



HAL
open science

Synthetic lines from historical manuscripts: an experiment using GAN and style transfer

Chahan Vidal-Gorène, Jean-Baptiste Camps, Thibault Clérice

► To cite this version:

Chahan Vidal-Gorène, Jean-Baptiste Camps, Thibault Clérice. Synthetic lines from historical manuscripts: an experiment using GAN and style transfer. Visual Processing of Digital Manuscripts: Workflows, Pipelines, Best Practices. ICIAP 2023 Workshops. ICIAP 2023., Sep 2023, Udine, Italy. hal-04178550v2

HAL Id: hal-04178550

<https://inria.hal.science/hal-04178550v2>

Submitted on 23 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Synthetic lines from historical manuscripts: an experiment using GAN and style transfer

Chahan Vidal-Gorène^{1,2}[0000-0003-1567-6508], Jean-Baptiste Camps¹[0000-0003-0385-7037], and Thibault Clérico³[0000-0003-1852-9204]

¹ Centre Jean Mabillon, PSL – École nationale des Chartes

² Califa

³ ALMAnaCH, Inria Paris

Abstract. Given enough data of sufficient quality, HTR systems can achieve high accuracy, regardless of language, script or medium. Despite growing pooling of datasets, the question of the required quantity of training material still remains crucial for the transfer of models to out-of-domain documents, or the recognition of new scripts and under-resourced character classes. We propose a new data augmentation strategy, using generative adversarial networks (GAN). Inspired by synthetic lines generation for printed documents, our objective is to generate handwritten lines in order to massively produce data for a given style or under-resourced character class. Our approach, based on a variant of ScrabbleGAN, demonstrates the feasibility for various scripts, either in the presence of a high number and variety of abbreviations (Latin) and spellings or letter forms (Medieval French), in a situation of data scarcity (Armenian), or in the instance of a very cursive script (Arabic Maghribi). We then study the impact of synthetic line generation on HTR, by evaluating the gain for out-of-domain documents and under-resourced classes.

Keywords: gan, handwritten text recognition, data augmentation

1 Introduction

Generative Adversarial Networks (GANs) have shown promising results in generating realistic images, audio, and video. In this paper, we explore the use of GANs to generate synthetic text lines for historical documents, particularly ancient and medieval manuscript in Latin scripts (French and Latin) and non-Latin scripts (Arabic and Armenian). Given a text input, our goal is to obtain the generation of one or several realistic lines. The ability to generate synthetic text lines can have multiple applications, such as data augmentation for handwritten text recognition systems, or generating fake lines under specific constraints such as style, script, or date. This can help in filling the gaps in ancient manuscripts, which are often incomplete or have missing sections, and can also serve as a tool for researchers to study and analyse different writing styles and scripts.

Here, we focus on the goal of data augmentation, to try to measure the improvements that can be obtained on some frequent issues in handwritten text

recognition. Data augmentation could potentially help alleviating the central problem in HTR for ancient documents that is the availability of training data and the cost (in terms of human time and expertise) needed to create them, which is often a bottleneck in workflows. In particular, we wish to investigate the success of data augmentation regarding:

1. the prediction of rare character classes, such as abbreviation signs, diacritics or variant letter forms;
2. the situation of extreme data scarcity due to the rarity of transcribed material;
3. very complex scripts, due to a high level of cursiveness, ligatures or variation.

More generally, we also wish to address the problem of generalisation and alleviate the issue of overfitting, since the generated lines can also be used to augment the training data with variations in writing style, script, and other factors that can help to improve the robustness of the recognition system.

In this work, we propose an approach based on a variant of ScrabbleGAN [10], trained on datasets containing Medieval manuscripts and incunabula in Medieval French and Latin [7, 15, 17, 21, 23], as well as Arabic manuscripts [20] and finally Armenian manuscripts [19], which we compare with a style transfer approach based on cycleGAN[24]. We use very different datasets to check the feasibility and validity of the method on a wide variety of cases: Latin and non-Latin spellings, right-left (e.g. Arabic) and left-right (e.g. Latin, Medieval French and Armenian) reading directions, complex abbreviation system (e.g. Latin and Medieval French) or minimal abbreviation system (e.g. Armenian and Arabic), large variety of hands (e.g. Latin, Medieval French and Armenian) or more restricted dataset (e.g. Arabic).

2 Background and Related Work

2.1 Background

Compared to printed documents, historical manuscripts present several challenges for training recognition models. The variation in individual letter shapes, even within the same hand, coupled with larger variations between scribes, script types, regions and periods, complicate the task of training more general models, as well as creating synthetic data. Moreover, historical manuscripts often present a large spectrum of non standard signs, that include abbreviative markers, diacritics, correction marks, allographs, or ligatures and letter fusions [2, 4, 5]. In addition, the creation of training data is made more complicated by the required level of expertise for transcribing such documents. Even when transcription exists, the pooling of datasets is made very difficult because there often does not exist a single unambiguous way to encode non standard characters and abbreviations, necessitating many normalisation and mapping efforts between datasets before being able to handle the diversity of the data [3]. Finally, physical damage to documents as well as different digitisation methods add a supplementary layer of variation and complexity for recognition tasks (Figure 1).



Fig. 1: Example of lines in Medieval French manuscripts.

To experiment with challenges of varying natures, we choose to work on three different datasets. The Latin and Medieval French dataset offers a large chronological (8th to 15th century) and linguistic span, contains different scripts (Caroline, Gothic scripts) and even early types (incunabula in Gothic types). Moreover, it contains important allographic and graphematic variation: Latin manuscripts can be heavily abbreviated with a diversity of abbreviations (contractions, suspensions, tachograph signs, etc.), while Medieval French presents important linguistic variation in space and time. Both contain allographs for many letters. The Armenian case is characterised by a strong scarcity of available data, while Maghrebi Arabic dataset presents very cursive scripts, with many ligatures.

2.2 Related Work

Creating synthetic data to improve training of recognition systems has already been done with success for printed materials [9]. GANs appear as a relevant answer to this challenge, as they reach very convincing results in different scenarios. State-of-the-art results rely on a style-transfer approach [14], in which a generator tries to map on a targeted dataset unsupervised learned features from another. Such an approach has already been successfully applied to historical manuscripts to create realistic Latin manuscripts [22] with very constrained styles in training (page-level), or for contemporary cursive hand-writings [11] (line-level with a semi-supervised approach). For contemporary scripts, the IAM dataset is mainly used to evaluate and to compare line or word generation, using ScrabbleGAN at character level [10], or similar approaches at the word level [13], even with transformers [1]. This dataset nevertheless suffers from a low variation in the handwritten styles, and from the low level of noise in the images, in comparison with historical documents. Scrabble’s latest enhancements also allow control over the style produced by the generator, which increases the benefits of generating artificial handwritten lines for HTR in order to target a specific hand [12]. GANs have also been successfully applied to other tasks on historical documents and manuscripts, such as artificial colourisation [6].

3 Proposed Method

We evaluate two approaches for line synthesis, ScrabbleGAN and CycleGAN.

3.1 ScrabbleGAN

The first one is based on ScrabbleGAN, whose relevance has already been demonstrated for non-cursive modern scripts [10] and for contemporary Arabic lines [8],

on the condition that the script doesn't contain too much ligatures. ScrabbleGAN is based on the simultaneous training of a generator/discriminator, and a CNN used as a recogniser. Basically, this GAN is trained to generate each character independently, and the line is then recomposed by putting each character image end to end. This splitting, which forces the GAN to use the different possible input and output contexts of a character (upper and lower), certainly only makes sense in the context of a very unequal distribution of character classes. The GAN therefore has two contextual informations: the character and its neighborhood, and within the character the possible inputs and outputs. In addition to the traditional discrimination of a GAN between a real and false image, the network must also confirm that the generated image is well readable, beyond its plausible aspect. Our dataset (see *below*) includes a large number of cursive cases and superscript/subscript signs (particularly Latin abbreviations), leading us to introduce a small change in the behaviour of Scrabble: after a first classic training without modification allowing the network to identify the space occupied by a character in a line of text, each character is split into four sub-images (see Figure 2) in order to force the network to focus on ‘paleographic’ features shared by letters or specific to some ligatures / abbreviations. In addition, in order to prevent overfitting of the recogniser on the dataset (leading to its ability to recognise a character despite a bad generation), we reset the weights of the recogniser every 100 epochs. Main ScrabbleGAN hyperparameters are kept. We favour the decomposition into strokes because of the large number of ambiguities and ligatures that our datasets present, in particular in Arabic and contemporary Armenian scripts (see below).

One of the foreseeable limits, that we wish to investigate, is that this strategy will perhaps find itself also limited regarding the production of under-resourced classes: it is indeed possible that the GAN will perform badly on those classes, similarly to HTR systems.

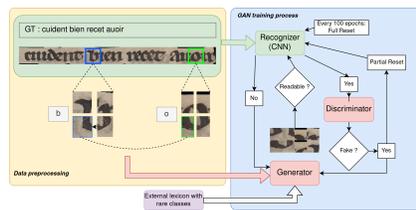


Fig. 2: Training strategy of ScrabbleGAN using a recogniser and a constrained discriminator

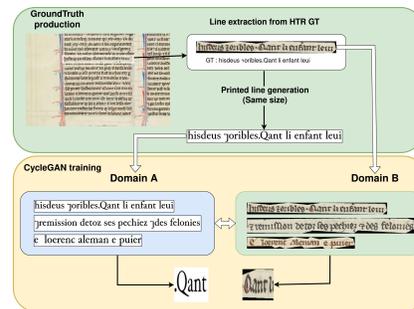


Fig. 3: Training strategy of CycleGAN using non-aligned pair images

3.2 CycleGAN

To try to overcome this problem of under-resourced classes, our second approach is based on CycleGAN [24], which adopts a “style transfer” strategy, in order

to apply to target images (domain B) the characteristics of a training dataset (domain A). It is an image-to-image translation strategy, usually applied to unpaired data on generation tasks of a particular style, for example artistic. The use of CycleGAN has already been proven on historical documents, at the page level to generate a look-like Latin manuscript page, with real text [22]. In particular, the authors trained CycleGAN by providing images of manuscripts (domain B) and natively printed PDF texts (domain A), to train GAN to convert from one to the other by giving the PDF a handwritten look. The results produced are quite convincing. The authors use in particular a recogniser in addition to ensure, as for ScrabbleGAN, that the result produced has real relevance and is not simply a decorum added to the PDF. In our case, we limit ourselves to the level of the line of text, by creating a dataset of paired printed and handwritten lines (without correspondence at the pixel level). Our assertion is that it may not be necessary to use a recogniser as backup, and that the task of generating lines could be a task of transferring a handwritten style to a line of printed text. We want to check at this stage if this strategy would make it possible to easily generate lines with under-resourced classes in order to compensate for the lack of data and improve their recognition.

4 Experimental Setup

Training and evaluation of models are performed with Kraken (4.1.2). Training used architecture with CTC Decoding (Figure 4) with a learning rate of 0.001 halved every 5 epochs of plateau.

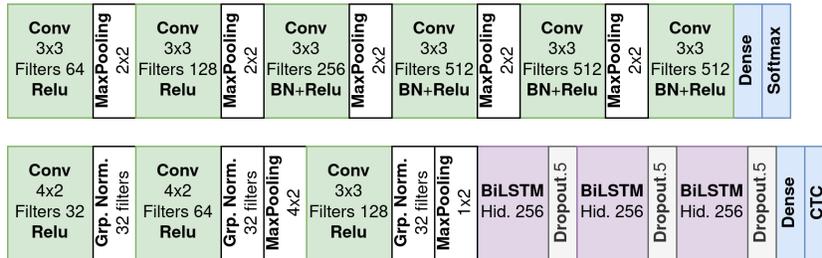


Fig. 4: Differences of architectures used for HTR. Top: Architecture used during GAN training to help discriminator. Bottom: Architecture used for HTR experiments

Latin and Medieval French Experiment We consider two training subsets for Latin and Medieval French, based on type (print vs. manuscript) and periods (Table 1). Both subsets follow the same “graphetic” transcription guidelines [16]. A graphetic transcription implies the non resolution of abbreviation

Table 1: Composition of Medieval Latin and French (left) and Classical Armenian (right) datasets

Subset	Ref.	Chars*	Lines	Period	Lang	Manuscript	Period Script	Lines
A	[15]	593k	22269	12-15th	fro	M1	1633 <i>bolorgir</i>	512
A	[7]	236k	6040	12-15th	lat	M6	16th <i>bolorgir</i>	819
A	[17]	169k	5937	15th	fro	M982	1460 <i>bolorgir</i>	3035
B	[23]	17k	439	9-12th	lat	MAF52	15-16th <i>bolorgir</i>	954
B	[21]	87k	2826	9th	lat	MAF54	16th <i>bolorgir</i>	1253
B	[18]	245k	7607	15th	fro	MAF62	18th <i>bolorgir</i>	522
Total		1149K	45118			TBI122	17th <i>bolorgir</i>	1209
						V. Melkonian Arch.	19th <i>Shghagir</i>	500
						V. Trianz Arch.	19th <i>Shghagir</i>	500
Total						Total		9304

* Space characters excluded.

and the use of common characters for different shape of signs, *e.g.* long s (f) are transcribed as **s**, or more commonly for abbreviated manuscripts, combining horizontal abbreviating marks are transcribed using a tilde (\sim).

Subset A combines three different dataset which have a large amount of data around the end of the Middle Ages in both Latin and French (1100–1500). Latin represents around 23.4% of the subset’s characters count but has a far higher abbreviation rate than Medieval French, which implies more rare classes with higher frequency. Subset A is used for generating synthetic data, as it is far more cohesive in terms of written shapes than subset B and more diverse in terms of characters and manuscripts.

Subset B serves as a secondary dataset to augment Subset A and analyze if data augmentation of out-of-domain data improves results. Subset B consists mostly of early medieval Latin data written in Caroline Minuscule as well as one dataset of early print (*incunabula*) written in Gothic types close to the manuscripts of Subset A but far more uniform than its manuscript counterparts.

Four manuscripts (not included in Table 1) are used for evaluation: they present similar writing style as seen in Subset A, three of them are in Latin, one in Medieval French⁴. They are used to test the models capacities (see Figure 5). The amount of data in Medieval French and Latin allows us to test more conditions than in Arabic and Armenian.

The sigma dataset was generated using out-of-domain transcription of Latin and Medieval French manuscripts, randomly abbreviating 30% of the words using ad-hoc rules. We generated 8,000 lines with character length of 5 to 40 characters.

Arabic For Arabic, we used the RASAM dataset [20], which contains 7,540 lines and 483,725 characters. The dataset is built from three very different BU-LAC manuscripts, offering either very dense scripts, or on the contrary with

⁴ Folders `Ms_KBR_9232_Examen_Moraux` in CREMMA Medieval; folders `SBB_PK_Hdschr25`, `UBL758` and `BG0-511` in CREMMA Medii Aevi.

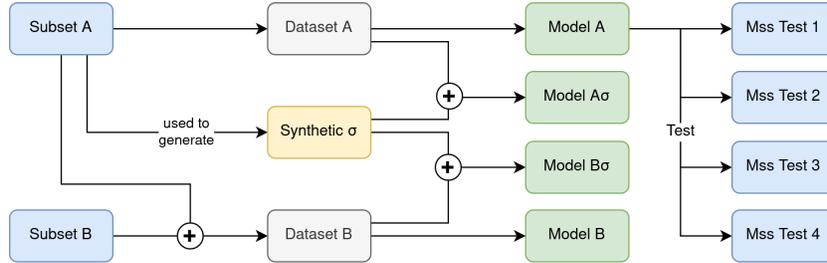


Fig. 5: Roles of subsets and models for the Medieval French and Latin experiment.

very large loops. The dataset nevertheless presents a limited variability, each manuscript being represented equally, guaranteeing for the training of GANs a fair distribution of classes for each targeted style. The number of classes is also limited to 54 (including 7 vocalisation signs). For the training of the GANs, we carried out an equal distribution of the data, ensuring the same number of lines for each manuscript (1725 lines for each).

Armenian The Armenian dataset (Table 1) is composed of 8 Armenian manuscripts in a very regular *bolorgir* script [19], from the 15th to the 18th century, to which we add two manuscripts in cursive *Shghagir* script from the 19th century.

5 Experiments

5.1 Task 1: Line Generation

Medieval Latin and French Line generation using ScrabbleGAN for Medieval Latin and French provided encouraging results, as seen in Figure 6. The model reproduced in part the fact that training data contained black mask and has a tendency to leave very little horizontal padding at the end or beginning of lines. Extremely rare classes of characters have a tendency to be less well represented, such as the sign \div (Figure 6b), used as an abbreviation for *est* in Latin. The model learnt some form of medieval decoration habits, such as colour-shading of *litterae notabiliores* transcribed as uppercase (Figure 6c).

Line generation using CycleGAN gave however rather bad results. The model focused more on the manuscript colours (back- and foreground) and on the mask than on the shape of the input (See Figure 7). Lines can sometimes be read but they are mostly indecipherable.

Armenian and Arabic The GAN generates completely believable lines in Armenian and Arabic (Figure 8). The lines generated show little variety in the styles produced but appear to have been written by a human. A blurred effect on the Arabic lines nevertheless betrays the GAN. For the Armenian, the

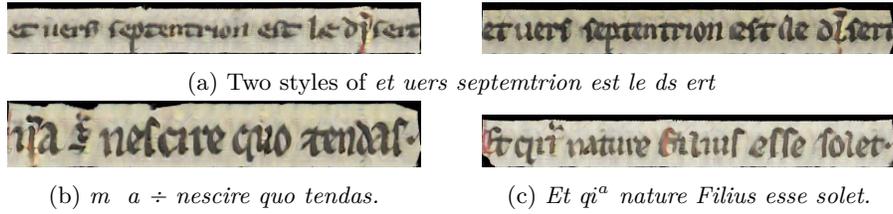


Fig. 6: Example of generated lines using ScrabbleGAN in Medieval French (a) and Latin (b-c).



Fig. 7: Example of generated lines using CycleGAN.

abusive and erroneous use of the abbreviation sign (a superscript horizontal line) also indicates that the GAN randomly generates the latter, which a human would not do. As a general rule, the Arabic GAN has largely specialized on the MS.ARA.417 manuscript.

5.2 Task 2: HTR usage

Medieval Latin and French The impact of the HTR fake lines can be considered as non-existent, or a result of local minimums (see Table 2). On the small dataset A & A σ , the model shows a decrease of 0.1% of the score on the Medieval French dataset, which might be due to an already high score (97.80%). Testing on the Latin dataset, which is the least represented language in the training set A and the most represented one in the σ subset, shows an increase of 2.2 percentage points of the accuracy using fake lines. However, this result is contrasted in the context of the B* models, where the addition of the σ dataset always results in a decrease of the score, including for the Latin testing set.

The analysis of the training curves (Figure 9a) do not provide us with any indication of a real impact of the fake lines on the model training, specifically given the fact that the *sigma* subset was only used in the training split of the



Fig. 8: Example of generated lines using ScrabbleGAN in Armenian (top 2) and Arabic (bottom 2).

Table 2: Accuracy on out of domain test set depending and the language and the training set.

Test set	A	$A\sigma$	B	$B\sigma$
Med. French	97.78	97.66	97.87	97.70
Latin	87.80	89.00	89.84	88.61

model. Analysis of the rolling average of the difference between non-sigma and sigma accuracy (Figure 9b) shows a very limited impact at training time (less than 0.5% of accuracy) on the first 5 to 10 epochs depending on the size of the training set, and a null or negative impact afterwards.

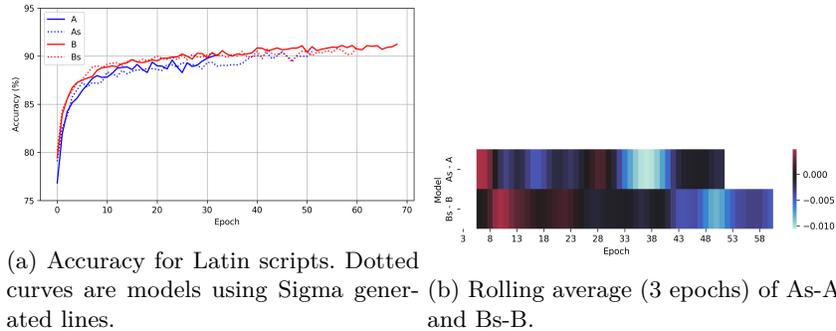


Fig. 9: Validation accuracy for Latin

Armenian At this stage, generated images with their corresponding transcription can already be used as a relevant data augmentation. We have led a very short experiment with the Armenian dataset. We have divided our dataset in 4 subsets : train (827 real lines), val (1.703 real lines), test in-domain (3.772 real lines) and test out-of-domain (2.002 lines). In-domain data are composed of *bolorgir* manuscripts and out-of-domain data composed of *shghagir* scripts. We have performed 4 training:

1. default: train set only
2. default+gan250: train set + 250 generated lines
3. default+gan500: train set + 500 generated lines
4. default+gan1000: train set + 1.000 generated lines

Training curves (Figure 10) show that we significantly speed-up training with fake images, as we could have with a bigger training dataset. CER achieved by default model and default+gan500 model are respectively 6.13% and 5.63% on in-domain test, and 22.72% and 10.79% on out-of-domain test.

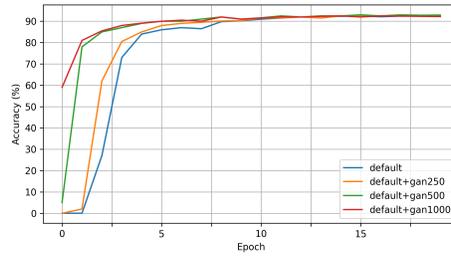


Fig. 10: Validation accuracy on the Armenian dataset.

6 Conclusion

In our experiments, the ScrabbleGan approach proved a lot more effective than the CycleGan approach. The lines generated by the former produce synthetic lines in Medieval French, Latin, Arabic and Armenian that are both usually very legible and similar to lines found in authentic documents. Though we lack, for now, a proper evaluation metric for verisimilitude, a qualitative expert assessment shows that some synthetic lines are hard to differentiate from actual lines encountered in historical documents (N.B.: this type of expert based evaluation could be systematised in the future with a stricter protocol).

Concerning the improvements of our generation as data augmentation techniques, the results tend to show that:

1. in the case of rare character classes in large datasets, the GAN seems as much in trouble to handle them as the HTR engines, and somehow fails to generate plausible outputs that can help in learning said classes. In consequence, the improvements in HTR seem for now insignificant, if any.
2. On the other hand, improvements in case of small datasets (e.g., Armenian dataset) is more substantial. In particular, models tend to learn in less epochs and, more importantly, generalise better, if a small training set (e.g. around 1000 lines) is up to doubled with artificial lines.

Further research could focus on defining metrics for the assessment of generated lines, based on expert consensus and in improvements in automated tasks. In addition, we plan to work on controlling the different ‘styles’ of the output [12], in order to be able to generate several outputs for the same line, while specifying a given script (e.g. Caroline, Gothic *Textualis*,...), a specific hand, etc. The use of diffusion models to generate a precise style is a possible field of experimentation[25].

Beyond mere increase in the scores of automated tasks, artificial lines generation could be useful to paleographers and philologists for goals such as creating virtual restoration of damaged manuscripts or studying the evolution of scripts and their variation.

Acknowledgement. This research was funded by the project “Artificial Pasts: lost texts and manuscripts that never were” (Paris Artificial Intelligence Research Institute – PRAIRIE, SHS call).

References

1. Bhunia, A.K., Khan, S., Cholakkal, H., Anwer, R.M., Khan, F.S., Shah, M.: Handwriting transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1086–1094 (2021)
2. Camps, J.B.: La ‘Chanson d’Otinél’: édition complète du corpus manuscrit et prolégomènes à l’édition critique. thèse de doctorat, dir. dominique boutet, Paris-Sorbonne (2016). <https://doi.org/10.5281/zenodo.1116735>, <https://halshs.archives-ouvertes.fr/tel-01664932>
3. Camps, J.B., Vidal-Gorène, C., Stutzmann, D., Vernet, M., Pinche, A.: Data diversity in handwritten text recognition. challenge or opportunity? In: Digital Humanities 2022. pp. 160–165 (2022)
4. Cappelli, A.: Dizionario di abbreviature latine ed italiane: usate nelle carte e codici specialmente nel medio-evo riprodotte con oltre 13000 segni incisi. Hoepli (1899)
5. Careri, M., Hasenohr, G., Féry-Hue, F., Gasparri, F., Labory, G., Lefèvre, S., Leurquin, A.F., Ruby, C.: Album de manuscrits français du XIII^e siècle. Viella (2001)
6. Clérice, T., Pinche, A.: Artificial colorization of digitized microfilms: a preliminary study (2022), <https://hal-lirmm.ccsd.cnrs.fr/MOM/hal-03335326v2>, preprint
7. Clérice, T., Vlachou-Efstathiou, M., Chagué, A.: CREMMA Medii Aevi: Literary manuscript text recognition in Latin (2023), <https://hal-enc.archives-ouvertes.fr/hal-03828353>, preprint
8. Eltay, M., Zidouri, A., Ahmad, I., Elarian, Y.: Generative adversarial network based adaptive data augmentation for handwritten arabic text recognition. PeerJ Computer Science **8**, e861 (2022)
9. Etter, D., Rawls, S., Carpenter, C., Sell, G.: A synthetic recipe for ocr. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 864–869. IEEE (2019)
10. Fogel, S., Averbuch-Elor, H., Cohen, S., Mazor, S., Litman, R.: Scrabblegan: Semi-supervised varying length handwritten text generation. In: CVPR 2020, WiDS TLV 2020 (2020), <https://www.amazon.science/publications/scrabblegan-semi-supervised-varying-length-handwritten-text-generation>
11. Fogel, S., Averbuch-Elor, H., Cohen, S., Mazor, S., Litman, R.: Scrabblegan: Semi-supervised varying length handwritten text generation (2020). <https://doi.org/10.48550/ARXIV.2003.10557>, <https://arxiv.org/abs/2003.10557>
12. Gan, J., Wang, W., Leng, J., Gao, X.: Higan+: Handwriting imitation gan with disentangled representations. ACM Trans. Graph. **42**(1) (2022). <https://doi.org/10.1145/3550070>, <https://doi.org/10.1145/3550070>
13. Kang, L., Riba, P., Wang, Y., Rusinol, M., Fornés, A., Villegas, M.: 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 16. pp. 273–289. Springer (2020)
14. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks (2018). <https://doi.org/10.48550/ARXIV.1812.04948>, <https://arxiv.org/abs/1812.04948>

15. Pinche, A.: Generic HTR Models for Medieval Manuscripts The CREMMALab Project, <https://hal.science/hal-03837519>, preprint
16. Pinche, A.: Guide de transcription pour les manuscrits du X^e au XV^e siècle (2022), <https://hal.archives-ouvertes.fr/hal-03697382>, working paper
17. Pinche, A., Gabay, S., Leroy, N., Christensen, K.: Données HTR manuscrits du 15e siècle, <https://github.com/Gallicorpora/HTR-MSS-15e-Siecle>, dataset
18. Pinche, A., Gabay, S., Leroy, N., Christensen, K.: Données HTR incunables du 15e siècle (2022), <https://github.com/Gallicorpora/HTR-incunable-15e-siecle>, dataset
19. Vidal-Gorène, C., Decours-Perez, A.: A computational approach of armenian paleography. In: Document Analysis and Recognition–ICDAR 2021 Workshops: Lausanne, Switzerland, September 5–10, 2021, Proceedings, Part II 16. pp. 295–305. Springer (2021)
20. Vidal-Gorène, C., Lucas, N., Salah, C., Decours-Perez, A., Dupin, B.: Rasam – a dataset for the recognition and analysis of scripts in arabic maghrebi. In: Barney Smith, E.H., Pal, U. (eds.) Document Analysis and Recognition – ICDAR 2021 Workshops. pp. 265–281. Springer International Publishing, Cham (2021)
21. Vlachou-Efstathiou, M.: Voss.lat.o.41 - eutyches "de uerbo" glossed (2022), <https://github.com/malamatenia/Eutyches>, dataset
22. Vöglin, L., Drazyk, M., Pondenkandath, V., Alberti, M., Ingold, R.: Generating synthetic handwritten historical documents with ocr constrained gans. In: International Conference on Document Analysis and Recognition. pp. 610–625. Springer (2021)
23. White, N., Karaisl, A., Clérice, T.: Caroline minuscule by rescribe (2022), <https://github.com/rescribe/carolineminuscule-groundtruth>, dataset
24. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Computer Vision (ICCV), 2017 IEEE International Conference on (2017)
25. Zhu, Y., Li, Z., Wang, T., He, M., Yao, C.: Conditional text image generation with diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14235–14245 (2023)