbols, most of them invented. Some examples of these manuscripts are shown in Figure 1. A summary of the particularities of each cipher is shown in Table 1.

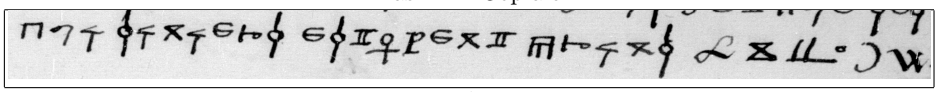


Task 1: Digits

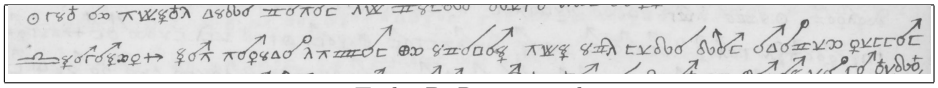
Task 2A: Borg



Task 2B: Copiale



Task 3A: BNF



Task 3B: Ramanacoil

Fig. 1. Samples from the encrypted manuscripts used in this competition.

Dataset	Total #pages	Training #lines	Test #lines	Alphabet type (size)	Classes #	Language	Century	Observations
T1-Vatican	323	5519	1250	Digits (10)	533	Italian/Latin/Spanish	15-18th	with cleartext
T2A-Borg	171	2594	576	Symbols (34)	174	Latin	17th	touching symbols
T2B-Copiale	105	1502	313	Symbols (100)	133	German	18th	
T3A-BNF	53	788	172	Symbols (39)	39	French	16th	
T3B-Raman.	39	1291	302	Symbols (24)	182	Dutch	17th	skewed lines

Table 1. Datasets specifications. For each task, we show the total number of pages, the number of lines used for training and testing, the type and size of the alphabet, as well as the language and century of the encrypted manuscript. Please note that the number of classes can be higher than the size of the alphabet.

To facilitate the participation of the maximum number of researchers, we have proposed the transcription of different ciphers with increasing level of difficulty. Thus, each task consists of transcribing a particular cipher document. Participants could freely decide to participate in one or more tasks. Each task is fully described below.

Task 1: Cipher with alphabet of digits. These numerical ciphers are found in the Secret Archive of the Vatican and originate from different centuries. These Vatican ciphers are written using 76 different symbols, based mostly on digits with multiple diacritics. The plain-text language - before encryption - is Latin, Italian or Spanish. This task is considered to be at an easy level because the alphabet is mainly composed of digits, and there is a sufficient amount of pages for training. However, there are two aspects to consider. First, there are more than 500 different classes due to the variations in diacritics and low-confident transcriptions (e.g. the transcriber used ? whenever doubting); and second, these documents typically include cleartext (not encrypted text) mixed with ciphertext. These two aspects can make the transcription harder than expected.

Task 2A: Borg Cipher, with alphabet of symbols. The Borg cipher is a long encrypted manuscript, composed of 408 pages, from the 17th century. The entire manuscript is encoded except for the first and last two pages and some headings in Latin. The cipher consists of 34 different symbols, from graphic signs to Latin letters and some diacritics. The cipher has touching symbols. The plaintext language - before encryption - is Latin (and partly Italian) describing how to treat various kinds of symptoms and diseases and reveals other pharmaceutical knowledge or secrets of that time. This task is considered to be of medium difficulty, since it uses an alphabet of symbols, but there is a sufficient number of pages for training.

Task 2B: Copiale Cipher, with alphabet of symbols. The Copiale cipher is a 105-page encrypted manuscript from the mid 18th century. The cipher consists of 100 different symbols, including symbols from Latin and Greek alphabets, as well as some ideograms (graphic symbols, such as Zodiac and alchemical signs) that represent important words (e.g. special entities). The plaintext language - before encryption - is German. This manuscript was created by an 18th century secret society, namely the "oculist order". This task is also considered to be of medium difficulty, because it uses an alphabet of symbols with a sufficient number of pages for training.

Task 3A: BNF Cipher, with alphabet of symbols. These are encrypted documents from the Bibliothèque Nationale de France, specifically Français 3029 and Français 3092. These 16th-century letters are entirely in cipher, so there is no mention of the sender or recipients. However, they are part of collections of letters from various French nobles from the time of French King François I.

There are approximately 39 distinct types of graphical symbols, generally non-touching. The plaintext - before encryption - is in Middle French. This task is considered to be of high difficulty because it uses an alphabet of symbols and there are few pages for training.

Task 3B: Ramanacoil Cipher, with alphabet of symbols. The Ramanacoil manuscript is a 46-page Dutch East India Company (VOC) document from 1674, comprising 39 pages of ciphertext and a key page. The document is preserved in the National Archives of the Netherlands (The Hague). It employs 24 unique symbols for the Latin alphabet (excluding V and J), additional special symbols for double letters (EE, FF, LL, OO, and PP), and also logograms for seven important words (e.g., "Ramanacoil"). The symbols generally do not touch, but lines are very skewed. The plaintext language - before encryption - is Dutch. This task is also considered difficult because it uses an alphabet of symbols with a very low number of pages for training.

2.2 Evaluation

Since this competition focuses on the transcription stage, we opted for converting the color images into grayscale, and manually segmenting the manuscript pages into text lines, so that participants can avoid the difficulties in the segmentation of these manuscripts. For this purpose, we have developed our own desktop application. Then, each line has been manually transcribed by several experts.

Therefore, the evaluation has been carried out at the line level. Given that cipher texts avoid grouping symbols into words to make the deciphering more difficult, the evaluation cannot be based on Word Error Rate. Instead, we have assessed the performance of each method using the Character Error Rate (CER) metric, defined as the number of insertion, deletion, and substitution edits required to obtain the ground truth from the prediction divided by the length of the ground truth. Formally:

$$CER = \frac{S + D + I}{N} \tag{1}$$

where S is the number of substitutions, D is the number of deletions, I is the number of insertions and N is the number of symbols in the ground truth. Please note that a Character can be a letter (e.g. a, b, ...), a digit (e.g. 1,2,...), a letter with a diacritic, or a symbol such as Zodiac or alchemical signs (e.g. "Taurus", "Cancer", Conjunction", "Saturn", etc.).

The evaluation has been carried out through the Robust Reading Competition (RRC) platform (https://rrc.cvc.uab.es/). Thus, participants just had to upload their results, and then, the server evaluated the performance of each method on each test set. Moreover, this competition will remain active in the long term through the RRC competition server.

3 Methods

The competition has attracted considerable interest, although the final number of submissions has been moderate. With the aim of providing greater variety in the methodologies and a more complete analysis of the results, the organizers of this competition have submitted some reference baselines using state-of-the-art methods.

According to the competition platform, 43 users registered in the competition, with a total of 236 downloads and 24 submissions by the different teams. The statistics for each task are shown in Table 2.

	Task 1	Task 2	Task 3	Task 4	Task 5	Total
Number of Downloads	69	50	41	39	37	236
Number of Submitted results	9+3	6+3	5+3	2+3	2+2	38

Table 2. Competition Statistics. For each task, we show the number of downloads and the number of submitted results (participant's results + submitted baselines).

Next, we describe the participants and their methods, as well as the baseline methods.

3.1 Team 1 (T1)

Participants: Hui Zheng.
Description of the method:
This team has used a hybrid neural network of convolutional neural network (CNN) and transformer network to recognize the characters. The backbone of the CNN network has been a Resnet34. Features from multi-scale stages are concatenated as input to the encoder of a transformer. Specifically, the CNN part in the network is Resnet34, with three blocks outputting feature maps. Features are down-sampled by max-pools with stride 3, and then converted to 3*3 patches. The patches are converted into 1D vector embeddings (from flattened 2D patches). As input to the encoder, each patch embedding is added with positional encoding. The transformer includes encoder and decoder, each with 3 layers.

In the training stage, images are normalized to 48*960, features extracted with the Resnet34 at three resolutions (12*240, 6*120, 3*60). After max-pools and flattening, features are converted to 420 length 1D vector embeddings. KL-loss is adopted in our network. Different configurations of this architecture (e.g. different number of layers) have been trained and the results have been uploaded to the competition platform. In the results, each variant is denoted with a letter (e.g. T1-a, T1-b, etc.).

3.2 Team 2 (T2)

Participants: Raphael Baena, Syrine Kalleli , Mathieu Aubry.
Affiliation: ENPC Imagine.
Description of the method: *Character Detection and Classification with Transformer Architecture.*
The team has employed a transformer-based architecture that detects characters in parallel, ensuring fast and accurate predictions. For each character, it provides a bounding box and its likelihood, which are then used for Optical Character Recognition (OCR). Notably, this approach does not rely on any language prior.

The team first pre-trained the architecture on synthetic data consisting of text lines with characters from various fonts. They used standard classification and bounding box positioning losses for this process.

Subsequently, they fine-tuned the architecture on real datasets. Unlike synthetic data, these datasets do not include ground truth bounding boxes, but only text transcriptions. Therefore, they were enable to use the same training losses as before. Instead, they used the pre-trained model to detect the characters' bounding boxes. Following this, they organized the characters based on these bounding boxes and computed the Connectionist Temporal Classification (CTC) loss. During fine-tuning, this approach demonstrated the ability to learn the bounding boxes of new characters.

3.3 Team 3 (T3)

Participants: Wenbo Hu, Xinchen Ma, Yue Lu.
Affiliation: East China Normal University.
Description of the method:
The team analyzed the data and found that the text lines they need to process are very long, have complex spatial structures, and include superscripts. This often leads to inaccurate attention when using attention-based recognition algorithms for such complex text lines. Therefore, their method enhances the model's awareness of text symbol positions by jointly optimizing the symbol counting task and the text recognition task (i.e., adding a counting module).

Their network is a unified end-to-end trainable framework that includes the application of DenseNet as the backbone to obtain a two-dimensional feature map. Next, a multi-scale counting module is used to predict the number of each symbol class and generate a counting vector representing the counting results. The obtained feature map and counting vector are then input into the attention decoder to produce the predicted output. In the loss function part, they use the cross-entropy loss function to calculate the loss regarding the predicted probability relative to the ground truth. Additionally, they use a smooth L1 loss function for the counting of each symbol class.

This implementation considers the effectiveness of counting as an auxiliary task to enhance the model's perception of symbol positions, thereby improving the overall accuracy of text recognition. By leveraging the advantages of multi-task learning, the joint optimization of counting and recognition tasks allows for

the sharing of feature extraction parameters, increasing training efficiency and model performance. This approach is particularly beneficial for handling long text lines with complex structures and superscripts, providing higher accuracy and robustness in recognition.

The authors note that it is their initial attempt at historical symbol recognition. Although their current method [6] has been optimized for long text lines, the model may still struggle to maintain attention and contextual information for extremely long text lines. Furthermore, while they have introduced a counting module, inaccurate counting results could negatively impact the overall recognition performance. They believe that adding a semantic module to the current method could help achieve better results.

3.4 Team 4 (T4)

Participants: Simon Corbillé, Elisa H Barney Smith.
Affiliation: Machine Learning, Luleå Tekniska Universitet.
Description of the method: *Sequence-to-Sequence model trained on multiple datasets for Handwriting Recognition of Historical Ciphers.*
The team uses a Sequence-to-Sequence model, one of the state-of-the-art architectures for handwriting recognition. It is composed of an encoder, an attention component and a decoder. The encoder uses a CRNN architecture. It is composed of convolutional layers to extract spatial features and LSTM layers to extract temporal features. The attention module focuses the decoders on a specific part of the features extracted by the encoder to predict character by character. The model is trained with a hybrid loss (CTC loss for the encoder and cross-entropy loss for the decoder).

Regarding data specification, the images are resized and padded to a fixed size of pixels (based on the mean height and width values). The training data is divided randomly into a training set (80%) and a validation set (20%). During the training, they use affine augmentation on the training data for data augmentation.

They found empirically that the use of a combination of cipher datasets improves the recognition performance. For task 2A, they train the model on a combination of Borg and BNF datasets, whereas for task 2B, they use a combination of Borg, Copiale, and BNF for training. For task 3A, they train the model on a combination of Copiale and BNF, whereas for task 3B, they train by combining the Borg, Copiale, and Ramanacoil datasets and consider classes where the number of samples in the training set is greater than 10.

3.5 Baseline 1 (B1)

As commented before, we have submitted some reference baselines using state-of-the-art methods. The first baseline is based on Long Short-Term Memory Recurrent Neural Networks (LSTMs) proposed in [2], and applied to the recognition of handwritten music scores. For this baseline, input images are normalized (resized) and fed into the model, which corresponds to a bi-directional LSTM,

so that we can reduce ambiguities when recognizing some symbols. We have provided two variants of this architecture:

- Variant B1-a: input images are normalized to 92*900, with 4 lstm layers configuration.
- Variant B1-b: input images are normalized to 64*500, with 1 lstm layer configuration.

3.6 Baseline 2 (B2)

The second baseline is based on Sequence to Sequence models. Concretely, we have implemented an attention-based Sequence to Sequence model, based on the work of [8], which was applied to the recognition of handwritten text. More specifically, the architecture has three main parts: an encoder, consisting of a CNN (a VGG-19-BN pretrained on ImageNet) and a bi-directional Gated Recurrent Unit (BGRU); an attention mechanism that focuses on the important features; and the decoder, consisting of one-directional multi-layered GRUs, which outputs symbol by symbol. Both the encoder and the decoder are a stack of 4 layers respectively.

4 Results

Table 3 shows the submitted results for each task (1, 2A, 2B, 3A, and 3B) in the columns and methods in the rows by the four teams (T1, T2, T3, T4) and their variants, along with the two baseline models (B1 and B2) and their variants. The results are arranged in increasing order of CER; i.e., for each task, the best results are listed first and followed by lower models' performance.

Overall, the competition showcases a range of methods with varying success across different tasks, highlighting both the strengths of certain approaches and the challenges inherent in handwritten text recognition of historical ciphers. This is not surprising, given the different characteristics of the encrypted sources in the different tasks with respect to i) the symbol set size ii) the symbol types iii) the type of writing in terms of touching vs segmented handwriting styles, iv) the variation of the handwriting styles with respect to the number of scribes, and v) the size of the training data; the very reason these sources were chosen for the various tasks.

In the competition, the best performers for each task were as follows:

Task 1: With the numerical ciphers as input in clearly segmented digits but many different handwriting styles and lots of training data, the baseline method with LSTMs (B1-a) achieved the lowest Character Error Rate (CER) of 7.83%, followed by Team 1's first model (T1-a) with 9.25%. Surprisingly, the CER is higher than expected, probably due to the high number of classes and the variations in the handwriting styles.

Task 1		Task 2A		Task 2B		Task 3A		Task 3B	
Method	CER	Method	CER	Method	CER	Method	CER	Method	CER
B1-a	7.83	T2	6.76	T4	1.62	T4	0.89	T2	5.61
T1-a	9.25	T1-i	7.10	T2	2.73	B1-a	1.09	B1-a	6.07
T2	11.88	B1-b	7.42	T1-l	3.22	B1-b	1.25	T4	6.08
B1-b	11.91	T4	7.60	T1-m	3.44	T2	1.73	B1-b	6.25
B2	12.75	B1-a	7.91	B1-a	4.33	B2	15.69		
T1-b	16.55	B2	9.56	B1-b	4.69				
T1-c	17.68	T1-j	15.48	T1-n	12.09				
T1-d	19.43	T3	25.55	B2	17.37				
T1-e	19.79	T1-k	25.62						
T1-f	20.21								
T1-g	22.64								
T1-h	23.61								

Table 3. Competition Results. CER ranges from 0% to 100%. The lower value, the better performance.

Task 2A: With the Borg cipher as the basis with 34 graphical touching symbols in rather similar hand-writing styles, T2 achieved the best performance with a CER of 6.76%, closely followed by T1-i at 7.10%. In this case, the performance is relatively moderate due to the touching symbols, and also by the fact that there are some symbols (e.g "Afsicanns") with very low frequency. Also, some images have a fold at the beginning or at the end of the line, making difficult to obtain just the transcription of the current line.

Task 2B: With the Copiale cipher as a large, eclectic, graphical symbol set with segmented and meticulously written handwriting style by one scribe and a large training set, T4 achieved the lowest CER of 1.62%, with T2 and T1-l also performing well at 2.73% and 3.22%, respectively.

Task 3A: With the BNF cipher as base consisting of 39 graphical signs clearly written in non-touching symbols but with a few number of training pages, T4 performed the best with a CER of 0.89%, followed by B1-a at 1.09%.

Task 3B: With the Ramanacoil cipher as the base with 24 symbols including a few special signs, with partly connected lines, different handwriting styles and a few number of training pages, T2 led with a CER of 5.61%, followed by B1-a at 6.07%. In this case, lines are skewed, so they could not be clearly segmented (some lines can contain parts of the previous and/or the next line), which has an impact in the final models' performance.

Given the results, we observe that the easiest tasks are documents with clearly segmented non-touching symbols, with good paper quality, and written by one or

a few scribes. This is the case of Task 3A (CER of 0.89) with 39 symbol types, and Task 2B (CER of 1.62) with over 100 different symbols. It is noteworthy that the training data size is of less relevance for these models than expected. Obviously, when the segmentation becomes harder, as in the case of the Ramanacoil cipher (Task 3B), with skewed lines, and the Borg cipher (Task 2B), with touching symbols, the error rates increase (5.61 and 6.76 respectively).

Surprisingly, despite the large set of training data with clearly segmented symbols with a small set of symbol types (0-9), the most difficult task is Task 1, the digit-base ciphers, with the best CER of 7.89. We believe that the high variation among scribes poses a problem. Also, there is a mix of cleartext and ciphertext, which makes the transcription more difficult. In addition, the high amount of classes (more than 500) increases the difficulty, especially because many of them are digits with different diacritics, so the confusion is higher. Indeed, there are cases in which the transcriber was doubting between 2 digits, specifying this issue with question marks (e.g. 7/1?, 5/6?, 2?, 1?). In some other cases, the same symbol was transcribed using a similar transcription label (e.g. "0_*Loop*" versus "0_*Loop*$\bigwedge$*Aries*", "0$\bigwedge v$" versus "0$\bigwedge$*Aries*").

In general, we have observed that there is no significant differences regarding the performance of the Baseline 1 (LSTMs), Team 2's method (Transformers) and Team 4's method (Seq2Seq). This aspect suggests that, although bigger architectures are more powerful, in case of few labeled data to train, more classical and shallower networks (such as LSTMs) could be preferred. Indeed, Baseline 1-LSTMs achieves the best performance in Task 1.

We have also observed that there are significant differences in the methods submitted by team 1 (see all variations ranging from T1-a to T1-n). This suggests that the selection of parameters (e.g. number of layers) is crucial, and must be appropriately selected for each dataset, otherwise the performance decreases significantly.

Finally, it must be noted that there is no method that obtains the best results in all tasks. Indeed, Team 2 obtains the best performance in Tasks 2A and 3B, whereas Team 4 obtains the best performance in Tasks 2B and 3A.

5 Conclusions

The ICDAR 2024 competition on handwriting recognition of historical handwritten ciphers aims to raise interest in handwritten text recognition in low-resource scenarios. We strongly believe that it is an interesting scientific problem for the research community in document analysis and reading systems. Furthermore, ciphers documents are manuscripts with a high degree of interest from a historical perspective.

From the analysis of the participants' results, we have observed that there is no significant difference in methods' performances, especially if we compare with the baseline based on the well-known classical LSTMs. This aspect suggests that, even though the latest deep learning based architectures, such as Transformer

Networks, are very powerful, their performance is similar to simpler architectures when the training data is limited.

In conclusion, we believe that this competition can serve as a benchmark for researchers in HTR in low-resource scenarios. Indeed, as this competition will remain open and continuous via the Robust Reading Competition platform, researchers can contribute by uploading new results at any time.

Acknowledgments. We thank all participants in this competition. This work has been partially supported by the Swedish Research Council (grant 2018-06074, DECRYPT), the Spanish projects PID2021-126808OB-I00 (GRAIL) and CNS2022-135947 (DOLORES) from the Ministerio de Ciencia e Innovación, the Departament de Cultura of the Generalitat de Catalunya, and the CERCA Program / Generalitat de Catalunya. Pau Torras is funded by the Spanish FPU Grant FPU22/00207.

References

1. Ao, X., Zhang, X.Y., Liu, C.L.: Cross-modal prototype learning for zero-shot handwritten character recognition. Pattern Recognition **131**, 108859 (2022)
2. Baró, A., Riba, P., Calvo-Zaragoza, J., Fornés, A.: Optical music recognition by long short-term memory networks. In: Graphics Recognition. Current Trends and Evolutions: 12th IAPR International Workshop, GREC 2017, Kyoto, Japan, November 9-10, 2017, Revised Selected Papers 12. pp. 81–95. Springer (2018)
3. Bhunia, A.K., Chowdhury, P.N., Yang, Y., Hospedales, T.M., Xiang, T., Song, Y.Z.: Vectorization and rasterization: Self-supervised learning for sketch and handwriting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5672–5681 (2021)
4. Gan, J., Wang, W., Leng, J., Gao, X.: Higan+: handwriting imitation gan with disentangled representations. ACM Transactions on Graphics (TOG) **42**(1), 1–17 (2022)
5. Hu, W., Zhan, H., Liu, C., Yin, B., Lu, Y.: Ots: A one-shot learning approach for text spotting in historical manuscripts. arXiv preprint arXiv:2304.00746 (2023)
6. Hu, W., Zhan, H., Ma, X., Lu, Y., Suen, C.Y.: Spotting the unseen: Reciprocal consensus network guided by visual archetypes. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 12556–12564 (2024)
7. Kang, L., Riba, P., Rusinol, M., Fornes, A., Villegas, M.: Content and style aware generation of text-line images for handwriting recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(12), 8846–8860 (2021)
8. Kang, L., Toledo, J.I., Riba, P., Villegas, M., Fornés, A., Rusinol, M.: Convolve, attend and spell: An attention-based sequence-to-sequence model for handwritten word recognition. In: Pattern Recognition: 40th German Conference, GCPR 2018, Stuttgart, Germany, October 9-12, 2018, Proceedings 40. pp. 459–472. Springer (2019)
9. Megyesi, B., Esslinger, B., Fornés, A., Kopal, N., Láng, B., Lasry, G., Leeuw, K.d., Pettersson, E., Wacker, A., Waldispühl, M.: Decryption of historical manuscripts: the decrypt project. Cryptologia **44**(6), 545–559 (2020)
10. Nikolaidou, K., Retsinas, G., Christlein, V., Seuret, M., Sfikas, G., Smith, E.B., Mokayed, H., Liwicki, M.: Wordstylist: styled verbatim handwritten text generation with latent diffusion models. In: International Conference on Document Analysis and Recognition. pp. 384–401. Springer (2023)

11. Penarrubia, C., Garrido-Munoz, C., Valero-Mas, J.J., Calvo-Zaragoza, J.: Spatial context-based self-supervised learning for handwritten text recognition. arXiv preprint arXiv:2404.11585 (2024)
12. Pippi, V., Cascianelli, S., Cucchiara, R.: Handwritten text generation from visual archetypes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22458–22467 (2023)
13. Siglidis, I., Gonthier, N., Gaubil, J., Monnier, T., Aubry, M.: The learnable typewriter: A generative approach to text analysis. arXiv preprint arXiv:2302.01660 (2023)
14. Souibgui, M.A., Biswas, S., Mafla, A., Biten, A.F., Fornés, A., Kessentini, Y., Lladós, J., Gomez, L., Karatzas, D.: Text-diae: A self-supervised degradation invariant autoencoder for text recognition and document enhancement. In: proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 2330–2338 (2023)
15. Souibgui, M.A., Fornés, A., Kessentini, Y., Megyesi, B.: Few shots are all you need: a progressive learning approach for low resource handwritten text recognition. Pattern Recognition Letters **160**, 43–49 (2022)
16. Souibgui, M.A., Torras, P., Chen, J., Fornés, A.: An evaluation of handwritten text recognition methods for historical ciphered manuscripts. In: Proceedings of the 7th International Workshop on Historical Document Imaging and Processing. pp. 7–12 (2023)
17. Zhu, Y., Li, Z., Wang, T., He, M., Yao, C.: Conditional text image generation with diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14235–14245 (2023)