MUNI
FI

Application of Superresolution Models in Optical Character Recognition of Czech Medieval Texts

Mikuláš Bankovič **Vít Novotný** **Petr Sojka**
{456421,witiko}@mail.muni.cz,
sojka@fi.muni.cz

Faculty of Informatics, Masaryk University

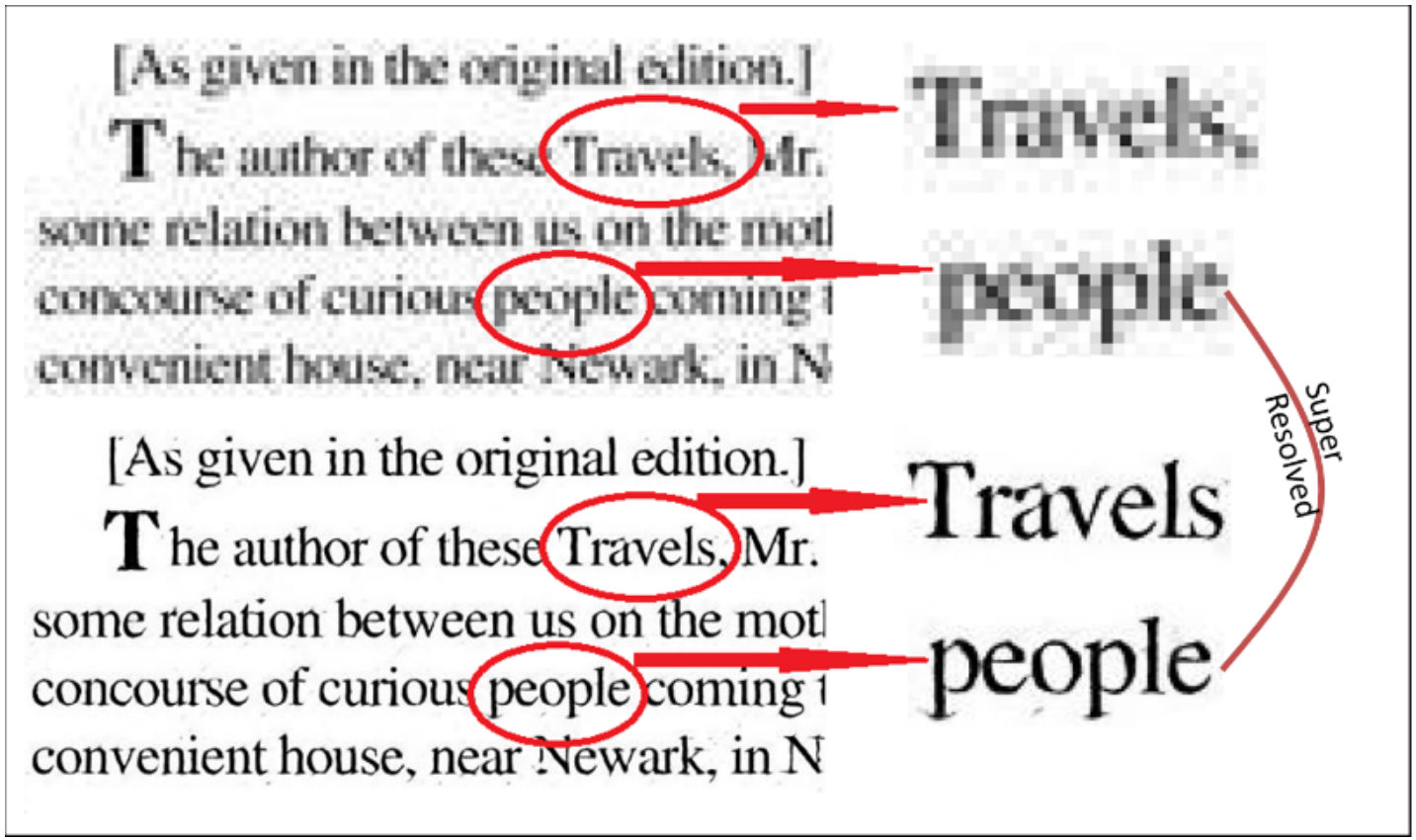
December 10, 2021

Introduction

- Given by Lat and Jawahar [1], accuracy[1] of Tesseract on English documents dataset can be increased by more than 20% with SRGAN model.
- Can we use it in AHISTO project and increase accuracy of Tesseract on scans historical documents?

[1] Levenshtein distance between ground truth document and generated document

Expectations

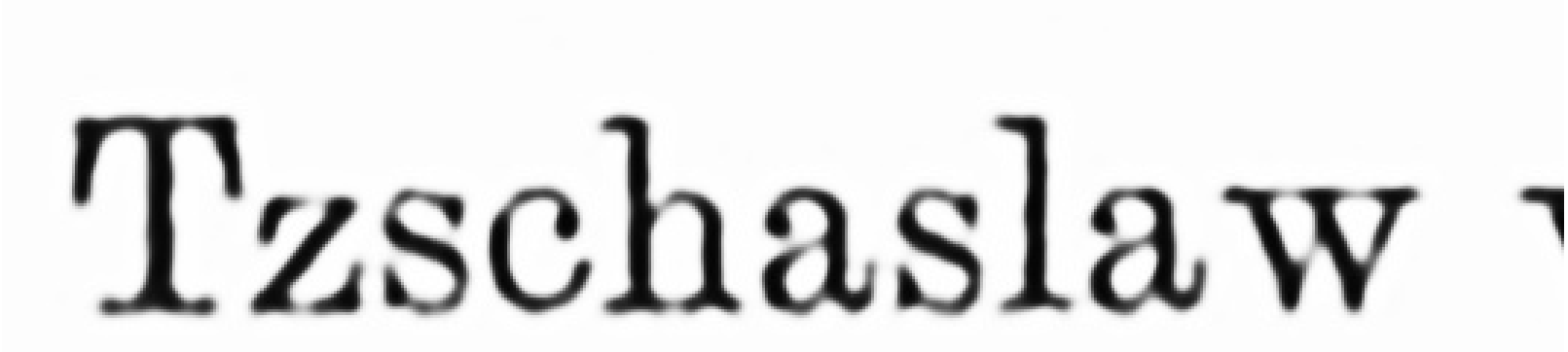


Baselines

- Bilinear interpolation
- Potrace - vectorization
- AHISTO historical documents: low and high-res pairs (134 annotated pages)

AHISTO project example

(a) Low resolution image patch



(b) High resolution image patch

Super-resolution models

SR Model - Waifu

- Convolution neural network (CNN) trained on collection of anime images augmented with JPEG noise.

(a) Waifu low resolution image

(b) Waifu high resolution resolution

SR Model - SRGAN [2]

- Generative adversarial network (GAN) - discriminator and generator.
- Extracted generator part of network is called SRResNet.

Datasets

- Automatically latex typesetted ground truth text from AHISTO documents - columns, fonts, spaces, text location.
- AHISTO scans - some are damaged, even high-res is still scan with potential problems.
- External typeset book with source code: Codex Diplomaticus Bohemiae Tomus VI.

Data Augmentations

- Salt & pepper adds random black and white pixels simulating unwanted dots.
- JPEG noise is inspired by waifu model.
- Rotate corrects slight rotations of scans.

Data Augmentations - Examples

(a) Original image

(b) Image with JPEG noise

(c) Rotated image

(d) Salted and peppered image

Figure: Data augmentations of low-resolution images

Results

Figure: Low-resolution image vs. high-resolution image

Results

Figure: Bilinear interpolation image vs. high-resolution image

Results

Figure: Waifu2x image vs. high-resolution image

Results

Figure: SRResNet without any modifications vs. high-resolution image

Results

Figure: SRResNet image rotated by angle 2° vs. high-resolution image

Results

Table: Impact of super-resolution on ocr accuracy. Best results are bold.

Architecture	Modification	Epochs	wer (%)
Low-resolution			14.75
Bilinear			7.77
Potrace			9.29
High-resolution			8.74
SRGAN		20 + 1	9.63
SRResNet		20	8.95
SRResNet	binarize	20	9.72
SRResNet	grayscale	20	8.79
SRResNet	rotate 2°	20	8.19
SRResNet	rotate 2° + greyscale	20	8.32
Waifu2x			7.46
Waifu2x	jpeg noise		**7.45**

Conclusion

- The resolution of the image matters for the Tesseract OCR engine.
- Even bilinear interpolated images work significantly better than original low-resolution images.
- The gray-scaling of weights can be used to decrease the size and training time of image super-resolution models without any adverse effect on OCR accuracy.
- The victory of the Waifu2x models, which were pre-trained on data from different domain, shows that the size of our training dataset was insufficient for training larger models such as SRGAN and SRResNet.
- Salt and pepper augmentation did not reflect real scan damage.

Future Work

- We should collect more training data, for example by typesetting the OCR texts produced from scanned document pages.
- We should focus at more realistic damaged scan augmentations, such as modified salt and pepper resembling ink droplets and blank spots after ink has peeled off the paper and flaked away.

Future Work

- The future work should also experiment with modified loss functions: such as weighted MSE loss function, that has higher weight at the edges of letters.
- Su et al. [5] showed that adding ℓ_1 loss to the SRGAN model helps maintain detail in letter forms.
- Ray et al. [4] and Randika et al. [3] added the gradient loss of the OCR algorithm to the image super-resolution model, creating an end-to-end deep learning framework.

Summary

- We applied Super-Resolution models and baselines on image before OCR engine.
- We trained our own Super-Resolution model to remove different kind of noise from scan images and smooth the letter forms.
- We experimented with different Super-Resolution modifications like gray-scale weights.
- We are planning to expand Super-Resolution loss function and connect it to the gradient flow or loss function in OCR engine and experiment with tailor-made loss functions for recovering letter forms.

Bibliography I

[1] Ankit Lat and C. V. Jawahar. “Enhancing OCR Accuracy with Super Resolution”. In: *2018 24th International Conference on Pattern Recognition (ICPR)*. 2018, pp. 3162–3167. doi: 10.1109/ICPR.2018.8545609.

[2] Christian Ledig et al. “Photo-realistic single image super-resolution using a generative adversarial network”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4681–4690. url: https://arxiv.org/abs/1609.04802v5.

[3] Ayantha Randika et al. *Unknown-box Approximation to Improve Optical Character Recognition Performance*. [cit. 2021-11-03]. url: https://arxiv.org/abs/2105.07983v1 (visited on 11/03/2021).

Bibliography II

[4] Anupama Ray et al. “An End-to-End Trainable Framework for Joint Optimization of Document Enhancement and Recognition”. In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. 2019, pp. 59–64. doi: 10.1109/ICDAR.2019.00019.

[5] Xiangdong Su et al. “Improving Text Image Resolution using a Deep Generative Adversarial Network for Optical Character Recognition”. In: *2019 International Conference on Document Analysis and Recognition (ICDAR)*. 2019, pp. 1193–1199. doi: 10.1109/ICDAR.2019.00193.

MUNI
FACULTY
OF INFORMATICS