Multi-Modal Super-Resolution with Deep Guided Filtering

Bernhard Stimpel[1,2], Christopher Syben[1,2], Franziska Schirrmacher[1], Philip Hoelter[2], Arnd Dörfler[2], Andreas Maier[1]

[1]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
[2]Department of Neuroradiology, University Clinics Erlangen
bernhard.stimpel@fau.de

Abstract. Despite the visually appealing results, most Deep Learning-based super-resolution approaches lack the comprehensibility that is required for medical applications.
We propose a modified version of the locally linear guided filter for the application of super-resolution in medical imaging. The guidance map itself is learned end-to-end from multi-modal inputs, while the actual data is only processed with known operators. This ensures comprehensibility of the results and simplifies the implementation of guarantees. We demonstrate the possibilities of our approach based on multi-modal MR and cross-modal CT and MR data. For both datasets, our approach performs clearly better than bicubic upsampling. For projection images, we achieve SSIMs of up to 0.99, while slice image data results in SSIMs of up to 0.98 for four-fold upsampling given an image of the respective other modality at full resolution. In addition, end-to-end learning of the guidance map considerably improves the quality of the results.

1 Introduction

Spatial resolution is subject to trade-offs in many medical imaging applications. For example in Magnetic Resonance Imaging (MRI), spatial resolution must be weighed against the signal-to-noise ratio and acquisition time. A retrospective increase in resolution by post-processing measures could alleviate this problem. To this end, a vast amount of super-resolution (SR) methods have been proposed and proven in the past [1]. In general, a differentiation can be made between single and multiple image SR methods. The latter is of particular interest for medical imaging, as non-existent information in one image can be derived from another image of the same patient. Especially in diagnostics, the presence of several scans of the same patient is common. In processing these data, Deep Learning (DL) has recently developed the state of the art in SR towards a previously unknown image quality [2]. Thereby, most learning-based methods apply high-dimensional non-linear transformations that are very difficult or impossible to comprehend. If only additional information is generated, e.g. in segmentation, the lack of comprehensibility can be tolerated, as blatant errors can be quickly identified. However, if the image is modified, as is the case with super-resolution, a failure of the method cannot be detected trivially. This is a limiting factor in medical applications where no less than the lives of patients are at stake. Despite these downsides, DL-based methods are highly promising if used in an appropriate way. In combination with well understood known operators, the advantages of DL can be combined with the necessary comprehensibility [3,4]. To transfer this to the task of super-resolution, we present the combination of the guided filter, which applies a local linear transformation to the input image, with a guidance map that is learnd end-to-end from multi-modal input.

2 Methods

2.1 The Guided Filter

First proposed by He et al. [5], the guided filter has been applied to a variety of tasks. Simply put, a given input image $\boldsymbol{I}$ is processed by incorporating structural information from a guidance image $\boldsymbol{G}$. The filtering operation in this case assumes a locally linear model between the guidance and the input image. In general, the guidance map can be any given image, even the input image itself. To fully leverage the power of the guided filter, a more appropriate guide is needed. Given multiple input images from the same object, a combination of these is beneficial. However, this raises the question of how the combined guidance map is composed.

2.2 End-to-End Trainable Guided Filter

Based on the wide range of possible applications of the guided filter as well as the ongoing success of Deep Learning, Wu et al. [6] incorporated the guided filter into a DL framework as a differentiable layer. This allows to backpropagate gradients through the filter to previous layers. We employ a convolutional neural network with the task to generate a guidance map for the guided filter based on the multi-modal input. Being able to train this generator in an end-to-end fashion enables for an optimal selection of features from all input modalities directly by the network.

The proposed pipeline consists of a guidance map generator network, for which we use a U-net-like architecture [7] with two separate encoding and a single decoding path, and the guided filtering layer.

Starting from two images, the low-resolution image $\boldsymbol{I}_{\mathrm{lr}}$, which is to be raised to higher resolution, and a higher-resolution image $\boldsymbol{L}_{\mathrm{hr}}$ that serves as a guide. First, the input image $\boldsymbol{I}_{\mathrm{lr}}$ is upsampled by bicubic interpolation to the desired output resolution as an initialization, further denoted as $\boldsymbol{I}_{\mathrm{up}}$. Second, $\boldsymbol{I}_{\mathrm{lr}}$ and $\boldsymbol{L}_{\mathrm{hr}}$ are fed into the generator network in order to extract the best possible combined representation $\boldsymbol{G}$. Finally, the learned guidance map $\boldsymbol{G}$ and the upsampled input $\boldsymbol{I}_{\mathrm{up}}$ are processed by the guided filter, resulting in the high-resolution output $\boldsymbol{I}_{\mathrm{hr}}$. By this, only (locally) linear processing steps are applied to the computed output image. A graphical representation of the pipeline is shown in Fig. 1.

Optimization is performed using a feature matching (FM) loss [8] based on the VGG-19 network [9].

3 Experiments

For evaluation, on the one hand multi-modal MRI data is used in the form of 8 tomographic T1 and T2 flair datasets with a spatial resolution of 256 x 256. On the other hand, 13 cone-beam MR and X-ray projections at a resolution of 512 x 512 are processed. All experiments are performed on clinical patient datasets provided by the Department of Neuroradiology, University Clinics Erlangen (MR: 1.5 T MAGNETOM Aera / CT: SOMATON Definition, Siemens Healthineers, Erlangen / Forchheim, Germany). Of each combination of modalities two corresponding patient dataset pairs are reserved for final testing. Image registration is done using 3D slicer [10]. The forward projections are taken from the work on hybrid MR/CT imaging by [11,4,12] and are generated using the CONRAD framework [13]. The low resolution images are artificially created by nearest neighbor downsampling by a factor of 4, resulting in a resolution of 64 x 64 for the tomographic and 128 x 128 for the projection data. For quantitative evaluation, we compute the mean squared error (MSE) and multi-scale structural similarity (MS-SSIM) measures. To avoid optimistic bias by the large homogeneous air regions, all background pixel are ignored for the evaluation metrics.

4 Results

The proposed approach was evaluated in comparison with bicubic upsampling and guided filtering using only the high-resolution image $\boldsymbol{L}_{\mathrm{hr}}$ as guidance. The results are presented in Tab. 1. Exemplary qualitative results with their respec-

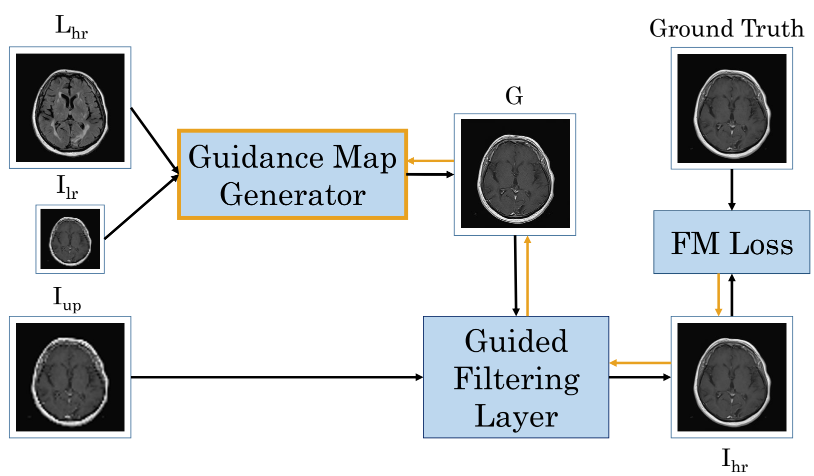


Fig. 1. The proposed guided filtering pipeline. Black arrows indicate the order of processing steps and orange arrows the gradient flow.

Table 1. Evaluation metrics of the proposed multi-modal guided filter

CT & MRI Projection Images ($128 \times 128 \rightarrow 512 \times 512$)

	Bicubic	GF w/o learned guidance	GF w/ learned guidance
MSE	0.0019 ± 0.001	0.0873 ± 0.0398	0.0005 ± 0.0001
MS-SSIM	0.97 ± 0.00	0.74 ± 0.04	0.99 ± 0.003

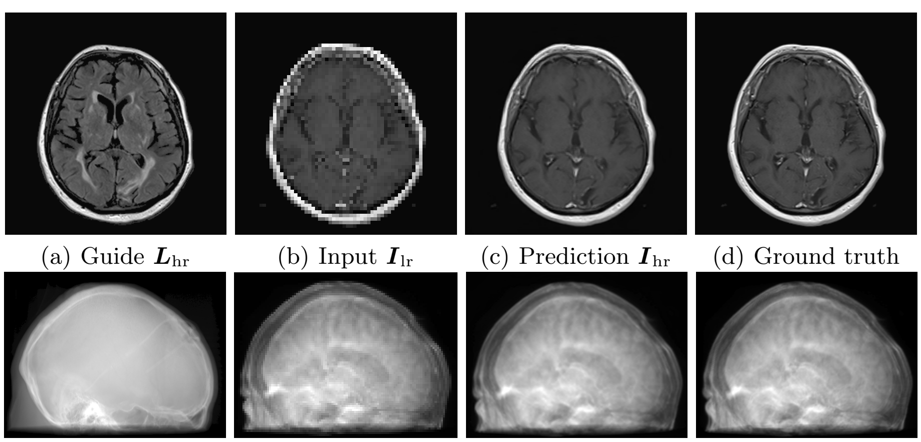
Tomographic T1 & T2 MRI images ($64 \times 64 \rightarrow 256 \times 256$)

	Bicubic	GF w/o learned guidance	GF w/ learned guidance
MSE	0.0997 ± 0.0526	0.1506 ± 0.0493	0.0138 ± 0.0077
MS-SSIM	0.89 ± 0.04	0.63 ± 0.13	0.98 ± 0.01

tive inputs are shown in Fig. 2. Furthermore, in Fig. 3 a region of interest can be seen to better observe the differences in the fine details.

5 Discussion

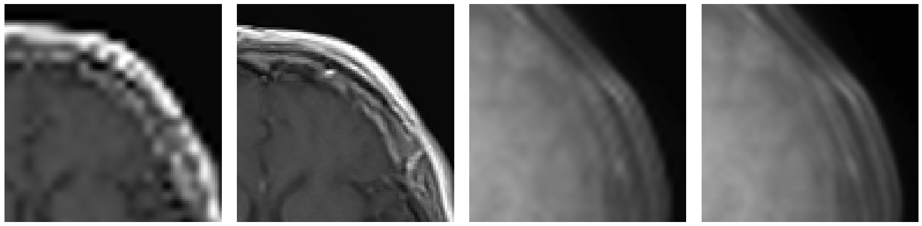
The quantitative and qualitative results show clear improvement of the proposed guided filtering pipeline compared to the reference method. Especially when observing the differences between the bicubic and the proposed upsampling method for the tomographic images in Fig. 3(a) and 3(b), respectively, the improved performance of the guided filter upsampling becomes apparent. The proposed framework captures fine details that are present in the guidance map which can not be estimated from the low-resolution input alone. Furthermore,



(a) Guide $\boldsymbol{L}_{\mathrm{hr}}$ (b) Input $\boldsymbol{I}_{\mathrm{lr}}$ (c) Prediction $\boldsymbol{I}_{\mathrm{hr}}$ (d) Ground truth

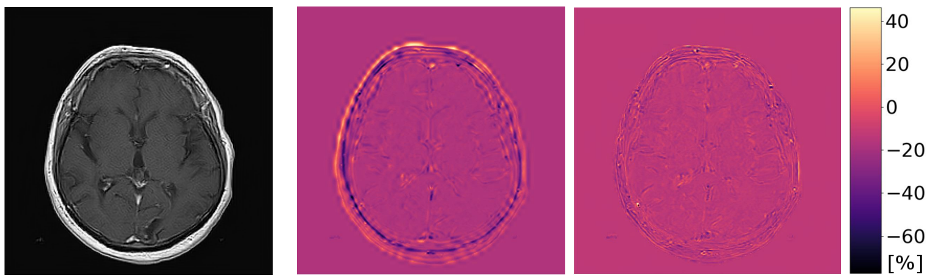
(e) Guide $\boldsymbol{L}_{\mathrm{hr}}$ (f) Input $\boldsymbol{I}_{\mathrm{lr}}$ (g) Prediction $\boldsymbol{I}_{\mathrm{hr}}$ (h) Ground truth

Fig. 2. Results of the guided filtering process. T1 & T2 MRI image pairs (a)-(d) and CT & MRI projection images (e)-(h).



(a) Bicubic (b) Ours (c) Bicubic (d) Ours

Fig. 3. Comparison of the proposed GF results with bicubic upsampled images.



(a) Guidance map (b) Bicubic (c) Ours

Fig. 4. An exemplary guidance map (a). Difference maps of the bicubic (b) and guided filtering (c) upsampled images w.r.t the ground truth.

the end-to-end learned guidance map clearly benefits the processing, as indicated in Table 1. This comes with the additional advantage that the optimal guidance map can be learned individually for each task and each combination of inputs. The learned guidance maps are already close to the desired high-resolution output images (see Fig. 4(a)). However, due to the high-dimensional transforms applied in the computation of these, the required comprehensibility is not given. In contrast, when only used as guidance, the modifications to the input images can be reduced to locally linear operations.

For future work, the proposed pipeline needs to be evaluated more thoroughly against a variety of comparable methods. In addition, we want to compare our method with state-of-the-art deep learning super-resolution methods, although these are not in line with our fundamental considerations regarding comprehensibility of the results. Furthermore, we would like to apply the proposed approach to other tasks that can be addressed by the guided filter, e.g., denoising.

6 Conclusion

We presented a guided filtering pipeline for multi-modal medical image super-resolution. The proposed approach has two key points. First, it solves the problem of the unknown best combination of multi-modal inputs by learning a task-optimal guidance map directly from the data. Second, the actual data is only processed with known operators, which ensures comprehensibility of the results and simplifies the implementation of guarantees. The achieved results closely resemble the ground truth data which is substantiated by the low error and high structural similarity measures.

Acknowledgment

This work has been supported by the project P3-Stroke, an EIT Health innovation project. EIT Health is supported by EIT, a body of the European Union. Additional financial support for this project was granted by the Emerging Fields Initiative (EFI) of the Friedrich-Alexander University Erlangen-Nürnberg (FAU). Furthermore, we thank the NVIDIA Corporation for their hardware donation.

References

1. Köhler T. Multi-Frame Super-Resolution Reconstruction with Applications to Medical Imaging. Friedrich-Alexander-Universität Erlangen-Nürnberg; 2018.
2. Wang Y, Perazzi F, McWilliams B, Sorkine-Hornung A, Sorkine-Hornung O, Schroers C. A Fully Progressive Approach to Single-Image Super-Resolution. In: IEEE Conf. Comput. Vis. Pattern Recognit. Work.; 2018. p. 864–873.
3. Maier A, Schebesch F, Syben C, Würfl T, Steidl S, Choi JH, et al. Precision Learning: Towards Use of Known Operators in Neural Networks. In: Int. Conf. Pattern Recognit. 2018; 2017. Available from: http://arxiv.org/abs/1712.00374.
4. Syben C, Stimpel B, Lommen J, Würfl T, Dörfler A, Maier A. Deriving Neural Network Architectures using Precision Learning: Parallel-to-fan beam Conversion. Proc Pattern Recognition, 40th Ger Conf. 2018;.
5. He K, Sun J, Tang X. Guided image filtering. IEEE Trans Pattern Anal Mach Intell. 2013;35(6):1397–1409.
6. Wu H, Zheng S, Zhang J, Huang K. Fast End-to-End Trainable Guided Filter. In: IEEE Conf. Comput. Vis. Pattern Recognit.; 2018. .
7. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015. 2015; p. 234–241.
8. Johnson J, Alahi A, Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. arXiv:160308155. 2016;.
9. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:14091556. 2015;.
10. Pieper S, Halle M, Kikinis R. 3D Slicer. In: 2nd IEEE Int. Symp. Biomed. Imaging Macro to Nano. vol. 2. IEEE; 2004. p. 632–635.
11. Stimpel B, Syben C, Würfl T, Mentl K, Dörfler A, Maier A. MR to X-Ray Projection Image Synthesis. In: Proc. Fifth Int. Conf. Image Form. X-Ray Comput. Tomogr.; 2017. .
12. Lommen JM, Syben C, Stimpel B, Bayer S, Nagel AM, Fahrig R, et al. MR-projection imaging with perspective distortion as in X-ray fluoroscopy for interventional X/MR-hybrid applications. In: Proc. 12th Interv. MRI Symp.; 2018. p. 54.
13. Maier A, Hofmann HG, Berger M, Fischer P, Schwemmer C, Wu H, et al. CONRAD - A software framework for cone-beam imaging in radiology. Med Phys. 2013;40(11).