Single Depth Image Super-Resolution Using Convolutional Neural Networks

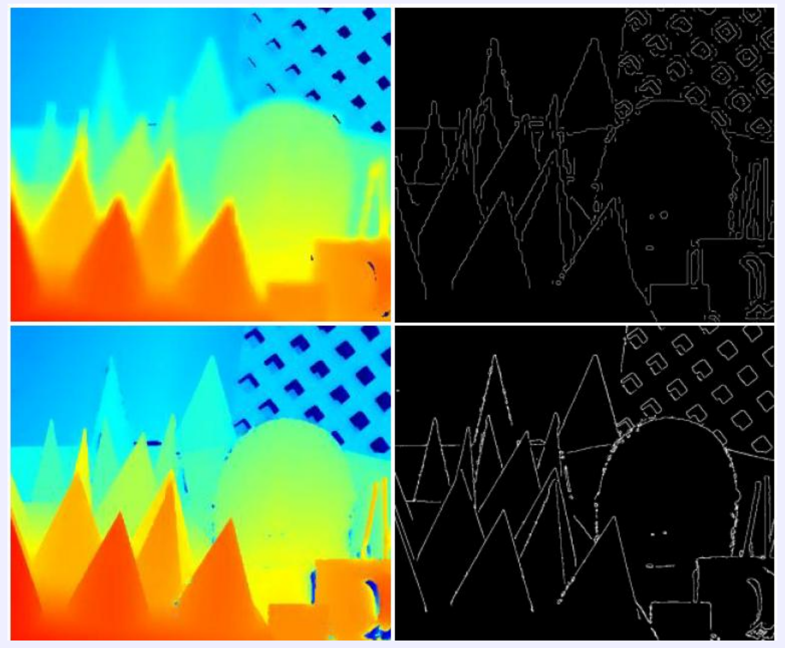
Baoliang Chen and Cheolkon Jung

School of Electronic Engineering, Xidian University, China

ICASSP 2018

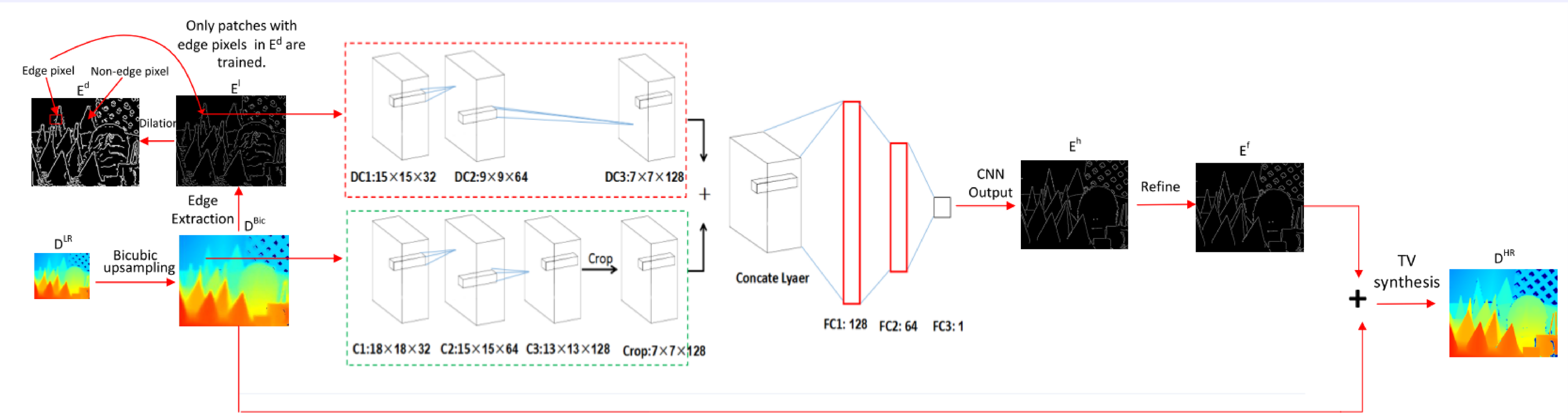
Motivation

- **Edges are of particular importance in the textureless depth image.**
- **We have addressed depth image super-resolution (DISR) by high-resolution (HR) edge prediction, instead of HR texture prediction.**
- **Trained network makes the low-quality edge map more sharp while removing jagged artifacts.**
- **Guided by the high-quality edge map, we perform DISR.**



- **Top:** Interpolated (×4) depth map by bicubic and its edges.
- **Bottom:** Our final result and edge map learned from our CNN.

Proposed Method



Proposed Framework

We extract initial edge map from interpolated LR depth map and repair it using CNN. Since the CNN output often contains broken edges and holes, we refine it. Guided by the high-quality edge map, we do upsampling using a TV model.

Training dataset

Middlebury stereo dataset and Laser Scan dataset.

First input: Patch of size 21×21 around each pixel in E^d.

Second input: Its corresponding patch region in D^{Bic}.

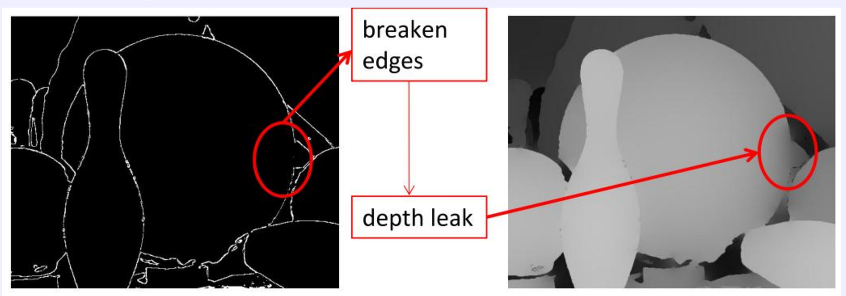
Loss Fuction:

$$\min_{w} -[y_{gt}\log p(y_{gt},w)+(1-y_{gt})\log(1-p(y_{gt},w))]$$

where y_{gt} is the binary label value, $p(y_{gt})$ is the output that indicates the probability to be edge point.

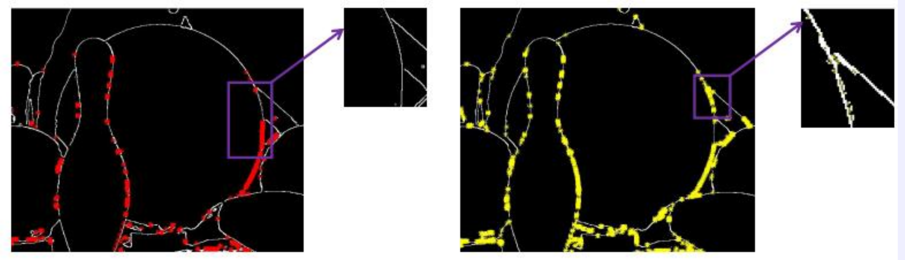
Edge Map Refinement

Broken edges and holes between edges cause depth leak in the SR reconstruction:



1.We first detect the broken edges and then connect them using E^l :

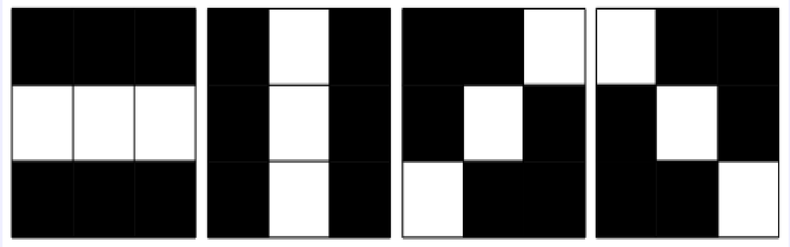
$$E^f(p)=\begin{cases} E^l(p) & \text{if } E^l(p)=1 \text{ and } Sum(Ml\,\&Mh)=0 \\ E^h(p) & \text{else} \end{cases}$$



Left: Connected edges. Right: Filled holes

2. We extract 3 × 3 patch centered by each non-edge pixel in E^h and perform AND operation using four patterns:

$$E^f(p)=\begin{cases} 1 & \text{if } E^h(p)=0 \text{ and } ((R1>2) or (R2>2) or (R3>2) or (R4>2)); \\ E^h(p) & else; \end{cases}$$



Four edge patterns.

Edge-Guided Upsampling

Guided by the edge map E^f, we use a variational approach to get the depth SR image D^R:

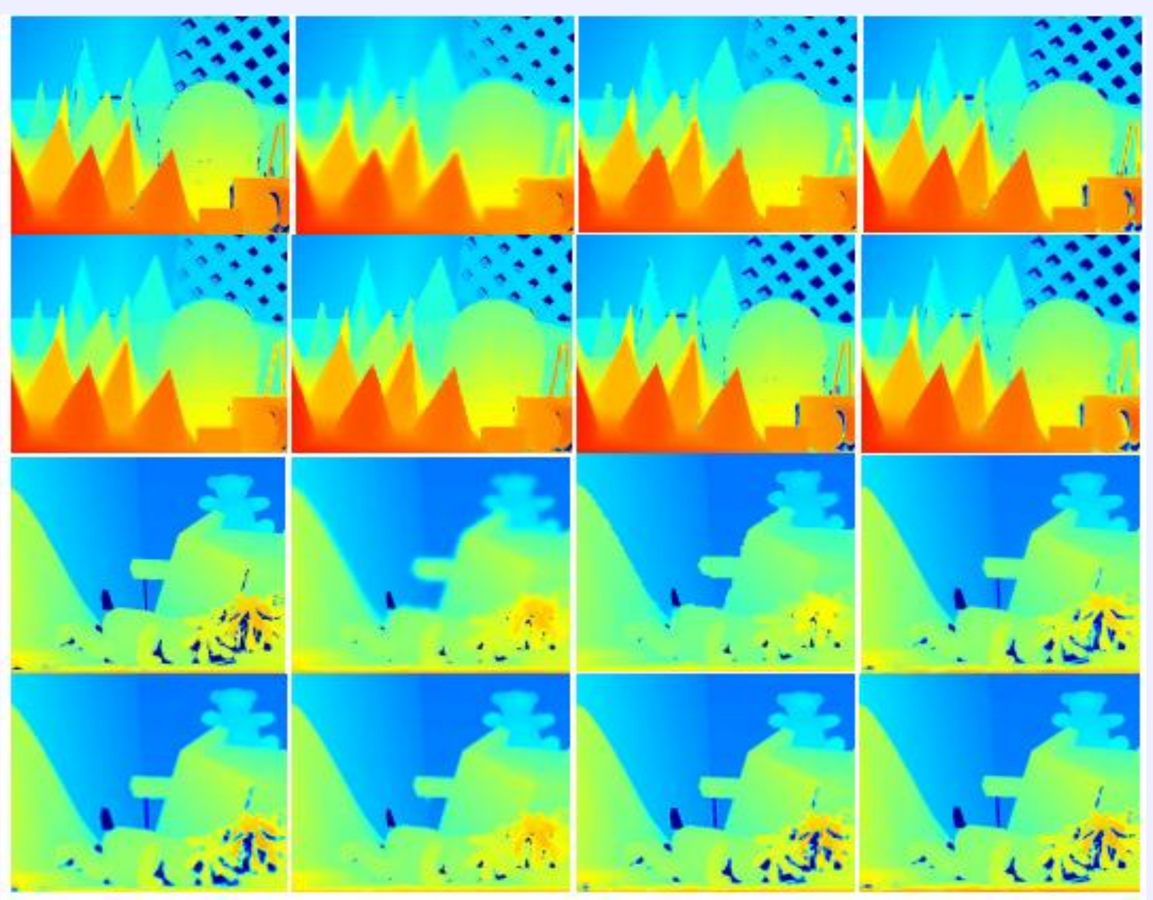
$$\min_{D^R} E(D^R)=E(D^{Bic})+\lambda R_{smooth}$$

$$E(D^{Bic})=\sum_p (D^R(p)-D^{Bic}(p))^2$$

$$R_{smooth}=\sum_p E^f(p)[|\partial_x(D^R)|^2+|\partial_y(D^R)|^2]$$

Experimental Results

- **Proposed method produces more visually pleasing results: Object boundaries are sharper along edges, and thus the scene structure are successfully preserved.**



Depth SR reconstruction results in *Cones* and *Teddy* when the upsampling factor is 4. Up to down, left to right: Ground truth, Bicubic interpolation, Park et al. [1], ANR [2], SRCNN [3], Xie et al. [4], Ferstl et al. [5], and Proposed method.

Depth SR reconstruction results in Laser Scan dataset. Upsampling factor is 4. Left to right: Ground Truth, Bicubic interpolation, ANR [2] , SRCNN [3], Xie et al. [4], Ferstl et al. [5], and proposed method.

Method	x4				x4		
	Cones	Teddy	Tsukuba	Venus	Scan21	Scan30	Scan42
NN	6.0054	4.5466	12.9083	2.9333	2.6474	2.5196	5.6044
Bicubic	3.8635	2.893	8.7103	1.9403	2.0324	1.9764	4.5813
Park et al.[3]	6.5447	4.3366	12.1231	2.2595	*N/A*	*N/A*	*N/A*
Yang et al.[18]	5.139	4.066	13.1748	2.7559	*N/A*	*N/A*	*N/A*
Ferstl et al.[4]	3.9968	2.808	10.0352	1.6643	*N/A*	*N/A*	*N/A*
NE+NNLS	3.4362	2.4887	7.5344	1.6291	1.7313	1.6849	3.5733
SRCNN [19]	4.219	2.456	8.6643	1.9717	1.6732	1.5141	2.783
Aodha et al.[8]	12.6938	4.1113	12.6938	2.6497	2.5983	2.6267	6.1871
Hornacck ct al.[9]	5.4898	5.0212	11.1101	3.5833	2.8585	2.7243	4.5074
Ferstl et al.[11]	3.568	2.6474	7.5356	1.7771	1.4349	1.4298	3.141
Xie et al.[12]	4.4087	3.2768	9.7765	2.3714	1.3993	1.4101	2.691
Proposed	**3.1742**	**2.1357**	**6.3472**	**0.9955**	**1.2453**	**1.0022**	**1.5431**

RMSE Evaluation Results

Conclusions

- **We propose a new CNN based method to acquire the high-quality edge map from the low-quality one.**
- **We utilize the low-quality edge map to connect broken edges and fill holes in the edge map.**
- **We use the high-quality edge map to adjust the weight of the regularization term in total variation (TV).**
- **Various experiments on Middlebury stereo dataset and Laser Scan dataset demonstrate the superiority of the proposed method to state-of-the-arts in both qualitative and quantitative measurements.**

References

[1] Jaesik Park, Hyeongwoo Kim, Yu Wing Tai, Michael S.Brown, and Inso Kweon, "High quality depth map upsampling for 3d-tof cameras," in *Proceedings of the IEEE International Conference on Computer Vision*, 2012, pp. 1623–1630.

[2] Radu Timofte, Vincent De, and Luc Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1920–1927.

[3] Chao Dong, Change Loy Chen, Kaiming He, and Xiaoou Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. European Conference on Computer Vision*, 2014.

[4] Jun Xie, Rogerio Schmidt Feris, and Ming Ting Sun, "Edge guided single depth image super resolution," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 428–438, 2015.

[5] David Ferstl, Matthias Ruther, and Horst Bischof, "Variational depth superresolution using example-based edge representations," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 513–521.