BOOSTING HIGH-LEVEL VISION WITH JOINT COMPRESSION ARTIFACTS REDUCTION AND SUPER-RESOLUTION



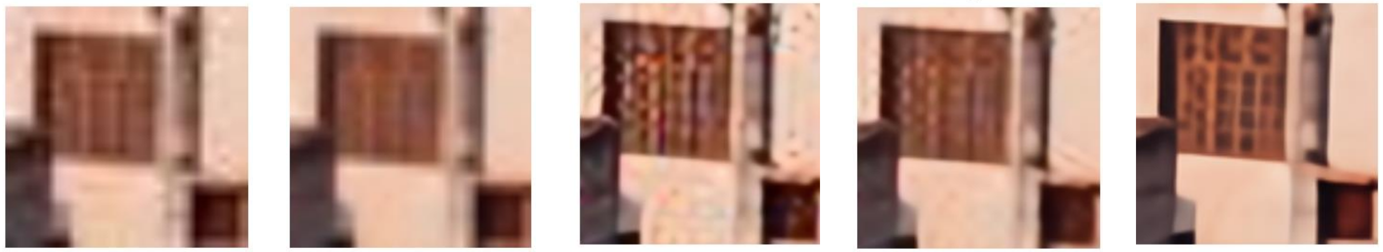
Xiaoyu Xiang[1], Qian Lin[2], Jan Allebach[1]

[1]School of Electrical and Computer Engineering, Purdue University
[2]HP Inc.

Motivation

High Resolution, high quality Image → Low Resolution, low quality Image

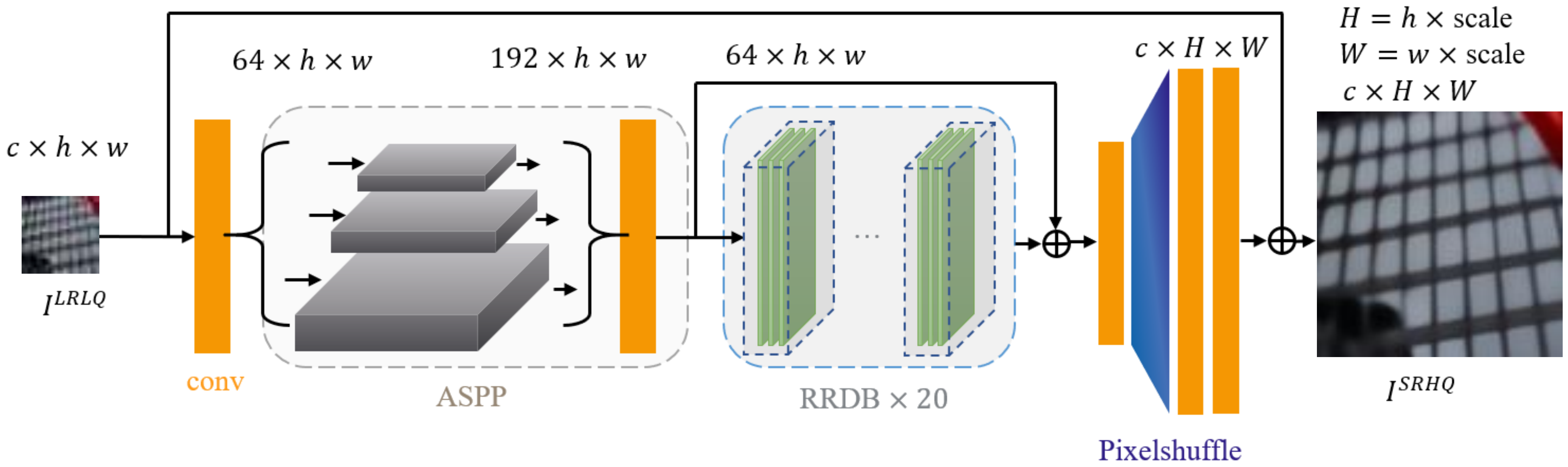
???



Input | CAR | SR | CAR+SR | Ours

- **Problem**: Reconstruct high-resolution and high-quality image
- **Previous methods**: CAR + SR
  - Over-smooth / amplify the artifacts
  - Fail to make full use of the locally related features between CAR and SR tasks
  - 2 reconstruction networks increase the model size and inference cost
- **Our solution**:
  - We propose a context-aware framework that <u>jointly</u> solves the <u>CAR and SR</u> problems for real-world images with <u>unknown quality factors</u>

Single-Stage Network for Joint Compression Artifacts Reduction and Super Resolution



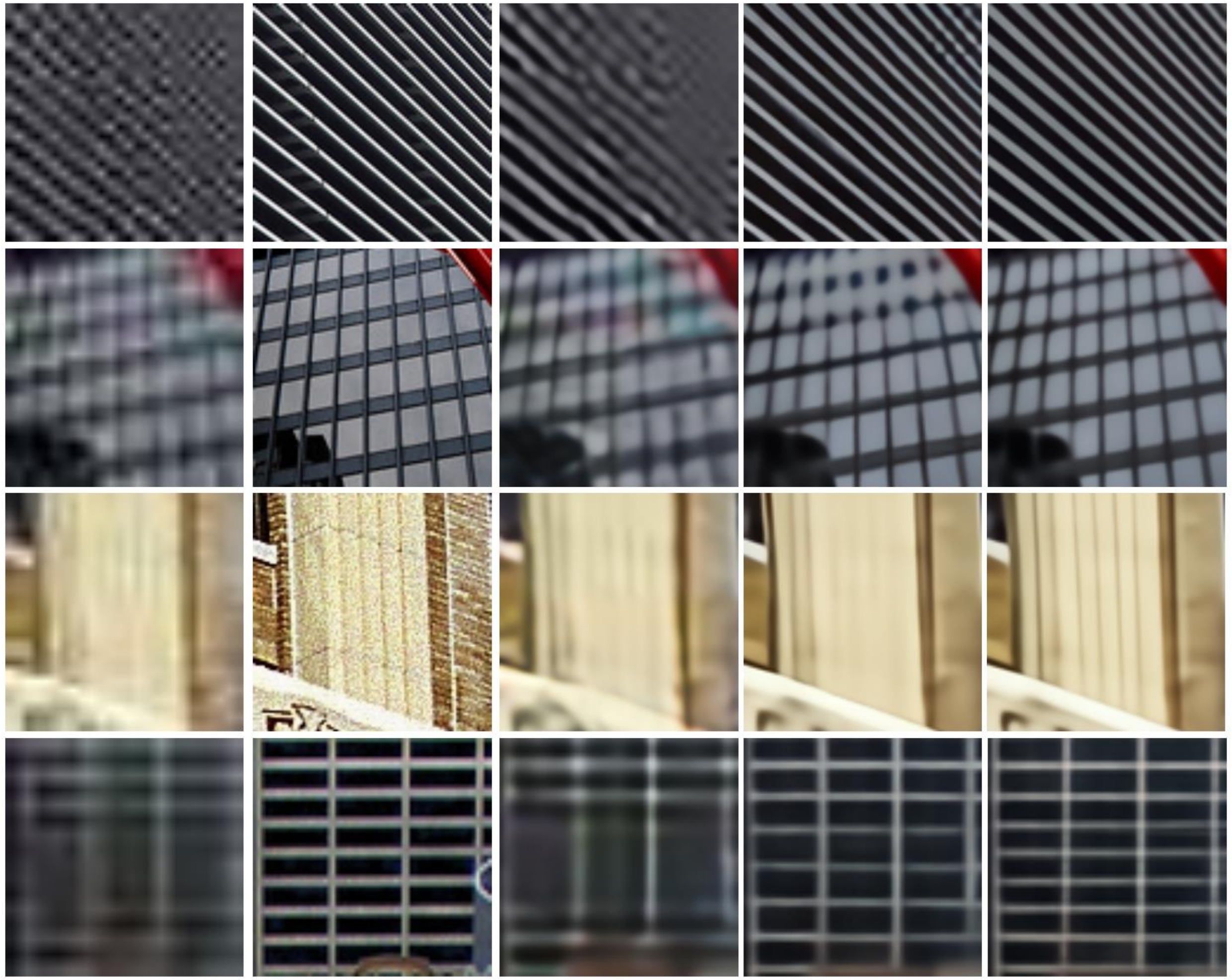
The network architecture of our proposed CAJNN. It directly reconstructs artifact-free HR images from the LR low-quality images. Atrous Spatial Pyramid Pooling (ASPP) is adopted to utilize the inter-block features and intra-block contexts for the joint CARSR task. The reconstruction module turns the features into a deep feature map, which is converted to a high-quality SR output by the upsampling and enhancement module.

COMPARISON WITH PREVIOUS METHODS

QF	Method	Network	Runtime (s)	Parameters (Million)	Set5		Set14		BSD100		Urban100		Manga109	
					PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
10	SR	Bicubic	-	-	23.99	0.6329	22.94	0.5513	23.33	0.5303	20.95	0.5182	21.94	0.6383
		EDSR	1.94	43.1	23.41	0.6019	22.48	0.5272	22.96	0.5098	20.57	0.5006	21.53	0.6151
		RCAN	2.04	16	23.14	0.5733	22.29	0.5064	22.78	0.4984	20.36	0.4819	21.21	0.5878
		RRDB	0.65	16.7	22.43	0.5223	22.86	0.5051	20.43	0.4940	20.43	0.4940	21.34	0.6075
	CAR+SR	ARCNN+RRDB	3.20+0.65	0.56+16.7	24.21	0.6699	23.38	0.5774	23.63	0.5474	21.28	0.5466	22.36	0.6856
		DnCNN+RRDB	0.38+0.65	0.06+16.7	24.07	0.6434	23.13	0.5582	23.37	0.5324	21.04	0.5305	22.10	0.6532
	Joint CAR&SR	CAJNN (ours)	0.48	14.8	25.04	0.7169	23.95	0.6028	23.84	0.5598	21.97	0.5977	23.29	0.7333
		CAJNN (ours, self-ensembled)	2.50	14.8	25.14	0.7202	24.03	0.6052	23.88	0.5610	22.18	0.6051	23.44	0.7377
20	SR	Bicubic	-	-	25.32	0.6761	23.85	0.5870	24.14	0.5611	21.66	0.5526	22.84	0.6724
		EDSR	1.94	43.1	24.76	0.6490	23.59	0.5707	23.88	0.5482	21.38	0.5427	22.58	0.6549
		RCAN	2.04	16	24.44	0.6226	23.40	0.5502	23.65	0.5351	21.12	0.5234	22.14	0.6253
		RRDB	0.65	16.7	24.65	0.6450	23.57	0.5661	23.79	0.5442	21.25	0.5365	22.38	0.6474
	CAR+SR	ARCNN+RRDB	3.20+0.65	0.56+16.7	25.40	0.7082	24.30	0.6091	24.39	0.5755	22.02	0.5811	23.52	0.7172
		DnCNN+RRDB	0.38+0.65	0.06+16.7	25.55	0.6946	24.24	0.6001	24.28	0.5679	21.90	0.5732	23.24	0.6961
	Joint CAR&SR	CAJNN (ours)	0.48	14.8	26.59	0.7604	25.03	0.6391	24.70	0.5924	23.06	0.6482	24.81	0.7783
		CAJNN (ours, self-ensembled)	2.50	14.8	26.65	0.7633	25.10	0.6404	24.74	0.5936	23.28	0.6550	24.98	0.7820
40	SR	Bicubic	-	-	26.38	0.7154	24.55	0.6201	24.77	0.5898	22.26	0.5877	23.66	0.7081
		EDSR	1.94	43.1	26.01	0.6972	24.48	0.6120	24.62	0.5836	22.18	0.5893	23.73	0.7003
		RCAN	2.04	16	25.70	0.6726	24.30	0.5936	24.36	0.5704	21.86	0.5690	23.13	0.6673
		RRDB	0.65	16.7	25.99	0.6958	24.50	0.6079	24.54	0.5804	22.10	0.5851	23.50	0.6918
	CAR+SR	ARCNN+RRDB	3.20+0.65	0.56+16.7	26.65	0.7495	25.16	0.6424	25.06	0.6053	22.82	0.6235	24.68	0.7578
		DnCNN+RRDB	0.38+0.65	0.06+16.7	26.87	0.7403	25.15	0.6373	25.00	0.5995	22.78	0.6194	24.42	0.7404
	Joint CAR&SR	CAJNN (ours)	0.48	14.8	28.05	0.7981	25.96	0.6729	25.43	0.6240	24.09	0.6962	26.25	0.8177
		CAJNN (ours, self-ensembled)	2.50	14.8	28.16	0.7993	26.03	0.6742	25.46	0.6251	24.31	0.7011	26.44	0.8211

Quantitative comparison of our one-stage CAJNN with other methods. The best two results are highlighted in red and blue colors, respectively. Our method outperforms all two-stage methods in terms of PSNR and SSIM while having a relatively small model size and fast inference speed.

QUALITATIVE COMPARISON ON TEST SET (URBAN100)



Input | Ground Truth | ARCNN +RRDB | RRDB (finetune) | **Ours**

Qualitative comparison of different methods with upscale factor = 4 on the Urban100 dataset. As shown in the images, our model can reconstruct artifact-free high-resolution results with sharper edges and more accurate patterns compared with other methods.

RESULTS OF INPUTS WITH DIFFERENT QUALITY FACTORS

JPEG quality factor	75	80	85	90	95	100

Low Resolution (Input)

Input size	2.14 KB	2.35 KB	2.64 KB	3.10 KB	4.10 KB	6.89 KB

Super Resolution (Our Output)

* Original PNG Size: 116 KB

RESULTS FOR EXTREMELY COMPRESSED IMAGES

JPEG quality factor	20	30	40	50	60	70

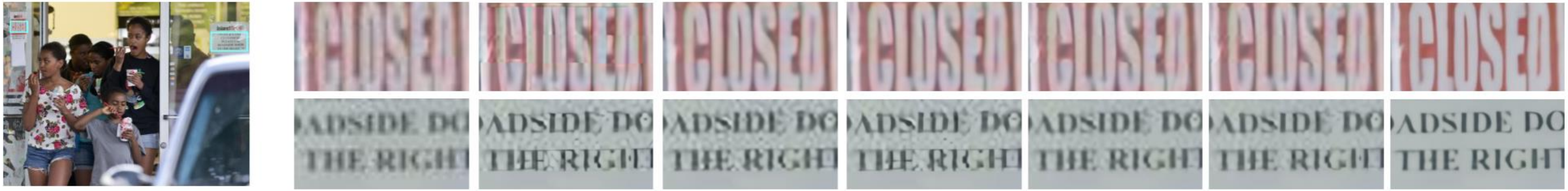
Low Resolution (Input)

Input size	1.21 KB	1.39 KB	1.53 KB	1.67 KB	1.82 KB	2.01 KB

Super Resolution (Our Output)

* Original PNG Size: 116 KB

RESULTS ON IN-THE-WILD IMAGES



Full-size input | Input | RCAN | ARCNN+RRDB | ARCNN+RCAN | DnCNN+RRDB | DnCNN+RCAN | **Ours (CAJNN)**

Observation: The texts become more recognizable.

IMPROVING LOW-RESOLUTION TEXT RECOGNITION

- We compare the total accuracy of generic text detection on the ICDAR2013 Focused Scene Text dataset with TPS-ResNet-BiLSTM-Attn as the text recognition method.

Method	Accuracy	Runtime (s)
Baseline [4]	85.30 %	31.22
Ours + Baseline [4]	**85.75%**	41.56
Ours + Downsample + Baseline [47]	85.57%	31.22

Text recognition accuracy results on the ICDAR 2013 Focused Scene Text dataset. Compared with the baseline method, the introduction of our CARSR method improves the detection performance by 0.45 %. The third row improves 0.27% compared to the baseline due to the reduction of compression artifacts, which indicates that our model is capable of extracting and maintaining critical features of input images.

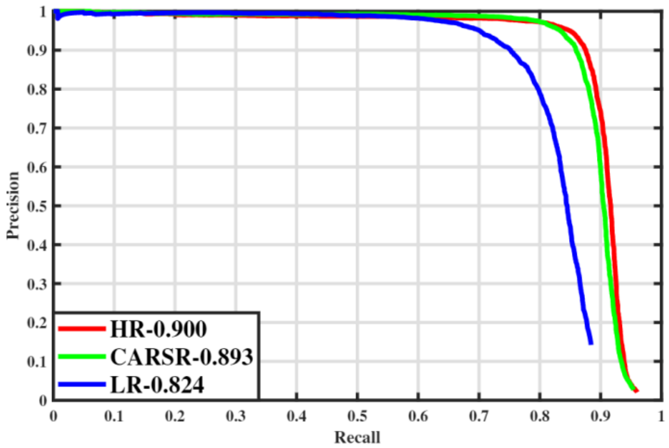
GT Ours(×4) Ours(downsampled)

IMPROVING EXTREMELY TINY FACE DETECTION

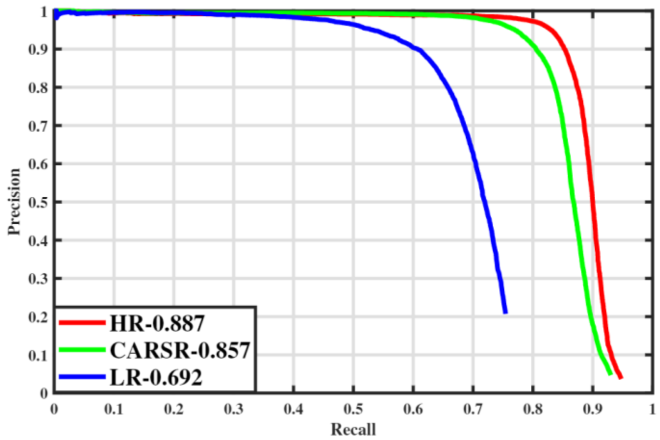
- We compare the average precision and precision-recall curve of face detection on WIDER FACE dataset with Hu et al. as the baseline detector.

	Average precision		
Input Data	Easy	Medium	Hard
GT	0.900	0.887	0.792
LR	0.824	0.692	0.317
LR + Ours	0.893	0.857	0.611

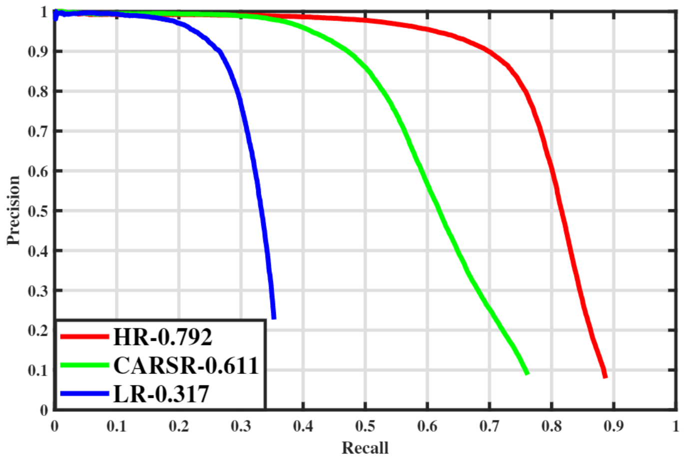
Average precision of three data types on the WIDER FACE validation set with the identical face detector. The application of our CARSR method greatly improves the detection performance with LR images on all three subsets.



Easy



Medium



Hard

CONCLUSION

- We propose a single-stage network for the joint CARSR task to directly reconstruct an artifact-free high-resolution image from a compressed low-resolution input.
- To address the CARSR problem, we make use of the contextual information by introducing a specially designed ASPP that integrates both intra- and inter-block features.
- Our experiments illustrate the effectiveness and efficiency of our method with both standard test images and real-world images.
- Moreover, the extensive experimental results reveal a high potential for enhancing the performance of current methods for various high-level computer vision tasks, e.g. real-scene low resolution text recognition, and extremely tiny face detection.

THANKS

Questions are invited.

Paper: https://arxiv.org/abs/2010.08919
Demo Video: https://tinyurl.com/boost2020icpr
Contact: xiang43@purdue.edu