Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Task-Aware-Downscaling
Improving Super Resolution and Colorization in Image and Video Domain

Semester Project

Advisor: Dr. Radu Timofte, Shuhang Gu
Supervisor: Prof. Dr. Luc van Gool
Computer Vision Laboratory, ITET ETH

May 21, 2019

Abstract

The abstract gives a concise overview of the work you have done. The reader shall be able to decide whether the work which has been done is interesting for him by reading the abstract. Provide a brief account on the following questions:

- What is the problem you worked on? (Introduction)
- How did you tackle the problem? (Materials and Methods)
- What were your results and findings? (Results)
- Why are your findings significant? (Conclusion)

The abstract should approximately cover half of a page, and does generally not contain citations.

Acknowledgements

Contents

List of Figures

1 Introduction

With the rise of deep learning in image processing super-resolution (SR) and image colorization (IC) in both the image and the video domain have received significant attention [20]. While SR aims to reconstruct a high-resolution (HR) image from a low-resolution (LR) image, image colorization deals with the transformation from an uncolored, grayscale (GR) image to a RGB colored (COL) image. However, in most of the recent works (e.g. [19], [18], [8], [17]) the problem of downscaling and upscaling or decolorization and colorization are regarded as seperate problems although upscaling often is preceded by downscaling, leading to a loss of information from the downscaling process which makes the inverse problem of SR highly ill-posed [9]. Despite of the large progress in SR in the last years ([20]) very specific details therefore often cannot be reconstructed, when interpolation is used for downsampling. However, as shown in Fig. 1 the downsampling method has a large impact on the performance of the subsequent upscaling task.

Figure 1: Comparison between an upscaled image based on bicubic downsampled (left) and task-aware downsampled (right) LR image applied on the same model with upscaling factor 4, for the image-domain, SET 14 dataset, and video domain, CALENDAR dataset.

As can be seen above a task-aware approach can dramatically improve the performance of existing super-resolution models. However, the research on task-aware downscaling methods is a very new field and therefore there still are a lot of unresolved issues such as the effect of noise or the feasibility of applying it in other domains.

1.1 Focus of this Work

For this reason this work focuses on Task-Aware-Downscaling (TAD) for several standard computer vision problems such as super-resolution or colorization in both the image and video domain, as recently purposed by Heewon Kim et. alt. ([9]) for the image domain only. Therefore the goals of this work are the following

- reimplement and evaluate the TAD framework purposed in ([9])
- improve the TAD framework especially with regards on accuracy (PSNR) and speed in the image domain
- evaluate the effect of external effects such as noise on the TAD framework
- extend the TAD framework to the video domain

By that to the best of our knowledge this work is the first one using deep learning for downscaling in the video domain.

1.2 Thesis Organization

After the problem statement Chapter 1 related works are introduced for both the image and video domain Chapter 2. Chapter 3 explains the methods that are used in order to achieve the goals described above and which are evaluated in Chapter 4. A final discussion of the results as well as an outlook on further work can be found in Chapter 5 and Chapter 6. Further visualization and experiments are shown in the abstract.

2 Related Work

In the following previous work in super-resolution, colorization and task-aware-downscaling are presented. At the end of each section the models used for comparison and evaluation of the the underlying approach are further explained in detail. Thereby the models were selected based on several criterias performance compared to the state-of-the-art, the use as benchmark in related papers and availability of (pretrained) models.

2.1 Super-Resolution in Image Domain

The problem of SR in the image domain is called Single-Image-Super-Resolution (SISR) and is shown in Fig. 2. A lot of approaches have been tried in order to cope with the SISR problem. While early approaches such as bicubic and Lanczos [5] tackle the problem using simple deterministic filters which are computational cheap but produce blurry results and lack in high frequency details, more recent approaches approach the problem using example-based methods such as sparse encoding or deep learning methods.

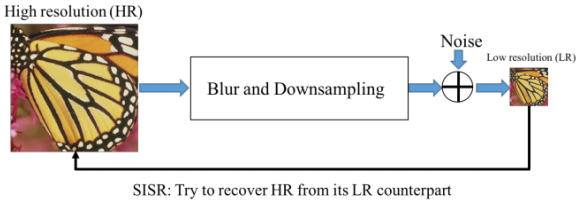

Figure 2: General SISR problem according to [22].

Sparsity-based techniques assumes the LR image to be transformable in another domain (usually a dictionary of image atoms [6]) and tries to find correspondences between the LR and HR patches in the transformed space, as implemented in [4]. However, these techniques usually are very computationally expensive. Among other learning based approaches such as the use of random forests [14], in-place example regression models [21] or adjusted anchored neighborhood regression [16], in terms of accuracy applying CNN based approaches have shown the largest success. [1] Dong et al. [2] trained a shallow CNN end-to-end to build the HR image based on a bicubicly upscaled LR image. This approach was improved by Kim et al. [10] (VDSR) using a deeper network (20 layers) and cascading small filters many times in a deep network structure to exploit contextual

[1] An overview of various other deep learning based approaches for SISR can be found in [22].

information over large image regions in an efficient way. By advancing the network model VDSR was further improved by Lim et al. [11] which got the best results in the NTIRE2017 Super-Resolution Challenge [1].

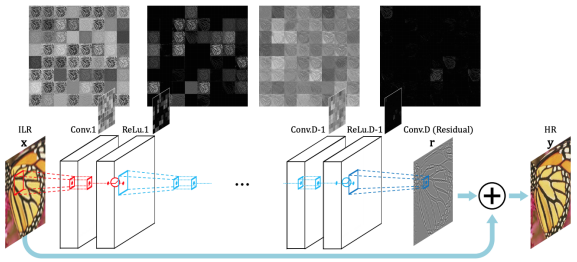

Figure 3: Overview of VDSR network design [10].

2.2 Super-Resolution in Video Domain

Video Super-Resolution combines information from multiple adjacent LR frames to take temporal information into account, leading to higher quality results. Takeda et al. [15] apply a 3D kernel regression on a patch of adjacent LR frames to implicitly encounter temporal information. Since purposed by Caballero et al. [3] end-to-end approaches including motion compensation such as the CNN framework from [3] have large success in the VSR area. Liu et al. [12] added temporal addaptivity to the framework to be able to aggregate the resulting HR frame based on a weighted sum of several estimates as well as a varying number of input LR frames. Sajjadi et al. [13] purposed a frame-recurrent architecture iteratively using the previously inferred HR frames for the subsequent prediction. Wang et al. [17] (SOFVSR) implemented an end-to-end trainable approach to predict both, the HR frame as well as the HR optical flow. Therefore, first the HR optical flow is inferred in a coarse-to-fine manner, then motion compensation is performed according to the HR optical flows and finally, the compensated LR inputs are fed to a super-resolution network to generate the HR frame estimate (comp. Fig. 4).

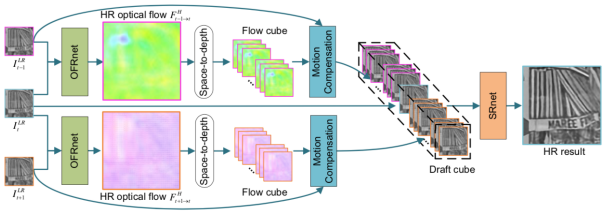

Figure 4: Overview of SOFVSR pipeline [17].

2.3 Colorization

Image colorization methods can be categorized in two categories: Non-parametric approaches, such as [7], model the correspondence between the grayscale and the colored image by finding analogeous regions in reference image(s), while parameteric models learns this correspondence from large datasets, transforming the colorization problem into a regression problem. Zhang et al. [23] (CIC) purpose posing colorization as a classification task and use class-rebalancing at training time to increase the diversity of colors in the result, using the CNN shown in Fig. 5 and not requiring any user-interaction.

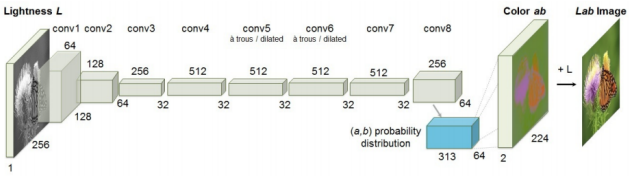

Figure 5: Overview of CIC network design [23].

2.4 Task-Aware-Downscaling

Over all of the problems stated above most of the approaches merely take into account one side of the process, e.g. by fixing the transformation HR to LR to bicubic interpolation in order to large amount of training data and focusing on estimating the inverse transformation. Kim et al. [9] (TAID) purpose taking into account the downscaling method in order to improve the upscaling performance, by training an autoencoder in an end-to-end manner while the latent space representation again is an image of same size as the LR image. The loss function thereby contains both the difference between the decoded SHR and the original HR image as well as the difference between the encoded SLR and the bicubic interpolated LR image, such that the SLR image is a humanly understandable representation. Next to SISR the approach is shown to be applicable for large scale factor up to 128 as well as for colorization.

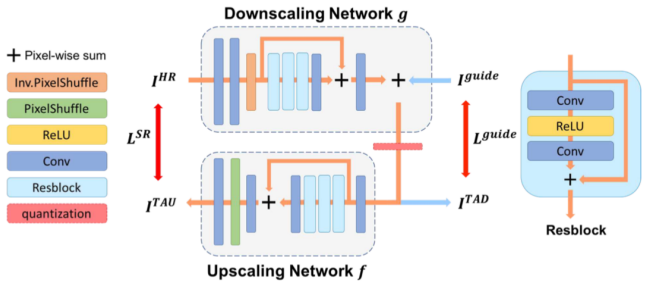

Figure 6: Overview of TAID autoencoder network design [9].

3 Approach

The objectives of the "Materials and Methods" section are the following:

- *What are tools and methods you used?* Introduce the environment, in which your work has taken place - this can be a software package, a device or a system description. Make sure sufficiently detailed descriptions of the algorithms and concepts (e.g. math) you used shall be placed here.
- *What is your work?* Describe (perhaps in a separate section) the key component of your work, e.g. an algorithm or software framework you have developed.

4 Experiments and Results

Describe the evaluation you did in a way, such that an independent researcher can repeat it. Cover the following questions:

- *What is the experimental setup and methodology?* Describe the setting of the experiments and give all the parameters in detail which you have used. Give a detailed account of how the experiment was conducted.
- *What are your results?* In this section, a *clear description* of the results is given. If you produced lots of data, include only representative data here and put all results into the appendix.

5 Discussion

6 Conclusion

List the conclusions of your work and give evidence for these. Often, the discussion and the conclusion sections are fused.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1122–1131, 2017.

[2] Dong C., Loy C.C., He K., and Tang X. Learning a deep convolutional network for image super resolution. *ECCV 2014*, 2014.

[3] Jose Caballero, Christian Ledig, Andrew P. Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and motion compensation. *CoRR*, abs/1611.05250, 2016.

[4] W. Dong, L. Zhang, G. Shi, and X. Wu. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Transactions on Image Processing*, 20(7):1838–1857, July 2011.

[5] Claude E. Duchon. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology*, 18(8):1016–1022, 1979.

[6] M. Elad. Sparse and redundant representations: From theory to applications in signal and image processing. *Springer Publishing Company*, 2010.

[7] Raj Kumar Gupta, Alex Yong-Sang Chia, Deepu Rajan, Ee Sin Ng, and Huang Zhiyong. Image colorization using similar images. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pages 369–378, New York, NY, USA, 2012. ACM.

[8] Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita. Recurrent back-projection network for video super-resolution. *CoRR*, abs/1903.10128, 2019.

[9] Heewon Kim, Myungsub Choi, Bee Lim, and Kyoung Mu Lee. Task-aware image downscaling. In *ECCV*, 2018.

[10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. *CoRR*, abs/1511.04587, 2015.

[11] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. *CoRR*, abs/1707.02921, 2017.

[12] D. Liu, Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, and T. Huang. Robust video super-resolution with learned temporal dynamics. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2526–2534, Oct 2017.

[13] Mehdi S. M. Sajjadi, Raviteja Vemulapalli, and Matthew Brown. Frame-recurrent video super-resolution. *CoRR*, abs/1801.04590, 2018.

[14] S. Schulter, C. Leistner, and H. Bischof. Fast and accurate image upscaling with super-resolution forests. pages 3791–3799, June 2015.

[15] H. Takeda, P. Milanfar, M. Protter, and M. Elad. Super-resolution without explicit subpixel motion estimation. *IEEE Transactions on Image Processing*, 18(9):1958–1975, Sep. 2009.

[16] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *ACCV*, 2014.

[17] Longguang Wang, Yulan Guo, Zaiping Lin, Xinpu Deng, and Wei An. Learning for video super-resolution through HR optical flow estimation. *CoRR*, abs/1809.08573, 2018.

[18] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. ESRGAN: enhanced super-resolution generative adversarial networks. *CoRR*, abs/1809.00219, 2018.

[19] Yifan Wang, Federico Perazzi, Brian McWilliams, Alexander Sorkine-Hornung, Olga Sorkine-Hornung, and Christopher Schroers. A fully progressive approach to single-image super-resolution. *CoRR*, abs/1804.02900, 2018.

[20] Zhihao Wang, Jian Chen, and Steven C. H. Hoi. Deep learning for image super-resolution: A survey. *CoRR*, abs/1902.06068, 2019.

[21] J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. pages 1059–1066, June 2013.

[22] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, and Jing-Hao Xue. Deep learning for single image super-resolution: A brief review. *CoRR*, abs/1808.03344, 2018.

[23] Richard Zhang, Phillip Isola, and Alexei A. Efros. Colorful image colorization. *CoRR*, abs/1603.08511, 2016.

A Appendix

In the appendix, list the following material:

- Data (evaluation tables, graphs etc.)
- Program code
- Further material