SN Computer Science (2024) 5:258
https://doi.org/10.1007/s42979-023-02583-6

REVIEW ARTICLE

Data Augmentation for Offline Handwritten Text Recognition: A Systematic Literature Review

Arthur Flor de Sousa Neto[1] · Byron Leite Dantas Bezerra[1] · Gabriel Calazans Duarte de Moura[1] · Alejandro Héctor Toselli[2]

Received: 28 October 2023 / Accepted: 26 December 2023

Abstract
Offline Handwritten Text Recognition (HTR) systems concern the automatic recognition and transcription of handwritten text from scanned images to digital media. Recently, HTR research field has become increasingly important due to the growing need for digitizing documents and automating data entry across various industries. However, achieving satisfactory results depend on the amount of available samples to train an optical model. Creating and labeling large enough datasets for this purpose often require significant time and effort, that in some situations may be impractical. To address this problem, data augmentation approaches are commonly used as an essential component of HTR systems. In this way, the present work aims to identify, explore, and analyze the scope of data augmentation approaches for offline HTR systems. Furthermore, we detailed our research protocol and answered four pertinent research questions, which enabled us to discuss trends and possible gaps. A search was conducted across five scientific databases, focusing on papers published between 2012 and 2023. The search yielded 976 primary papers, with 32 meeting the criteria for inclusion in this review. Our results indicate that handwriting synthesis is an emerging research field, and we observed that Digital Image Processing (DIP) is still widely used as an image generator. Nevertheless, the application of Generative Adversarial Networks (GAN) has gained traction in recent years owing to its impressive ability to synthesize images of handwritten text with arbitrary style and content. In addition, we explored and analyzed the most commonly used datasets and text recognition levels in the selected works.

Keywords Systematic literature review · Data augmentation · Handwriting synthesis · Handwritten text generation · Handwritten text recognition

Introduction

Writing has been a significant milestone for humanity. It has allowed us to record and preserve knowledge, which can be passed on from one generation to the next. In modern times, writing has become indispensable, particularly in the industrial/commercial sector, where it performs a role in facilitating business transactions, such as administrative, legal, and financial documents [1].

In the digital age and with process automation, handwriting has lost its strength with the preference for online platforms. However, there are still documents that require validation through handwriting, such as forms [2], medical prescriptions [3], and bank checks [4]. As a consequence, modern systems must adjust and incorporate offline documents, which may include partially or entirely handwritten material. This is essential for migrating of historical data to digital environments.

In this context, the research field of offline Handwritten Text Recognition (HTR) has gained notoriety due to its objectives of identifying, recognizing, and transcribing cursive texts from images to digital media (ASCII, Unicode) [1, 5]. However, this is not a simple task, and the most difficult challenge in this research area is related to the complexity

✉ Alejandro Héctor Toselli
ahector@prhlt.upv.es

Arthur Flor de Sousa Neto
afsn@ecomp.poli.br

Byron Leite Dantas Bezerra
byron.leite@upe.br

Gabriel Calazans Duarte de Moura
gcdm@ecomp.poli.br

[1] Universidade de Pernambuco, Recife, Pernambuco, Brazil

[2] Universitat Politècnica de València, Valencia, Spain

Published online: 10 February 2024

and variability of human handwriting. Unlike printed text, which has a homogeneous pattern, cursive writing has several factors that alter the pattern, due mainly to differences in style between writers and in style of a same writer within the same text. Consequently, handwriting recognition attracts attention in its development in both academic and industrial areas [1–4].

In recent decades, HTR systems have evolved significantly. The first approach with a great impact on handwriting recognition was based on Hidden Markov Model (HMM) models [6–8]. This approach led to the combination of Long Short-Term Memory (LSTM) neural networks for feature extraction, along with HMM for text decoding. This architecture also yielded good results with the use of Multidimensional LSTM (MDLSTM) layers [9, 10], and Connectionist Temporal Classification (CTC) as a way to calculate the loss function of the optical model [11].

Although MDLSTM-based systems have shown promise, their high computational cost has encouraged alternatives such as Convolutional Neural Networks followed by Bidirectional LSTM layers (CNN-BLSTM) [12–14]. Currently, optical models aim for low computational cost alongside high recognition performance, such as architectures based solely on convolutional networks, compact architectures, or networks based on attention mechanisms [14–17]. It is noteworthy that even with the evolution of optical models, language models have been used to complement the text decoding step, which provides better results than using the optical model alone [18, 19].

Despite achieving good results in the academic field, optical models are still unsatisfactory for many industrial use cases. One of the problems is the restriction in the volume of labeled data for training an optical model, which occurs with each new type of document to be transcribed. In other words, it is necessary to invest time and effort in labeling a minimal relevant amount of data to train an optical model, which is often impossible due to the lack of data [4, 20–23].

To minimize the issue of data restriction, some data augmentation approaches are employed in the workflow of an optical model to create synthetic text images. The first and most common approach is to apply random transformations in image preprocessing; the second is to increase the knowledge through transfer learning; and the third is to combine the preprocessing transformations with a robust model for low data volume. However, these approaches still have their limitations, which eventually lead to premature overfitting in the training of optical models.

Therefore, the present work presents a systematic literature review on data augmentation applied to offline handwritten text recognition. The aim is to identify, explore, and analyze the existing approaches for generating synthetic images of handwritten text, which can enhance the training of optical models. Thus, we intend to map the progress achieved in the last decade and discuss possible trends based on the knowledge obtained from the state-of-the-art. This period of time was selected due to the most significant recent advances in machine learning, which are interesting for the context of computer vision.

The rest of this paper is structured as follows: The section "Protocol Mapping" describes the methodology applied for the systematic review. Next, in the section "Results", the results of the review are presented. The section "Discussion" answers the research questions and discusses the state-of-the-art. Finally, the section "Conclusion" presents the conclusions of the current work.

Protocol Mapping

Following the guidelines and protocols proposed in the works of [24, 25], we developed our method to plan and execute our study. Our goal was to define a specific scope for collecting a set of scientific papers that would make the review useful, accessible, and reproducible for the academic community. To achieve this, we considered detailed documentation of the process.

Based on the established scope, the papers that fit into it are selected, analyzed, and scored according to the research questions defined by the authors before the review. In addition, we used a combination of Zotero[1] with a custom spreadsheet for paper management, tracking, and analysis.

The first step in a systematic review is to define its objective, followed by research questions, and only then the scope. Next, a search strategy is constructed, which includes search keywords, selection and exclusion criteria, and lastly quality assessment.

Research Objective

Data augmentation approaches generate synthetic data, which aim to minimize the problem of sample limitation in the training of deep learning models. Then, the main objective of the systematic review proposed in this work was to explore the literature and identify data augmentation approaches applied to offline handwritten text recognition.

It is important to note that handwriting recognition works often use some form of data augmentation, but it is not the central focus of the work and therefore not considered. On the other hand, works that solely focus on data augmentation of cursive text images are considered, as they have a central theme aligned with the objective of the review, and also offer great potential for application in the research area of handwriting recognition in general.

[1] https://www.zotero.org.

It is also worth mentioning that we have restricted the scope of this work to the specific research field of offline handwritten text recognition. In other words, we do not consider other research areas of text recognition, such as online and printed, nor sub-areas such as scene text, digits, signatures, and mathematical expression recognition.

Research Questions and Strategy

Research Questions (RQs) are defined to guide the review, direct the reading, and discuss specific aspects of the papers. Thus, we defined our research questions, aiming to identify and discuss the state-of-the-art regarding data augmentation approaches applied to offline handwritten text recognition. Four questions were formulated for this purpose:

- **RQ1**: What are the most commonly used recognition levels for data augmentation applied to offline handwritten text recognition?
- **RQ2**: What are the most commonly used datasets for data augmentation applied to offline handwritten text recognition?
- **RQ3**: What is the current state of data augmentation research field applied to offline handwritten text recognition?
- **RQ4**: What are the current challenges in data augmentation applied to offline handwritten text recognition?

The first question was defined to explore and understand the text structures most commonly used by data augmentation approaches. The second question aims to explore the most commonly used datasets and the languages with which they were built. The third question was defined to explore and understand the field of image data augmentation applied to offline handwritten text recognition systems. This involves understanding and analyzing different approaches within the research field to solve the same recognition problem. Finally, the fourth question was defined to discuss the gaps and trends in the research field of data augmentation of handwritten text images.

Based on our research objective and questions, we defined a set of keywords, period of time, and inclusion and exclusion criteria for our search strategy. Accordingly, we selected five academic databases to compose the review protocol: (i) ACM Digital Library[2]; (ii) IEEE Digital Library[3]; (iii) Science Direct[4]; (iv) Scopus[5]; and (v) Springer Link.[6]

[2] https://dl.acm.org.

[3] https://ieeexplore.ieee.org.

[4] https://www.sciencedirect.com.

[5] https://www.scopus.com.

[6] https://link.springer.com.

Table 1 Keywords and synonyms used to generate the search string

Keywords	Synonyms
Handwritten text recognition	Handwriting recognition, htr
Data augmentation	Image augmentation, generator
Synthetic	Synthesis, synthesize

We selected these databases due to their extensive coverage in the field of technology, comprising a vast collection of scientific papers widely recognized in the academic area. It is important to note that we selected only direct academic databases, which enable the reproducibility of the work through a search string. Additionally, the search was conducted using the advanced search mechanism of each platform mentioned, taking into consideration all metadata and full-text papers as search sources. In this way, the keywords were defined with the following objectives: (i) to be directed toward the research area of handwritten text recognition; (ii) to have a comprehensive search of data augmentation approaches; (iii) to capture the most relevant works. Moreover, we analyzed different terms and variations to ensure that the search string is precise and not generic, with the specific focus on the research objective. The keywords are described in Table 1.

Based on the defined keywords, the following search string has been determined: ("*handwritten text recognition*" OR "*handwriting recognition*" OR "*htr*") AND ("*data augmentation*" OR "*image augmentation*" OR "*generator*") AND ("*synthetic*" OR "*synthesis*" OR "*synthesize*").

The research field of offline handwritten text recognition has evolved significantly through deep learning, which is a relatively recent area of study. Thus, to better understand the studies developed over time, we have defined the period between 2012 and 2023, which is more than 10 years from the start of this review. This time frame is adequate to comprehend recent changes and advancements, as well as trends of interest.

The Exclusion Criteria (EC) are characteristics that disqualify the works from the review, which will not be included for the next step. The exclusion criteria are defined as follows:

- **EC1**: Works that are not in the computer science subject area;
- **EC2**: Works that are poster, tutorial, editorial, call for papers, shorts papers, book, book chapter, or thesis;
- **EC3**: Ongoing works;
- **EC4**: Works that are not in English;
- **EC5**: Works that are not within the established period of time for the review;
- **EC6**: Works that are literature revisions or surveys;
- **EC7**: Duplicate works;

- **EC8**: Works that are not available in full. A work is considered unavailable only after contacting the corresponding authors and receiving no response;
- **EC9**: Works that are not within the scope of optical character recognition;
- **EC10**: Works that do not present, or focus on, the offline handwritten text recognition problem;
- **EC11**: Works that do not present, or focus on, a data augmentation approach applied to the offline handwritten text recognition problem;
- **EC12**: Works that reached less than five points on the Quality Criteria.

It is worth mentioning that the five-point threshold was defined to cover works that present high quality in both the technical and descriptive aspects [26]. Finally, the Inclusion Criteria (IC) are characteristics that qualify the works for the next step. The inclusion criteria are defined as:

- **IC1**: Works that present, or focus on, a data augmentation approach applied to the offline handwritten text recognition problem.

Research Steps and Information Extraction

With the research scope defined, we followed a four-step pipeline for study selection consisting of: (i) primary studies collection; (ii) preliminary selection; (iii) final selection; and (iv) quality assessment. For the first step, we applied the search string to the database sources. At this stage, some exclusion filters are applied directly from the search engines of each database. This refers to the period of time, subject area, document type, publication stage, and language. In addition, some platforms offer more filters than others, and hence, the same filters are considered and manually applied in the next step.

The titles and abstracts of the primary papers were briefly read. This first reading was performed by pairs of reviewers, where each reviewer included or excluded each paper, defining at least one inclusion or exclusion criteria. Any paper accepted by at least one reviewer advanced to the next step.

The papers selected in the previous step were fully read during the third step to look for false positives. Each paper was read by both reviewers and was again evaluated for inclusion or exclusion.

During the quality assessment step, the included papers were scored by the reviewers using the Quality Criteria (QC) defined below:

- **QC1**: Is there a detailed description of the motivations, objectives, and contributions of the research? (weight 0.5)
- **QC2**: Is there a detailed description of the dataset used? (weight 0.5)
- **QC3**: Are the datasets used publicly available? (weight 1.0)
- **QC4**: Is there a detailed description of the optical models used? (weight 0.5)
- **QC5**: Is there a detailed description of the data augmentation approach? (weight 1.5)
- **QC6**: Is the proposed approach applied in different recognition levels, such as words, lines, and paragraphs? (weight 1.5)
- **QC7**: Is the proposed approach applied in different languages, such as English, French, and Spanish? (weight 1.5)
- **QC8**: Is there a detailed description of the results achieved? (weight 1.0)
- **QC9**: Do the results contribute to handwritten text recognition research area? (weight 1.0)
- **QC10**: Is the source code of the approach publicly available? (weight 1.0)

For each question, the following scale was used: No (N) = 0.0 points; Partially (P) = 0.5 points; and Yes (Y) = 1.0 points. Additionally, each question has its corresponding weight, which contributes to the weighted average at the end of the assessment. Finally, we applied the EC12 (see the section "Research Questions and Strategy") based on the final score of each paper.

To facilitate the discussion of the papers, the adopted strategy for information extraction was to collect the following data in each study:

1. Search engine/base;
2. Publication year;
3. Authors' names;
4. Paper title;
5. Datasets;
6. Datasets languages;
7. Recognition level;
8. Model type;
9. Results.

Results

The initial corpus of the research comprised 976 primary papers, which 341 obtained from the IEEE Digital Library, 307 from Scopus, 194 from Springer Link, 80 from Science Direct, and 54 from the ACM Digital Library. Figure 1 shows the distribution of the primary studies among the databases in the first step of selection.

It is important to emphasize that the selection obtained in the first only used the filters provided by the

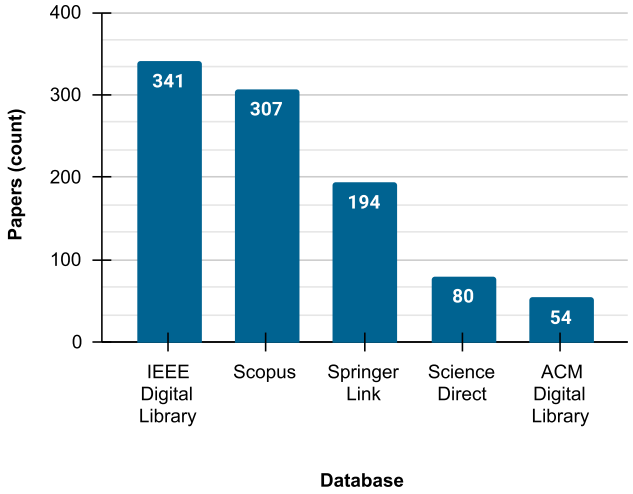

Fig. 1 Distribution of primary studies among academic databases—last access in September 2023

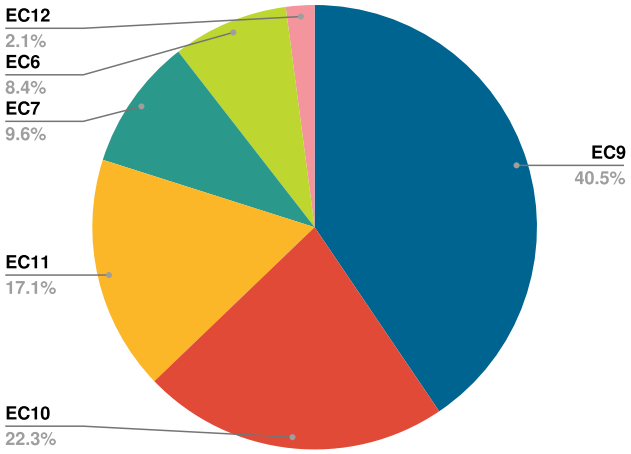

Fig. 2 Proportion of exclusion criteria (EC) applied

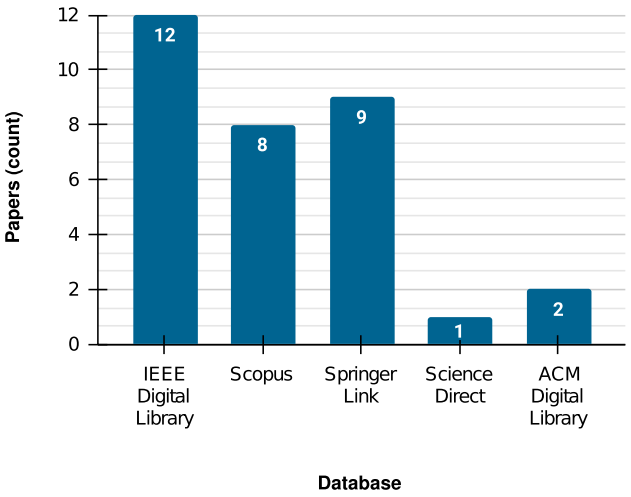

Fig. 3 Distribution of studies selected among databases

databases themselves (see the section "Research Steps and Information Extraction"). Nonetheless, the same filter criteria will be applied in the next step of selection.

Following the protocol defined, the papers were given to the reviewers during the second step, or preliminary selection. The reading consisted of titles and abstracts, seeking any information that could fit the paper into one of the 12 exclusion criteria. Out of the 976 primary studies, only 125 were selected through this first reading. It is worth noting that several papers were excluded due to the absence of filters in the research platform itself, as well as due to the duplication of studies among databases.

In the third step, the remaining papers were read in full. This secondary reading aimed to identify possible false positives that were not noticed before. This step was more demanding on the part of the reviewers and their considerations. Finally, the results of the evaluations served as the basis for selecting the works for the next step.

The full reading showed that many of the papers that were previously believed to be within the scope of the review were not related to the central theme. For example, works on handwritten text recognition with no focus on data augmentation, and works on scene text recognition and handwritten digit string recognition were the most common. Among the 125 papers previously selected, only 50 remained accepted for quality evaluation.

The last step of the review process, quality assessment, was conducted once again in pairs. The reviewers made notes on the selected papers based on the quality criteria described in the section "Research Steps and Information Extraction". In this step, EC12 was applied based on the final score of each paper, which only 32 reached the minimum required score.

Regarding the application of ECs, their results are reported next. Overall, out of the 976 studies obtained from research platforms, the most common exclusion criterion was EC9, which corresponds to works that are not related to the scope of optical character recognition, reaching in 351 exclusions (40.5%). The second criteria most common was EC10 with 193 exclusions (22.3%), which corresponds to studies with different contexts applications, such as online, printed, and scene text recognition. The third criterion most common was EC11, with 148 exclusions (17.1%), which indicate studies of offline handwritten text recognition, but with no focus on data augmentation as a central theme of the work. More data on the exclusion criteria are shown in Fig. 2.

After completing the selection process, we observed the new distribution regarding the databases used: 12 from IEEE Digital Library, 8 from Scopus, 9 from Springer Link, 2

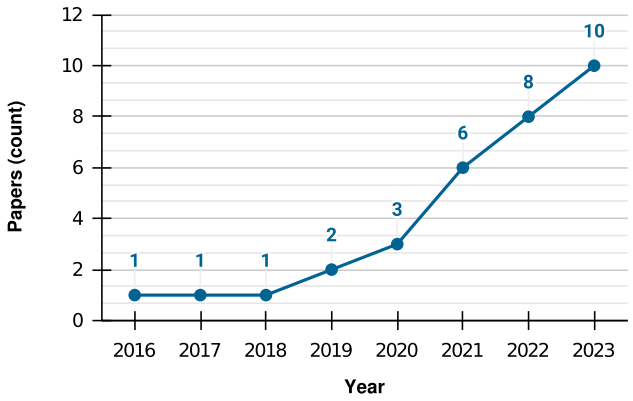

Fig. 4 Distribution of selected studies over time based on year of publication. No papers were selected with a publication date prior to 2016

from ACM Digital Library, and 1 from Science Direct. Figure 3 shows the distribution of studies among the databases after the four steps.

Another important data to analyze are the distribution of selected studies over time through the year of publication. As shown in Fig. 4, we observe a trend in the research area of data augmentation applied to offline handwritten text recognition. The publication of studies that fit within the scope of this review began in 2016 with only 1 publication. Since then, the number of publications has been increasing, reaching 8 publications in 2022. It is worth noting that 2023 is not yet complete, providing only partial data; however, it already shows a promising trend with 10 papers. In addition, no papers published before 2016 were selected.

Additionally, Fig. 5 presents the overview of the systematic review conducted in this work.

Finally, the quality assessment scores of each work are presented in Table 2 in descending order.

The following subsections are structured to present different aspects of the selected papers. Among these aspects, we present the recognition tasks and the datasets used in each work. We also present the types of approaches and the contribution to the state-of-the-art of each work.

Recognition Tasks and Datasets

The reviewed works present a wide range of applications in the research field of offline handwritten text recognition. This corresponds to our expectations, as there are different recognition levels and different application contexts.

We consider recognition levels as different text structural components: (i) characters; (ii) words; (iii) lines; and (iv) paragraphs. It is important to note that the paragraph level automatically considers the other three structural components (line, word, and character); the line level also includes the word and character structural components; and the word level involves the character one. This is also evaluated, as there is a degree of complexity in the recognition process for each text structure.

For the application contexts, we consider the characteristics and challenges that the datasets offer. In this way, sets of images that involve multiple writers tend to be more challenging than sets with only one writer, due to the high variability in writing patterns. On the other hand, sets of images of historical documents tend to be more challenging than images of form documents due to the high level of noise in the documents themselves. Furthermore, languages also influence recognition as they directly impact on the charset

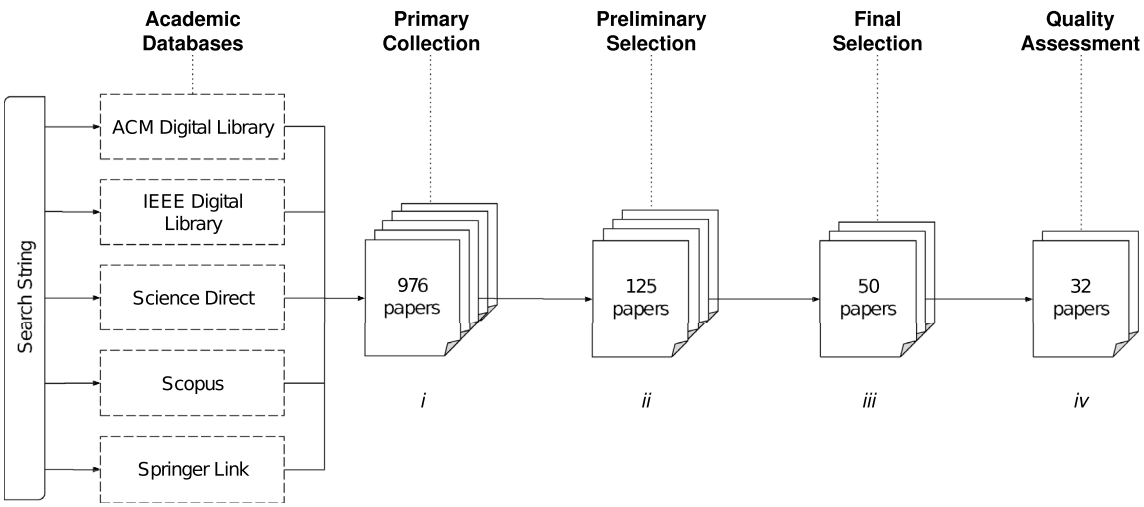

Fig. 5 Overview of the systematic review conducted. (i) A search string was used in five academic databases to collect primary studies. (ii) A preliminary screening by reading titles and abstracts identified relevant studies. (iii) A full-text reading removed false positives in the final selection. (iv) A quality assessment selected studies that fit the scope defined

Table 2 Quality scores of approved works in descending order

Year	Paper	References	Score
2020	ScrabbleGAN: Semi-Supervised Varying Length Handwritten Text Generation	[27]	7.75
2022	HiGAN+: Handwriting Imitation GAN with Disentangled Representations	[28]	7.25
2021	Handwritten Text Generation via Disentangled Representations	[29]	7.00
2017	Data Augmentation for Recognition of Handwritten Words and Lines Using a CNN-LSTM Network	[30]	6.88
2022	Content and Style Aware Generation of Text-Line Images for Handwriting Recognition	[31]	6.75
2021	JokerGAN: Memory-Efficient Model for Handwritten Text Generation with Text Line Awareness	[32]	6.75
2020	GANwriting: Content-Conditioned Generation of Styled Handwritten Word Images	[33]	6.50
2019	Manifold mixup improves text recognition with CTC loss	[34]	6.50
2022	Script-Level Word Sample Augmentation for Few-Shot Handwritten Text Recognition	[35]	6.50
2023	Handwritten Text Generation from Visual Archetypes	[36]	6.38
2023	Conditional Text Image Generation with Diffusion Models	[37]	6.25
2023	Handwritten Text Generation with Character-Specific Encoding for Style Imitation	[38]	6.25
2023	Generation of a synthetic handwritten Bangla compound character dataset using a modified conditional GAN architecture	[39]	6.00
2022	Active Transfer Learning for Handwriting Recognition	[40]	5.75
2022	Application of Deep Convolutional Generative Adversarial Network for Russian Handwritten Text Recognition	[41]	5.75
2020	Learn to Augment: Joint Data Augmentation and Network Optimization for Text Recognition	[42]	5.75
2023	Improving OCR Accuracy for Kazakh Handwriting Recognition Using GAN Models	[43]	5.75
2023	WordStylist: Styled Verbatim Handwritten Text Generation with Latent Diffusion Models	[44]	5.75
2021	Handwriting Transformers	[45]	5.63
2019	Adversarial Generation of Handwritten Text Images Conditioned on Sequences	[46]	5.50
2022	Improving Handwriting Recognition for Historical Documents Using Synthetic Text Lines	[47]	5.50
2021	Multilingual GAN: A Multilingual GAN-based Approach for Handwritten Generation	[48]	5.50
2022	SLOGAN: Handwriting Style Synthesis for Arbitrary-Length and Out-of-Vocabulary	[49]	5.50
2016	A Method of Synthesizing Handwritten Chinese Images for Data Augmentation	[50]	5.38
2023	Crosslingual Handwritten Text Generation Using GANs	[51]	5.38
2023	Content-Aware Urdu Handwriting Generation	[52]	5.25
2023	Zero-shot Generation of Training Data with Denoising Diffusion Probabilistic Model for Handwritten Chinese Character Recognition	[53]	5.13
2021	A Convolutional Neural Network-based Ancient Sundanese Character Classifier with Data Augmentation	[54]	5.00
2018	A study of data augmentation for handwritten character recognition using deep learning	[55]	5.00
2022	Generative adversarial network based adaptive data augmentation for handwritten Arabic text recognition	[56]	5.00
2021	Improving Handwritten Arabic Text Recognition Using an Adaptive Data Augmentation Algorithm	[57]	5.00
2023	AFFGANwriting: A Handwriting Image Generation Method Based on Multi-feature Fusion	[58]	5.00

used by the optical model. Thus, we evaluated the applicability of the data augmentation approach for text images on different datasets.

As shown in Table 3, the different recognition levels found in the studies are displayed. In general, the most utilized recognition level was at word level, followed by

Table 3 Recognition levels utilized in the studies

References	Character	Word	Line	Paragraph
[27]		✓		
[28]			✓	✓
[29]		✓		
[30]		✓	✓	
[31]			✓	
[32]		✓	✓	
[33]		✓		
[34]			✓	
[35]		✓		
[36]		✓		
[37]		✓		
[38]		✓	✓	
[39]	✓			
[40]		✓		
[41]			✓	
[42]		✓		
[43]		✓		
[44]		✓		
[45]			✓	✓
[46]		✓		
[47]			✓	
[48]			✓	
[49]		✓	✓	
[50]			✓	
[51]		✓		
[52]		✓		
[53]	✓			
[54]	✓			
[55]	✓			
[56]		✓		
[57]	✓	✓		
[58]		✓		

text-line level. This indicates that these two levels share similar challenges, where data augmentation applied to words can expand to text lines, and data augmentation applied to text lines can contract to words. The other levels, characters and paragraphs, were the least used among the studies.

Languages are intrinsic to the datasets, and the selected studies show a good diversity in this regard, in which English language is the starting point for nearly all studies. In that way, the dataset of the "Institut für Informatik und Angewandte Mathematik" (IAM) [59] was utilized by the works of [27–31, 34, 36–38, 40, 42, 44, 45, 48, 49, 51, 58], and is the most famous dataset within the study area, comprising 1,539 pages written by 657 different writers. Another dataset is the "Computer Vision Lab" (CVL-Database) [60], which is designed for writer retrieval and identification, including 311 different writers. It was employed in the works of [27, 32, 35, 37, 40, 49]. The historical Bentham dataset [61], which consists of images of letters by the English philosopher Jeremy Bentham (1748–1832), was utilized in the work of [47]. Furthermore, the English subset of the Maurdor dataset [62] was used in [34] and contains heterogeneous images of different types of documents. Finally, the dataset "GoodNotes Handwriting Kollection" (GNHK) [63] comprises unrestricted camera-captured images of English handwritten text from various regions, characterized by diverse styles and increased noise, was used in the work of [38].

French was the second most commonly used language in the selected studies. However, only two datasets were available in this language. The "Reconnaissance et Indexation de données Manuscrites et de fac similÉS" (RIMES) dataset [64], which comprises handwritten letters from various writers, is considered simple due to its image quality and uniformity. The RIMES dataset was used in the studies [27, 29–31, 34, 37, 40, 42, 46, 49]. The second dataset was the French subset of Maurdor, which was utilized in the study [34].

Regarding German language, the CVL-Database [60] was the most frequently used dataset [27, 32, 35, 40, 49], and the READ dataset [65], containing historical German documents, was utilized in the study of [30]. Finally, the Bullinger dataset, presented and utilized in the study [47], comprises historical letters written in German sent to the reformer Heinrich Bullinger (1504–1575).

The Arabic language was also significantly utilized, offering substantial variation across datasets. A first remarkable dataset in Arabic is the Maurdor [62] subset, which has approximately 13,000 text-line samples and was used in the study of [34]. OpenHaRT [66], boasting a large database of approximately 710,000 images, was utilized in the study of [46]. The dataset from the "Institute for Communications Technology/Ecole Nationale d'Ingénieurs de Tunis" (IFN/ENIT) [67], employed in the studies conducted by [35, 56, 57], offers character and word recognition capabilities and encompasses around 411 different writers. The "Arabic Handwriting Data Base" (AHDB) [68], used in the studies conducted by [56, 57], consists of characters and words derived from numerical values and bank check filling. Finally, the "Multilingual Automatic Document Classification Analysis and Translation" (MADCAT) database [69–71] was used by the work [51], and consists of handwritten Arabic documents scanned at high resolution, totaling 750,000 images of segmented lines.

The "Handwritten Kazakh and Russian" (HKR) dataset [72], representing Kazakh and Russian languages, has been utilized in the studies by [35, 41, 43]. Furthermore, the Chinese language is also represented by the dataset from the "Chinese Academy of Sciences' Institute of Automation" (CASIA) [73], that was used in the works of [34, 37, 50, 53]. CASIA offers online and offline recognition versions. In the

offline version, the studies worked with almost 1.4 million labeled characters.

Some datasets were less utilized, either because they were a subset of another dataset or because they were proposed for a specific competition. For the Spanish language, the "Spanish Numbers" dataset [74] was employed in the work of [31]. This dataset comprises handwritten numerals written by 30 different writers. For Vietnamese language, a small dataset was introduced in the Cinnamon AI Marathon competition: "the Cinnamon Handwritten OCR for Vietnamese Address Challenge" [75], which contains handwritten address images and was utilized in the work of [48]. Another Vietnamese dataset is the "Vietnamese Online Handwritten Text Recognition" (VNonDB) [76], used by [51], which is an online handwritten Vietnamese dataset released as a challenge for ICFHR2018 and converted into an offline version, comprising 100,000 images of word-level segmentations. In the case of Latin, a subset of the Bullinger dataset [47] was used in the work of [47]. The languages of Bangla and Mongolian are represented via the "CMATERdb" [77] and "Mongolian-Database" [35] datasets, respectively, as utilized in the studies by [39] and [35]. For the Sudanese language, the historical "Sundanese Palm Leaf Manuscript" (HSPLM) dataset [78] was used by [54] for character recognition. For the Urdu language, the work of [52] utilized two databases: the "Center of Language Engineering" (CLE) [79] and UCOM [80]. The CLE database contains 18,000 Urdu ligatures in Unicode format, while the UCOM database comprises 48 distinct lines of Urdu text authored by 100 different writers. For the Japanese language, the simulated "Japanese Handwriting Dataset" (JHD) [32] was adopted in the work of [32], along with the "ETL Character Database" (ETL) [81] in the work of [55]. These last two datasets contain handwritten character images.

Finally, Table 4 shows the distribution of all selected studies according to datasets and respective languages.

Table 4 Distribution of the selected studies across the datasets and respective languages (alphabetical order)

Language	Dataset	References
Arabic	AHDB [68]	[56, 57]
	IFN/ENIT [67]	[35, 56, 57]
	MADCAT [69–71]	[51]
	Maurdor [62]	[34]
	OpenHaRT [66]	[46]
Bangla	CMATERdb [77]	[39]
Chinese	CASIA [73]	[34, 37, 50, 53]
English	Bentham [61]	[47]
	CVL-Database [60]	[27, 32, 35, 37, 40, 49]
	GNHK [63]	[38]
	IAM [59]	[27–31, 34, 40]
		[36–38, 45, 48]
		[44, 49, 51, 58]
	Maurdor [62]	[34]
French	Maurdor [62]	[34]
	RIMES [64]	[27, 29–31, 34]
		[37, 40, 42, 46, 49]
German	Bullinger [47]	[47]
	CVL-Database [60]	[27, 32, 35]
		[40, 49]
	READ [65]	[30]
Japanese	ETL [81]	[55]
	JHD [32]	[32]
Kazakh	HKR [72]	[35, 43]
Latin	Bullinger [47]	[47]
Mongolian	Mongolian-Database [35]	[35]
Russian	HKR [72]	[35, 41]
Spanish	Spanish Numbers [74]	[31]
Sundanese	HSPLM [78]	[54]
Urdu	CLE [79]	[52]
	UCOM [80]	[52]
Vietnamese	Cinnamon [75]	[48]
	VNonDB [76]	[51]

Data Augmentation

In this subsection, we delve into data augmentation approaches, which have been classified into three main categories in the offline handwriting recognition research field. The first, Digital Image Processing, comprises traditional methods with lower computational requirements and a stand-alone functionality. The second, Transfer Learning, encompasses strategies that utilize pre-existing datasets to enhance the training of optical models. Finally, Deep Learning refers to advanced techniques using deep learning architectures to augment data through image synthesis.

Additionally, it is noteworthy that some studies have used an end-to-end solution, creating their own data augmentation methods and testing them with standard handwriting recognition metrics. Others used metrics typically for image generation. In either case, these methods could be beneficial for the offline handwritten text recognition research field.

Digital Image Processing

The Digital Image Processing (DIP) involves algorithms to apply transformations to digital images. DIP is a traditional approach to data augmentation in the field of offline handwritten text recognition research, and its algorithms are widely used with optical models [30, 34, 35, 42, 50, 54, 55, 57]. This allows to apply randomness in text images transformations, which in turn helps to prolong the training process of optical models and prevent premature overfitting.

Initially, Shen and Messina [50] explored character-level segmentation to evaluate various strategies for generating synthetic text lines from isolated characters. These strategies range from simple processing, such as placing characters one after the other; to more complex processing, such as using the coordinates of characters in annotated lines to create images of text lines with a more realistic appearance. A key strength of this approach was that generating full pages rather than single lines led to more realistic images, preserving the placement and relative positioning of characters in text lines. However, the strategy still encounters challenges with deformations caused by varying height and width ratios of characters, suggesting a need for refinement in the page synthesis methodology. Subsequently, an optical model was trained using a balanced combination of synthetic and real images, which contributed to a relative improvement of 10.4% on CASIA dataset. This underscores the potential of augmenting the training data with synthetic images.

Wigington et al. [30] noted that Shen and Messina's [50] proposal is promising, but highlighted its dependency on a character-level dataset to be effective. Thus, they proposed new methods for image normalization and deformation. The suggested normalization method is adaptive, accommodating variations in handwriting scale, which consequently improves the optical model's tolerance to writing differences. Further, they introduced a distortion grid implementing random deformations to apply slight scale and inclination variations, character by character, within each word. This, however, might be computationally intensive and may require more optimization. Remarkably, they achieved Character Error Rates (CER—the lower, the better) of 3.0%, 1.4%, and 5.0% on the IAM, RIMES, and READ datasets.

Hayashi et al. [55] presented another approach for Kanji character recognition, based on a novel data augmentation technique involving statistical character structure models. The goal was not only to generate Kanji character images of diverse cursive writing styles, but also to do so using a unique probability distribution of character strokes that were immune to the influence of the original image. However, complete control over the generated characters' structure remained a challenge, causing instability in the character images. Despite this, the approach contributed to a notable Character Accuracy Rate (CAR—the higher, the better) of approximately 93.1% on the ETL-9B dataset.

Using manifold mixup as basis, Moysset and Messina [34] proposed a new training strategy for offline handwritten text recognition systems. This strategy involves merging two input images or their corresponding feature maps and serves as a regularizer in the optical model training. The study did not compare their technique directly with other advanced data augmentation methods, leaving a gap in understanding its relative performance. The strategy also presented potential implementation complexities due to the need to adapt it to varying image sizes. Overall, they achieved a CER of 23.9%, 3.3%, 4.6%, 8.9%, 14.8%, and 10.5% on the CASIA, RIMES, IAM, and Maurdor French, English, and Arabic subsets.

Following, Luo et al. [42] proposed an adaptive data augmentation method for optical model training. This approach adaptively adjusts transformation functions based on the model's learning progress, thereby gradually increasing the difficulty of images. The method demonstrated broad applicability, enhancing text recognition performance across diverse settings, as evidenced by the achieved CER of 2.4% and 5.1% for the IAM and RIMES datasets, respectively. On the other hand, the use of custom fiducial points and joint learning may add to the complexity of implementation, potentially making it difficult to use in general. The work is accessible in a public repository.[7]

Eltay et al. [57] proposed a data augmentation method based on the frequency distribution of characters across the dataset. In this way, the method gives more weight to less frequent characters in a word, aiming to balance the character distribution across the dataset. While this approach effectively manages class imbalances, it does lean heavily on the character occurrence probabilities, which might make it less adaptable to datasets with different character distributions. Furthermore, the approach, which primarily solves class imbalance, might have limited effectiveness in situations where imbalance is not a key issue. Using the Word Accuracy Rate (WAR—the higher, the better) metric, the method achieved 99.0%, 95.1%, and 93.6% for the *abc-d*, *abcd-e*, and *abcde-f* subsets of the IFN/ENIT dataset, respectively. For the AHDB dataset, they achieved 98.1%.

Meanwhile, Hidayat et al. [54] addressed the challenge of limited and historical data samples in the HSPLM dataset, using data augmentation techniques. They did not just use geometric transformations, but also added noise to the image background, and adjusted brightness to generate new samples, enhancing the variety of their data. Their approach, however, was not just about generating more data; they also carefully balanced the data at the character level. It is worth noting that while these methods showed promising results with a CAR of 97.4%, they were specifically tailored to the ancient Sundanese characters of this dataset. It is unclear how these techniques would fare with other languages or character sets. Moreover, the paper did not touch on potential downsides or limitations of their augmentation methods, such as the possibility of overfitting with too much augmentation.

Finally, Chen et al. [35] presented a rule-based handwritten word augmentation method at the script level. The method initially divides the handwritten word into curve components, applies deformations, and then joins them

[7] https://github.com/Canjie-Luo/Text-Image-Augmentation.

back together. The authors put their approach to the test, proving it outperforms the traditional augmentation methods in experiments. In this way, they used the WAR metric and achieved 30.5%, 81.5%, 73.0%, and 72.6% for the Mongolian-Database, CVL-Database, IFN/ENIT, and HKR datasets, respectively. However, this method has some potential drawbacks. For one, it involves a more complex process than traditional methods, which can make its use difficult. Another challenge is that it relies on prior knowledge of the languages being used. In addition, the work is available in a public repository.[8]

Transfer Learning

Transfer Learning proposals in the field of offline handwritten text recognition research involve storing the knowledge given by an optical model acquired in one dataset, and applying it to recognized samples from another dataset. Thus, two or more sets of document images that have good similarity in writing pattern can benefit from training a joint optical model. Even so, this approach was unusual and did not have much exploration as a data augmentation method, which only two selected studies used it [40, 52, 55].

In this context, Hayashi et al. [55] presented an additional feature of their Digital Image Processing (DIP) method by conducting a Transfer Learning experiment on the ETL-9B dataset, which they divided into three parts. In their comparisons of different training approaches, they found that sharing knowledge between subsets resulted in a modest but noteworthy improvement in character recognition. Specifically, the Character Accuracy Rate (CAR) increased by 1%, achieving an overall score of 94.5%.

Burdett et al. [40] proposed a combination of Transfer Learning and Active Learning as a solution for offline handwritten text recognition. In their work, the authors designed a training pipeline that leveraged pre-trained optical models within an Active Learning framework, thereby enhancing the learning outcomes. They tested their approach on IAM, RIMES, and CVL-Database datasets, and achieved CER values of 4.2%, 4.3%, and 4.8%, respectively, demonstrating the effectiveness of their method.

Recently, Memon et al. [52] developed a handwriting generation model trained initially on ligatures images and later fine-tuned via transfer learning on handwritten images from the CLE to UCOM database. Their results showed significant progress, achieving a Fréchet Inception Distance (FID—the lower, the better) score of 38.03, a Geometry Score (GS—the lower, the better) of 8.81×10^{-4}, and a recognition accuracy of 72.6%. The authors also highlighted the promise of transfer learning in handwriting tasks, especially when training data are limited, suggesting potential improvements when blending real and rendered handwriting data.

Deep Learning

Deep Learning is a machine learning technique that mimics how humans acquire certain types of knowledge. Nowadays, it enables optical models to achieve good results, which is why it is the most widely used approach [27–29, 31–33, 36–39, 41, 43–49, 51–53, 56, 58].

In this context, Alonso et al. [46] proposed using Generative Adversarial Networks (GANs) to generate synthetic images of handwritten words and integrate an optical model into the architecture. However, despite their innovative approach, there were still some artifacts visible in the generated images, indicating that the image quality might need further improvement. Thus, unlike other works that focused on reducing error rates, the authors aimed to measure the improvement obtained through the synthetic text images of the generative model. To this end, the authors used Fréchet Inception Distance (FID—the lower, the better) and Geometry Score (GS—the lower, the better) metrics, which achieved 23.94 and 8.58×10^{-4}, respectively. The FID is the current standard metric for assessing the quality of generative models, which compares the statistics of two distributions (generated and real images) calculating the distance between them. On the other hand, the GS metric involves comparing the geometrical properties of the underlying data manifold with those of the generated data. This method provides both qualitative and quantitative measures for evaluating the GAN's performance. Additionally, their model achieved a Word Error Rate (WER—the lower, the better) of 11.9% on the RIMES dataset.

Based on this same idea, Fogel et al. [27] introduced the ScrabbleGAN model to generate synthetic images of handwritten words. This generative model not only generates a wide range of images, but it also has the ability to adapt to new styles, enhancing its versatility. One of the main issues is that it assumes all characters have the same width, limiting the diversity and realism of the generated images. Furthermore, while the model can create diverse styles, there is a lack of finer control over text style parameters. The ScrabbleGAN model achieved an FID of 23.78, GS of 7.60×10^{-4}, and Inception Score (IS—the higher, the better) of 1.33. The model also delivered WER values of 11.3%, 23.61%, and 22.9% on the RIMES, IAM, and CVL-Database datasets, respectively. The work is available in a public repository.[9]

Following this research line, Kang et al. [33] presented the GANwriting model to generate also synthetic images of handwritten words. This model generates realistic images and can mimic specific writing styles, allowing it to create different handwritten styles for the same text content.

[8] https://github.com/IMU-MachineLearningSXD/script-level_aug_ICFHR2022.

[9] https://github.com/amzn/convolutional-handwriting-gan.

However, the model, a Sequence-to-Sequence (Seq2Seq) [82], has a limitation in that it can only synthesize short words. This limitation can sometimes reduce the model's flexibility and utility. Nevertheless, this innovative model achieved an FID score of 125.23 and an IS score of 1.33, demonstrating its ability to generate realistic text images that resemble those in the IAM dataset. This work is also available in a public repository.[10]

Bhunia et al. [45] introduced a new approach called Handwriting Transformers (HWT) that creates synthetic images of handwritten text using the Transformer model [83]. The HWT captures long- and short-range contextual relationships within the writing style sample through a self-attention mechanism [83]. Unlike other models, HWT can work with text of any length and any style, giving it a lot of flexibility. However, it is also complex and computationally expensive. Even with these challenges, HWT performed quite well, achieving FID of 19.40, GS of 1.01×10^{-2}, and IS of 1.36 on the IAM dataset. The work is available in a public repository.[11]

Liu et al. [29] developed the HTG-GAN model, which can synthesize text images with arbitrary length. The authors redefined the structural relationship between characters in a sequence by breaking the bond between style and content. This allows generating images with new styles and chosen content. However, HTG-GAN has the challenge of dealing with languages with many independent characters, such as Chinese or Japanese. This is because it uses an encoding strategy that considers the character's place in the alphabet, which does not work well with these languages. The generative model achieved an FID of 12.18 and a GS of 2.23×10^{-3}. When used for handwriting recognition, they achieved WERs of 10.2% and 20.5% in the RIMES and IAM datasets, respectively.

Huu et al. [48] developed the Multilingual-GAN model for synthesizing text images. This new model is distinguished by its ability to work efficiently in multiple languages without additional training. Moreover, it is capable of generating diverse character styles, which enhances the versatility of the output. An aspect of their approach is the application of perceptual loss, which ensures content consistency between the input and the generated images. However, it is not without shortcomings. The model currently yields results that may exhibit blur and insufficient stroke precision. Additionally, the generated images can contain artifacts, affecting the overall realism. Despite these limitations, the authors explored and emphasized the importance of both adversarial and perceptual losses for producing realistic handwritten images. The study is publicly available in a repository.[12]

Zdenek and Nakayama [32] proposed the JokerGAN, a new GAN architecture for offline handwritten text recognition. The model stands out due to its ability to use character sequences of variable lengths as conditional input, making it flexible and adaptable. It is also memory efficient, remaining largely unaffected by the size of the character set. This makes it possible to handle languages with a large number of characters, such as Japanese and Latin, simultaneously. An innovative feature of the model is its awareness of the vertical alignment of characters, which enhances the quality of generated handwritten text. However, the study does not delve into other computational costs such as training time, which could be a significant factor for large datasets. Despite these considerations, the model managed to surpass state-of-the-art models, achieving an FID of 9.18.

To enhance the recognition of the Arabic language, Eltay et al. [56] combined their previous work on adaptive algorithms with a GAN model. Their method managed the inherent issue of class imbalance in text data, a prevalent concern in this field. However, it is worth noting that the current work is restricted to generating individual words, not entire lines of text. Moreover, their efforts resulted in a WAR of 97.2%, 95.9%, and 93.8% for *abc-d*, *abcd-e*, and *abcde-f* subsets of the IFN/ENIT dataset. They also achieved a notable 99.3% for the AHDB dataset.

In their subsequent research, Kang et al. [31] showed that employing realistic synthetic texts during training is advantageous for enhancing the performance of handwritten text recognition. The authors switched the Seq2Seq model [82], a recurrent neural network (RNN), with the Transformer model [83], which is notable for its self-attention mechanism. This change enabled the generation of images with longer lines of text. However, the approach has limitations when handling special characters like accents, making it less effective for certain languages. Furthermore, to adapt to new handwriting styles, the model needs access to unlabeled text-line images, which could pose challenges in some situations. In addition, they achieved a CER of 8.62% and a WER of 26.69% on the IAM dataset. Similarly, on the RIMES dataset, they managed to attain a CER of 6.45% and a WER of 19.56%.

Luo et al. [49] improved their previous research by proposing the SLOGAN model, which synthesizes handwritten text images of arbitrary length. In their recent study, they synthesized writing data by parameterizing the style and controlling the parameters to generate new cursive writing styles. However, this system relies heavily on identifying the writer from the original images, which could limit its ability to cope with entirely new styles. While they can create new

[10] https://github.com/omni-us/research-GANwriting.

[11] https://github.com/ankanbhunia/Handwriting-Transformers.

[12] https://github.com/HoSyTuyen/MultilingualGAN.

words or sentences, it might have difficulty with particularly rare or complex ones. The generative model achieved an FID of 12.06 and a GS of 5.59×10^{-4}. The optical model reached a CER of 3.4%, 5.9%, and 14.1% on the RIMES, IAM, and CVL-Database datasets, respectively.

In their research, Spoto et al. [47] employed GANs to facilitate the recognition of handwriting in historical documents. This was achieved through the integration of authentic and synthetically generated handwriting samples. The effort was considerably successful, leading to a significant reduction in character error rate (CER) ranging from 3% to a notable 60%. Nonetheless, this study had its constraints. The dependency on large amounts of training data could pose difficulties with smaller datasets. Furthermore, the synthetic samples, while reflecting the targeted style, lacked the inherent variability of natural handwriting.

Gan et al. [28] proposed HiGAN+, a novel generative model based on disentangled representations. HiGAN+ enables the synthesis of realistic handwritten text images conditioned on arbitrary textual content and diverse cursive writing styles, allowing for the generation of paragraphs with different styles. However, it is worth noting that humans handwriting can be highly detailed and intricate, posing challenges for HiGAN+ in synthesizing text that captures all these intricacies. Nevertheless, the model achieved FID of 9.65 and IS of 1.41 on IAM dataset. In addition, the research is publicly available in a repository.[13]

Recently, Kudaibergen and Hamada [41] have focused on Russian handwritten text recognition. They employed GANs and used a model trained on synthetic data generated by ScrabbleGAN [27], resulting in a significant improvement in the optical model's performance. However, it's essential to note that the study was limited by its exploration of only one GAN architecture and a relatively low achieved accuracy. Moreover, the issue of optimal data size for training was raised but not fully investigated. Nonetheless, the experiment yielded promising results on HKR dataset, with a WAR increase up to 24.1% when combining different types of synthetic data.

Yeleussinov et al. [43] proposed a novel use of GANs for handwriting recognition. The GAN model consists of a handwriting word image generator and an image quality discriminator. In this way, the model is trained with multiple losses to learn the structural properties of texts and produce high-quality images of handwritten text. The study reached a CER of 11.15% and a WER of 25.65% on HKR dataset.

Das et al. [39] developed a GAN model to create synthetic handwritten Bangla compound characters. Their improved model, inspired by the Auxiliary Classifier GAN (AC-GAN), demonstrated an enhanced FID score compared to the original AC-GAN, which reached 7.81 on CMATERdb dataset. However, the study lacks comparative analysis with other state-of-the-art datasets, which may help provide a more comprehensive view of the model's performance and position within the research field. The study is publicly available in the repository.[14]

Wang et al. [58] proposed the AFFGANwriting model, which employs a VGG19-based style encoder to extract multiscale handwriting features and generate realistic handwriting images. The approach captures both global and local characteristics of handwriting, reaching an FID score of 28.65 on the IAM dataset.

Recently, Gui et al. [53] proposed a Denoising Diffusion Probabilistic Model (DDPM). This model transforms font library-based Chinese character images into handwritten samples. When tested on the CASIA dataset, the model trained with synthesized samples showed comparable recognition accuracy to training with real samples. In general, the DDPM-based approach achieved a 98.6% accuracy, outperforming other methods even when using fewer synthesized samples. On the other hand, the authors highlighted potential improvements, such as refining synthesis quality and the extended DDPM training time, suggesting more exploration.

Memon et al. [52] discussed the challenges of recognizing multiple cursive scripts due to limited labeled training data. The work proposed a content-controlled training approach for Urdu handwriting generation combined with a pre-trained recognizer loss. This model, trained on diverse ligatures images and further fine-tuned through transfer learning, is distinct from the predominant GAN-focused research. In this way, reached an FID score of 69.01 and an accuracy of 77% on CLE dataset and FID score of 23.24 and an accuracy of 69.7% on UCOM dataset.

Chang et al. [51] presented a method using GANs to generate handwritten content across different languages, with the goal of enhancing handwriting recognition in low-resource contexts. They reported FID scores for the VNonDB Vietnamese dataset as 27.46, 77.10, and 142.08 for printed, crosslingual, and semi-supervised GANs, respectively. For the MADCAT Arabic dataset, the scores were 23.28, 70.56, and 111.74. Moreover, they observed a notable improvement of up to 5 percentage points when doubling the data for augmentation. Nonetheless, the study might benefit from a more extensive evaluation across languages and a comparison with other existing approaches.

Nikolaidou et al. [44] introduced a method using a conditional Latent Diffusion Model to generate realistic word image samples across various writer styles. This approach leverages class index styles and text content prompts, eliminating the need for adversarial training, writer recognition, or handwriting recognition. On the other hand, the method requires a large amount of training data to learn the

[13] https://github.com/ganji15/HiGANplus.

[14] https://github.com/hachiro-2001/Bengali_Compound_Characters.

distribution of different writer styles and is computationally expensive. The evaluation on the IAM dataset reached an FID score of 22.74, a CER of 4.67%, and a WER of 13.28%. The study is publicly available in the repository.[15]

Zhu et al. [37] presented the Conditional Text Image Generation with Diffusion Models (CTIG-DM) for generating handwritten text images. The model effectively synthesizes diverse text images, adaptable to specifics like content, font, and background. Although promising for real-world scenarios, including scene text and diverse handwritten scripts, CTIG-DM demands substantial training data and computational power, and may face challenges with unseen text styles. In their evaluations, they reported an FID score of 25.52 on the IAM dataset. Furthermore, when trained on the IAM dataset, the recognition model achieved CER and WER scores of 10.89% and 26.24%, respectively, on the CVL-Database.

Pippi et al. [36] introduced the Visual Archetypes-based Transformer (VATr), a model designed for generating synthetic handwritten text, emphasizing on capturing writer-specific styles, especially when faced with unseen styles or rare characters. The unique approach of VATr uses standard GNU Unifont glyphs to represent textual content, making it efficient in handling characters seen less during training. Furthermore, using pre-training on a large synthetic dataset, the model becomes adept at focusing on writing styles without getting distracted by backgrounds or ink textures. Experimental results were promising, with an FID score of 17.79 on the IAM dataset. The study is publicly available in the repository.[16]

Finally, Zdenek and Nakayama [38] presented an extension of their previous work called JokerGAN++. In this study, the model uses a Vision Transformer (ViT)-based style encoder to generate handwritten text images, which can replicate specific handwriting styles from reference images and produce random styles as well. A unique feature is its ability to provide character-specific style encodings using the target character sequence. The authors registered FID scores of 2.13 on the IAM and 5.99 on the GNHK datasets. Furthermore, was recorded a WER of 25% on the IAM dataset with an additional 100,000 synthetic data.

Discussion

The papers identified in this systematic review satisfied our search criteria, showcasing a range of approaches, methods, and applications in the field of offline handwritten text recognition. Consequently, we were able to identify several research gaps, which were not adequately explored in the presented works. In the remainder of this section, we discuss specific and relevant topics, and provide answers to the research questions defined.

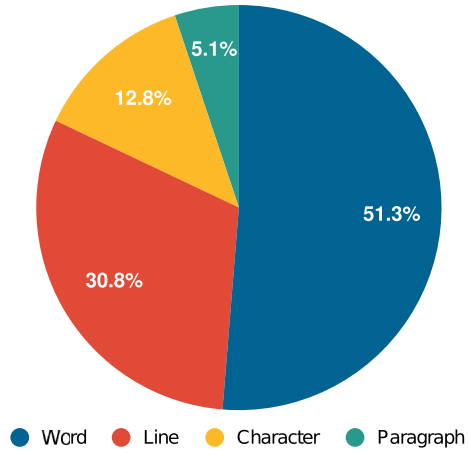

Fig. 6 Proportion of recognition levels used by studies. Each work may have more than one type of recognition associated

RQ1: What are the most commonly used recognition levels for data augmentation applied to offline handwritten text recognition?—Through our analysis of selected studies, we observed that approximately 51.3% employed word-level recognition for data augmentation in the research field of offline handwriting recognition [27, 29, 30, 32, 33, 35–38, 40, 42–44, 46, 49, 51, 52, 56–58]. This represents a broader application of a word-focused data augmentation approach.

Currently, word sequencing into lines is still considered a trend in generative models, representing an advancement in the field. However, the challenges of generating line structures are associated with the limited availability of data for training deep learning models and the computational costs involved. Although approximately 30.8% of the reviewed studies have focused on line-level applications [28, 30–32, 34, 38, 41, 45, 47–50], this presents significant opportunities for the development of handwriting recognition systems.

This leads us to reflect on character and paragraph scenarios, often unexplored, appearing in 12.8% and 5.1% of the reviewed studies, respectively. In paragraph scenarios, we encounter greater complexity than line structures, considering the sequencing of words and then stacking lines. This was a less explored approach due to its application and high cost. On the other hand, character scenarios mainly correspond to their application in glyph-based languages, such as Chinese and Japanese. Finally, Fig. 6 shows the proportion of recognition levels used by studies.

RQ2: What are the most commonly used datasets for data augmentation applied to offline handwritten text recognition?—In general, we noticed that the papers presented prioritized using the IAM, and RIMES datasets, with

[15] https://github.com/koninik/WordStylist.

[16] https://github.com/aimagelab/VATr.

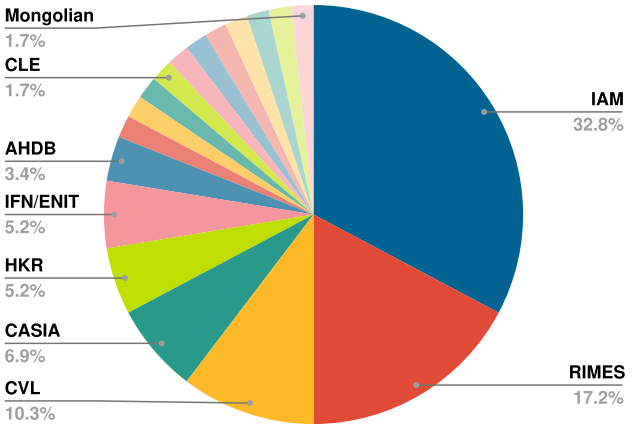

Fig. 7 Proportion of datasets used by studies. Each work may have more than one dataset associated

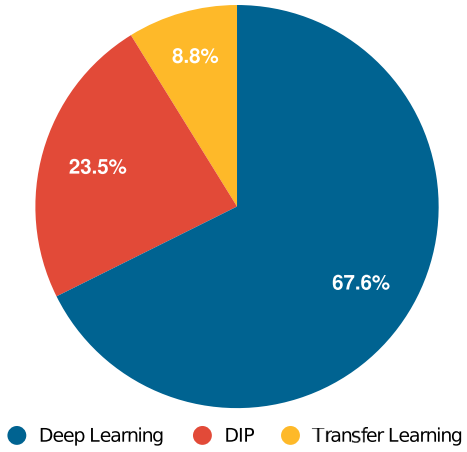

Fig. 8 Proportion of data augmentation approaches used by studies. Each work may have more than one approach associated

approximately 32.8%, and 17.2% of the papers [27–31, 34, 36–38, 40, 42, 44–46, 48, 49, 51, 58]. These datasets hold significant prominence in the field and serve as well-established benchmarks. Following closely, the CVL-Database is popular for its focus on historical handwritten documents, while CASIA is valuable for research involving glyph-based language, making them the second most frequently employed, appearing in about 10.3% and 6.9% of the papers [27, 32, 34, 35, 37, 40, 49, 50, 53]. The third group of frequently used datasets includes IFN/ENIT and HKR, with 5.2% and 5.2% of the papers [35, 41, 43, 56, 57]. Finally, the least used datasets reached less than 5% of the papers. This refers to specific studies on a particular dataset, many of which explore a particular language or even introduce a new dataset. Figure 7 shows the distribution of dataset usage among the reviewed studies.

RQ3: What is the Current State of Data Augmentation Research Field Applied to Offline Handwritten text recognition?—We found that approximately 23.5% utilized different techniques in Digital Image Processing (DIP) in addition to the optical model [30, 34, 35, 42, 50, 54, 55, 57]. This approach provides great flexibility for usage across various datasets while maintaining low computational costs. However, it is important to note that the presented methods have limitations due to the text structure they are applied to. That is, the higher the recognition level, the more difficult it is for transformation functions to generate new images without losing the text's content or structure.

Transfer Learning was another field of study, but less explored (8.8%) [40, 52, 55], since it is not the main objective to use it as data augmentation. In any case, the challenge with this approach is to leverage the previous knowledge of the optical model to retrain it on another dataset. Initially, the proposal proved effective in simple scenarios, or at least with the same degree of image text pattern similarity. However, this approach became challenging to implement under restricted dataset, requiring a larger volume of data to improve performance. On the other hand, a recent study showed an improvement in the results achieved when applied to datasets in the Urdu language [52]. It is worth noting that the authors highlighted the potential of transfer learning in handwriting tasks, especially with limited training data.

In the end, the most widely used approach among the reviewed studies was applying deep learning to synthesize handwritten text images, accounting for roughly 67.6%. Initially, the generative models had some limitations, including text length and cursive style, and required high computational costs. However, the models have undergone significant improvements as text image generators and can presently generate text images of arbitrary size, content, and cursive writing style [27–29, 31–33, 36–39, 41, 43–49, 51–53, 56, 58]. Figure 8 shows the proportion of data augmentation approaches used by studies.

In the Deep Learning domain, we observed three types of models applied to synthesize handwriting images. The Transformer model was one of the least employed approaches over time, with 13.0% of the reviewed studies [31, 36, 46]. In general, these models were used by offline handwritten text recognition works to boost data augmentation. In addition, we consider this as an initial approach within the field of handwritten text synthesis research. In contrast, Diffusion models, although only emerging in the year of 2023, already represent 13.0% of the reviewed studies [37, 44, 53]. This recent and significant growth indicates its potential for future applications. Finally, Generative Adversarial Networks (GANs) were extensively explored and developed over time [27–29, 32, 33, 38, 39, 41, 43, 45, 47–49, 51, 52, 56, 58], which represent 74.0% of the reviewed studies. Through the reviewed studies and the extensive adoption of GANs, we have observed a growing trend toward more realistic synthesis of handwriting images,

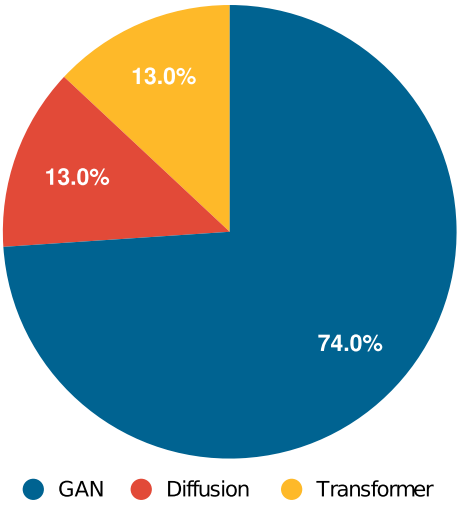

Fig. 9 Proportion of deep learning models applied to synthesize handwriting images in reviewed studies

with a simultaneous focus on reducing computational costs. This trend motivates for further research in the field and refinements in its application. Figure 9 shows the proportion of Deep Learning models applied in synthesizing handwriting images.

In general, DIP and Transfer Learning approaches are limited to the content of the dataset itself, either by applying transformations to an existing image, or using the knowledge learned from an optical model in another. On the other hand, works based on deep learning involve synthesizing images of handwritten text from scratch using the cursive style learned from the dataset. This versatility makes its application more comprehensive (Fig. 10).

RQ4: What are the Current Challenges in Data Augmentation Applied to Offline Handwritten Text Recognition?—Our analysis focused on the challenges faced in studies related to generative models, as DIP methods have already been extensively explored in the field of offline handwriting recognition research. Therefore, we

Fig. 10 Examples of synthetic handwritten text images created by the reviewed works. **a** Visual comparison of the results obtained from the studies by Alonso et al. [46], Fogel et al. [27], Liu et al. [29], and Luo et al. [49]. **b** Handwritten words generated by Gan et al. [28] using reference-guided synthesis. **d** Handwriting style interpolation in the work of Kang et al. [31]. **c** Images generated by Gui et al. [53] of Chinese character samples with different content and writer guidance scales

identified three main gaps in the current literature: (i) low computational cost; (ii) integration between the synthesizer model with the optical model; and (iii) application to restricted datasets.

In this regard, the computational cost has been a less explored feature. A few studies have examined the performance provided by the generator model, particularly when applied to an offline handwriting recognition system. This kind of analysis has only started recently in the latest studies, but in an isolated manner, that is, without considering the optical model.

The second gap identified is related to focus on the integration between the generator model and the optical model in an end-to-end system. A few explorations have been done on continuous and adaptable integration between the two models, often resulting in two independent workflows. In other words, the pipeline for learning and generating synthetic handwritten text images is executed first, and only then, the optical model makes use of the generated data.

Finally, the third gap identified was the applicability of proposed models to restricted datasets, which offer a limited volume of data as a challenge in handwriting recognition. Deep learning models face significant challenges in such scenarios due to the absence of large-scale data samples. This is the type of situation that tends to benefit the most from data augmentation, but has been under-addressed.

Conclusion

Data augmentation is a topic that currently presents various nuances, varying according to the application domain. Furthermore, data augmentation techniques have the potential to be applied in various related fields, such as handwriting recognition, writer identification, keyword spotting, and more. Each of these areas, although sharing some similarities, has its own peculiarities and specific requirements. Thus, we presented a systematic literature review on data augmentation applied to offline handwritten text recognition. We consider the following main contributions:

- Scope definition of a systematic literature review on data augmentation applied to offline handwritten text recognition;
- Exploration of the used datasets and recognition levels to synthesize handwriting images;
- Analysis of data augmentation approaches and the synthesis of handwritten text images over the past decade in the offline handwritten text recognition research field;
- Identification of current gaps and challenges in the literature, which led us to suggest future research directions to address them.

Initially, 976 papers were collected from five academic databases using relevant keywords for the research field. After a four-step exclusion process, 32 papers were selected and reviewed. Additionally, the quality evaluation scored the papers between 0 and 10 points, in which the highest score obtained was 7.75 [27].

Through the selected works, we explored and described relevant aspects of each study. We mapped the datasets and levels of handwriting recognition most commonly used, and consequently, the most used languages as well. This allowed us to relate and analyze each proposed method within its specific application context.

Based on the study conducted, it can be concluded that Digital Image Processing methods are practical and improve optical models in the training process. However, the data augmentation approach through Generative Adversarial Networks is the new trend in the synthesis of handwritten text images realistically. This approach has the potential to open new research, and its use with optical models is highly promising.

It should be emphasized that the field of offline handwritten text recognition with a central focus on data augmentation is still relatively new. Nevertheless, we have observed a trend in this research area in recent years, accompanied by significant progress in the application of Generative Adversarial Networks as generators of synthetic images of handwritten text. This trend indicates an increasing interest of the academic community in the benefits of combining these research lines.

In conclusion, a future work perspective is related to using low-volume datasets, where generating synthetic images of handwritten text can benefit optical model training. Another relevant aspect is associated with the development of generative models integrated with optical models, following a self-supervised learning approach.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This study was financed in part by the founding public agencies: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)—Finance Code 001; Fundação de Amparo a Ciência e Tecnologia de PE (FACEPE) (APQ-1216-1.03/22); and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (315251/2018-2, 141721/2023-5).

Data Availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of Interest The authors declare that there is no conflict of interest. They have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Bezerra B, Zanchettin C, Toselli A, Pirlo G. Handwriting: recognition. Development and analysis-computer science: technology and applications. New York: Nova Science Pub Inc; 2017.
2. Palehai D, Fanany MI. Handwriting recognition on form document using convolutional neural network and support vector machines (CNN-SVM). In: 5th International conference on information and communication technology (ICoIC7) (2017). https://doi.org/10.1109/ICoICT.2017.8074699.
3. Dhar D, Garain A, Singh P, Sarkar R. Hp_docpres: a method for classifying printed and handwritten texts in doctor's prescription. Multimed Tools Appl. 2021;80:1–34. https://doi.org/10.1007/s11042-020-10151-w.
4. Neto AFS, Bezerra BLD, Lima EB, Toselli AH. HDSR-Flor: a robust end-to-end system to solve the handwritten digit string recognition problem in real complex scenarios. IEEE Access. 2020;8:208543–53. https://doi.org/10.1109/ACCESS.2020.3039003.
5. Muehlberger G, et al. Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. J Doc. 2019. https://doi.org/10.1108/JD-07-2018-0114.
6. Bunke H, Roth M, Schukat-Talamazzini EG. Off-line cursive handwriting recognition using hidden Markov models. Pattern Recognit. 1995;28:1399–413. https://doi.org/10.1016/0031-3203(95)00013-P.
7. Doetsch P, Kozielski M, Ney H. Fast and robust training of recurrent neural networks for offline handwriting recognition. In: Proceedings of international conference on frontiers in handwriting recognition, ICFHR, pp. 279–284 (2014). https://doi.org/10.1109/ICFHR.2014.54.
8. Toselli AH, Vidal E. Handwritten text recognition results on the Bentham collection with improved classical N-Gram-HMM methods. In: *Proceedings of the 3rd international workshop on historical document imaging and processing*, pp. 15–22 (2015). https://doi.org/10.1145/2809544.2809551.
9. Graves A, Fernández S, Schmidhuber J. Multi-dimensional recurrent neural networks. In: International conference on artificial neural networks, pp 549–558 (2007). https://doi.org/10.1007/978-3-540-74690-4_56.
10. Voigtlaender P, Doetsch P, Ney H. Handwriting recognition with large multidimensional long short-term memory recurrent neural networks. In: 15th International conference on frontiers in handwriting recognition (ICFHR), pp. 228–233 (2016). https://doi.org/10.1109/ICFHR.2016.0052.
11. Graves A, et al. A novel connectionist system for unconstrained handwriting recognition. IEEE Trans Pattern Anal Mach Intell. 2009;31:855–68. https://doi.org/10.1109/TPAMI.2008.137.
12. Bluche T, Messina R. Gated convolutional recurrent neural networks for multilingual handwriting recognition. In: 14th IAPR international conference on document analysis and recognition (ICDAR), pp. 646–651 (2017). https://doi.org/10.1109/ICDAR.2017.111.
13. Puigcerver J. Are multidimensional recurrent layers really necessary for handwritten text recognition? In: *14th IAPR international conference on document analysis and recognition (ICDAR)*, pp. 67–72 (2017). https://doi.org/10.1109/ICDAR.2017.20.
14. Neto AFS, Bezerra BLD, Toselli AH, Lima EB. A robust handwritten recognition system for learning on different data restriction scenarios. Pattern Recognit Lett. 2022;1:1–7. https://doi.org/10.1016/j.patrec.2022.04.009.
15. Ingle RR, Fujii Y, Deselaers T, Baccash J, Popat AC. A scalable handwritten text recognition system. In: 2019 International conference on document analysis and recognition (ICDAR), pp. 17–24 (2019). https://doi.org/10.1109/ICDAR.2019.00013.
16. Kass D, Vats E. Attentionhtr: handwritten text recognition based on attention encoder-decoder networks. In: Document analysis systems, pp. 507–522 (2022). https://doi.org/10.1007/978-3-031-06555-2_34.
17. Kang L, Riba P, Rusiñol M, Fornés A, Villegas M. Pay attention to what you read: non-recurrent handwritten text-line recognition. Pattern Recognit. 2022;129: 108766. https://doi.org/10.1016/j.patcog.2022.108766.
18. Scheidl H, Fiel S, Sablatnig R. Word beam search: a connectionist temporal classification decoding algorithm. In: 2018 16th International conference on frontiers in handwriting recognition (ICFHR), pp. 253–258 (2018). https://doi.org/10.1109/ICFHR-2018.2018.00052.
19. Neto AFS, Bezerra BLD, Toselli AH. Towards the natural language processing as spelling correction for offline handwritten text recognition systems. Appl Sci. 2020;10(21):1–29. https://doi.org/10.3390/app10217711.
20. Jayasundara V, et al. Textcaps: handwritten character recognition with very small datasets. In: 2019 IEEE winter conference on applications of computer vision (WACV), pp 254–262 (2019). https://doi.org/10.1109/WACV.2019.00033.
21. Bhunia AK, Das A, Bhunia AK, Kishore PSR, Roy PP. Handwriting recognition in low-resource scripts using adversarial learning. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 4762–4771 (2019). https://doi.org/10.1109/CVPR.2019.00490.
22. Pham H, et al. Robust handwriting recognition with limited and noisy data. In: 2020 17th International conference on frontiers in handwriting recognition (ICFHR), pp. 301–306 (2020). https://doi.org/10.1109/ICFHR2020.2020.00062.
23. Souibgui MA, Fornés A, Kessentini Y, Megyesi B. Few shots are all you need: a progressive learning approach for low resource handwritten text recognition. Pattern Recognit Lett. 2022;160:43–9. https://doi.org/10.1016/j.patrec.2022.06.003.
24. Kitchenham B, Charters S. Guidelines for performing systematic literature reviews in software engineering—technical report EBSE-2007-01—School of Computer Science and Mathematics (2007). https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf.
25. Kitchenham B, et al. Systematic literature reviews in software engineering—a systematic literature review. Inf Softw Technol. 2009;51(1):7–15. https://doi.org/10.1016/j.infsof.2008.09.009.
26. Kitchenham B, et al. Systematic literature reviews in software engineering—a tertiary study. Inf Softw Technol. 2010;52(8):792–805. https://doi.org/10.1016/j.infsof.2010.03.006.
27. Fogel S, Averbuch-Elor H, Cohen S, Mazor S, Litman R. Scrabblegan: semi-supervised varying length handwritten text generation. In: 2020 IEEE/CVF conference on computer vision and

pattern recognition (CVPR), pp. 4323–4332 (2020). https://doi.org/10.1109/CVPR42600.2020.00438.
28. Gan J, Wang W, Leng J, Gao X. Higan+: handwriting imitation gan with disentangled representations. ACM Trans Graph. 2022. https://doi.org/10.1145/3550070.
29. Liu X, Meng G, Xiang S, Pan C. Handwritten text generation via disentangled representations. IEEE Signal Process Lett. 2021;28:1838–42. https://doi.org/10.1109/LSP.2021.3109541.
30. Wigington C, et al. Data augmentation for recognition of handwritten words and lines using a cnn-lstm network. In: 2017 14th IAPR International conference on document analysis and recognition (ICDAR), pp. 639–645 (2017). https://doi.org/10.1109/ICDAR.2017.110.
31. Kang L, Riba P, Rusiñol M, Fornés A, Villegas M. Content and style aware generation of text-line images for handwriting recognition. IEEE Trans Pattern Anal Mach Intell. 2022;44(12):8846–60. https://doi.org/10.1109/TPAMI.2021.3122572.
32. Zdenek J, Nakayama H. Jokergan: memory-efficient model for handwritten text generation with text line awareness. In: Proceedings of the 29th ACM international conference on multimedia, pp. 5655–5663 (2021). https://doi.org/10.1145/3474085.3475713.
33. Kang L, et al. Ganwriting: content-conditioned generation of styled handwritten word images. In: Computer vision—ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII, pp. 273–289 (2020). https://doi.org/10.1007/978-3-030-58592-1_17.
34. Moysset B, Messina R. Manifold mixup improves text recognition with CTC loss. In: 2019 International conference on document analysis and recognition (ICDAR), pp. 799–804 (2019). https://doi.org/10.1109/ICDAR.2019.00133 .
35. Chen W, Su X, Zhang H. Script-level word sample augmentation for few-shot handwritten text recognition. In: 18th International conference on frontiers in handwriting recognition (ICFHR), pp. 316–330 (2022). https://doi.org/10.1007/978-3-031-21648-0_22.
36. Pippi V, Cascianelli S, Cucchiara R. Handwritten text generation from visual archetypes. In: 2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 22458–22467 (2023). https://doi.org/10.1109/CVPR52729.2023.02151.
37. Zhu Y, Li Z, Wang T, He M, Yao C. Conditional text image generation with diffusion models. In: 2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 14235–14244 (2023). https://doi.org/10.1109/CVPR52729.2023.01368.
38. Zdenek J, Nakayama H. Handwritten text generation with character-specific encoding for style imitation. Doc Anal Recognit ICDAR. 2023;2023:313–29. https://doi.org/10.1007/978-3-031-41679-8_18.
39. Das A, Choudhuri A, Basu A, Sarkar R. Generation of a synthetic handwritten bangla compound character dataset using a modified conditional gan architecture. Multimed Tools Appl. 2023;82(10):14775–97. https://doi.org/10.1007/s11042-022-13891-z.
40. Burdett E, et al. Active transfer learning for handwriting recognition. In: Frontiers in handwriting recognition: 18th international conference, ICFHR 2022, Hyderabad, India, December 4–7, 2022, Proceedings, pp. 245–258 (2022). https://doi.org/10.1007/978-3-031-21648-0_17.
41. Kudaibergen T, Hamada MA. Application of deep convolutional generative adversarial network for Russian handwritten text recognition. In: Proceedings of the 7th international conference on digital technologies in education, science and industry (DTESI), vol. 3382, pp. 1–11 (2022).
42. Luo C, Zhu Y, Jin L, Wang, Y. Learn to augment: joint data augmentation and network optimization for text recognition. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 13743–13752 (2020). https://doi.org/10.1109/CVPR42600.2020.01376.
43. Yeleussinov A, Amirgaliyev Y, Cherikbayeva L. Improving OCR accuracy for Kazakh handwriting recognition using gan models. Appl Sci. 2023. https://doi.org/10.3390/app13095677.
44. Nikolaidou K, et al. Wordstylist: styled verbatim handwritten text generation with latent diffusion models. Doc Anal Recognit ICDAR. 2023;2023:384–401. https://doi.org/10.1007/978-3-031-41679-8_22.
45. Bhunia AK, et al. Handwriting transformers. In: 2021 IEEE/CVF international conference on computer vision (ICCV), pp. 1066–1074 (2021). https://doi.org/10.1109/ICCV48922.2021.00112.
46. Alonso E, Moysset B, Messina R. Adversarial generation of handwritten text images conditioned on sequences. In: *2019 International conference on document analysis and recognition (ICDAR)*, pp. 481–486 (2019). https://doi.org/10.1109/ICDAR.2019.00083.
47. Spoto M, Wolf B, Fischer A, Scius-Bertrand A. Improving handwriting recognition for historical documents using synthetic text lines. In: Intertwining graphonomics with human movements, pp. 61–75 (2022). https://doi.org/10.1007/978-3-031-19745-1_5.
48. Huu M-K N, Ho S-T, Nguyen V-T, Ng, TD. Multilingual-gan: a multilingual gan-based approach for handwritten generation. In: 2021 International conference on multimedia analysis and pattern recognition (MAPR), pp. 1–6 (2021). https://doi.org/10.1109/MAPR53640.2021.9585285.
49. Luo C, Zhu Y, Jin L, Li Z, Peng D. Slogan: Handwriting style synthesis for arbitrary-length and out-of-vocabulary text. In: IEEE transactions on neural networks and learning systems, pp. 1–13 (2022). https://doi.org/10.1109/TNNLS.2022.3151477.
50. Shen X, Messina R. A method of synthesizing handwritten Chinese images for data augmentation. In: 2016 15th International conference on frontiers in handwriting recognition (ICFHR), pp. 114–119 (2016). https://doi.org/10.1109/ICFHR.2016.0033.
51. Chang CC, Perera LPG, Khudanpur S. Crosslingual handwritten text generation using gans. In: Document analysis and recognition—ICDAR 2023 workshops, pp. 285–301 (2023). https://doi.org/10.1007/978-3-031-41501-2_20.
52. Memon Z, Ul-Hasan A, Shafait F. Content-aware Urdu handwriting generation. Doc Anal Recognit ICDAR. 2023;2023:428–44. https://doi.org/10.1007/978-3-031-41685-9_27.
53. Gui D, Chen K, Ding H, Huo Q. Zero-shot generation of training data with denoising diffusion probabilistic model for handwritten Chinese character recognition. Doc Anal Recognit ICDAR. 2023;2023:348–65. https://doi.org/10.1007/978-3-031-41679-8_20.
54. Hidayat AA, Purwandari K, Cenggoro TW, Pardamean B. A convolutional neural network-based ancient Sundanese character classifier with data augmentation. In: 5th International conference on computer science and computational intelligence 2020, vol. 179, pp. 195–201 (2021). https://doi.org/10.1016/j.procs.2020.12.025.
55. Hayashi T, Gyohten K, Ohki H, Takami T. A study of data augmentation for handwritten character recognition using deep learning. In: 2018 16th International conference on frontiers in handwriting recognition (ICFHR), pp. 552–557 (2018). https://doi.org/10.1109/ICFHR-2018.2018.00102.
56. Eltay M, Zidouri A, Ahmad I, Elarian Y. Generative adversarial network based adaptive data augmentation for handwritten Arabic text recognition. PeerJ Comput Sci. 2022. https://doi.org/10.7717/peerj-cs.861.
57. Eltay M, Zidouri A, Ahmad I, Elarian Y. Improving handwritten Arabic text recognition using an adaptive data-augmentation algorithm. In: Document analysis and recognition—ICDAR 2021 workshops, pp. 322–335 (2021). https://doi.org/10.1007/978-3-030-86198-8_23.
58. Wang H, Wang Y, Wei H. Affganwriting: a handwriting image generation method based on multi-feature fusion. Doc Anal Recognit ICDAR. 2023;2023:302–12. https://doi.org/10.1007/978-3-031-41685-9_19.

59. Marti U-V, Bunke H. The IAM-database: an English sentence database for offline handwriting recognition. In: International journal on document analysis and recognition, vol. 5 (2002). https://doi.org/10.1007/s100320200071.
60. Kleber F, Fiel S, Diem M, Sablatnig R. Cvl-database: an off-line database for writer retrieval, writer identification and word spotting. In: *2013 12th International conference on document analysis and recognition*, pp. 560–564 (2013). https://doi.org/10.1109/ICDAR.2013.117.
61. Gatos B, et al. Ground-truth production in the transcriptorium project. In: 2014 11th IAPR international workshop on document analysis systems, pp. 237–241 (2014). https://doi.org/10.1109/DAS.2014.23.
62. Brunessaux S, et al. The Maurdor project: improving automatic processing of digital documents. In: 2014 11th IAPR international workshop on document analysis systems, pp. 349–354 (2014). https://doi.org/10.1109/DAS.2014.58.
63. Lee AWC, Chung J, Lee M. Gnhk: a dataset for English handwriting in the wild. In: Document analysis and recognition—ICDAR 2021: 16th international conference, Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part IV, pp. 399–412 (2021). https://doi.org/10.1007/978-3-030-86337-1_27.
64. Grosicki E, Carre M, Brodin J-M, Geoffrois E. Rimes evaluation campaign for handwritten mail processing. In: ICFHR 2008: 11th international conference on frontiers in handwriting recognition, pp. 1–6 (2008). https://doi.org/10.1109/ICDAR.2009.224.
65. Sánchez JA, Romero V, Toselli AH, Vidal E. ICFHR2016 competition on handwritten text recognition on the read dataset. In: 2016 15th International conference on frontiers in handwriting recognition (ICFHR), pp. 630–635 (2016). https://doi.org/10.1109/ICFHR.2016.0120.
66. National Institute of Standards and Technology (NIST). Open handwriting recognition and translation evaluation (OpenHaRT) (2010). https://www.nist.gov/system/files/documents/itl/iad/mig/OpenHaRT2010_EvalPlan_v2-8.pdf.
67. Pechwitz M, Margner V. Baseline estimation for Arabic handwritten words. In: Proceedings eighth international workshop on frontiers in handwriting recognition, pp. 479–484 (2002). https://doi.org/10.1109/IWFHR.2002.1030956.
68. Al-Ma'adeed S, Elliman D, Higgins C. A data base for Arabic handwritten text recognition research. In: Proceedings eighth international workshop on frontiers in handwriting recognition, pp. 485–489 (2002). https://doi.org/10.1109/IWFHR.2002.1030957.
69. Lee D, et al. MADCAT phase 1 training set. In: Linguistic Data Consortium (LDC) (2012). https://doi.org/10.35111/9bm5-nz55.
70. Lee D, et al. MADCAT phase 2 training set. In: Linguistic Data Consortium (LDC) (2013). https://doi.org/10.35111/044b-ah68.
71. Lee D, et al. MADCAT phase 3 training set. In: Linguistic Data Consortium (LDC) (2013). https://doi.org/10.35111/w1px-d922.
72. Nurseitov D, et al. Handwritten Kazakh and Russian (HKR) database for text recognition. Multimed Tools Appl. 2021. https://doi.org/10.1007/s11042-021-11399-6.
73. 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 (2011). https://doi.org/10.1109/ICDAR.2011.17.
74. Toselli AH, et al. Integrated handwriting recognition and interpretation using finite-state models. Int J Pattern Recognit Artif Intell (IJPRAI). 2004;18:519–39. https://doi.org/10.1142/S0218001404003344.
75. Cinnamon AI Labs. Cinnamon Handwritten OCR for Vietnamese Address Challenge Dataset – Cinnamon AI Marathon (2018). https://it.tdtu.edu.vn/thong-tin-cuoc-thi-cinnamon-ai-marathon.
76. Nguyen HT, Nguyen CT, Nakagawa M. ICFHR 2018—competition on Vietnamese online handwritten text recognition using hands-VNOnDB (VOHTR2018). In: 2018 16th International conference on frontiers in handwriting recognition (ICFHR), pp. 494–499 (2018). https://doi.org/10.1109/ICFHR-2018.2018.00092.
77. Das N, et al. A genetic algorithm based region sampling for selection of local features in handwritten digit recognition application. Appl Soft Comput. 2012;12(5):1592–606. https://doi.org/10.1016/j.asoc.2011.11.030.
78. Suryani M, Paulus E, Hadi S, Darsa UA, Burie J-C. The handwritten Sundanese palm leaf manuscript dataset from 15th century. In: 2017 14th IAPR international conference on document analysis and recognition (ICDAR), vol. 01, pp. 796–800 (2017). https://doi.org/10.1109/ICDAR.2017.135.
79. Khattak IU, Siddiqi I, Khalid S, Djeddi C. Recognition of Urdu ligatures—a holistic approach. In: 2015 13th International conference on document analysis and recognition (ICDAR), pp. 71–75 (2015). https://doi.org/10.1109/ICDAR.2015.7333728.
80. Ahmed S, et al. Ucom offline dataset—an Urdu handwritten dataset generation. Int Arab J Inf Technol. 2017;14:239–45. https://api.semanticscholar.org/CorpusID:1019515.
81. Japan Electronics and Information Technology Industries Association. ETL Character Database—National Institute of Advanced Industrial Science and Technology (AIST) (2011). http://etlcdb.db.aist.go.jp/.
82. Cho K, van Merriënboer B, Bahdanau D, Bougares H, Fethi Schwenk, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: 2014 Conference on empirical methods in natural language processing (EMNLP), pp. 1724–1734 (2014). https://doi.org/10.3115/v1/D14-1179.
83. Vaswani A, et al. Attention is all you need. In: Proceedings of the 31st international conference on neural information processing systems, pp. 6000–6010 (2017). https://doi.org/10.5555/3295222.3295349.