FRIEDRICH-ALEXANDER-UNIVERSITÄT ERLANGEN-NÜRNBERG
SCHOOL OF ENGINEERING

# Master’s Thesis: Making offline handwriting editable

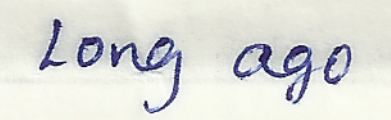
Martin Stumpf, Pattern Recognition Lab (CS 5), September 9, 2019

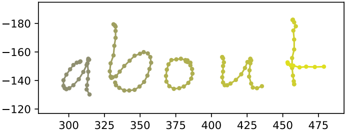
# Introduction

# Introduction

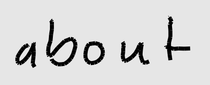
**”Making offline handwriting editable”**

- offline vs online handwriting:

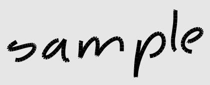



- ‘editable’:




+ "sample" =



# Motivation

*Why?*

- Combining benefits of digital and handwritten text
- Automated handwriting generation
- Because we can (hopefully)
- Getting new insights into neural networks

## General Approach

**Goal:**

- Full offline-to-offline handwriting style transfer algorithm

**Related Work:**

- *Alex Graves*: ”Generating Sequences With Recurrent Neural Networks” (2013)
- *Aksan et al.*: ”DeepWriting: Making Digital Ink Editable via Deep Generative Modeling” (2018)
- Disadvantage: online data, no pen or background style

**Idea:**

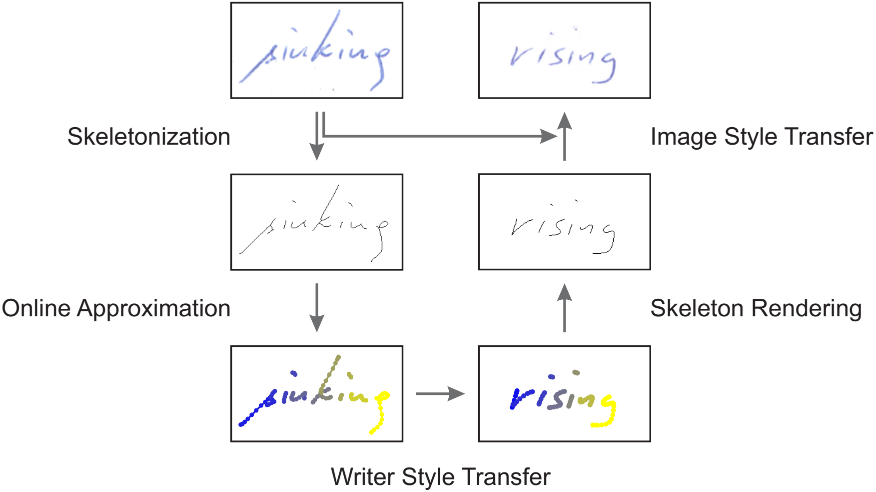
- Conversion of offline to online handwriting
- Usage of existing online handwriting synthesis algorithm
- Additional style transfer for pen style and background

# Goals

- Full offline-to-offline handwriting style transfer algorithm
- Finding a robust algorithm for handwriting skeletonization
- Is an offline to online handwriting conversion sufficient to use online algorithms?
- Finding a way to transfer the pen style to the output image
- *Optional:* Finding a way to transfer the background style to the output image

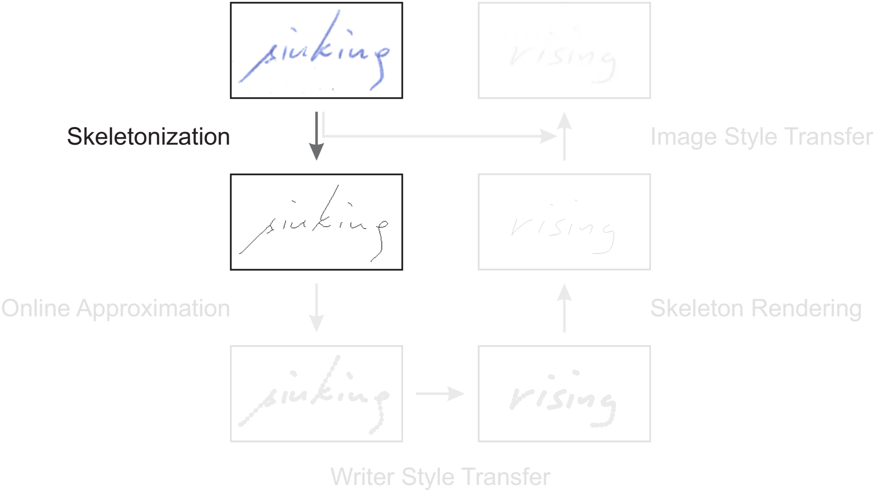
# Pipeline Overview

## Pipeline Overview



# Stage 1: Skeletonization

# Skeletonization



# Background - Pix2Pix (2016, Phillip Isola et al.)

Day to Night

input output

BW to Color

input output

Edges to Photo

input output

**Capabilities:**

- mapping from one visual representation to another

**Requirements:**

- pairwise annotations in training dataset

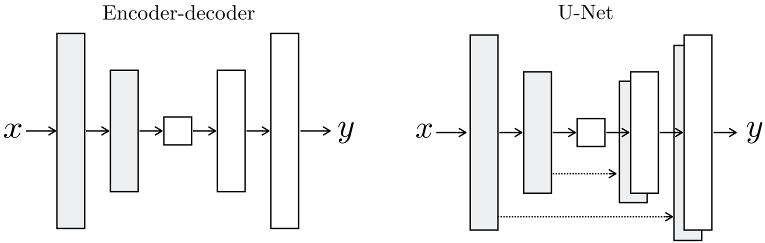
---

Phillip Isola et al. “Image-to-Image Translation with Conditional Adversarial Networks”. In: *arxiv* (2016).

# Background - Pix2Pix (2016, Phillip Isola et al.)

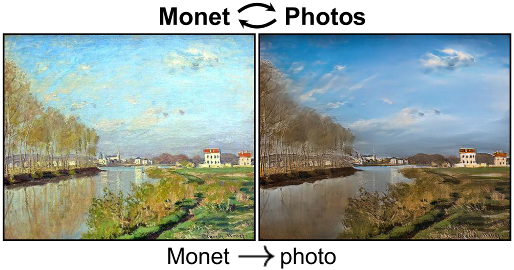
## Innovations of Pix2Pix

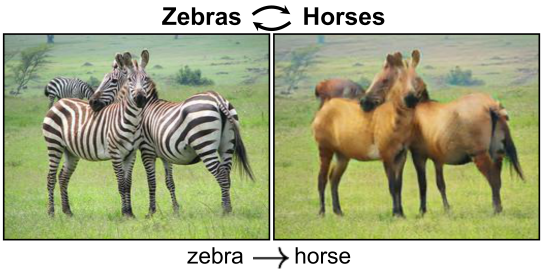
- U-Net




- Loss Function: PatchGAN + L1

# Background - CycleGAN (2017, Jun-Yan Zhu et al.)




**Capabilities:**

- mapping between two visual representations

**Requirements:**

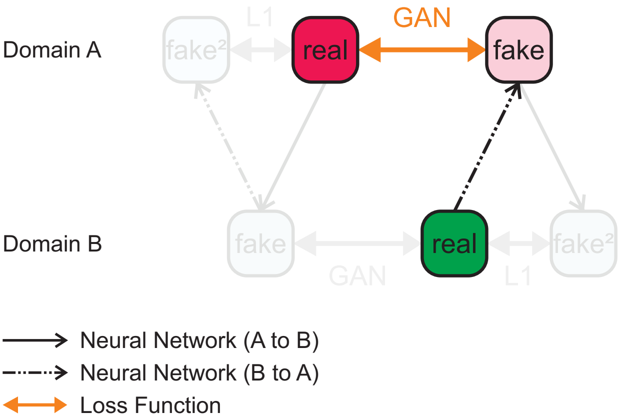
- one dataset of each domain *without* pairwise annotation

---

Jun-Yan Zhu et al. “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”. In: *Computer Vision (ICCV), 2017 IEEE International Conference on*. 2017.

# Background - CycleGAN (2017, Jun-Yan Zhu et al.)

## Cycle Consistency Training

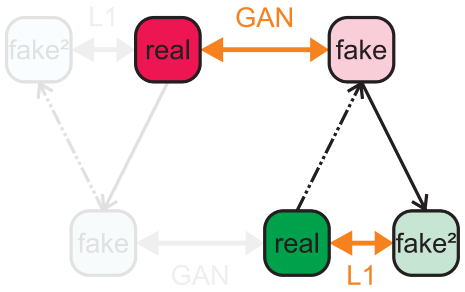


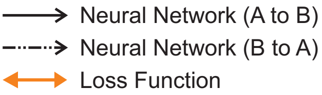
# Background - CycleGAN (2017, Jun-Yan Zhu et al.)

## Cycle Consistency Training

Domain A

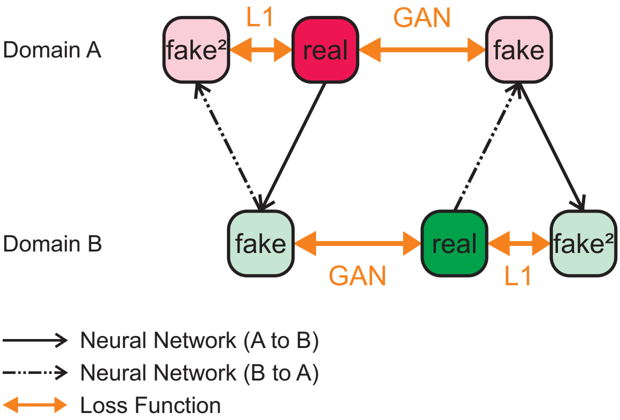
Domain B



# Background - CycleGAN (2017, Jun-Yan Zhu et al.)

## Cycle Consistency Training



# Skeletonization - Datasets



**CVL**

- Offline handwriting, white background
- 310 writers
- 7 texts (ca. 65 words per text)
- ca. 15000 lines of text

**IAM On-Line**

- Online handwriting
- 221 writers
- 1700 forms
- 13049 lines of text

**But:** no pairwise annotations between them

Florian Kleber et al. "CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting". In: *Proceedings of the 2013 12th International Conference on Document Analysis and Recognition*. ICDAR '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 560–564. ISBN: 978-0-7695-4999-6. DOI: 10.1109/ICDAR.2013.117.

Marcus Liwicki and Horst Bunke. "IAM-OnDB - an On-Line English Sentence Database Acquired from Handwritten Text on a Whiteboard". In: *8th Intl. Conf. on Document Analysis and Recognition*. Vol. 2. 2005, pp. 956–961.

# Skeletonization - First attempt: CycleGAN

(a) skeleton

(b) synthetic image, produced by CycleGAN

**Problems:**

- Datasets are created by different writers
- CycleGAN adapts by changing the writer style

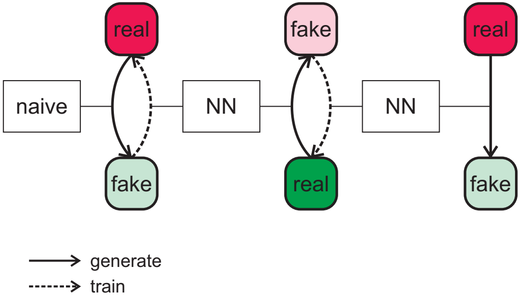
# Skeletonization - Second attempt: Pix2Pix

**Problem:**

- No dataset with pairwise annotations available

**Solution:**

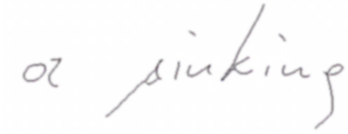
- Iterative knowledge transfer from a naive skeletonization



# Skeletonization - Second attempt: Pix2Pix

(a) CVL input data

(b) primitive skeletonization



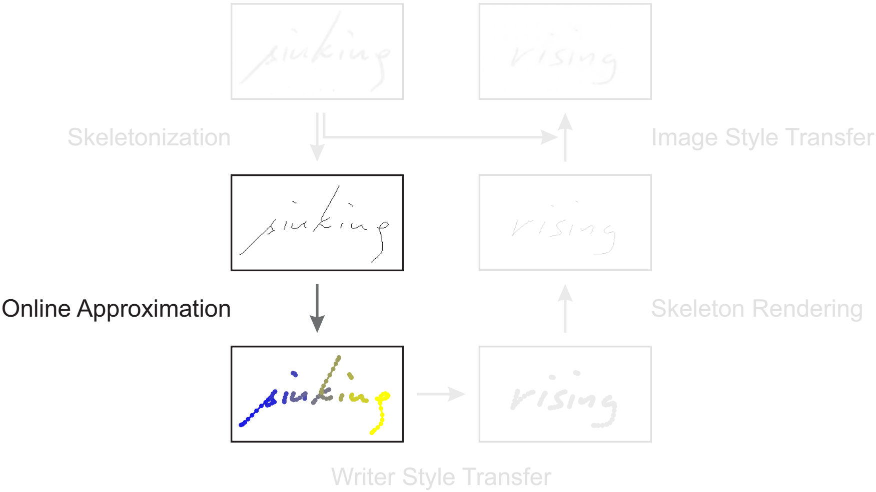
(c) learned skeletonization

**Improvements by knowledge transfer:**

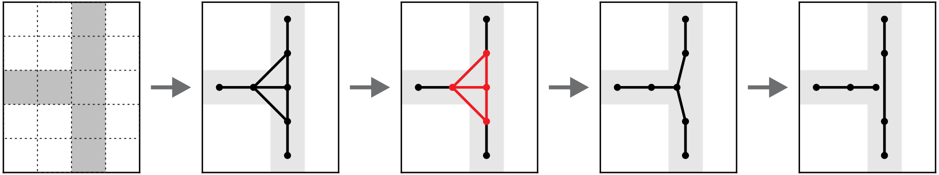
- Less artifacts
- Robustness towards contrast and brightness
- Better handling of noise on the background

# Stage 2: Conversion to Online Handwriting

# Conversion to Online Handwriting



# Graph Creation and Refinement



(a) (b) (c) (d) (e)

**Additional Steps** (not shown):

- Breaking of Cycles
- Reordering

# Resampling

**Reasons for resampling:**

- Less memory requirements
- Closer to real handwriting
- Better predictability in next stage

**Methods:**

- *Constant Velocity Resampling*
  - Simple and easy to implement
  - Still unrealistic and restrictive
- *Maximum Acceleration Resampling*
  - Closer to real handwriting, better understandable for predictions
  - More complex

# Maximum Acceleration Resampling

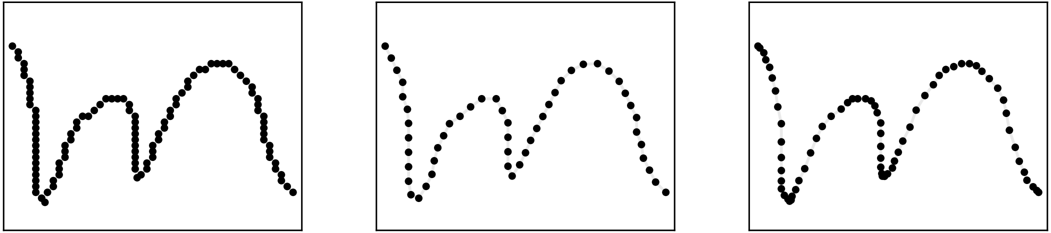
**Algorithm:**

- Basic Principle is a *4D Dijkstra Shortest Path Search*
- Every node consists of:
  - $p_x, p_y$ - Point Coordinates
  - $v_x, v_y$ - Incoming Velocity, defined as $\vec{v} = \vec{p} - \vec{p_{prev}}$
- Velocities at first and last node are defined to be $\vec{0}$
- Maximum acceleration constraint: $\|\vec{v} - \vec{v}_{prev}\| < a$
- Further constraint necessary to prevent shortcuts

**Performance Optimizations:**

- Dynamic programming
- Implementation in *C*

# Resampling Comparison



(a) no resampling

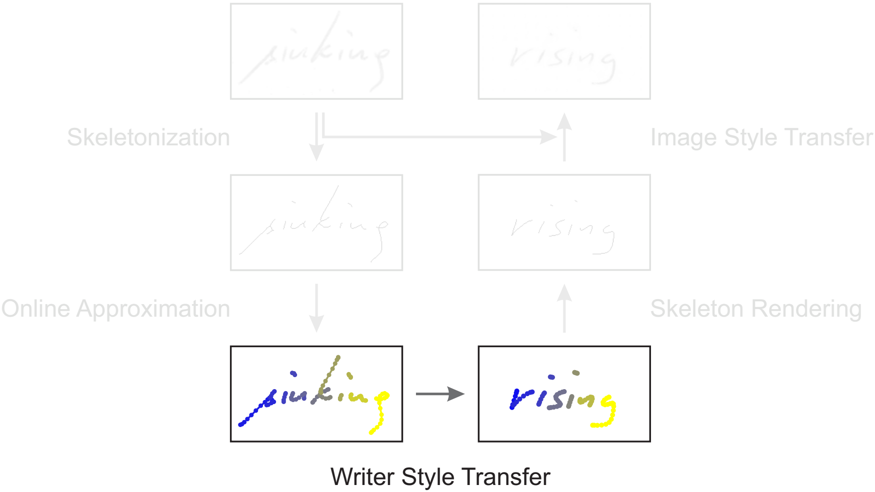
(b) constant velocity resampling

(c) maximum acceleration resampling

FRIEDRICH-ALEXANDER-
UNIVERSITÄT
ERLANGEN-NÜRNBERG
SCHOOL OF ENGINEERING

# Stage 3: Writer Style Transfer

# Writer Style Transfer



# Writer Style Transfer

- Has already been solved for real data
- *Challenge*: Extend to synthetic data

**Algorithms:**

- Graves' Prediction Network
- DeepWriting


Alex Graves. "Generating Sequences With Recurrent Neural Networks". In: *CoRR* abs/1308.0850 (2013). arXiv: `1308.0850`.

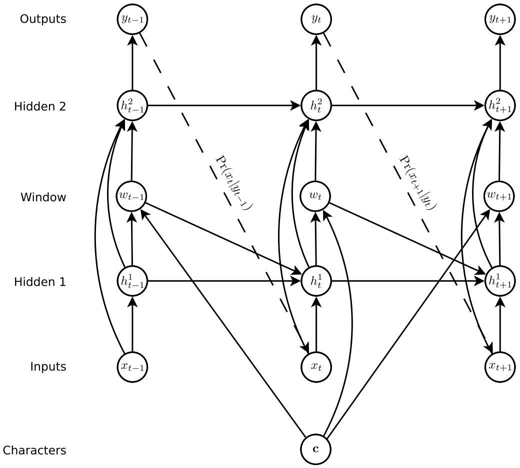
Emre Aksan, Fabrizio Pece, and Otmar Hilliges. "DeepWriting: Making Digital Ink Editable via Deep Generative Modeling". In: *SIGCHI Conference on Human Factors in Computing Systems*. CHI '18. Montréal, Canada: ACM, 2018.

# Background: Graves' Network (2013, Alex Graves)

- Originally just a demonstration of *LSTM* cell capabilities
- Works with online handwriting
- Predicts future pen position
- Takes the text content as additional input
- Attention mechanism to read the text content

**Handwriting Synthesis:**

- *Prime* network with target style
- Continue with target content, feed predicted positions back into the network



# Background: DeepWriting (2018, Emre Aksan et al.)

- *Conditional Variational Recurrent Neural Network*
- Splits content and style into separate latent random variables
- Network architecture requires *EOC* and *BOW* symbols

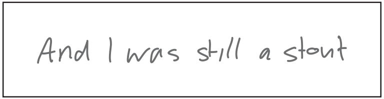
**Advantages:**

- Style and content are stored explicitly

**Disadvantages:**

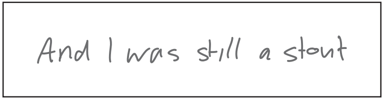
- Requires *EOC* and *BOW* annotations
- Only looks at one content character, switches at *EOC*-signal
  $\Rightarrow$ No character look-ahead

# Writer Style Transfer: Experiments

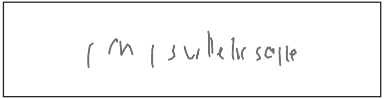



(a) Style Input

# Writer Style Transfer: Experiments

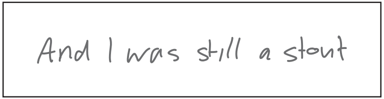



(a) Style Input

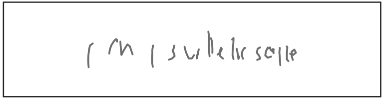


(b) DeepWriting + Constant Velocity

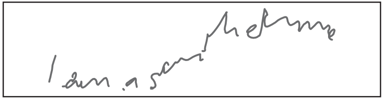
# Writer Style Transfer: Experiments




(a) Style Input

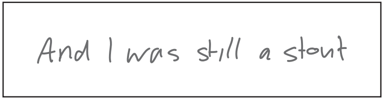


(b) DeepWriting + Constant Velocity

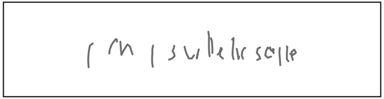


(c) DeepWriting + Maximum Acceleration

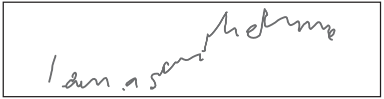
# Writer Style Transfer: Experiments



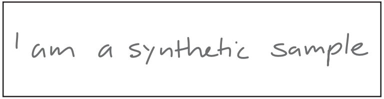

(a) Style Input



(b) DeepWriting + Constant Velocity



(c) DeepWriting + Maximum Acceleration




(d) Graves + Maximum Acceleration

# Writer Style Transfer: Training Data

**Question:**

- What should the network be trained with?

**Options:**

- *Real Skeletons*:
  - Contain less artifacts
  - Different from what it will see in the final pipeline
  - **Disadvantage**: Network might have trouble with style extraction
- *Real Offline Handwriting*:
  - Contain more artifacts
  - Match what it will see in the final pipeline
  - **Disadvantage**: Network might produce more artifacts

# Writer Style Transfer: Training Data

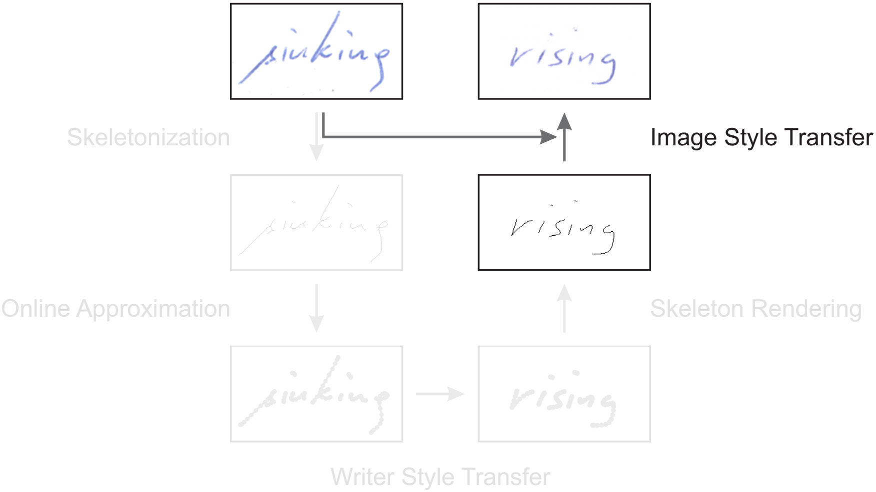
| Training Data | Style Input | Result |
|---|---|---|
| synthetic | synthetic | I am a synthetic sample |
| synthetic | real | I am a synthelic cample |
| real | synthetic | I am a synthetic sample |
| real | real | ' I am a synthetic sample |
| Reference Style | | The Prime Minister looked puzzled, |

# Writer Style Transfer: Qualitative Demonstration

| Style Input | Output |
|---|---|
| spineless!" "You think you could | I am a synthetic sample |
| Michael Caxton The father | I am a synthetic sample |
| they get'em. We've been alert | I am a synthetic sample |
| one for me?" James did so, and | I am a synthetic sample |
| any better? You know it was! He | I am a synthetic sample |
| and change the money | I am a synthetic sample |

# Stage 4: Pen Style Transfer

# Pen Style Transfer



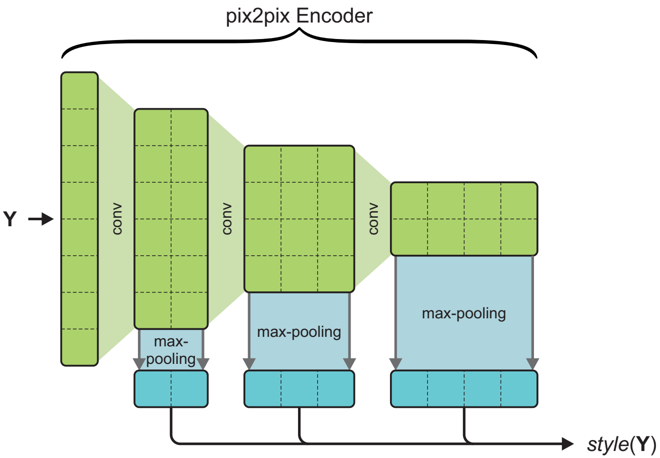
## Pen Style Transfer

Pix2Pix already proved to work (without style transfer)
$\Rightarrow$ promising, but modifications necessary

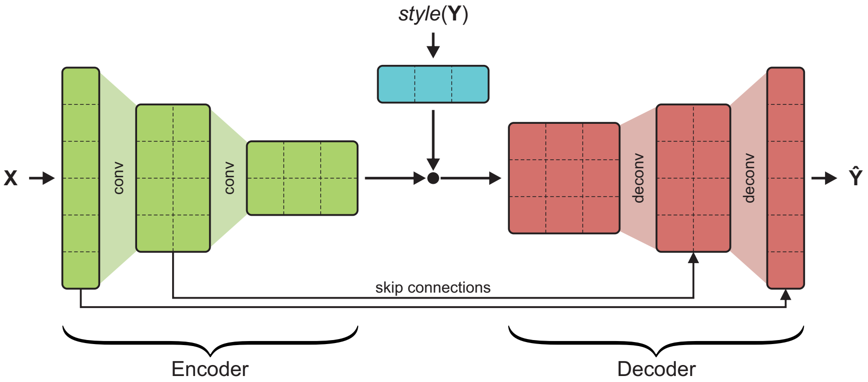
**Modifications:**

- Style extraction network
- Side input for Pix2Pix to inject style information
- Changes in training procedure to encourage style transfer
- (*Not shown*: Fine-tuning of depth and feature sizes)

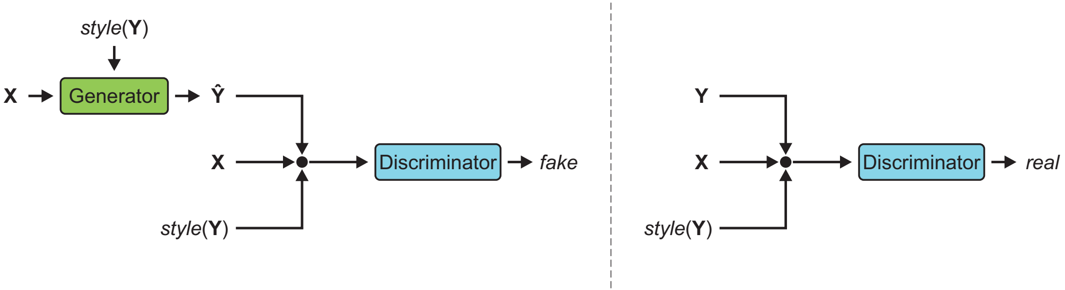
# Pen Style Transfer: Style Extraction



# Pen Style Transfer: Style Injection into Generator Network



# Pen Style Transfer: Conditional GAN Training



# Pen Style Transfer: Results

| Style Input | Output |
|---|---|
| closed suddenly | sinking |
| Lines, Triangles | sinking |
| the deep and dark | sinking |
| mich in Fesseln | sinking |
| a rebel's whore | sinking |

# Conclusion & Future Work

# Conclusion

**Goals of this thesis:**

- Full offline-to-offline handwriting style transfer algorithm
  $\Rightarrow$ Success
- Finding a robust algorithm for handwriting skeletonization
  $\Rightarrow$ Success
- Is an offline to online handwriting conversion sufficient to use online algorithms?
  $\Rightarrow$ It is!
- Finding a way to transfer the pen style to the output image
  $\Rightarrow$ Success
- *Optional:* Finding a way to transfer the background style to the output image
  $\Rightarrow$ Started, some success, but not shown here

# Future Work

- Further work on background style transfer
- Combined skeletonization and sampling
- Direct offline-to-offline style transfer

Thanks for your attention!

Thanks for your attention!.

Thanks for your attention!

# Appendix: Pix2Pix Knowledge Transfer Step 1


(a) real input image


(b) primitive skeletonization


(c) trained network result

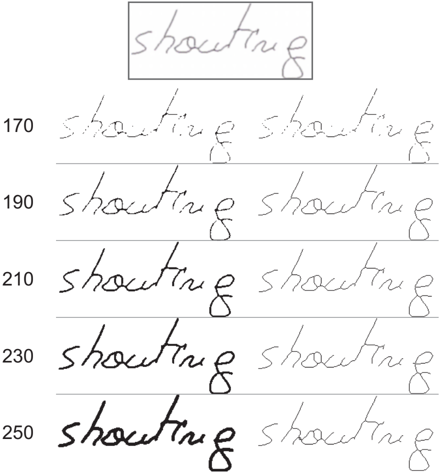
# Appendix: Pix2Pix Knowledge Transfer Step 2

(a) real skeleton

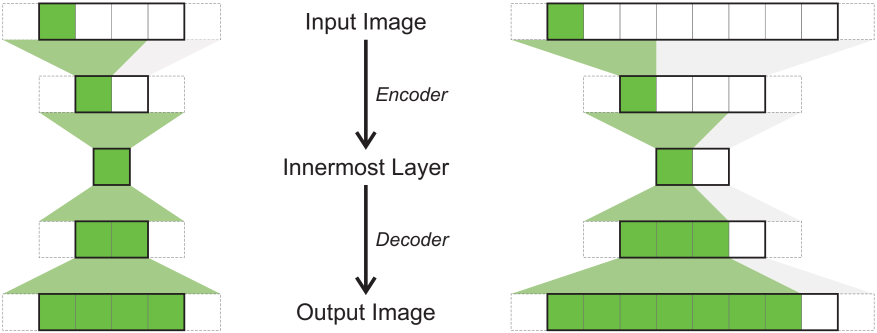
(b) synthetic image

(c) trained network result

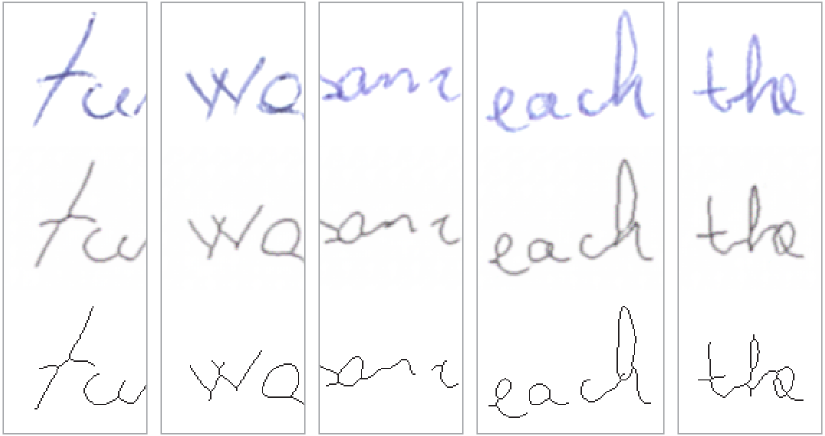
# Appendix: Thresholding Hyperparameter Study



# Appendix: Skeletonization of Variable-Sized Images

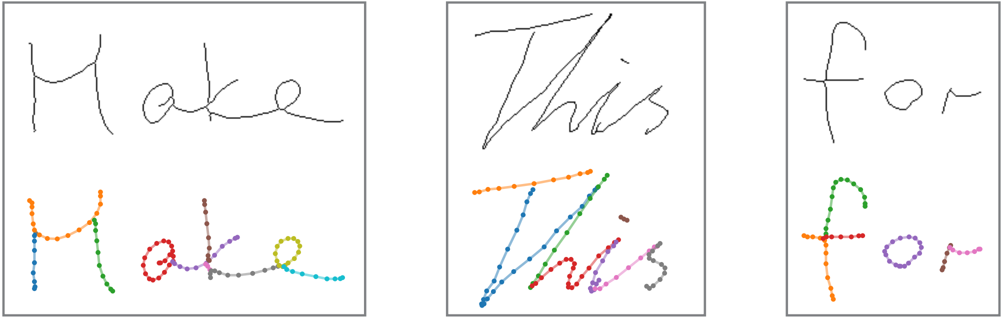


# Appendix: Existing Problems with Skeletonizaton



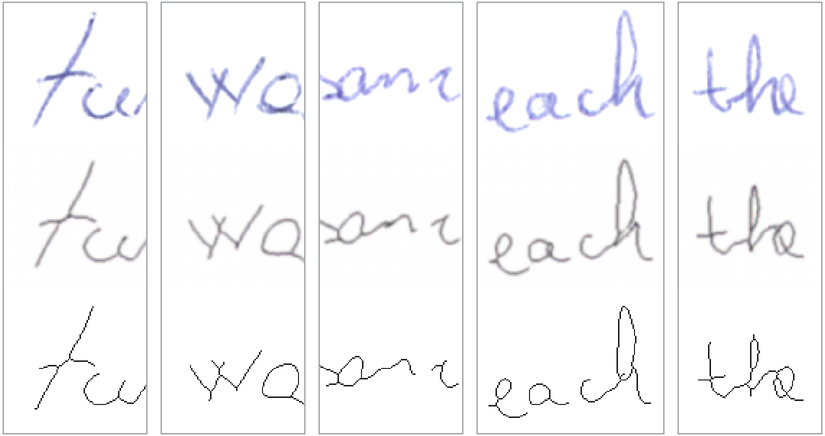
(a) (b) (c) (d) (e)

# Appendix: Existing Problems with Online Conversion



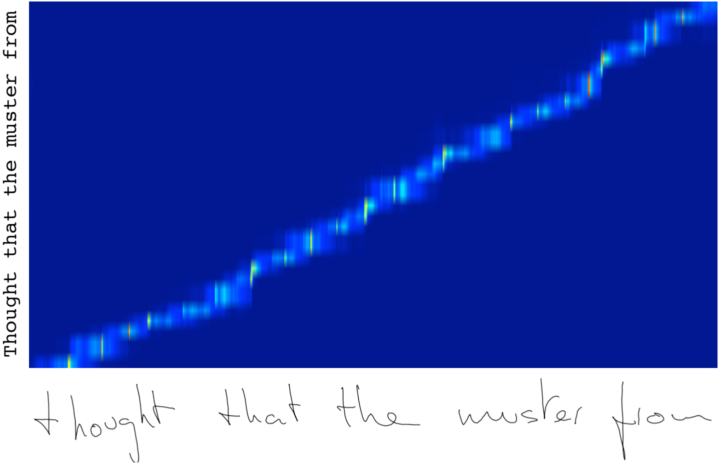
(a) (b) (c)

# Appendix: Existing Problems with the Online Conversion



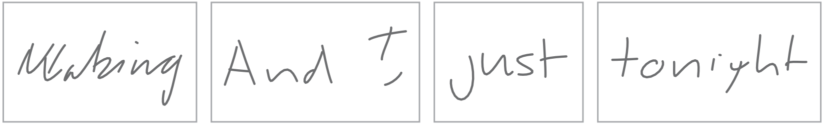
(a) (b) (c) (d) (e)

# Appendix: Graves' Attention Window




Alex Graves. "Generating Sequences With Recurrent Neural Networks". In: *CoRR* abs/1308.0850 (2013). arXiv: 1308.0850.

# Appendix: Existing Problems with Writer Style Transfer




(a) (b) (c) (d)

# Appendix: Pen Style Transfer Error Correction Properties


(a) skeleton input

(b) network output

# Appendix: Background Style Transfer - Graves vs SPADE

Style Input

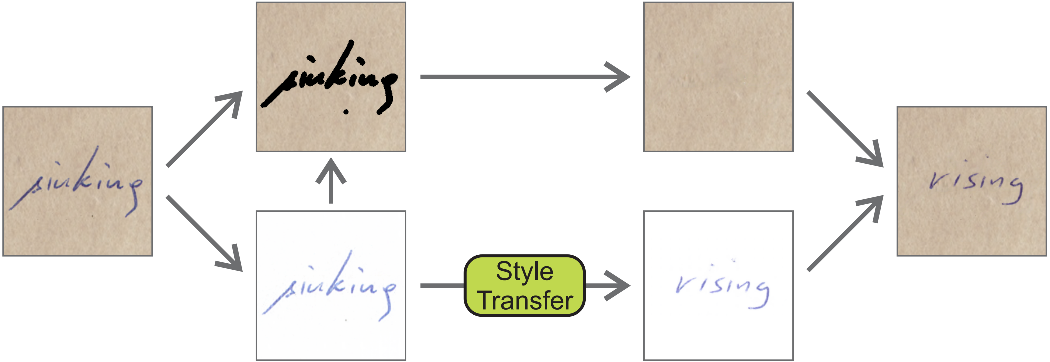
Graves

SPADE

Alex Graves. “Generating Sequences With Recurrent Neural Networks”. In: *CoRR* abs/1308.0850 (2013). arXiv: `1308.0850`.

Taesung Park et al. “Semantic Image Synthesis with Spatially-Adaptive Normalization”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.

# Appendix: Background Style Transfer - Pipeline



# References

[1] Emre Aksan, Fabrizio Pece, and Otmar Hilliges. “DeepWriting: Making Digital Ink Editable via Deep Generative Modeling”. In: *SIGCHI Conference on Human Factors in Computing Systems*. CHI ’18. Montréal, Canada: ACM, 2018.

[2] Alex Graves. “Generating Sequences With Recurrent Neural Networks”. In: *CoRR* abs/1308.0850 (2013). arXiv: `1308.0850`.

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. “Image-to-Image Translation with Conditional Adversarial Networks”. In: *arxiv* (2016).

[4] Florian Kleber, Stefan Fiel, Markus Diem, and Robert Sablatnig. “CVL-DataBase: An Off-Line Database for Writer Retrieval, Writer Identification and Word Spotting”. In: *Proceedings of the 2013 12th International Conference on Document Analysis and Recognition*. ICDAR ’13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 560–564. ISBN: 978-0-7695-4999-6. DOI: `10.1109/ICDAR.2013.117`.

[5] Marcus Liwicki and Horst Bunke. “IAM-OnDB - an On-Line English Sentence Database Acquired from Handwritten Text on a Whiteboard”. In: *8th Intl. Conf. on Document Analysis and Recognition*. Vol. 2. 2005, pp. 956–961.

[6] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. “Semantic Image Synthesis with Spatially-Adaptive Normalization”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.

[7] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks”. In: *Computer Vision (ICCV), 2017 IEEE International Conference on*. 2017.