GNHK: A Dataset for English Handwriting in the Wild

Alex W. C. Lee[1*], Jonathan Chung[2*], and Marco Lee[1]

[1] GoodNotes, Hong Kong
{alex, marco}@goodnotes.com
[2] Amazon Web Services, Vancouver
jonchung@amazon.com
* denotes equal contribution

Abstract. In this paper, we present the GoodNotes Handwriting Kollection (GNHK) dataset. The GNHK dataset includes unconstrained camera-captured images of English handwritten text sourced from different regions around the world. The dataset is modeled after scene text datasets allowing researchers to investigate new localisation and text recognition techniques. We presented benchmark text localisation and recognition results with well-studied frameworks. The dataset and benchmark results are available at https://github.com/GoodNotes/GNHK-dataset.

Keywords: Handwriting recognition · Benchmark dataset · Scene text

1 Introduction

Understanding handwriting from images has long been a challenging research problem in the document analysis community. The MNIST dataset, introduced by LeCun et al. [15] in 1998, was a widely known dataset containing 70,000 grayscale images with 10 classes of handwritten digits. In the age of deep learning, classifying individual digits is no longer a difficult task [2]. However, in reality, we see that handwriting can vary immensely in both style and layout (e.g. orientation of handwritten text lines compared with printed text lines). As such, detecting and recognizing sequences of handwritten texts in images still pose a significant challenge despite the power of modern learning algorithms.

With researchers moving from using Hidden Markov Models (HMM) [21] to using variants of Recurrent Neural Networks (RNN) (e.g. BiLSTM [11], MDLSTM [32] and CRNN [29]) with Connectionist Temporal Classification (CTC) loss, problems like word recognition and line recognition have quickly become less difficult. More recent challenges focus on trying to recognize a full page of handwriting without explicit segmentation. Two datasets that are commonly used in research are IAM Handwriting Database [22] and RIMES Database [3], consisting of distinct and mostly horizontally oriented text lines from scanned documents. Numerous attempts have been made to solve the problem, such as using attention mechanisms with encoder-decoder architecture [6,7,9] or approaches that exploit the 2D layout of the document images [28,35].

While research in handwriting recognition is making great progress, outside of the document analysis community, text localisation and recognition in the natural scene is becoming a more active research area [19]. As texts are the medium of communication between people, the ubiquity of cameras makes it possible for us to capture the text in the wild and store them in pixels. The existing scene-text datasets that are commonly used as benchmarks are all in printed text. Recently, Zhang et al. released SCUT-HCCDoc [37], an unconstrained camera-captured documents dataset containing Chinese handwritten texts, which is novel to the research community, as there has been a lack of data for handwritten text in the wild. However, such a dataset does not exist for offline English handwritten images and we believe there is a necessity to create a similar one.

Therefore, we created a dataset for offline English handwriting in the wild called GoodNotes Handwriting Kollection (GNHK), which contains 687 document images with 172,936 characters, 39,026 texts and 9,363 lines. Note that "texts" defined in this paper include words, ASCII symbols and math expressions (see Section 3.2 for a detailed explanation). As English is the leading global lingua franca, handwriting styles and lexicons can vary across regions. For example, words like "Dudu-Osun" will more likely appear in documents from Nigeria and "Sambal" may more likely be written in Malaysia. To capture the diversity of handwriting styles from different regions, we collected the images across Europe, North America, Asia and Africa. We also attempted to create a more representative "in the wild" dataset by including different types of document images such as shopping lists, sticky notes, diaries, etc.

In this paper, we make the following contributions to the field of document analysis:

1. We provide a benchmark dataset, GNHK, for offline English handwriting in the wild.
2. We present a text localisation and a text recognition model, both serving as baseline models for our provided dataset.

2 Related Work

Detecting and recognizing offline handwriting remains a challenging task due to the variety of handwriting and pen styles. Therefore, many researchers have created datasets that allow the document analysis community to compare against their results. For Latin scripts, the most popular datasets are IAM Handwriting Database (English) [22], the IUPR Dataset (English) [8], the IRONOFF dataset (English) [31], and RIMES Database (French) [3]. For other scripts, there are KHATT Database of Arabic texts [20], IFN/ENIT Database of Arabic words [24], CASIA Offline Chinese Handwriting Database [18], HIT-MW Database of Chinese text [30] and SCUT-EPT of Chinese text [38]. One common feature among the aforementioned datasets is that the images were scanned using flatbed scanners, so they do not have much noise and distortion compared to images captured using cameras. Recently, a Chinese dataset called SCUT-HCCDoc [37]

was released, which has a set of unconstrained camera-captured documents taken at different angles.

Table 1: Overview of different offline Latin-based handwriting datasets.

Dataset	# Texts	# Lines	Image Type
IAM	115,320	13,353	Flatbed-scanned
RIMES	66,982	12,093	Flatbed-scanned
Ours	**39,026**	**9,363**	**Camera-captured**

Our contribution in GNHK is similar to SCUT-HCCDoc, but for English texts. To the best of our knowledge, much work in offline Latin-based handwriting has been done using flatbed-scanned images. Table 1 shows a comparison between IAM, RIMES, and our dataset. We believe our dataset will bring in diversity and a modern take to the researchers working on offline English handwriting recognition.

3 Dataset Overview

3.1 Data collection

The images were collected by a data-labeling firm [1] upon our request. Since penmanship can vary from country to country [5], to make sure we have a diverse set of English handwriting, we sourced the images across Europe, North America, Africa and Asia (see Table 2). In addition, no more than 5 images were written by the same writer in the collection.

Table 2: Number of images per region in the dataset.

Region	# Images
Europe (EU)	330
North America (NA)	146
Africa (AF)	117
Asia (AS)	94
TOTAL	**687**

The dataset consists of 687 images containing different types of handwritten text, such as shopping lists, sticky notes, diaries, etc. This type of text tends to favor handwriting over typing because people found them to be more reliable when capturing fleeting thoughts [27]. Images were captured by mobile phone cameras under unconstrained settings, as shown in Fig. 1.

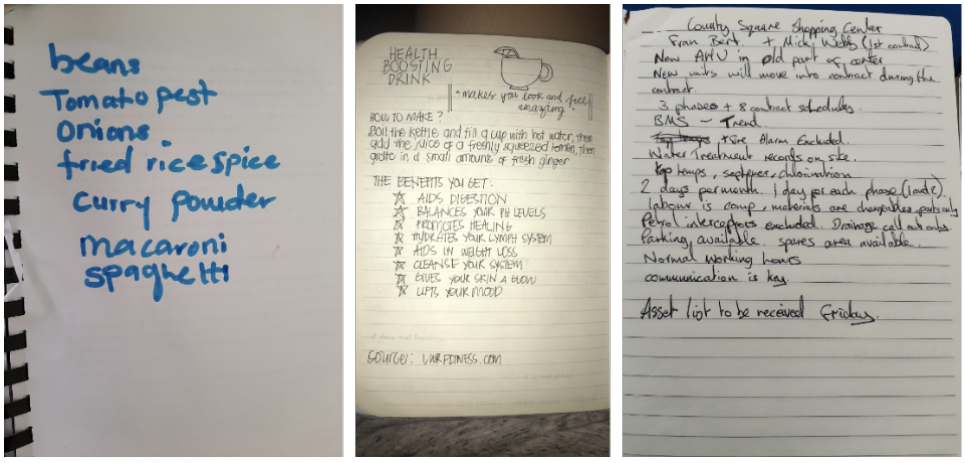


Fig. 1: Examples of the handwriting images captured under unconstrained settings.

3.2 Data annotation and format

For each handwritten-text image, there is a corresponding JSON file that contains the annotations of the image. Each JSON file contains a list of objects and each of them has a list of key-value pairs that correspond to a particular text, its bounding polygon and other associated information (see Fig. 2 for example):

text A sequence of characters belonging to a set of ASCII printable characters plus the British pound sign (`£`). Note that no whitespace characters are included in the value. In addition, instead of having a sequence of characters as value, sometimes we use one of the three special tokens for polygon annotations that meet some criteria (see Fig. 3 for examples):

- `%math%`: it includes math expressions that cannot be represented by ASCII printable characters. For example, ∞ and $\sum$.
- `%SC%`: it considered as illegible scribbles.
- `%NA%`: it does not contain characters or math symbols.

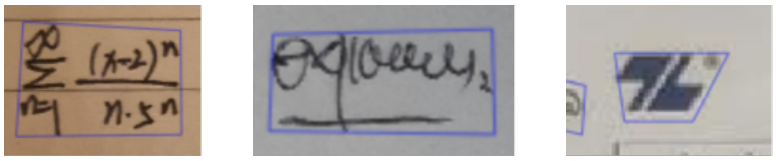
polygon Each polygon is a quadrilateral and it is represented by four (x, y) coordinates in the image. They are listed in clockwise order, with the top-left point (i.e `"x0"`, `"y0"`) being the starting point.

line index Texts belonging to the same line have the same index number.

type Either `H` or `P`, indicating whether the label is handwritten or printed.

```
{
   "text": "Proof",
   "polygon": {
       "x0": 845, "y0": 1592, "x1":  859, "y1": 1809,
       "x2: 1188, "y2": 1888, "x3": 1300, "y3": 1588
   },
   "line_idx": 4,
   "type": "H",
}, ...
```


Fig. 2: A sample object from one of the JSON files.



(a) `%math%` **(b)** `%SC%` **(c)** `%NA%`

Fig. 3: Examples of the three special tokens for the values of `"text"`. For (a), some of the math symbols in the polygon cannot be represented by ASCII characters. For (b), we can tell it is a sequence of characters inside the polygon, but it is hardly legible. For (c), the annotator labelled it as text, but it is not.

4 Dataset Statistics

Among the 687 images in the dataset, there are a total of 39,026 texts with 12,341 of them being unique. The median number of texts per image is 57 and the mean is 44. If we look at the text statistics by region, their distributions are roughly the same, as shown in Fig. 4. In Fig. 5, it shows the top 40 frequently used text in the dataset.

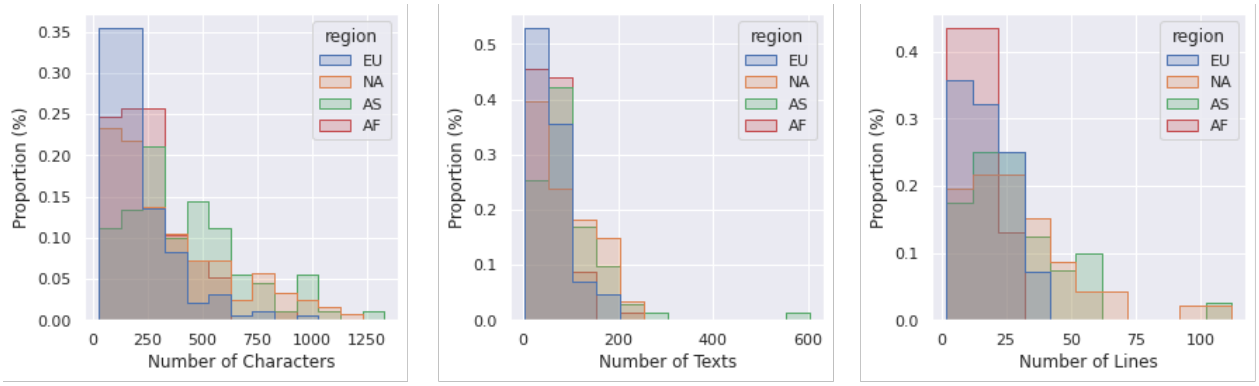


Fig. 4: Distributions of characters, texts, and lines for each region. Each plot includes four histograms with bin width of 100, 50, 10 for number of characters, texts, and lines respectively. Each histogram in a plot represents a region denoted by the color. For each histogram the counts are normalized so that the sum of the bar height is 1.0 for each region.

Fig. 6 illustrates the top 40 frequently used characters in the dataset. The dataset has 96 unique characters, with each of them showing up 3 to 17887 times. The median count is 486 and the mean is 1801.

In terms of total number of characters and lines, there are 172,936 and 9,363 respectively. Similar to texts, the distributions of character and lines are fairly similar, except for the EU region for characters (see Fig. 4). Detailed statistics are summarized in Table 3.

Table 3: Statistics of characters, texts and lines for each region.

Region	# Characters	# Texts	# Lines
Europe (EU)	58,982	13,592	3,306
North America (NA)	47,361	10,967	2,099
Asia (AS)	39,593	8,586	2,780
Africa (AF)	27,000	5,881	1,178
TOTAL	**172,936**	**39,026**	**9,363**

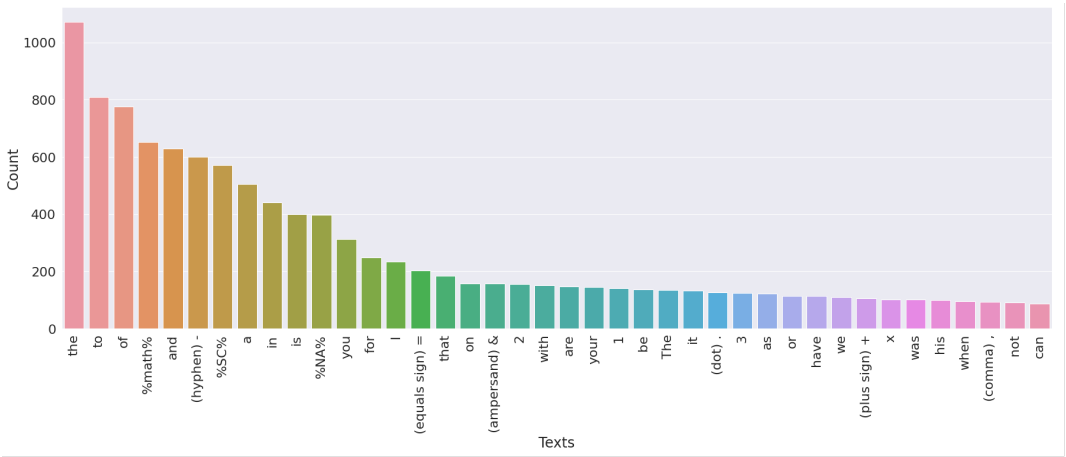


Fig. 5: Frequency of the top 40 common texts in the dataset.

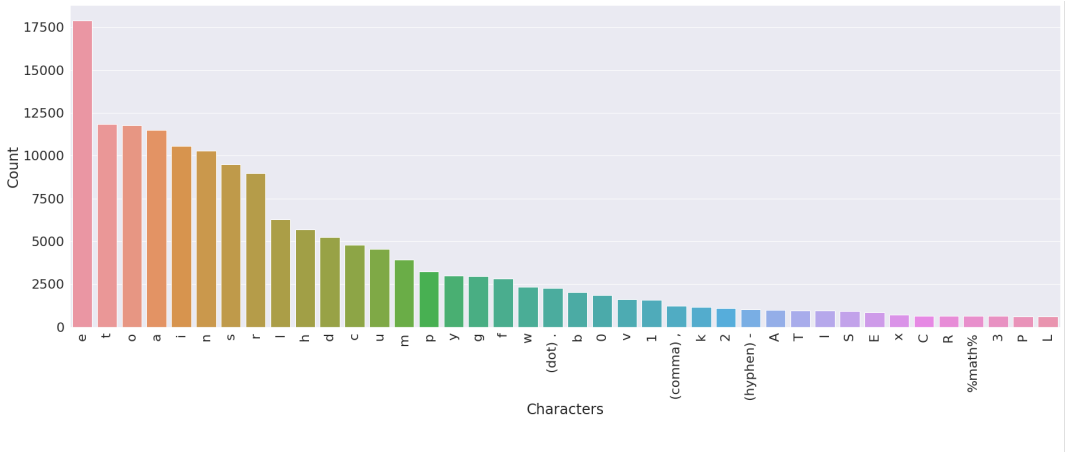


Fig. 6: Frequency of the top 40 common characters in the dataset.

To allow researchers to equally compare between different techniques, we randomly selected 75% of the data to be training data and the remaining 25% to be test data. Researchers are expected to determine their own train/validation split with the 75% of the training data.

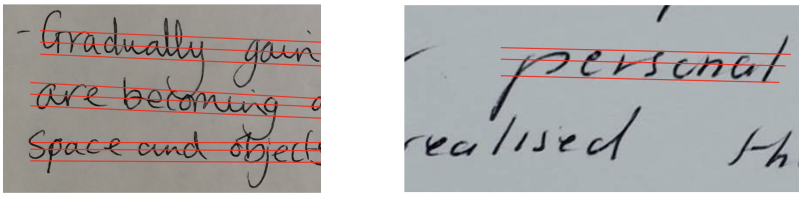
5 Benchmark

Building upon works in scene text, we evaluated the dataset with different state-of-the-art methods in localisation and recognition. Specifically, instance segmentation with Mask R-CNN [12] frameworks were used to localise handwritten text and the Clova AI deep text recognition framework [4] to perform text recognition. The benchmark and implementation details are provided at `https://github.com/GoodNotes/GNHK-dataset`.

5.1 Text localisation benchmark

Recent handwriting recognition frameworks investigating on the IAM Handwriting database typically forgo the text localisation steps [9,35]. More similar to a scene text dataset (e.g., [14,23]), the GNHK dataset poses several challenges: 1) the GNHK dataset includes camera captured images which are not limited to handwritten text (i.e., there are printed text, images, etc.), 2) the images are not perfectly aligned with varying degrees of brightness, luminosity, and noise, and 3) there are no set lines guides/spacing, and image size/resolution restrictions.

Unlike printed text, handwritten allographs widely differ between individuals. Especially with the absence of line guides, individuals' writing style on character ascenders and/or descenders (i.e., the portion of a character that extends above or below the baseline) complicate the assignment of text boxes (see Fig. 7 for examples).



(a) **(b)**

Fig. 7: Examples of the handwriting with problematic character ascenders and descenders. For example in (a), the "y" in *gradually* and "g" in *becoming* significantly increasing the bounding box size eventually overlapping with the regions of the words below.

We modelled the text localisation task using instance segmentation problem where the Mask R-CNN [12] framework was used as a baseline. Our dataset does

not contain pixel-level segmentation masks separate between word vs non-word. To circumvent this issue, given the polygon around each word, the maximum and minimum points in the x and y direction of each polygon were used as the bounding box for the R-CNN component of the Mask R-CNN framework. The polygon itself was then used as the segmentation mask.

The text localisation task was built upon the detectron2 [34] framework. The network consisted of a backbone of a ResNet-50 [13] with FPN [16] pretrained on ImageNet. Furthermore, the whole network was pretrained on the MS COCO [17] dataset for segmentation. We compared the results of the Mask R-CNN to Faster R-CNN [26] within the same detectron2 framework.

5.2 Text recognition benchmark

We modelled the text recognition benchmark as a segmented offline handwriting recognition problem [25,10]. That is, the ground truth bounding boxes for each word are used to crop word images and the word images are fed into a neural network to predict the corresponding text. Recent works have demonstrated segmentation free offline handwriting recognition [35,33,36,9], however considering that our dataset includes content other than handwritten text, we opted with a segmented offline handwriting recognition framework.

We leveraged Clova AI's deep text recognition framework for scene text to assess the handwriting recognition component. Clova AI's deep text recognition framework consists of four major components: transformation, feature extraction, sequence modelling, and prediction. Table 4 shows the settings that we considered for our benchmark (eight possible configurations in total). Note that we did not consider using the VGG image features as the performance was lower compared to ResNet in [4].

Table 4: Considered configurations of the text recognition benchmark. TPS - *thin-plate spline*, BiLSTM - *Bidirectional Long short-term memory*, CTC - *Connectionist temporal classification*

Component	Configurations
Transformations	{None, TPS}
Sequence	{None, BiLSTM}
Prediction	{CTC, Attention}

It is important to note that it was essential to pre-process the data that was fed into the network. Specifically, words that 1) are unknown, 2) contained scribbles, 3) contained mathematical symbols, and 4) contained only punctuation were removed. Please see Section 3.2 for more details.

6 Benchmark Results

6.1 Text localisation results

To evaluate text localisation, we used the criteria: recall, precision, and f-measure with an intersection-over-union (IOU) > 0.5 ratio of the segmentation mask or bounding boxes (Shown in Fig. 8 and Table 5).

Fig. 8: Examples of the Mask R-CNN results.

Fig. 8 presents three examples of the instance segmentation results. We can observe that the majority of text are detected, with the exception of several shorter texts (e.g., * in Fig. 8(a), *of* in Fig. 8(b)). Furthermore, in Fig. 8(c) even though the *f* in "Jennifer" overlaps with "mankind" in the line below, the network managed to differentiate between the two words. Table 5 presents the quantitative results of Mask and Faster R-CNN. Both Mask R-CNN and Faster R-CNN have greater than 0.86 f-measure scores. It is clear that the f-measure score is a result of high precision suggesting that the network has a smaller tendency to detect false positives.

Table 5: Text localisation results with IOU > 0.5.

Frameworks	Recall	Precision	f-measure
Mask R-CNN	0.8237	0.9079	0.864
Faster R-CNN	0.8077	0.9215	0.860

6.2 Text recognition results

Text recognition was performed on the individually cropped images from the ground truth bounding boxes and the results are shown in Table 6. The main assessment criteria are the character accuracy rate (CAR) and word accuracy rate (WAR):

$$CAR = \frac{1}{M}\sum_{i}^{M} 1 - \frac{dist(gt_i, preds_i)}{N_i} \tag{1}$$

$$WAR = \frac{1}{M}\sum_{i}^{M} [dist(gt_i, preds_i) == 0] \tag{2}$$

where gt = ground truth, $preds$ = predicted words, $dist(a, b)$ is the edit distance between the text a and b, M is the number of words in the dataset and N_i is the length of the word i.

Table 6 presents the results with different configurations of the deep text recognition framework:

Table 6: Text recognition results. TPS - *thin-plate spline*, BiLSTM - *Bidirectional Long short-term memory*, CTC - *Connectionist temporal classification*

Transformation	Sequence	Prediction	CAR	WAR
TPS	BiLSTM	Attention	0.861	0.502
None	BiLSTM	Attention	0.808	0.377
TPS	None	Attention	0.764	0.453
None	None	Attention	0.587	0.430
TPS	BiLSTM	CTC	0.416	≈0.000
None	BiLSTM	CTC	0.403	≈0.000
TPS	None	CTC	0.399	≈0.000
None	None	CTC	0.405	≈0.000

The benchmark recognition method achieved the highest CAR of 0.861 and WAR of 0.502. This configuration uses attention-based decoding, TPS for image transformation, and BiLSTM for sequential modelling. We can see that the attention-based decoding significantly outperforms CTC based decoding. The results of [4] also showed that attention-based decoders perform better than CTC

based decoders, but by a much lesser extent. When comparing the performance with and without TPS, we can see that TPS improves the CAR and WAR. For example, setting the sequence modelling to BiLSTM and prediction to attention decoding, the performance TPS improves the CAR by 5.3% and WAR by 12.5%. Similar results are found when the sequence modelling is not used (CAR increase=17.7%, WAR increase=2.3%). The results are likely suggesting that text alignment and input image normalisation assists in the recognition. Similarly, Table 6 shows that the BiLSTM consistently improves the CAR. When investigating the performance of including the BiLSTM without TPS, it is clear that the CAR substantially improves. However, the performance measured by the WAR decreases with the BiLSTM (0.377 vs 0.430).

7 Conclusion

To fill the gap in the literature for a new handwriting recognition dataset, we presented the GNHK dataset. The GNHK dataset consists of unconstrained camera-captured images of English handwritten text sourced from different regions around the world. The dataset is modelled after scene text dataset providing opportunities to novel investigate localisation with instance segmentation or object detection frameworks and text recognition techniques. In this paper, we demonstrated benchmark results of text localisation and recognition with well-studied architectures. Future works could explore end-to-end approaches to perform localisation and recognition sequentially within the same framework. The dataset and benchmark are available at `https://github.com/GoodNotes/GNHK-dataset` and we look forward to contributions from the researchers and developers to build new models with this dataset.

8 Acknowledgements

Thanks to Paco Wong, Felix Kok, David Cai, Anh Duc Le and Eric Pan for their feedback and helpful discussions. We also thank Elizabeth Ching for proofreading the first draft of the paper.

We thank Steven Chan and GoodNotes for providing a product-driven research environment that allows us to get our creativity ideas to the hands of millions of users.

References

1. Basicfinder. `https://www.basicfinder.com/`, accessed: 2021-01-20
2. Image classification on MNIST. `https://paperswithcode.com/sota/image-classification-on-mnist`, accessed: 2021-01-20
3. Augustin, E., Carré, M., Grosicki, E., Brodin, J.M., Geoffrois, E., Prêteux, F.: Rimes evaluation campaign for handwritten mail processing (2006)

4. Baek, J., Kim, G., Lee, J., Park, S., Han, D., Yun, S., Oh, S.J., Lee, H.: What is wrong with scene text recognition model comparisons? dataset and model analysis. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4715–4723 (2019)
5. Bernhard, A.: What your handwriting says about you. `https://www.bbc.com/culture/article/20170502-what-your-handwriting-says-about-you` (Jun 2017), accessed: 2021-01-20
6. Bluche, T.: Joint line segmentation and transcription for end-to-end handwritten paragraph recognition. In: Advances in Neural Information Processing Systems. pp. 838–846 (2016)
7. Bluche, T., Louradour, J., Messina, R.: Scan, attend and read: End-to-end handwritten paragraph recognition with mdlstm attention. In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). vol. 1, pp. 1050–1055. IEEE (2017)
8. Bukhari, S.S., Shafait, F., Breuel, T.M.: The iupr dataset of camera-captured document images. In: International Workshop on Camera-Based Document Analysis and Recognition. pp. 164–171. Springer (2011)
9. Coquenet, D., Chatelain, C., Paquet, T.: End-to-end handwritten paragraph text recognition using a vertical attention network. arXiv preprint arXiv:2012.03868 (2020)
10. Dutta, K., Krishnan, P., Mathew, M., Jawahar, C.: Improving cnn-rnn hybrid networks for handwriting recognition. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 80–85. IEEE (2018)
11. Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., Schmidhuber, J.: A novel connectionist system for unconstrained handwriting recognition. IEEE transactions on pattern analysis and machine intelligence **31**(5), 855–868 (2008)
12. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
13. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
14. Karatzas, D., Gomez-Bigorda, L., Nicolaou, A., Ghosh, S., Bagdanov, A., Iwamura, M., Matas, J., Neumann, L., Chandrasekhar, V.R., Lu, S., et al.: Icdar 2015 competition on robust reading. In: 2015 13th International Conference on Document Analysis and Recognition (ICDAR). pp. 1156–1160. IEEE (2015)
15. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
16. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
17. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
18. Liu, C.L., Yin, F., Wang, D.H., Wang, Q.F.: Casia online and offline chinese handwriting databases. In: 2011 International Conference on Document Analysis and Recognition. pp. 37–41. IEEE (2011)
19. Long, S., He, X., Yao, C.: Scene text detection and recognition: The deep learning era. International Journal of Computer Vision pp. 1–24 (2020)
20. Mahmoud, S.A., Ahmad, I., Al-Khatib, W.G., Alshayeb, M., Parvez, M.T., Märgner, V., Fink, G.A.: Khatt: An open arabic offline handwritten text database. Pattern Recognition **47**(3), 1096–1112 (2014)

21. Marti, U.V., Bunke, H.: Using a statistical language model to improve the performance of an hmm-based cursive handwriting recognition system. In: Hidden Markov models: applications in computer vision, pp. 65–90. World Scientific (2001)
22. Marti, U.V., Bunke, H.: The iam-database: an english sentence database for offline handwriting recognition. International Journal on Document Analysis and Recognition **5**(1), 39–46 (2002)
23. Mishra, A., Alahari, K., Jawahar, C.: Scene text recognition using higher order language priors (2012)
24. Pechwitz, M., Maddouri, S.S., Märgner, V., Ellouze, N., Amiri, H., et al.: Ifn/enit-database of handwritten arabic words. In: Proc. of CIFED. vol. 2, pp. 127–136. Citeseer (2002)
25. Puigcerver, J.: Are multidimensional recurrent layers really necessary for handwritten text recognition? In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). vol. 1, pp. 67–72. IEEE (2017)
26. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence **39**(6), 1137–1149 (2016)
27. Riche, Y., Riche, N.H., Hinckley, K., Panabaker, S., Fuelling, S., Williams, S.: As we may ink?: Learning from everyday analog pen use to improve digital ink experiences. In: CHI. pp. 3241–3253 (2017)
28. Schall, M., Schambach, M.P., Franz, M.O.: Multi-dimensional connectionist classification: Reading text in one step. In: 2018 13th IAPR International Workshop on Document Analysis Systems (DAS). pp. 405–410. IEEE (2018)
29. Shi, B., Bai, X., Yao, C.: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE transactions on pattern analysis and machine intelligence **39**(11), 2298–2304 (2016)
30. Su, T., Zhang, T., Guan, D.: Corpus-based hit-mw database for offline recognition of general-purpose chinese handwritten text. International Journal of Document Analysis and Recognition (IJDAR) **10**(1), 27 (2007)
31. Viard-Gaudin, C., Lallican, P.M., Knerr, S., Binter, P.: The ireste on/off (ironoff) dual handwriting database. In: Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318). pp. 455–458. IEEE (1999)
32. Voigtlaender, P., Doetsch, P., Ney, H.: Handwriting recognition with large multidimensional long short-term memory recurrent neural networks. In: 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 228–233. IEEE (2016)
33. Wigington, C., Tensmeyer, C., Davis, B., Barrett, W., Price, B., Cohen, S.: Start, follow, read: End-to-end full-page handwriting recognition. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 367–383 (2018)
34. Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: Detectron2. `https://github.com/facebookresearch/detectron2` (2019)
35. Yousef, M., Bishop, T.E.: Origaminet: Weakly-supervised, segmentation-free, one-step, full page text recognition by learning to unfold. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14710–14719 (2020)
36. Yousef, M., Hussain, K.F., Mohammed, U.S.: Accurate, data-efficient, unconstrained text recognition with convolutional neural networks. Pattern Recognition **108**, 107482 (2020)

37. Zhang, H., Liang, L., Jin, L.: Scut-hccdoc: A new benchmark dataset of handwritten chinese text in unconstrained camera-captured documents. Pattern Recognition **108**, 107559 (2020)
38. Zhu, Y., Xie, Z., Jin, L., Chen, X., Huang, Y., Zhang, M.: Scut-ept: a new dataset and benchmark for offline chinese text recognition in examination paper. IEEE Access (2018)