Handwritten Text Generation from Visual Archetypes
Supplementary Material

Vittorio Pippi, Silvia Cascianelli, Rita Cucchiara
University of Modena and Reggio Emilia
Via Pietro Vivarelli, 10, Modena (Italy)
{name.surname}@unimore.it

In this document, we present additional experimental analysis of our proposed approach for Few-Shot HTG, VATr. In particular, we show an ablation analysis on the role of each of the used loss terms in §1; additional details about the synthetic pre-training dataset in § 2; additional qualitative results in § 3; a comparison of the charsets generated in different styles in § 4; further analysis of the long-tail characters generation scenario in § 5; examples of the generation of some out-of-charset characters from other alphabets in §6; and some HTR results obtained by training both on real and VATr-generated images in §6.

1. Role of the Loss Function Terms

We analyze the effect of each loss term, both quantitatively and qualitatively. The results of this study are reported in Table 1 and show that our model needs the included regularization terms to be trained properly. Moreover, when none of the two terms enforcing handwriting style faithfulness (*i.e.*, the writer classification loss L_{class} and the cycle loss L_{cycle}) is used, VATr generates images of readable words but does not render the style of the reference images. Conversely, when the text recognition loss L_{HTR} is not employed, it renders "scrabbles" whose overall appearance resembles that of the reference style images.

2. Synthetic Pre-training Dataset

The synthetic dataset used to pre-train the convolutional backbone of the encoder is obtained by combining calligraphic fonts and words to generate word images. The process involves selecting $10\,400$ fonts from dedicated websites, ensuring that fonts with small caps and decorative elements such as hearts or stars are discarded. Next, $10\,400$ words of varying lengths are randomly chosen from the English vocabulary. All possible combinations between the selected fonts and words are then rendered, resulting in a total of $10\,400 \times 10\,400 = 108\,160\,000$ word images. Some exemplar images are reported in Figure 1.

Figure 1. Exemplar images from the synthetic dataset used for pre-training the style encoder.

3. Additional Styled HTG Qualitative Results

In Figures 2-5, we report qualitative examples of images generated with the proposed VATr compared to GANwriting [6] and HWT [1] in the four styled generation IV-S, IV-U, OOV-S, and OOV-U scenarios, respectively. It can be noticed that the style fidelity in the images produced by VATr does not deteriorate significantly when the style examples are unseen compared with the case in which they have been seen in training. It can be also noticed that VATr consistently disregards the background while reproducing the style, while the HWT competitor reproduces the background artifacts, especially in the unseen style scenarios.

4. Styled Charsets Generation

In Figure 6, we report qualitative examples of styled generation of the characters in the IAM dataset, sorted by the frequency in which they appear in the training set. Also in this case, we compare VATr against HWT and GANwriting, which have all been trained on the same data. It has to be noted that, by design, GANwriting cannot generate characters that are not letters. On the other hand, HWT includes also punctuation and digits in its charset. Nonetheless, it can be noticed that VATr generates images for a larger number of characters, including long-tail ones, with which both the competitors struggle.

IV-S scenario

Style examples		
The Labour movement in		Addressing the annual congress of the National Union of General
I he said he felt sure	VATr	*Addressing the annual congress of the National Union of General*
Of General and Workers	HWT	*Addressing the annual congress of the National Union of General*
	GANwriting	*Address the annual congress of the Nationa Union of General*

Style examples		
red mist veiled her All		But after that slight slip of the tongue on Nigels part she had
and it was as though a	VATr	*But after that slight slip of the tongue on Nigels part she had*
Nigel's past she had	HWT	*But after that slight slip of the tongue on Nigels part she had*
	GANwriting	*But after that slight slip of the tongue on Nigels part she had*

Style examples		
body of German in his		Banks have paid in a first instalment of almost in response to the
been frozen Mr. the	VATr	*Banks have paid in a first instalment of almost in response to the*
For nearly a year has	HWT	*Banks have paid in a first instalment of almost in response to the*
	GANwriting	*Banks have paid in a first instalm of almost in respons to the*

Style examples		
mind She younger		The girl is a knockout see pic ture of Julie Newmar for con She
has quite a mission in	VATr	*The girl is a knockout see pic ture of Julie Newmar for con She*
firmation) S he also	HWT	*The girl is a knockout see pic ture of Julie Newmar for con She*
	GANwriting	*The girl is a knocki See pic ture of Julie Newmar for con She*

Style examples		
Hour after hour it out		The dinner table is the best answer to the grumblers in Britain
I proclaimed the I and	VATr	*The dinner table is the best answer to the grumblers in Britain*
II THIS is colour day	HWT	*The dinner table is the best answer to the grumblers in Britain*
	GANwriting	*The dinner table is the best answer to the grumble in Britain*

Style examples		
defeat a Censure On II		The result of the vote was not in doubt For the Tories were massed
in answers to their to	VATr	*The result of the vote was not in doubt For the Tories were massed*
the Tories Were Massed	HWT	*The result of the vote was not in doubt For the Tories were massed*
	GANwriting	*The result of the vote was not in doubt For the Tories were massed*

Figure 2. Exemplar qualitative results of styled text generation for the IV-S setting on the IAM dataset.

IV-U scenario

Style examples		
a mistake about Me		Sometimes she took Harry around with her but never her husband a
I guess Sally made	VATr	Sometimes she took Harry around with her but never her husband a
accepted as inevitable	HWT	Sometimes she took Harry around with her but never her husband a
	GANwriting	Sometim she took Harry around with her ' but never her husband · a

Style examples		
establish any one text		THE Fourth Gospel was almost certainly written in Greek A modern
It is not possible to	VATr	THE Fourth Gospel was almost certainly written in Greek A modern
A modern text of the	HWT	THE Fourth Gospel was almost certainly written in Greek A modern
	GANwriting	Thi Fourth Gospel was almost certain written in Greek ' A modern

Style examples		
anyway NUF 121.		No she told Phil Gubbins a junior houseman it wasnt Nigel you saw
my lad.	VATr	No she told Phil Gubbins a junior houseman it wasnt Nigel you saw
it wasn't Nigel you	HWT	No she told Phil Gubbins a junior houseman it wasnt Nigel you saw
	GANwriting	' No ' · she told Phil Gubbins · a junior housema · · it wasnt Nigel you Saw

Style examples		
He peered as it were		Something s up said Lord Undertone carefully casual The servants
And the mules didn't	VATr	Something s up said Lord Undertone carefully casual The servants
carefully casual ?	HWT	Something s up said Lord Undertone carefully casual The servants
	GANwriting	· Somethi up ' · said Lord Underto · careful casual · · The servant

Style examples		
child his heart Was		Gay thanked him and walked out on to the terrace thinking that
a cold hand seeming to	VATr	Gay thanked him and walked out on to the terrace thinking that
drink and taken is	HWT	Gay thanked him and walked out on to the terrace thinking that
	GANwriting	Gay thanked him and walked out on to the terrace thinkin that

Style examples		
engaged. to the man		No other news except that Elaine is engaged and going to marry a
back in her bag Was	VATr	No other news except that Elaine is engaged and going to marry a
the letter it	HWT	No other news except that Elaine is engaged and going to marry a
	GANwriting	No other news except that Elaine is engaged and going to marry a

Figure 3. Exemplar qualitative results of styled text generation for the IV-U setting on the IAM dataset.

OOV-S scenario

Style examples

number to USL with :
These are the correct
To ensure the / it

finned voter Jacobs calmly hip clubs quintet blunts Grazie Barton

VATr — finned voter Jacobs calmly hip clubs quintet blunts Grazie Barton

HWT — finned voter Jacobs calmly hip clubs quintet blunts Grazie Barton

GANwriting — finned voter Jacobs calmly hip clubs quintet blunts Grazie Barton

Style examples

the scheme Profits
Guy Eden write : On
Wilson's offe ev in mind.

turtles audible Woodrow retreat Orders Conlow hobby skin tally

VATr — turtles audible Woodrow retreat Orders Conlow hobby skin tally

HWT — turtles audible Woodrow retreat Orders Conlow hobby skin tally

GANwriting — turtles audible Woodrow retreat Orders Conlow hobby skin tally

Style examples

tinkle and 9 thud from
good look round A
and you know how she a

weakly upriver Newsom Meeker weed fiscal Diane Errors Mig biz Drink

VATr — weakly upriver Newsom Meeker weed fiscal Diane Errors Mig biz Drink

HWT — weakly upriver Newsom Meeker weed fiscal Diane Errors Mig biz Drink

GANwriting — weakly upriver Newsom Meeker weed fiscal Diane Errors Mig biz Drink

Style examples

just as inevitably
these in July when)
tradition and arrives

nagged ribbing hunt AA Pohly triol saws popped aloof Ceramic thong

VATr — nagged ribbing hunt AA Pohly triol saws popped aloof Ceramic thong

HWT — nagged ribbing hunt AA Pohly triol saws popped aloof Ceramic thong

GANwriting — nagged ribbing hunt Af Pohly triol saws popped aloof Ceramic thong

Style examples

mind She younger
has quite a mission in
firmation) S she also

suffuse muffins Miners Cabrini weakly upriver Newsom Meeker weed

VATr — suffuse muffins Miners Cabrini weakly upriver Newsom Meeker weed

HWT — suffuse muffins Miners Cabrini weakly upriver Newsom Meeker weed

GANwriting — suffuse muffins Miners Cabrini weakly upriver Newsom Meeker weed

Style examples

An attempt to get more
There General de had
the British Prime of

upriver Newsom Meeker weed fiscal Diane Errors Mig biz Drink chop

VATr — upriver Newsom Meeker weed fiscal Diane Errors Mig biz Drink chop

HWT — upriver Newsom Meeker weed fiscal Diane Errors Mig biz Drink chop

GANwriting — upriver Newsom Meeker weed fiscal Diane Errors Mig biz Drink chop

Figure 4. Exemplar qualitative results of styled text generation for the OOV-S setting on the IAM dataset.

OOV-U scenario

Style examples		
When Achilles gave a A		sexy humming cheated lick grades heroic Clever foul mood warrior
roasting and boiling	VATr	sexy humming cheated lick grades heroic Clever foul mood warrior
cooking meat except by	HWT	sexy humming cheated lick grades heroic Clever foul mood warrior
	GANwriting	sely humming cheaked lick grades heroic foul mood warrior

Style examples		
A mistake about Me		Babin optimum Leyden enrage induces newel trim bolts frog cinder
I guess Sally made	VATr	Babin optimum Leyden enrage induces newel trim bolts frog cinder
accepted as inevitable	HWT	Babin optimum Leyden enrage induces newel trim bolts frog cinder
	GANwriting	Babin optimum Leyden enrage induces newel trim bolts frog cinder

Style examples		
King Hasla of Hark of		noon hoes Seafood yuh Mostly seeds bestow acetate jokers waning
cittens in the book	VATr	noon hoes Seafood yuh Mostly seeds bestow acetate jokers waning
natural springs do Or	HWT	noon hoes Seafood yuh Mostly seeds bestow acetate jokers waning
	GANwriting	noon hoes Seafood yuh Mostly seeds bestow acetate jokers waning

Style examples		
and here he was bland		sacker comest tense par fiend Soiree voted Putting pansy doormen
be met with violent	VATr	sacker comest tense par fiend Soiree voted Putting pansy doormen
vengeance She had to	HWT	sacker comest tense par fiend Soiree voted Putting pansy doormen
	GANwriting	sacker comest tense par fiend Soiree voted Putting pansy doormen

Style examples		
fringing of clumsy and		blunts Grazie Barton NAB specie Fonta narrow Swan denials Rawson
mantel-piece covers a	VATr	blunts Grazie Barton NAB specie Fonta narrow Swan denials Rawson
the overcrowded and	HWT	blunts Grazie Barton NAB specie Fonta narrow Swan denials Rawson
	GANwriting	blunts Grazie Barton NAB specie Fonta narrow Swan denials Rawson

Style examples		
They will think of the		Whisky super screwed Flower salads Glow Vapor Married recieve push
Help will not be for	VATr	Whisky super screwed Flower salads Glow Vapor Married recieve push
the midst of plenty	HWT	Whisky super screwed flower salads Glow Vapor Married recieve push
	GANwriting	Whisky super screwed Flower salads Glow Vapor Married recieve handle push

Figure 5. Exemplar qualitative results of styled text generation for the OOV-U setting on the IAM dataset.
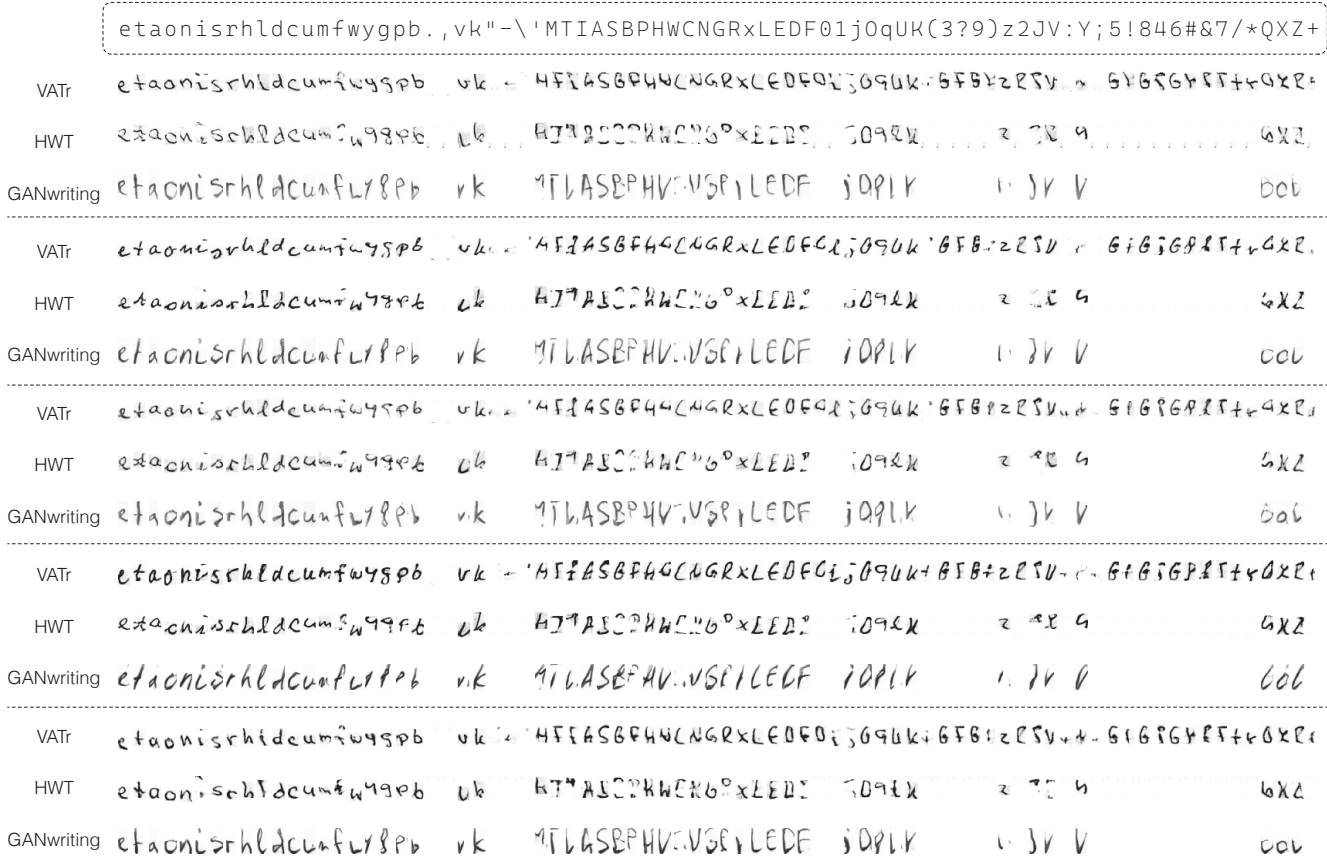
```
etaonisrhldcumfwygpb.,vk"-\'MTIASBPHWCNGRxLEDF01jOqUK(3?9)z2JV:Y;5!846#&7/*QXZ+
```

Figure 6. Exemplar IAM charsets generated in different styles sorted by frequency.

5. Additional Long-tail Character Generation Results

We deepen our analysis on the generation of test words containing rare characters in Figure 7, where we compare VATr against HWT in terms of the FID obtained by changing the Long-tail threshold. This is the value of the frequency of appearance in the training set for which a character can be considered as long-tail. Recall that in the experi-

Table 1. Ablation analysis on the loss terms. The qualitative examples refer to the generation of the word `that` in two different styles (indicated on top).

L_{adv}	L_{HTR}	L_{class}	L_{cycle}	FID	writer	aspect
✓				224.87		
✓			✓	261.23		
✓		✓		255.11		
✓		✓	✓	220.83		
✓	✓			46.61		
✓	✓		✓	41.68		
✓	✓	✓		18.70		
✓	✓	✓	✓	**17.79**		

ments in the main paper we set this threshold to 1000, which was determined by observing the character distribution in the IAM training set, reported both in logarithmic and linear scale in Figure 8. It can be noticed that the FID obtained by VATr is generally lower than that obtained by HWT, especially at lower threshold values. This demonstrates the robustness of our approach of exploiting dense representations of the characters, compared to resorting to one-hot vectors. It can be noticed an increase in the FID of both approaches when the threshold is around 3500. Setting the threshold to this value means excluding almost all the words containing only small letters. Afterward, the percentage of capital letters, digits, and punctuation becomes higher than that of small letters in the generated words. Thus, the capacity to faithfully generate also those symbols is more relevant for those words.

6. Additional Out-of-Charset results

In Figure 9, we report examples of the generation of out-of-charset characters from non-Latin alphabets (Greek and Coptic) with our proposed approach. Despite this setting being beyond the scope of our work, it can be noticed that VATr is able to reproduce symbols that are not present in
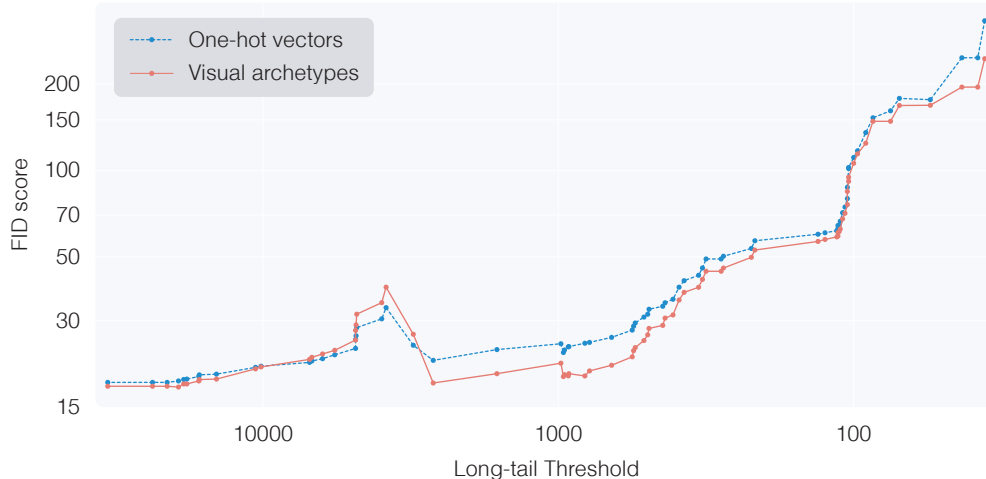
Figure 7. FID score on words containing long-tail characters with respect to the threshold set to consider a character as rare in the IAM dataset (the x-axis is in logarithmic scale).

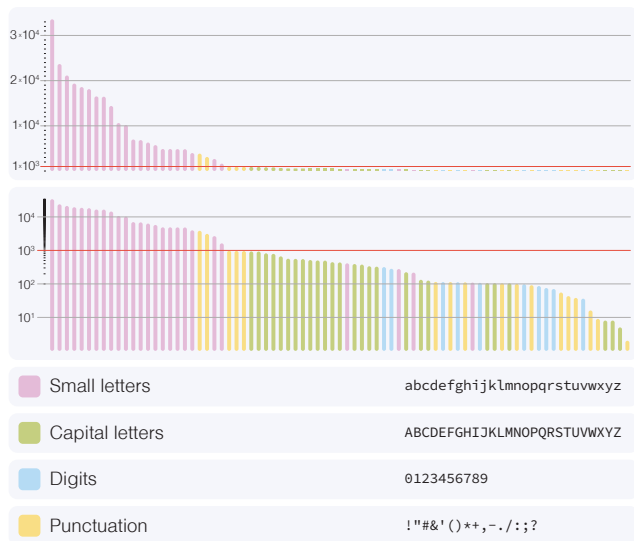| Small letters | abcdefghijklmnopqrstuvwxyz |
Capital letters	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Digits	0123456789
Punctuation	!"#&'()*+,-./:;?

Figure 8. Distribution and classification of the characters in the training set of the IAM dataset in linear scale (top) and logarithmic scale (bottom).

its training set, especially in the case these are geometrically similar to in-charset symbols. In fact, when generating Greek letters that are close to Latin ones starting from their respective visual archetypes, VATr can exploit learned geometric regularities. On the other hand, it struggles to generate characters from the Copto alphabet, whose visual archetypes are very different from those of Latin letters.

7. Styled HTG for HTR

One of the main applications of styled HTG is providing training data for HTR models to be applied to writer-

Table 2. Performance comparison of an HTR model trained on real data only and on a combination of real and generated styled text images (obtained with our approach and the SotA HWT approach) on two writer-specific datasets.

	Saint Gall		Washington	
	CER	WER	CER	WER
Real	**4.5**	32.5	3.4	15.9
HWT + Real	4.6	31.2	3.7	16.5
VATr + Real	**4.5**	**30.9**	**3.0**	**13.1**

specific manuscripts. To assess the potential benefits of this strategy when using VATr to improve an HTR model in this setting, we consider the interesting case of two low-resource single-author HTR datasets (Saint Gall [4] and Washington [5][1]) and generate synthetic training lines with VATr. We then use these as additional samples to train a SotA HTR model [2, 3] and compare the transcription results against a baseline not exploiting synthetic data and the baseline exploiting synthetic data generated with the one-hot vectors-based HWT [1]. These results are reported in Table 2, expressed in terms of Character Error Rate (CER) and Word Error Rate (WER). It can be observed that training on VATr-generated images reduces the errors *w.r.t.* both the compared approaches, especially in terms of WER. This suggests that the misrecognized characters are more concentrated in single words.
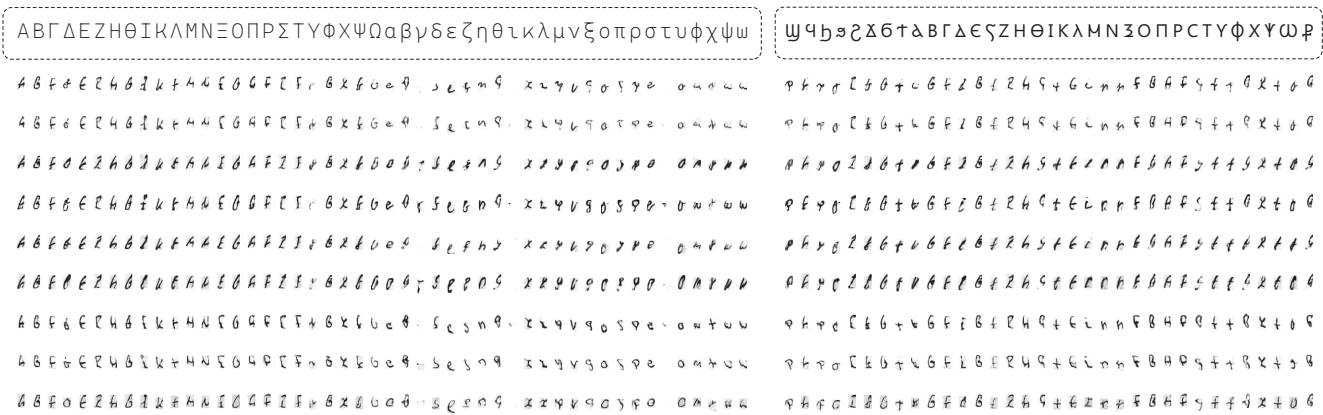
---

[1] https://fki.tic.heia-fr.ch/databases/iam-historical-document-database

Figure 9. Exemplar out-of-charset symbols generated in different styles. The alphabets used are Greek (left) and Coptic (right).

References

[1] Ankan Kumar Bhunia, Salman Khan, Hisham Cholakkal, Rao Muhammad Anwer, Fahad Shahbaz Khan, and Mubarak Shah. Handwriting Transformers. In *ICCV*, 2021. 1, 7

[2] Silvia Cascianelli, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara, et al. Boosting Modern and Historical Handwritten Text Recognition with Deformable Convolutions. *IJDAR*, pages 1–15, 2022. 7

[3] Iulian Cojocaru, Silvia Cascianelli, Lorenzo Baraldi, Massimiliano Corsini, and Rita Cucchiara. Watch your strokes: Improving handwritten text recognition with deformable convolutions. In *ICPR*, 2020. 7

[4] Andreas Fischer, Volkmar Frinken, Alicia Fornés, and Horst Bunke. Transcription alignment of Latin manuscripts using hidden Markov models. In *HIP*, 2011. 7

[5] Andreas Fischer, Andreas Keller, Volkmar Frinken, and Horst Bunke. Lexicon-free handwritten word spotting using character HMMs. *Pattern Recognition Letters*, 33(7):934–942, 2012. 7

[6] Lei Kang, Pau Riba, Yaxing Wang, Marçal Rusiñol, Alicia Fornés, and Mauricio Villegas. GANwriting: Content-Conditioned Generation of Styled Handwritten Word Images. In *ECCV*, 2020. 1