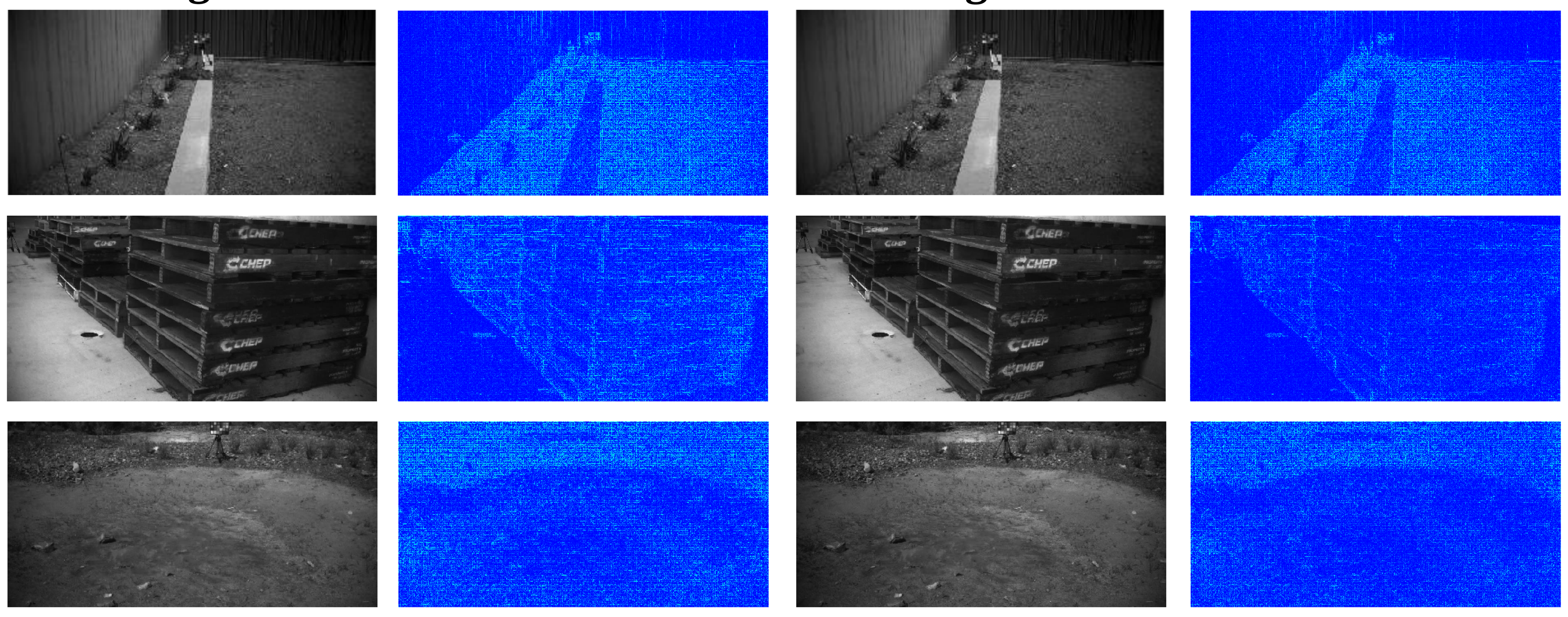
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
Image and Visual Representation Lab
ivrl.epfl.ch
Ruofan Zhou, Fayez Lahoud, and Sabine Süsstrunk

Multi-Modal Spectral Image Super-Resolution

Deep convolutional neural networks are powerful in learning the relation between low- and high-resolution images. However, they typically take a single-scale image as input and require large amounts of data to train. In this work, we use multi-modal inputs, both in spatial and spectral dimensions, to improve spectral super-resolution. Our approach produces state-of-the-art results and is economic in terms of parameters and computation time.

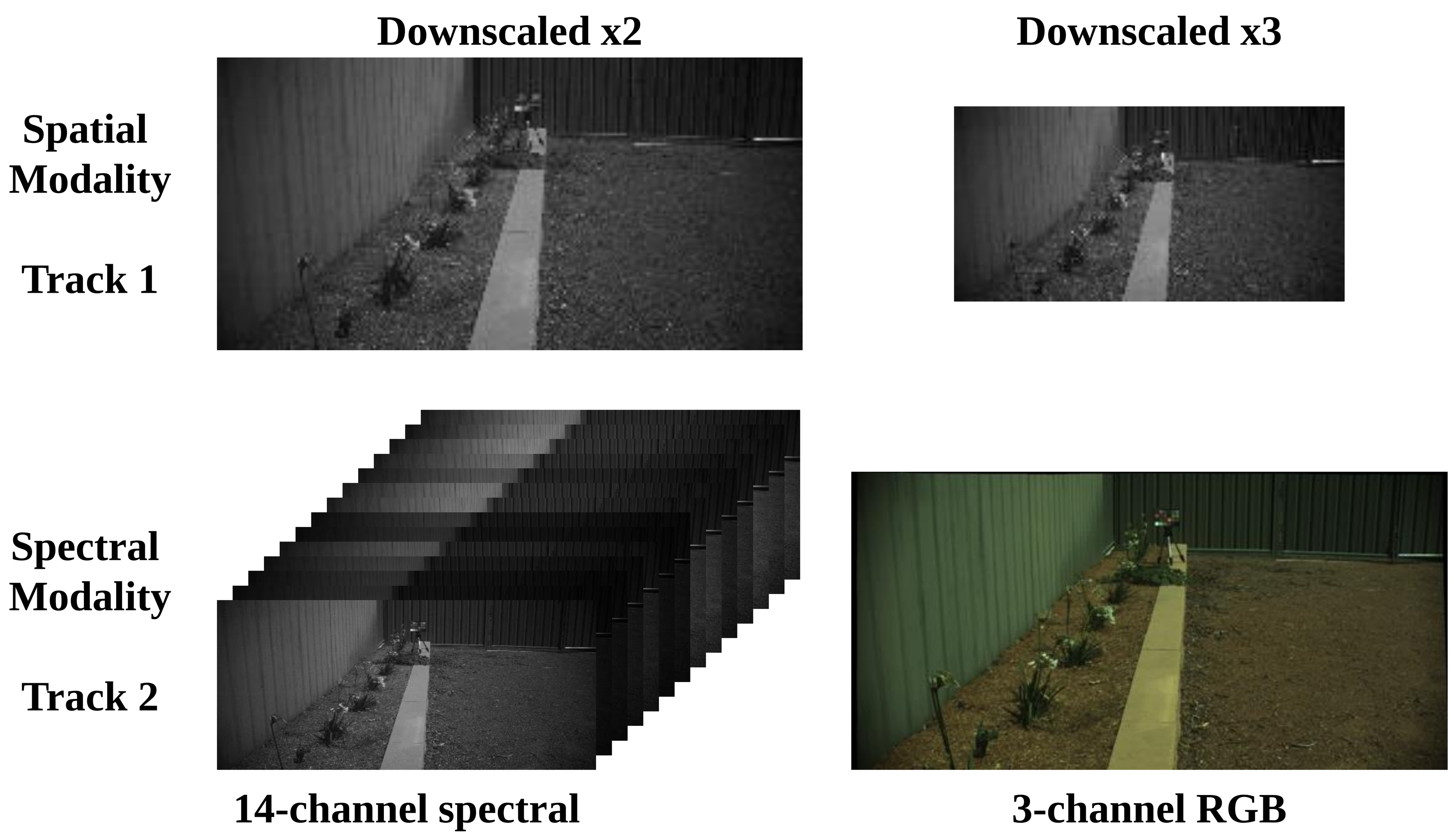
Stage-I | Error-I | Stage-II | Error-II



	Validation Set I			Validation Set II		
Metric	Bicubic x2	EDSR	**Ours**	Bicubic x2	EDSR	**Ours**
MRAE	0.11	0.10	**0.08**	0.13	0.16	**0.09**
SID	57.39	43.57	**43.48**	43.32	30.67	**24.51**
PSNR	36.07	37.27	**37.44**	36.48	37.13	**39.17**

Multi Modality

We use both spatial and spectral modalities to improve the performance of our super-resolution neural network.



Multi-Scale Upsampling

5	9	8	24	1	12
3	16	6	7	2	19
2	1	3	23	20	0
15	3	7	17	2	10
9	11	16	32	0	3
8	15	3	12	3	8

High Resolution Image

5	8	1
2	3	20
9	16	0

Downscaled x2

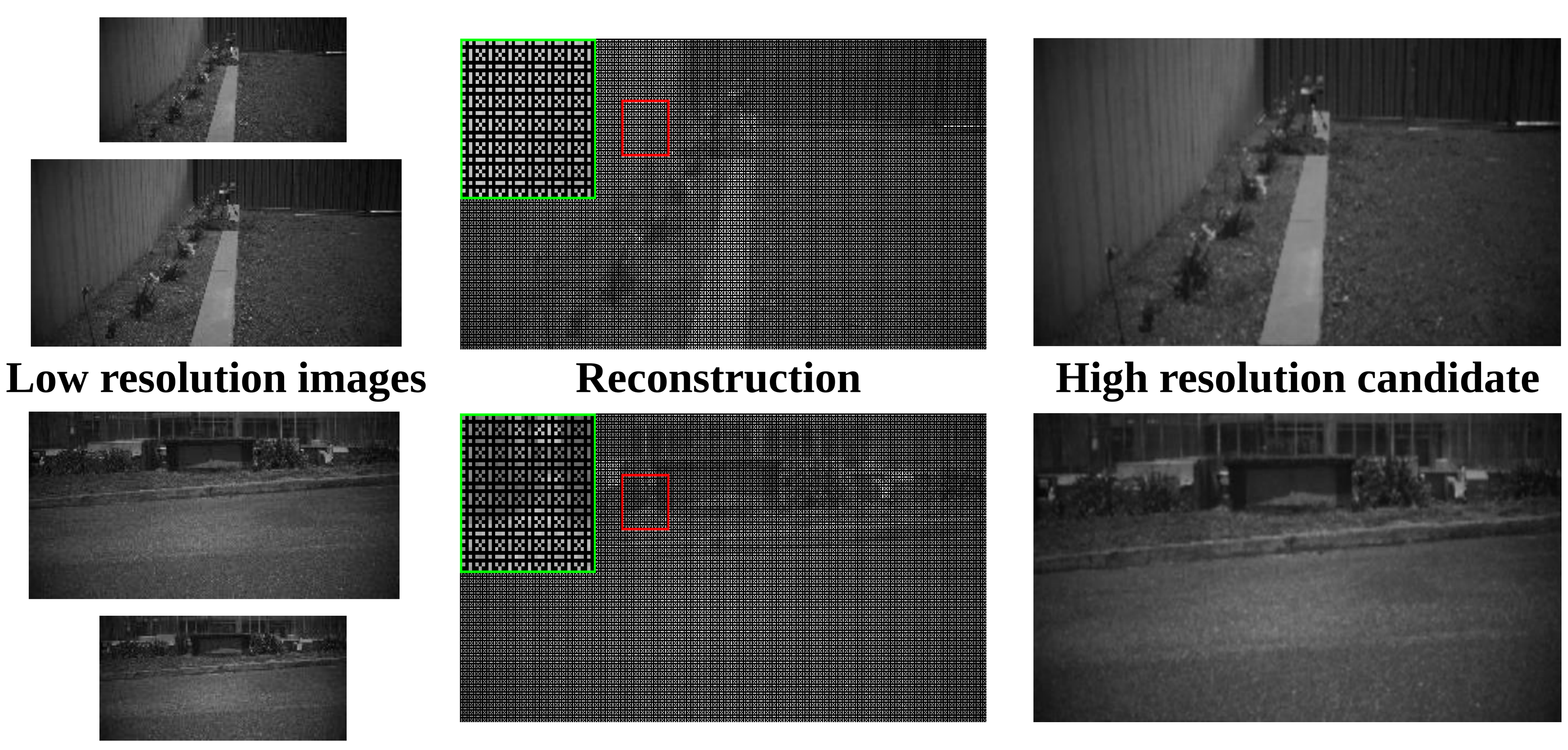
5	24
15	17

Downscaled x3

5		8	24	1	
2		3		20	
15			17		
9		16		0	

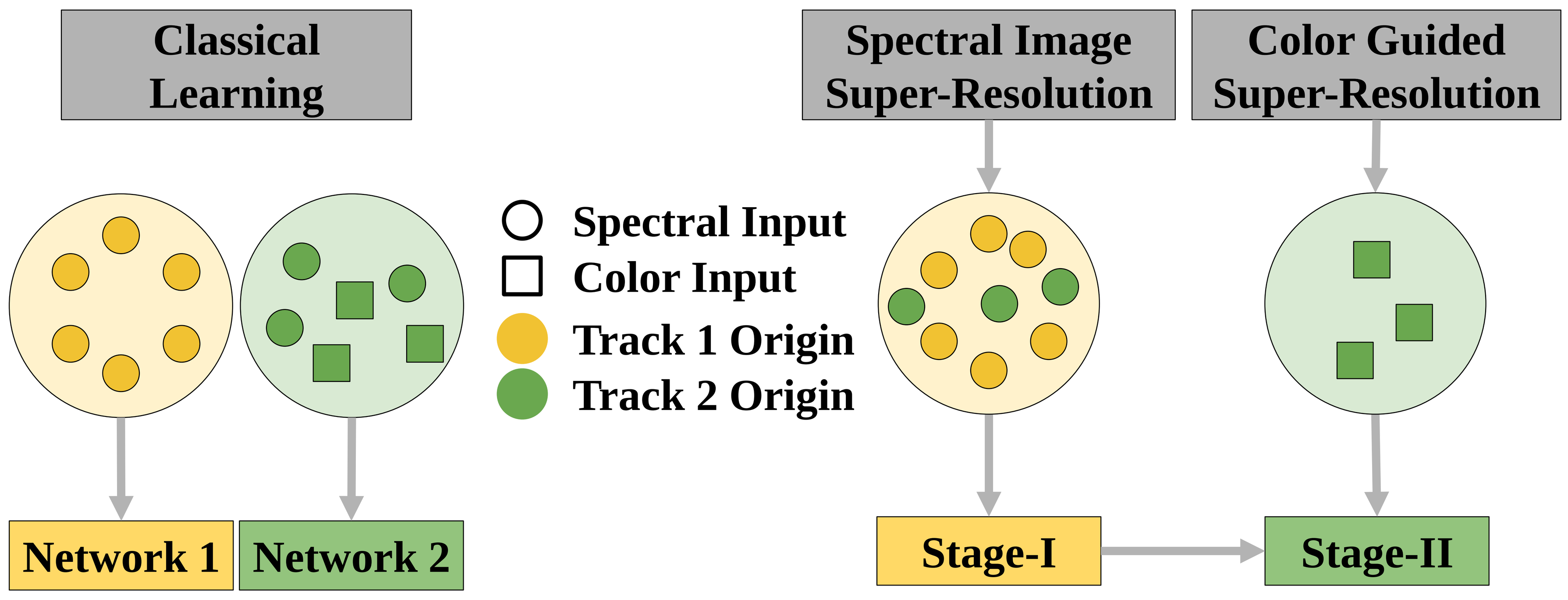
Reconstruction

Using nearest neighbor up-sampling for both x2 and x3 inputs, we reconstruct 1/3 of the original high resolution image, with 2/3 still missing. We reconstruct the missing pixels with the extreme image completion algorithm [1].

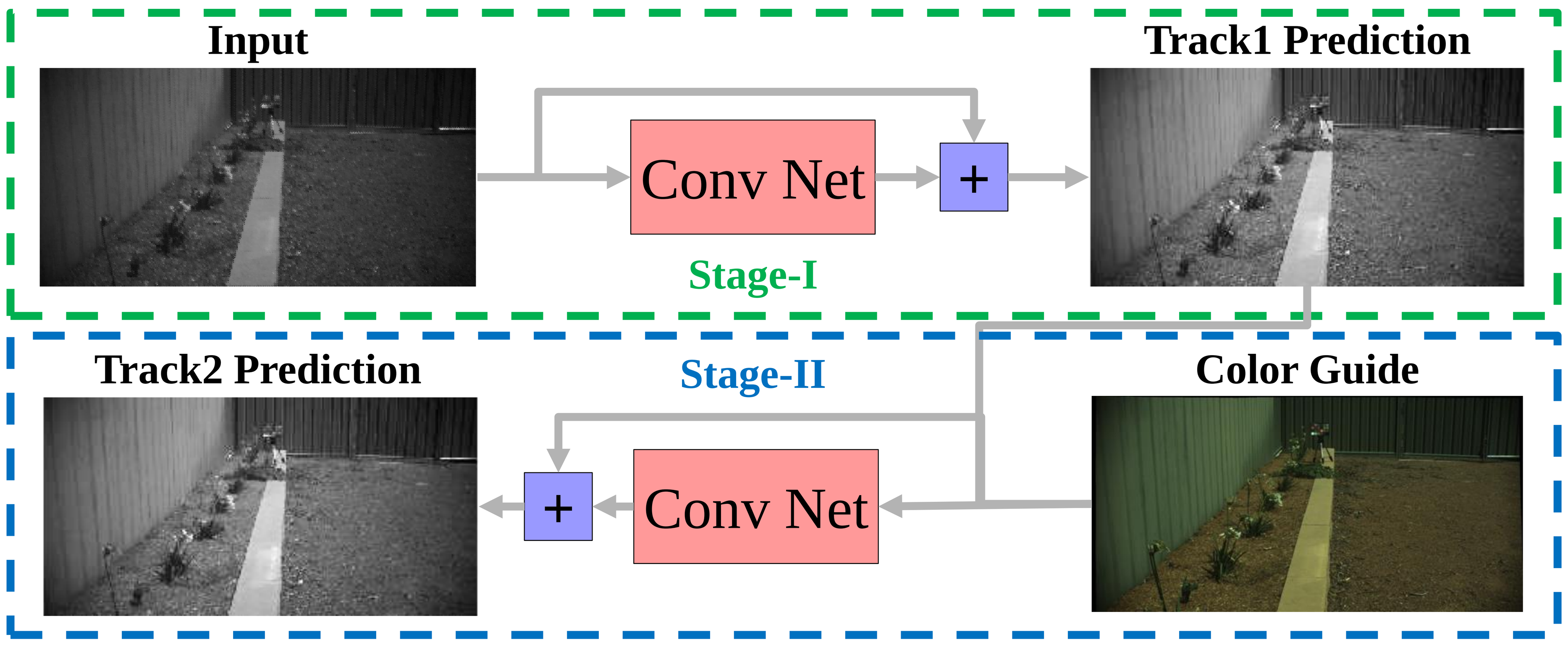


Transfer Learning

Due to the limited amount of available training data in Track2, we use transfer learning to adapt our network from Track1 with a new color modality. In Stage-I, the network learns to super-resolve spectral images. Stage-II uses Stage-I's input and an additional color guide image to fine-tune the super-resolution prediction.



It is more robust to train a network using residuals due to the sparser activations. We use the residual architecture to learn the degradation induced by the down-sampling operation.



We compare our up-sampling technique with the typically used bicubic interpolation. We also compare to the state-of-the art super-resolution network, EDSR [2], both for performance and efficiency.

	Ours	EDSR
Storage (MB)	**2.7**	153.8
Memory (MB)	**800**	8000
Inference time (sec)	**0.5**	1.1

Efficiency comparison

References

[1] R. Achanta, N. Arvanitopoulos, and S. Süsstrunk, "Extreme image completion," in the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017.
[2] Lim, B., Son, S., Kim, H., Nah, S. and Lee, K.M., "Enhanced Deep Residual Networks for Single Image Super-Resolution," in the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) workshops, 2017.