Tire tread superresolution

Daniel Bohuněk

Czech Technical University - Faculty of Information Technology

bohundan@fit.cvut.cz

December 31, 2022

1 Introduction

This work expands upon the author's bachelor thesis [1], in which a camera system for scanning tire treads was devised. This work aims to create a model capable of upscaling tire tread images to improve their sharpness. The model should also compensate for bad focus of the camera when scanning treads from greater distances than anticipated (the camera system is designed to scan just one closest tire but is capable of segmenting other tires present in the scene).

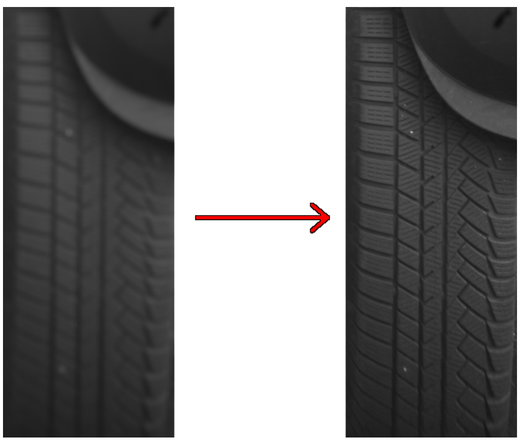


Figure 1: Goal of this work.

2 Training data

315 grayscale images of tire treads were created during the development and testing of the camera system. Postprocessing methods (gradient removal, histogram equalization) were implemented during the student summer research program (VýLet 2022). The equalized images can act as additional augmented data, but this might not be necessary, as the chosen model can learn to upscale accurately with less than a hundred images.

Of the 315 images, 59 high-resolution images were chosen as the training images. Nine other good-quality images plus five low-resolution images act as the testing data so that the model's accuracy can be measured.

3 Methods

SRCNN [2] model was chosen for its simplicity. It processes 64×64 subimages, so an image has to be split into so-called *patches* before being fed through the network.

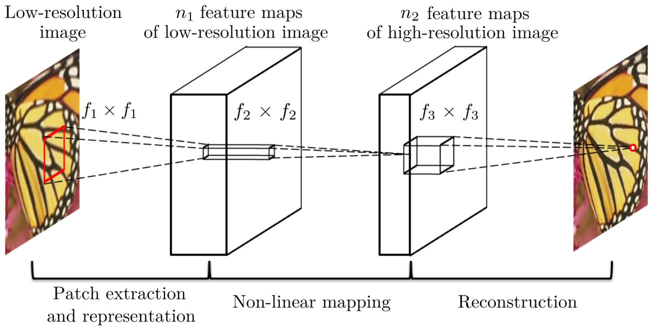


Figure 2: SRCNN pipeline [2].

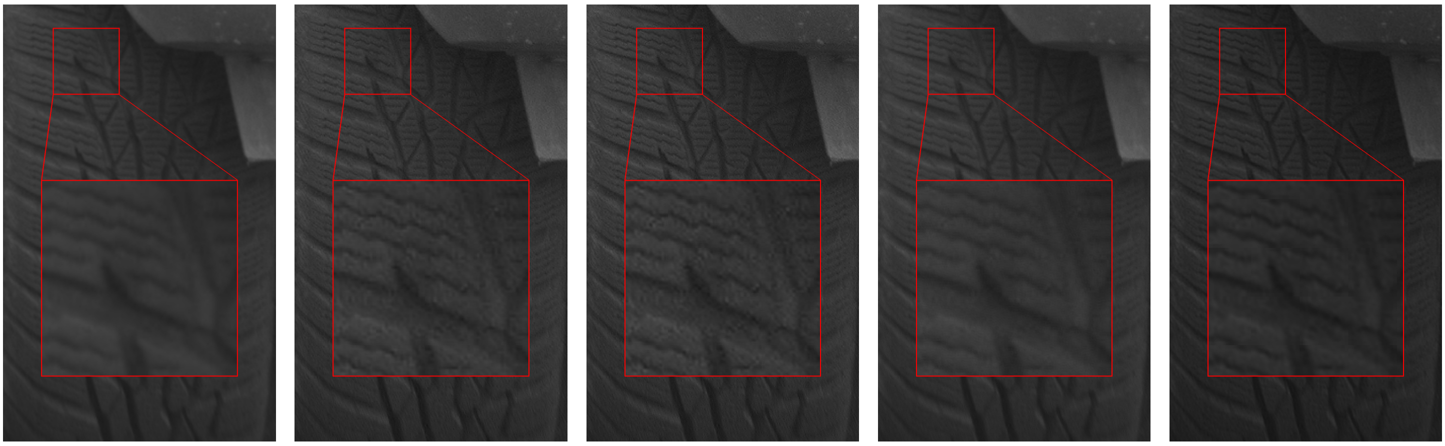
Because the tire tread scanning camera system works with grayscale images, the SRCNN model can process just one channel instead of three color channels. To measure the reconstruction accuracy, the peak signal-to-noise ratio (PSNR) can be used:

$$\mathrm{PSNR}(I_1, I_2) = 20 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I}{\sqrt{\mathrm{MSE}(I_1, I_2)}}\right),$$

where I_1 and I_2 are images, MAX_I is the maximum possible pixel value (usually 255), and $\mathrm{MSE}(I_1, I_2)$ is the mean square error of the two images.

4 Results

Splitting each training image into 64×64 patches with the stride of 4 yields a total of 9120 training patches. One training epoch on an RTX 3060 Mobile takes roughly 130 seconds to finish. The authors of the original paper trained their model with $8 \cdot 10^8$ backpropagations. Unfortunately, these many backpropagations are not doable with the available hardware. However, as the original authors noted, many backpropagations are not necessary to achieve good results.



(a) Original (b) 10 epochs (c) 250 epochs (d) 250 epochs with less downsampling (e) 50 epochs with dropout

Figure 3: Examples of upscaled images using trained SRCNN.

After just ten epochs, the upscaled images show promising detail. Going further to 250 epochs changes very little, but it does smooth out the rough edges caused by reconstructing the image from 64×64 patches. Fig. 3 shows this. It also indicates that upscaling introduces noise in the image. The amount of noise depends on the training data. When the downsampling of training images is toned down, the resulting model does not sharpen the image as much, but it also does not add so much noise, as seen in Fig. 3d. Another attempt to reduce noise was made by adding a dropout layer. Dropout works differently in convolutional layers than in fully connected layers [3]. While there is less noise in the upscaled image, it is also darker, as seen in Fig. 3e.

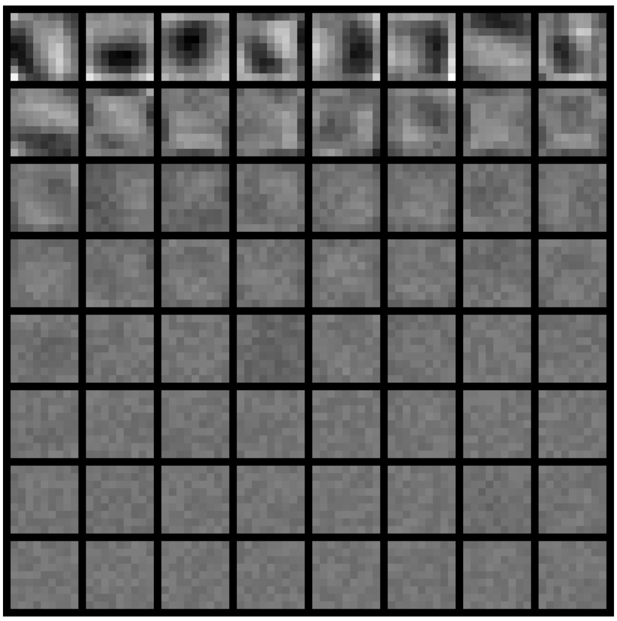


Figure 4: Learned filters of the first layer.

Fig. 4 contains learned filters of the first convolutional layer, sorted by their variance. Many of them seem to not be doing much. This aligns with a similar figure in the original SRCNN paper. Changing the amount of filters does not result in higher density, even using just 9 filters leads to 3 having high variance, while the rest have low variance.

Next, the accuracy of the trained SRCNN will be compared to SRGAN [4] and pretrained superresolution models implemented in OpenCV [5]:

- EDSR [6],
- ESPCN [7],
- FSRCNN [8],
- LapSRN [9].

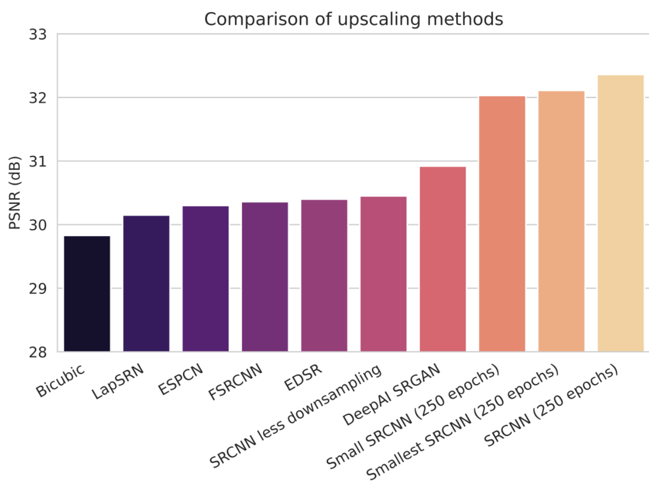


Figure 5: PSNR comparison of different models.

Fig. 5 shows the accuracy of different models on the testing images. The dropout model is not included in this figure, as it scored only 18.6 dB.

5 Conclusion

While the trained SRCNN is not perfect, as it does introduce some noise, it still provides better results than other models, which have been trained on a larger and more generalized dataset. Unfortunately, training more epochs to see if the noise would be reduced is not feasible on the available hardware (single laptop GPU). Online training services do exist but are either monetized or even slower at training than the RTX 3060 Mobile.

An image can be fed through the SRCNN multiple times (shown in Fig. 6). However, this further amplifies the added noise, as one would expect.

(a) Original (b) 4-times (c) 8-times

Figure 6: Multiple upsamplings.

Apart from more training and better training images (with higher resolution), other superresolution models could be evaluated in the future to determine the best one suited for this very narrow application of upscaling grayscale tire treads.

References

[1] D. Bohuněk, "Návrh kamerového systému pro snímání vzorku pneumatiky" [Design of a camera system for scanning a tire tread], *Bachelor thesis, FIT CTU*, May. 2022, Available: https://hdl.handle.net/10467/101628.

[2] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *Computer Vision and Pattern Recognition (cs.CV)*, Dec. 2014, doi: 10.48550/arXiv.1501.00092.

[3] J. Reinhold, "Dropout on convolutional layers is weird", *Towards Data Science*, Feb. 2019, Available: https://towardsdatascience.com/dropout-on-convolutional-layers-is-weird-5c6ab14f19b2.

[4] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", *Computer Vision and Pattern Recognition (cs.CV)*, Sep. 2016, doi: 10.48550/arXiv.1609.04802.

[5] OpenCV (Open Source Computer Vision Library) https://opencv.org/.

[6] B. Lim et al., "Enhanced Deep Residual Networks for Single Image Super-Resolution", *Computer Vision and Pattern Recognition (cs.CV)*, Jul. 2017, doi: 10.48550/arXiv.1707.02921.

[7] W. Shi et al., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", *Computer Vision and Pattern Recognition (cs.CV)*, Sep. 2016, doi: 10.48550/arXiv.1609.05158.

[8] C. Dong, C. C. Loy, X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network", *Computer Vision and Pattern Recognition (cs.CV)*, Aug. 2016, doi: 10.48550/arXiv.1608.00367.

[9] W. Lai, J. Huang, N. Ahuja and M. Yang, "Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks", *Computer Vision and Pattern Recognition (cs.CV)*, Oct. 2017, doi: 10.48550/arXiv.1710.01992.