Multiple-Image Super-Resolution

IFT6145 Project
Bowen Peng

Art assets might contain copyright and are included under fair use. All rights reserved to respective authors.

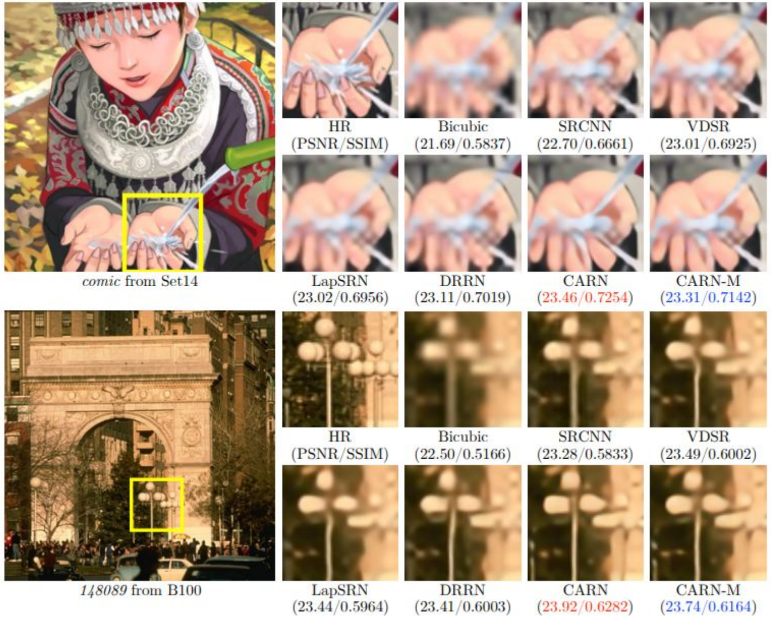
Why Super-Resolution ?

- Cheaper sensors
- Old image/video restoration
- Performance improvements (Nvidia DLSS)

PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models
https://arxiv.org/pdf/2003.03808v1.pdf

Why MISR?

- High quality SISR is slow! CARN-M uses 3.5B operations for a single image…
- There is only so much information in a single image… We are hitting the limits of SISR.
- Videos and image sequences have additional information that we can take advantage of.
- Taking multiple low resolution images can be cheaper than one single high resolution image. (Eg. astronomy, MRIs, 3D scans, …)



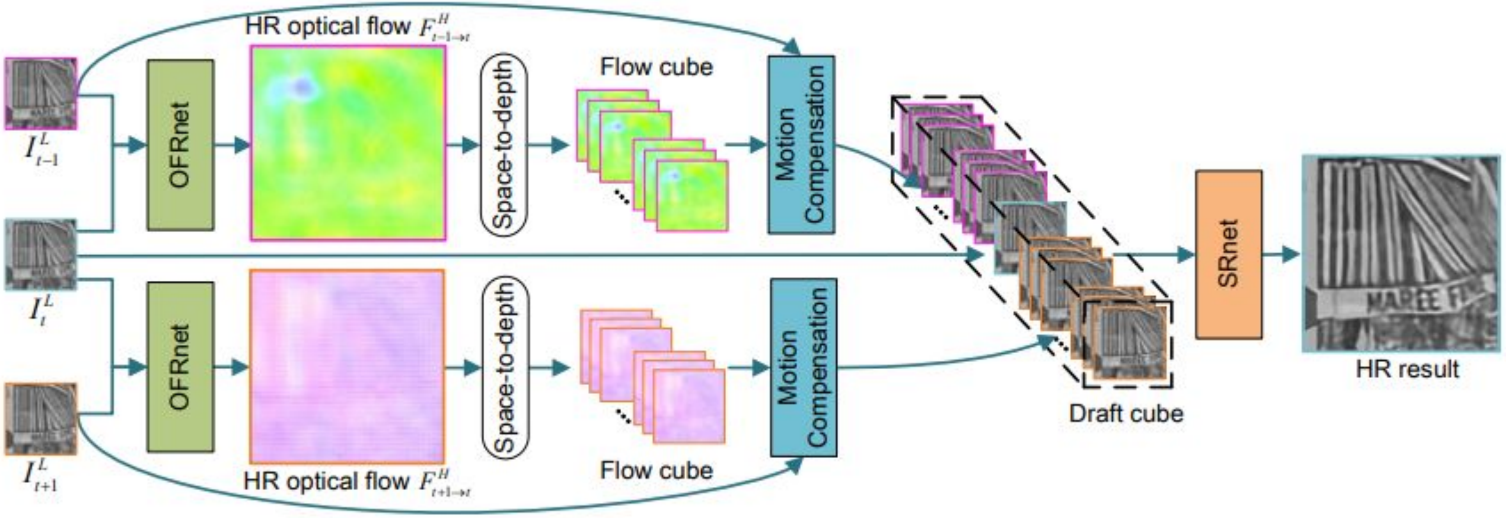
Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network
https://arxiv.org/pdf/1803.08664.pdf

Problems?

- If not taken at the same time, objects in multiple images might not be aligned.
  - Moving objects
  - Occlusions

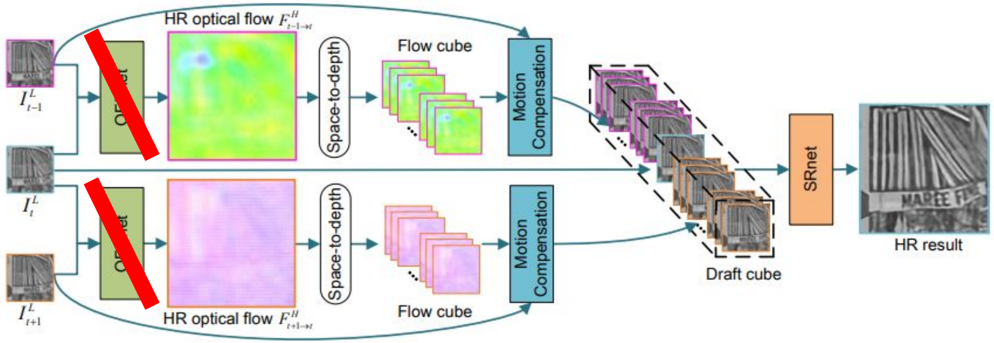
- Use optical flow!
  - SINTEL dataset
  - Fixes alignment
  - Occlusions still present...

Deep Video Super-Resolution using HR Optical Flow Estimation
Wang et al. 2020, IEEE



My Project

- Simpler implementation for the scope of the project.
- Focus on super resolution
- Assume existing optical flow information.
- Pre-process images, align them before passing to SRnet
- Compare against SISR model
- Ablation study



Dataset Generation

- SINTEL optical flow dataset
  - Only take luminance (Y) channel
- Forward warping
  - Splatting
- 8 input images per instance
  - 1 LR + 7 past LR images
  - Total of 880 instances
- x4 SR factor

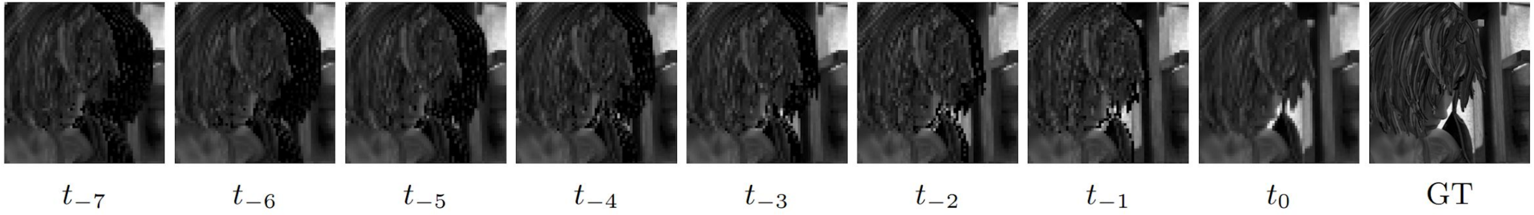
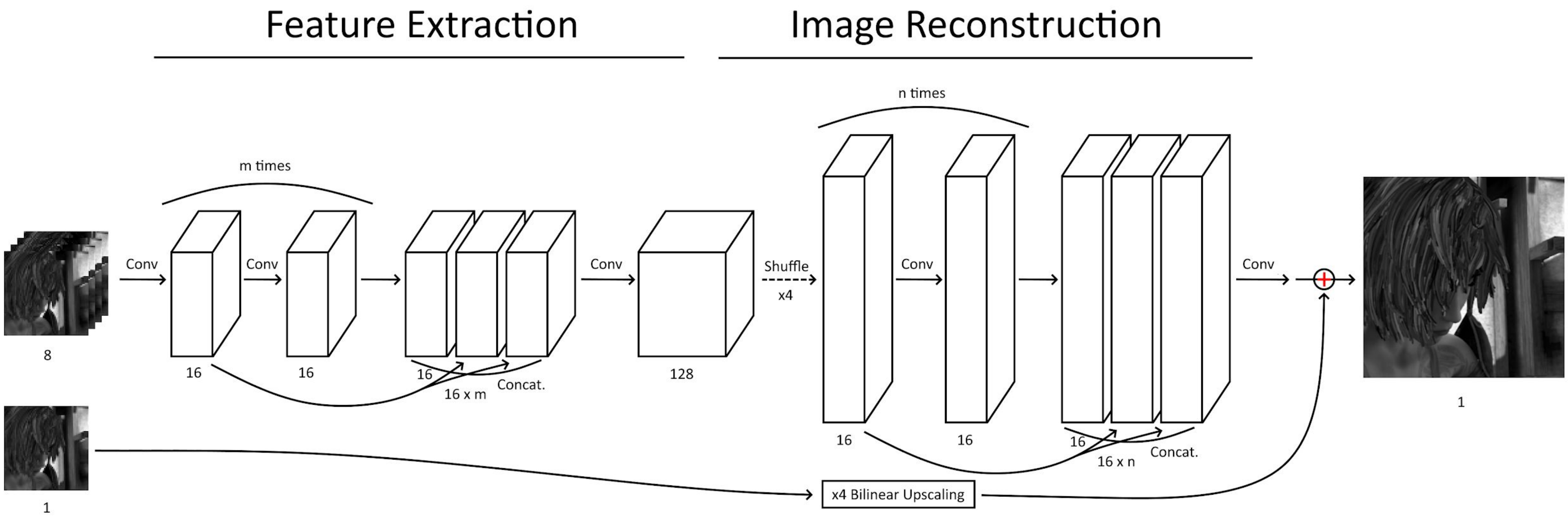


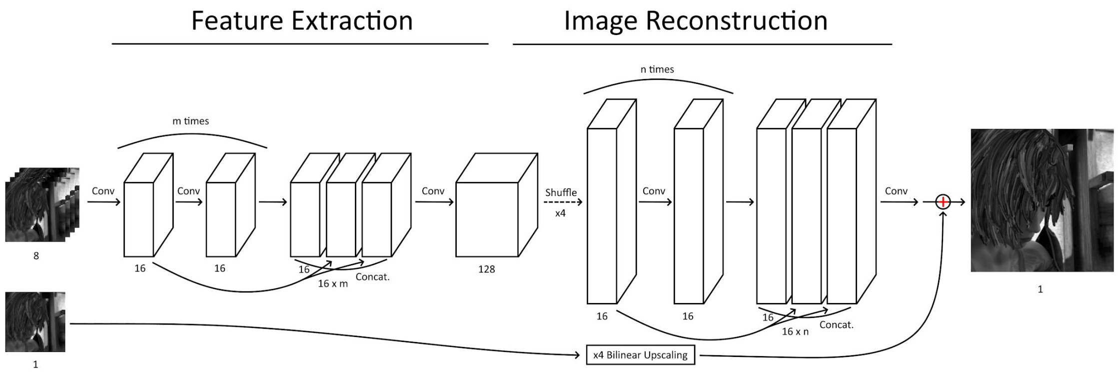
Figure 1: One example from the generated dataset. Occlusions are most noticeable for earlier frames.

Model Architecture



Model Architecture

- CReLU
- Global residual learning
- Densely connected layers
- Feature extraction and image reconstruction
- 17.6k parameters
  - Very light network!
- m = 3, n = 3



Training

- Gradient Descent using MSE loss
- Adam Optimizer
  - a = 0.001
  - b1 = 0.9
  - b2 = 0.999
- 780 + 780 training instances /w augmentation (x-flipping)
  - Note: Small mistake here
- 50 validation instances, 50 test
- Training for 60 epochs

- MISR vs SISR model comparison
  - Early stopping of MISR model for fair comparison against SISR model

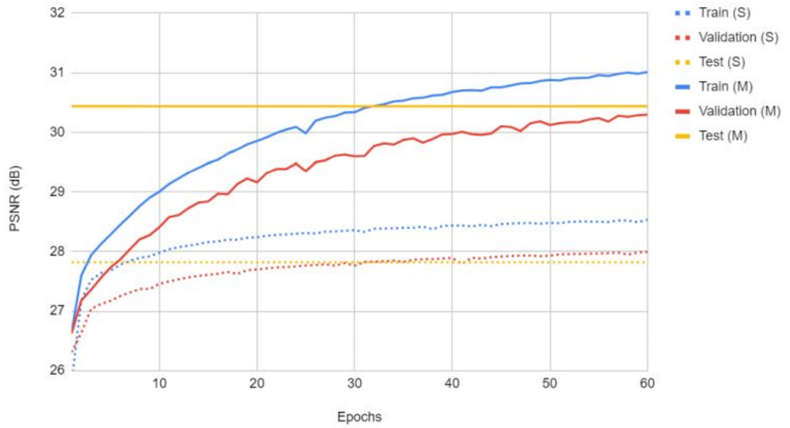
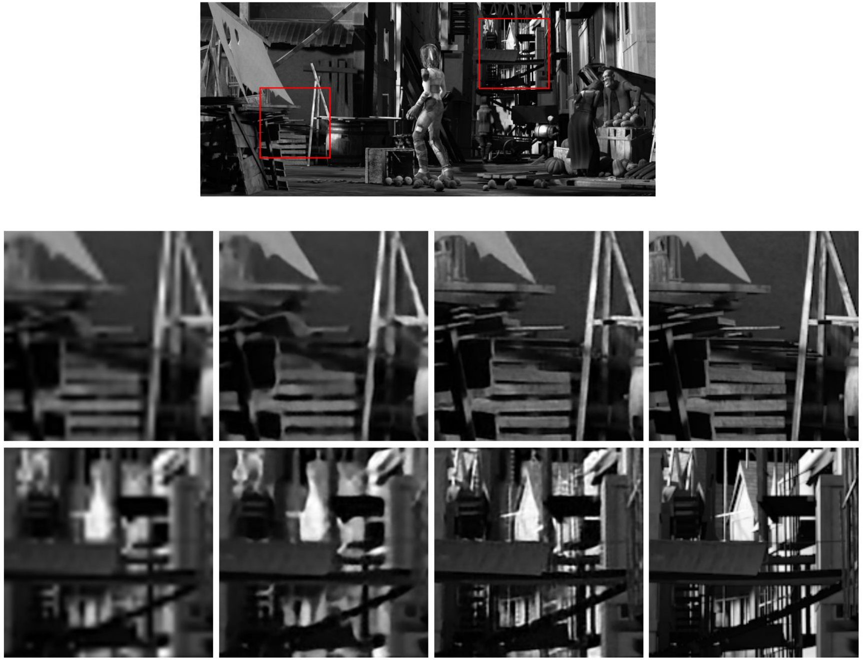


Figure 3: x4 SR PSNR during and after training. M represents multiple images and S means single image. The test result was computed at the end of training using the test set.

Results

- 30.44 dB PSNR for MISR on test set
- 27.82 dB PSNR for SISR on test set

- Significant visual quality improvements on MISR model compared to SISR model with same architecture and number of parameters.



Bicubic	SISR	MISR	Ground Truth
22.84	23.91	**28.14**	PSNR (dB)

Figure 4: x4 SR result on an image from the test set.

Failure Cases

- Self-occlusions and occlusions to other objects degrades the SR result on the affected objects.
- The degraded result is often not better than the result of a SISR model.
- Sometimes accompanied by artifacts.

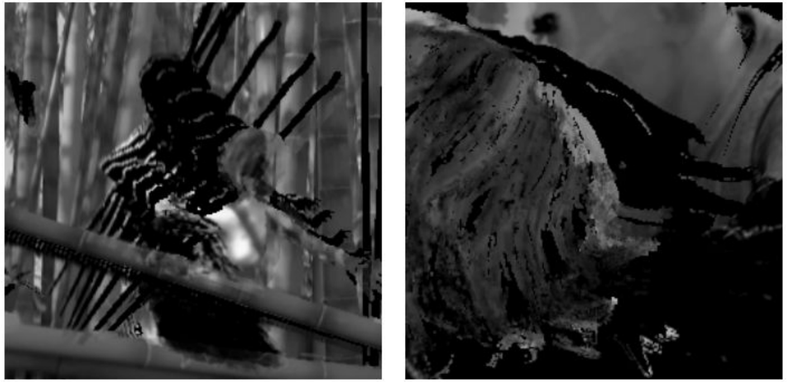


Figure 5: Examples of large amounts of occlusions and self-occlusions on fast moving objects.

LR | SISR | MISR | Ground Truth

Figure 6: Failure case due to occlusion. Note the blurriness and artifacts on and near the moving body on the MISR model. The foliage further away to the right is not affected.

Ablation Study

- Test whether the network is truly using all 8 input features.
    - Set some features to 0 as input.
- MISR network still surpasses the SISR model when only given 2 images as input.
    - But is worse with only 1 image.
- Due to time constraints, we were only able to test ablation of input features at prediction time.
- Future research can focus on the effects of feature ablation during training, or even training models that can accept a variable amounts of input features.

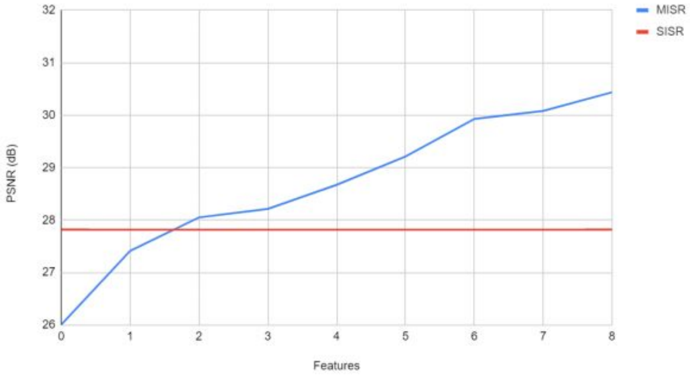
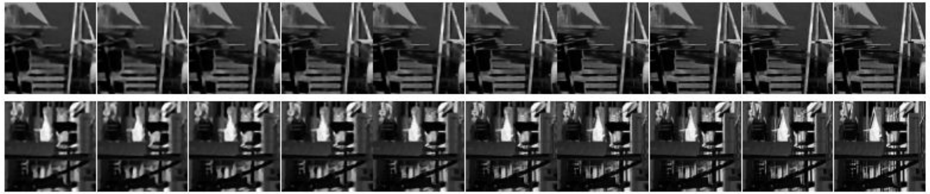
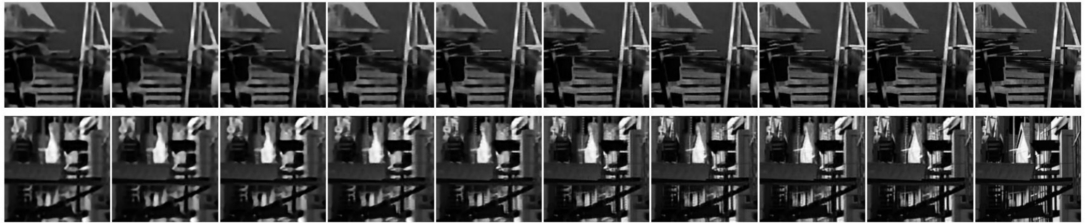


Figure 7: x4 SR PSNR ablation study at test time.



SISR	1	2	3	4	5	6	7	8	GT
23.91	23.59	24.16	24.33	25.23	26.21	27.29	27.61	**28.14**	PSNR (dB)

Figure 8: x4 SR ablation results on an image from the test set.



SISR	1	2	3	4	5	6	7	8	GT
23.91	23.59	24.16	24.33	25.23	26.21	27.29	27.61	**28.14**	PSNR (dB)

Figure 8: x4 SR ablation results on an image from the test set.

SISR	3	5	8	GT
23.91	24.33	26.21	**28.14**	PSNR (dB)

Figure 9: Same results as above, magnified for better visualization.

Future Work

- Generative Adversarial Networks can improve perceived image quality
- Give the neural network a larger receptive field?
- Explore different network architectures
- Ablation study during training
- Model that accepts a variable number of input features
- Implement trained network as GLSL Shader for real time video processing

Questions? & Thanks For Listening!

References

[1] Gao Huang et al. *Densely Connected Convolutional Networks*. IEEE, 2016.
[2] Longguang Wang et al. *Deep Video Super-Resolution using HR Optical Flow Estimation*. IEEE, 2020.
[3] Wenling Shang et al. *Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units*. ICML, 2016.
[4] Wenzhe Shi et al. *Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network*. IEEE, 2016.