max planck institut
informatik

Deep Learning based Multi Frame Super Resolution

HLCV Project Proposal

Team 43

12.06.2021

Introduction

Task

Perform multi-frame super resolution (MFSR) on images from smartphones, by doing domain adaptation on a model trained on satellite images.

MFSR: the problem of combining signals from multiple correlated images to get one high resolution image.

Motivation

- Many deep learning based approaches produce state of the art results for Super Resolution (SR). However, most of them focus on Single-Image Super-Resolution (SISR). Such approaches are limited to learning image priors in order to add high frequency details. In contrast, MFSR offers the possibility of reconstructing rich details by combining signal information from multiple *shifted* images. This has applications in remote sensing, medical imaging, burst photography, etc.
- MFSR problem has largely been approached with modelling based methods. However, with the 'Proba-V Super Resolution Challenge', deep learning based methods have also been explored recently. These methods have shown excellent performance, but have been trained only on satellite images.
- We intend to explore the performance of one of these methods for mobile phone photography. This is motivated by the fact that compared to DSLR cameras, smartphone cameras have smaller sensors, limiting their spatial resolution; smaller apertures, limiting their light gathering ability; and smaller pixels, reducing their signal-to-noise ratio. However, the natural hand tremor - typical in handheld photography - can be harnessed to acquire a burst of frames with small offsets. These frames can then be merged to form a single large resolution image.

Goals

- Transfer learning of MFSR network from satellite-data domain to our dataset.
- Generation of synthetic ‘multi-frame’ dataset for experimentation.
- Experiment over different SOTA training methods, loss functions in the scope of the course to observe and improve performance.
- Performance evaluation of proposed MFSR with existing works in the same, as well as SISR.

Goals for Interim Presentation:

- Study the model architecture and the research paper chosen for the task.
- Setup the baseline models required for the task.
- Generate the multi-frame dataset required (see the “Data” slide).
- Benchmark the pre-trained network on the generated dataset.

Methods

- We select model in “Multi-image Super Resolution of Remotely Sensed Images using Residual Feature Attention Deep Neural Networks”. This is the best-performing model as per the PROBA-V challenge leaderboard and will be used as our pre-trained MFSR model.
    - It was built on ‘Satellite Image Multi-Frame Super Resolution Using 3D Wide-Activation Neural Networks’ which used 3D CNN residual architecture with Wide Activation Super Resolution (WDSR) blocks. Fromthe various submissions, residual architectures with 3D Conv layers give good results in super resolution problems, taking advantage of the temporal dimension and small variations in frame positions.

- We will generate a synthetic dataset using the script from BURSTSR (Burst Image Super-Resolution Challenge, NTIRE which is part of CVPR 2021 Workshop and challenge). This will generate multiple images from ‘Holopix’ dataset with different lighting and coloring conditions as well as frame of view. This multi-frame dataset will be used for our experiments.

- Fine-tune the MFSR model on this synthetic dataset.

- Research and experiment with different Loss functions to improve performance.
    - Alongside, experiment with SOTA training methodologies to get see if results can be improved.

Data

- We will be using the Holopix dataset, which contains stereo pairs captured using mobile phones.
- The stereo nature of the dataset gives us a shifted-image-pair for each scene.
- More frames will be generated synthetically using the BURSTSR approach - A LR burst is generated by applying random translations and rotations on an SR image, followed by bilinear downsampling. The generated burst is then mosaicked and corrupted by random noise.
- The resultant LR stack is used for the problem, and the generated SR result is compared against one of the two original images.

Evaluation

- Quality metrics for performance evaluation:
    - PSNR: Peak signal-to-noise ratio - measures absolute differences.
    - SSIM: Structural similarity index - measures perceptual differences.

- We will be comparing our result against the following baseline techniques:
    - Single frame based:
        - Bicubic interpolation.
        - "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", Ledig, Christian, et al.
    - Multi frame based:
        - "Handheld Multi-Frame Super-Resolution", Wronski, Bartlomiej, et al (non deep learning based).
        - The original pre-trained model before transfer learning.

- The following types of scenes will be used for evaluation:
    - Typical well-lit photos from a smartphone.
    - Photos with structured texture, which are difficult to reconstruct, eg. focus chart, text, aliased structured patterns (i.e. with visible moire patterns).
    - Low-light captures with sensor noise.

References

- Salvetti, Francesco, et al. "Multi-image super resolution of remotely sensed images using residual attention deep neural networks." Remote Sensing 12.14 (2020): 2207.
- Dorr, Francisco. "Satellite Image Multi-Frame Super Resolution Using 3D Wide-Activation Neural Networks." Remote Sensing 12.22 (2020): 3812.
- Kim, Soo Ye, et al. "3dsrnet: Video super-resolution using 3d convolutional neural networks." arXiv preprint arXiv:1812.09079 (2018).
- Wang, Zhihao, Jian Chen, and Steven CH Hoi. "Deep learning for image super-resolution: A survey." IEEE transactions on pattern analysis and machine intelligence (2020).
- Bhat, Goutam, et al. "Deep Burst Super-Resolution." arXiv preprint arXiv:2101.10997 (2021).
- Yiwen Hua and Puneet Kohli and Pritish Uplavikar and Anand Ravi and Saravana Gunaseelan and Jason Orozco and Edward Li, Holopix50k: A Large-Scale In-the-wild Stereo Image Dataset
- https://www.frontiersin.org/research-topics/3055/modern-trends-and-applications-of-super-resolution-imaging
- Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- Wronski, Bartlomiej, et al. "Handheld multi-frame super-resolution." ACM Transactions on Graphics (TOG) 38.4 (2019): 1-18.

Thank You!