Application of Super-Resolution Models in Optical Character Recognition of Czech Medieval Texts

Mikuláš Bankovič, Vít Novotný, and Petr Sojka

Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic
{456421,witiko}@mail.muni.cz, sojka@fi.muni.cz
https://mir.fi.muni.cz/

Abstract. Optical character recognition (OCR) of scanned images is used in multiple applications in numerous domains and several frameworks and OCR algorithms are publicly available. However, some domains such as medieval texts suffer from low accuracy, mainly due to low resources and poor quality data. For such domains, preprocessing techniques help to increase the accuracy of OCR algorithms.
In this paper, we experiment with two super-resolution models: Waifu2x and SRGAN. We use the models to reduce noise and increase the image resolution of scanned medieval texts. We evaluate the models on the AHISTO project dataset and compare them against several baselines. We show that our models produce improvements in OCR accuracy.

Keywords: Super-resolution · Optical character recognition · Medieval texts

1 Introduction

The aim of the AHISTO project is to make documents from the Hussite era (1419–1436) available to the general public through a web-hosted searchable database. Although scanned images of letterpress reprints from the 19th and 20th century are available, accurate optical character recognition (OCR) algorithms are required to extract searchable text from the scanned images. However, the scanned images are noisy and low-resolution, which decreases OCR accuracy.

In our work, we develop image super-resolution models and data augmentation techniques for training these models. We use our image super-resolution models to increase the resolution of scanned pages and we evaluate the impact on the OCR accuracy on medieval texts.

In Section 2, we describe the related work in image super-resolution and the OCR of medieval texts. In Section 3, we describe our training and test datasets, data augmentation techniques, image super-resolution models, and baselines. In Section 4, we discuss the results of our evaluation. We conclude in Section 5 and offer directions for future work.

A. Horák, P. Rychlý, A. Rambousek (eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2021, pp. 11–18, 2021.

2 Related Work

Traditionally, the image processing techniques that improve the accuracy for OCR of medieval texts documents were primarily rule-based [3]. However, there has been a growing interest in using deep learning methods for OCR preprocessing.

In this section, we will present the recent work in deep learning methods for OCR preprocessing and in the OCR of medieval texts.

2.1 Super-Resolution Models

Walha et al. (2012) [17] showed that image super-resolution models based on learned dictionaries between low-resolution and high-resolution sparsely encoded patches improved performance on image upscaling. However, the computation demands for this algorithm were high. Nayef et al. (2014) [8] proposed selective patch processing, performing costly operations only on high variance patches and using bicubic interpolation otherwise.

As in many other domains, deep learning models that used convolution neural networks (CNNS) surpassed previous techniques for image super-resolution. These models included SRCNN [1] and more complex generative adversarial networks (GANS) such as SRGAN [6]. Nakao et al. (2019) [7] adapted the SRCNN loss function for text by maintaing sharp boundaries between letters.

Lat and Jawahar (2018) [5] used SRGAN to improve OCR accuracy. Su et al. (2019) [15] showed that adding ℓ_1 loss to the SRGAN model helps maintain detail in letterforms. Nguyen et al. (2019) [9] translated poorly visible letters to binarised letters using a variation of SRGAN and a weakly coupled dataset.

Fu et al. (2019) [2] suggested using cascaded networks consisting of CNN, improving OCR accuracy over both SRCNN and SRGAN. Ray et al. (2019) [12] and Randika et al. (2021) [11] added the gradient loss of the OCR algorithm to the image super-resolution model, creating an end-to-end deep learning framework.

2.2 Optical Character Recognition Engines

In 2020, the second author [10] showed that Tesseract 4 [4] gave the best trade-off between speed and OCR accuracy for medieval texts. Therefore, we only experiment with Tesseract 4 in our work.

3 Methods

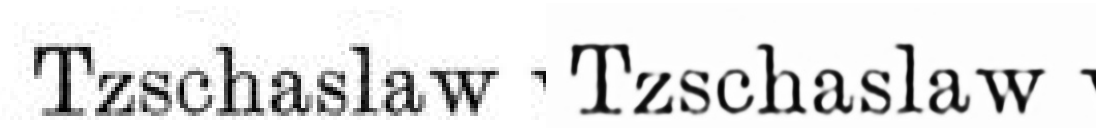
In this section, we discuss our training and test datasets, the data augmentations we used, and our super-resolution models and baselines.

3.1 Datasets

As our training dataset for the image super-resolution models, we used a born-digital PDF version of the sixth tome of the book *Codex Diplomaticus et Epistolaris*

Regni Bohemiae [16], which contains a collection of medieval texts (1278–1283) from the Kingdom of Bohemia.

As our test dataset for the OCR accuracy, we used the AHISTO dataset. The dataset contains 65,348 pairs of low-resolution and high-resolution scanned images [10, Section 3.1], see Figure 1. For 120 scanned images, the dataset contains human annotations with correct OCR texts. We used the human annotations with the word error rate (WER) measure [14] to evaluate the OCR accuracy. See another article from these proceedings on page 29 for more information about the human annotations and the WER evaluation measure.



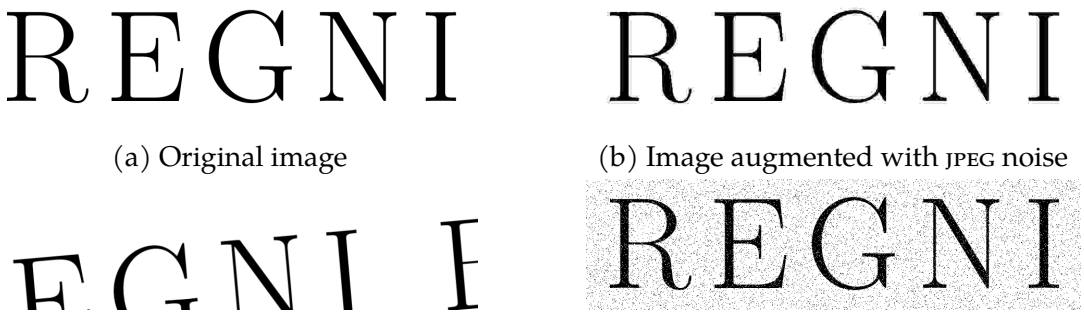
(a) Low resolution image patch (b) High resolution image patch

Fig. 1: Low-resolution and high-resolution image patches from our test dataset

3.2 Data Augmentations

We augmented images as shown in Figure 2 with the following methods:

- *Rotate* rotates by an angle, blank spaces are filled using bicubic interpolation.
- *JPEG noise* recompresses image to JPEG quality.
- *Salt and pepper* adds random black and white pixels to the image.



(a) Original image (b) Image augmented with JPEG noise

(c) Rotated image (d) Salted and peppered image

Fig. 2: Data augmentations of low-resolution images

3.3 Super-Resolution Models

For image super-resolution, we use the SRGAN and SRCNN models.

SRGAN has multiple hyperparameters to optimise: the number of epochs, the learning rate, and the size of the image patches. We augment SRGAN to work with greyscale weights, reducing the number of parameters approximately by a factor of 3. We also experiment with removing the discriminator part of an SRGAN network (further known as SRRESNET) [6]. Our code is available online.[1]

Due to time constraints, we do not train our own SRCNN model. Instead, we use public models[2] (*Waifu2x*) pre-trained on drawn images,[3] see Figure 3.

(a) Low resolution image (b) High resolution image

Fig. 3: Low-resolution and high-resolution images from the training dataset of pre-trained Waifu2x models. The image is licensed under CC BY-NC by piapro.

3.4 Baselines

As our baselines, we used the original low-resolution and high-resolution image pairs. Additionally, we also used the bilinear interpolation and the Potrace [13] rule-based image vectorizer to upscale the low-resolution images.

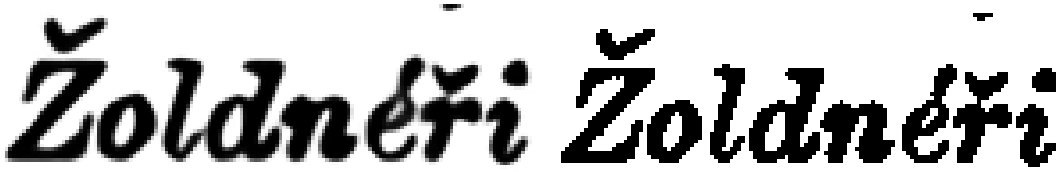
4 Results

Table 1 shows that high-resolution images have better performance than low-resolution images. Specific settings performed even better than original high-resolution images, which is unexpected in the case of bilinear interpolation baseline. Waifu2x with added JPEG noise achieved the best performance.

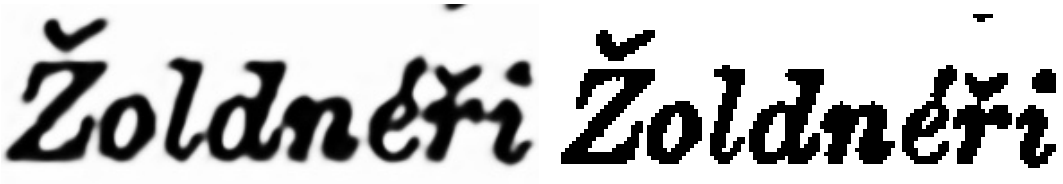
[1] https://github.com/xbankov/Fast-SRGAN
[2] https://github.com/nagadomi/waifu2x/tree/master/models/cunet/art
[3] https://github.com/nagadomi/waifu2x/issues/263

(a) Low-resolution image vs. high-resolution image

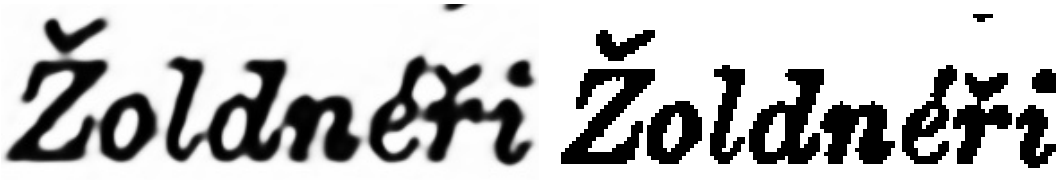
(b) Bilinear interpolation image vs. high-resolution image



(c) Waifu2x image vs. high-resolution image



(d) SRRESNET without any modifications vs. high-resolution image



(e) SRRESNET image rotated by angle 2° vs. high-resolution image

Fig. 4: The low resolution image in Fig. 4a is an input to other methods. In each figure in the left is tested example and on the right of each figure is the same original high resolution scan.

Table 1: 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**

We observed that SRRESNET bested SRGAN in every setting. Therefore, we only list a single result for SRGAN in Table 1 for comparison. The grayscale variant performs comparably with RGB. Most of the augmentations did not perform well, either due to wrong parameters or inappropriate design.

SRRESNET in Figure 4d looks intuitively better than bilinear interpolation in Fig. 4b. It is unclear why bilinear performs better within the Tesseract framework. In Figure 4c created by Waifu2x, the letters are separated and smoothed. Therefore the best result in OCR performance is justified. In contrast, in 4d, the letters *ř* and *i* are connected and can mislead the OCR engine.

5 Conclusion and Future Work

In our work, we have experimented with data augmentation methods for SRGAN. We tested the impact of super-resolution models on the OCR of medieval texts. We concluded that the resolution of the image matters for the Tesseract OCR engine. Even bilinear interpolated images work significantly better than original low-resolution images.

We also showed that the grayscaling 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. Future work should collect more training data, for example by typesetting the OCR texts produced from scanned document pages.

We realised that our salt and pepper augmentation did not reflect real scan damage. Future work 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 should also experiment with modified loss functions that improve the performance of image super-resolution techniques with text.

Acknowledgements. The South Moravian Centre graciously funded the second author's work for International Mobility as a part of the Brno PhD. Talent project. The research was also supported by TAČR Éta, project number TL03000365.

References

1. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014. pp. 184–199 (2014)
2. Fu, Z., Kong, Y., Zheng, Y., Ye, H., Hu, W., Yang, J., He, L.: Cascaded detail-preserving networks for super-resolution of document images. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 240–245. IEEE Computer Society (2019). https://doi.org/10.1109/ICDAR.2019.00047
3. Jon M. Booth, J.G.: Optimizing OCR accuracy on older documents: A study of scan mode, file enhancement, and software product, https://www.govinfo.gov/media/WhitePaper-OptimizingOCRAccuracy.pdf, [cit. 2021-11-06]
4. Kay, A.: Tesseract: An Open-Source Optical Character Recognition Engine. Linux Journal **2007**(159), 2 (Jul 2007), https://dl.acm.org/doi/10.5555/1288165.1288167
5. Lat, A., Jawahar, C.V.: Enhancing OCR accuracy with super resolution. In: 2018 24th International Conference on Pattern Recognition (ICPR). pp. 3162–3167. IEEE (2018)
6. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., 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. pp. 4681–4690 (2017), https://arxiv.org/abs/1609.04802v5
7. Nakao, R., Iwana, B.K., Uchida, S.: Selective super-resolution for scene text images. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 401–406 (2019). https://doi.org/10.1109/ICDAR.2019.00071
8. Nayef, N., Chazalon, J., Gomez-Krämer, P., Ogier, J.M.: Efficient example-based super-resolution of single text images based on selective patch processing. In: 2014 11th IAPR International Workshop on Document Analysis Systems. pp. 227–231 (2014). https://doi.org/10.1109/DAS.2014.25
9. Nguyen, K.C., Nguyen, C.T., Hotta, S., Nakagawa, M.: A character attention generative adversarial network for degraded historical document restoration. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 420–425 (2019). https://doi.org/10.1109/ICDAR.2019.00074
10. Novotný, V.: When Tesseract Does It Alone. In: Proceedings of the Fourteenth Workshop on Recent Advances in Slavonic Natural Language Processing, RASLAN 2020. pp. 3–12 (2020), https://nlp.fi.muni.cz/raslan/2020/paper1.pdf
11. Randika, A., Ray, N., Xiao, X., Latimer, A.: Unknown-box approximation to improve optical character recognition performance, https://arxiv.org/abs/2105.07983v1, [cit. 2021-11-03]

12. Ray, A., Sharma, M., Upadhyay, A., Makwana, M., Chaudhury, S., Trivedi, A., Singh, A., Saini, A.: An end-to-end trainable framework for joint optimization of document enhancement and recognition. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 59–64 (2019). https://doi.org/10.1109/ICDAR.2019.00019
13. Selinger, P.: Potrace: a polygon-based tracing algorithm (2003), http://potrace.sourceforge.net/potrace.pdf, [cit. 2021-11-07]
14. Soukoreff, R.W., MacKenzie, I.S.: Measuring errors in text entry tasks: An application of the levenshtein string distance statistic. In: CHI'01 extended abstracts on Human factors in computing systems. pp. 319–320 (2001)
15. Su, X., Xu, H., Kang, Y., Hao, X., Gao, G., Zhang, Y.: Improving text image resolution using a deep generative adversarial network for optical character recognition. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 1193–1199 (2019). https://doi.org/10.1109/ICDAR.2019.00193
16. Sviták, Z., Krmíčková, H., Krejčíková, J., Friedrich, G.: Codex diplomaticus et epistolaris Regni Bohemiae. Tomi VI, fasciculus primus, Inde ab A. MCCLXXVIII usque ad A. MCCLXXXIII. Academia (2006)
17. Walha, R., Drira, F., Lebourgeois, F., Alimi, A.M.: Super-resolution of single text image by sparse representation. In: Proceeding of the Workshop on Document Analysis and Recognition. p. 22–29 (2012). https://doi.org/10.1145/2432553.2432558