2022 하계 세미나

Stereo Super Resolution

Sogang University

Vision & Display Systems Lab, Dept. of Electronic Engineering

Presented by

조영수

Outline

- Introduction
  - Stereo Super Resolution
  - PAM (Pallax-attention Module) and epipolar line
- Method
  - PAM
    - Symmetric Parallax Attention for Stereo Image Super-Resolution (CVPRW 2021)
  - Cross-Attention module
    - NAFSSR: Stereo Image Super-Resolution Using NAFNet (CVPR 2022 Oral)
- Results
- Conclusion

Introduction

- Super Resolution
  - Restore High-Resolution (HR) image from Low-Resolution (LR) image
  - Ill-posed problem
    - Multiple solution could be obtained from a pixel of low-resolution image
  - According to the number of LR image
    - SISR (Single Image Super Resolution) / MISR (Multi Image Super Resolution)

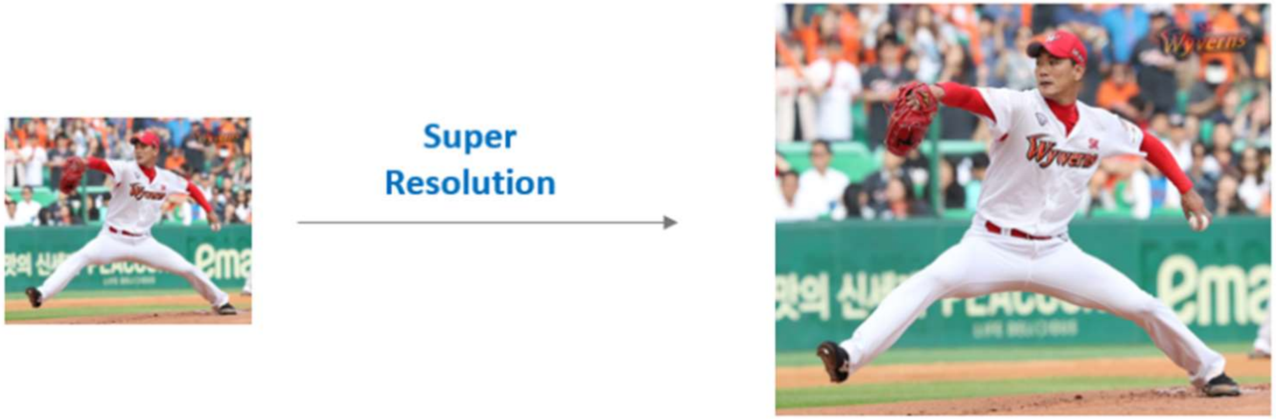


Figure 1: Example of Single Image Super Resolution

[1]Daniel S.Jeon,Seugn-Hwan Baek, et al. " Enhancing the Spatial Resolution of Stereo Images using a Parallax Prior" *CVPR* , 2018.

Introduction

- Stereo Super Resolution
  - Commonly used
    - Mobile phones and autonomous vehicles
  - Image SR and HR depth estimation
    - Jointly estimate the SR image and HR disparity
    - StereoSR limited with the large disparity variations

Figure 2: Example of dual camers

[2]Longguang Wang, et al. “Learning Parallax Attention for Stereo Image Super-Resolution" *CVPR* , 2019.

Background

- Parallax-attention Module (PAM)

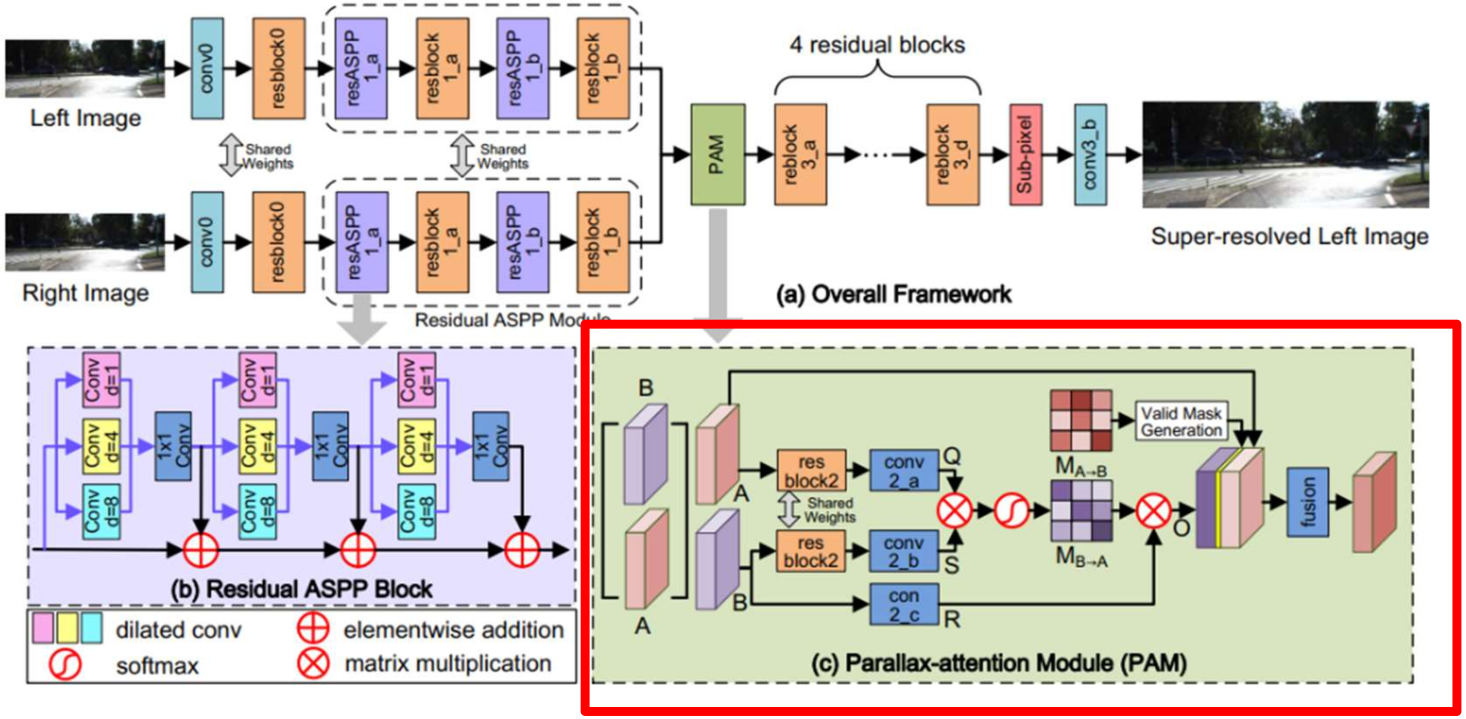


Figure 3: Overview of PASSRnet network

Background

- Parallax-attention Module (PAM)
    - Inspired by self-attention mechanism
    - Capture global correspondence
- Parallax-attention Mechanism
    - Attention map
        - Query feature map, Q and S generated
        - produce parallax attention map $M_{B\to A}$
    - Valid mask
        - $M_{A\to B}$ able to generated when $M_{B\to A}$

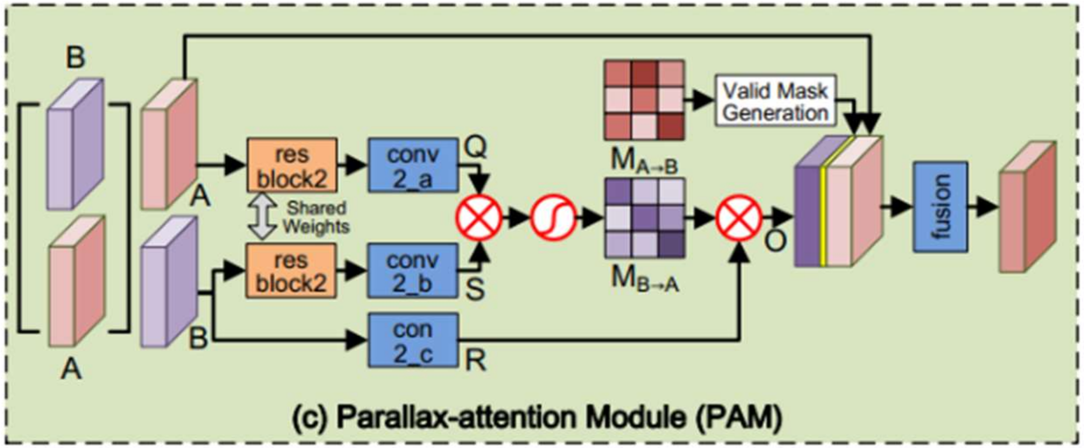


Figure 4: Parallax-attention module

Background

- Parallax-attention Module (PAM)
  - Focus on the most similar feature along the epipolar line
    - Rather than collecting all similar features
  - Parallax-attention map
    - Reflect the correspondence between stereo pairs
    - Encode disparity information

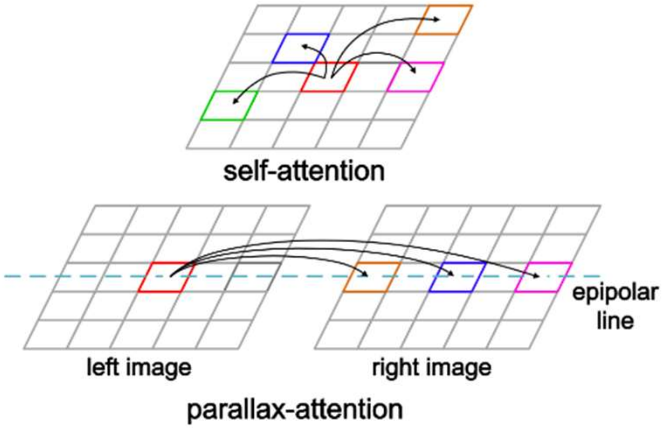


Figure 5: Parallax-attention and self-attention

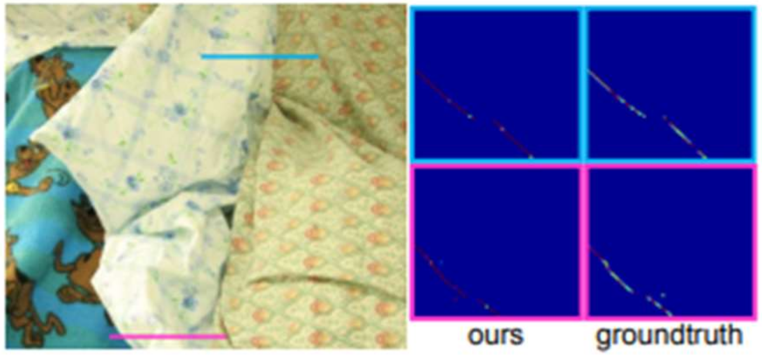


Figure 6: Parallax-attention maps $M_{right \rightarrow left}$

Background

- Epipolar geometry
  - The geometrical relationship between correspondences of image A and B
    - Images of same object or scene acquired from two different points
  - Epipolar line
    - The straight line of intersection of the epipolar plane with the image plane
    - Efficient for 1 D matching

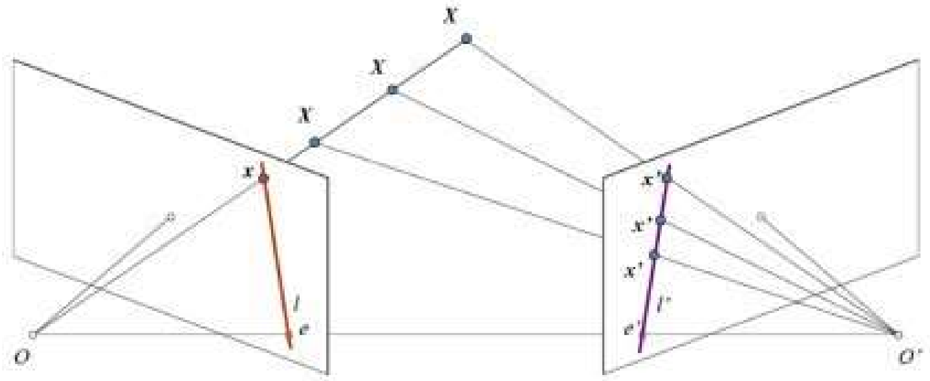


Figure 7: Epipolar Geometry

Background

- Left-right Consistency
  - Obtained if PAM captures accurate correspondence

$$\begin{cases} I^L_{left} = M_{right\to left} \otimes I^L_{right} \\ I^L_{right} = M_{left\to right} \otimes I^L_{left} \end{cases}$$

- Cycle consistency

$$\begin{cases} I^L_{left} = M_{left\to right\to left} \otimes I^L_{right} \\ I^L_{right} = M_{right\to left\to right} \otimes I^L_{left} \end{cases}$$

  - Cycle-attention map

$$\begin{cases} M_{left\to right\to left} = M_{right\to left} \otimes M_{left\to right} \\ M_{right\to left\to right} = M_{left\to right} \otimes M_{right\to left} \end{cases}$$

Background

- Valid masks
  - Occlusion detection method
    - Occluded regions represented with small weights
      - since occluded pixels in the left image not found their correspondence in the right image
  - Guide feature fusion
    - Occluded regions in the left image not able to obtain additional information from the right image

$$V_{left\rightarrow righ}(i,j) = \begin{cases} 1, & if \sum_{k\in[1,W]} M_{left\rightarrow right}\,(i,k,k) > \tau \\ 0, & otherwise \end{cases}$$

[3]Yingqian Wang,Xinyi Ying, et al. “Symmetric Parallax Attention for Stereo Image Super-Resolution“, *CVPRW*. 2021.

iPASSR[3]

- Contributions
  1. Exploit symmetric cues for stereo image SR
  2. A symmetric and bi-directional parallax attention module

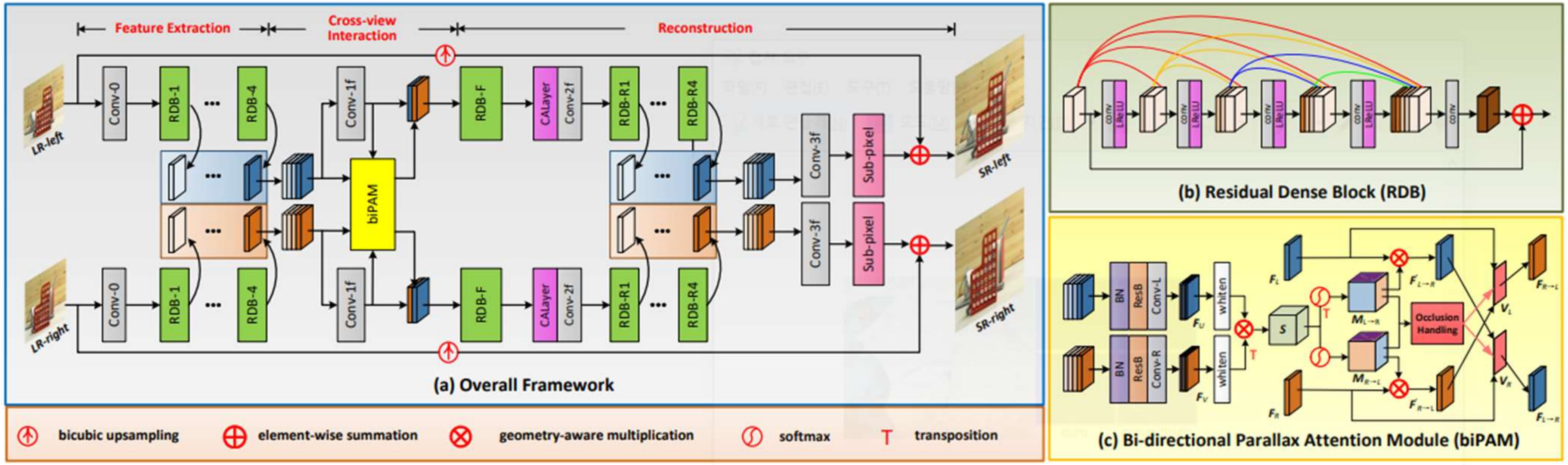


Figure 8: An overview of iPASSR network

iPASSR

- Methods
  - Feature Extraction
    - Features from all the layers concatenated and fed to a 1x1 convolution
      - Generate fused features for local residual connection

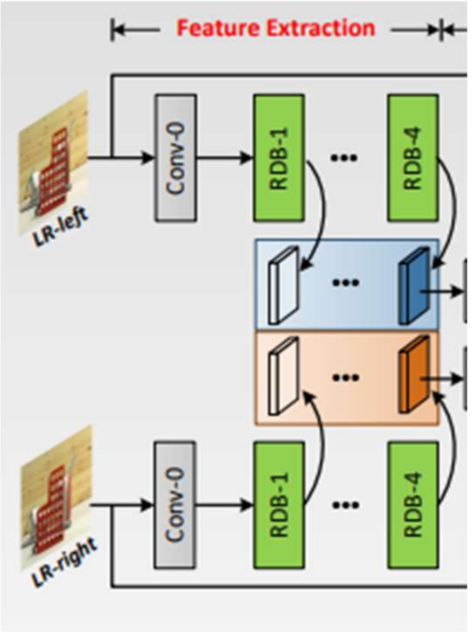


Figure 9: Feature extraction

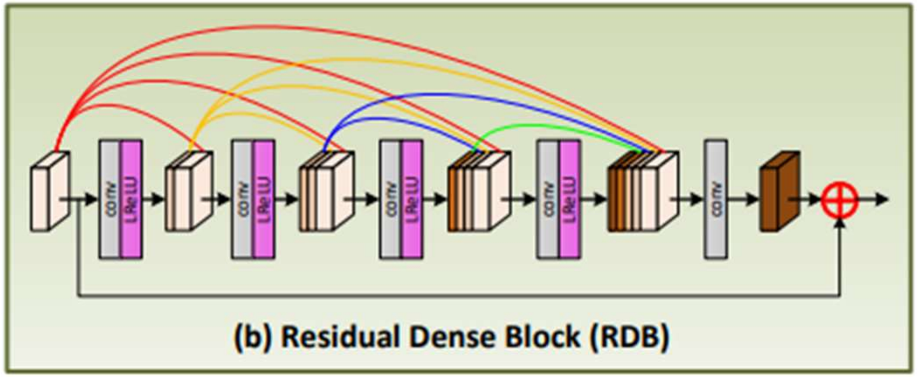


Figure 10: Residual Dense Block

iPASSR

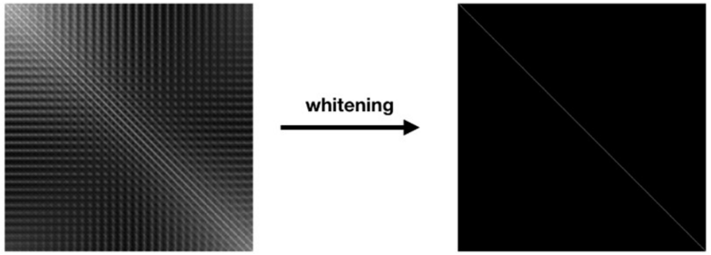


Figure 11: Whitening procedure

- Cross-view interaction
  - Generated F_U, F_V with the input stereo features
    - Batch-normalization (BN) layer, transition residual block, and separately fed to a 1x1 convolution
  - Whiten layer
    - Obtain normalized features F'_U, F'_V

$$\mathbf{F}'_{\mathrm{U}}(h,w,c) = \mathbf{F}_{\mathrm{U}}(h,w,c) - \frac{1}{W}\sum_{i=1}^{W} \mathbf{F}_{\mathrm{U}}(h,i,c),$$

$$\mathbf{F}'_{\mathrm{V}}(h,w,c) = \mathbf{F}_{\mathrm{V}}(h,w,c) - \frac{1}{W}\sum_{i=1}^{W} \mathbf{F}_{\mathrm{V}}(h,i,c).$$

  - Attention map
    - Initial score map S produced
      - F'_V transposed then batch-wise multiplied with F'_U
    - Attention maps $M_{R\to L,}M_{L\to R}$
      - Softmax normalization applied to S and S^T

iPASSR

- Cross-view interaction
  - Achieved cross-view interaction
    - Batch-wise multiplication with the corresponding attention maps

$$\mathbf{F'}_{\mathrm{R}\to\mathrm{L}} = \mathbf{M}_{\mathrm{R}\to\mathrm{L}} \otimes \mathbf{F}_{\mathrm{R}},$$

$$\mathbf{F'}_{\mathrm{L}\to\mathrm{R}} = \mathbf{M}_{\mathrm{L}\to\mathrm{R}} \otimes \mathbf{F}_{\mathrm{L}},$$

- Inline occlusion handling scheme
  - Avoid unreliable correspondence in occlusion and boundary regions
    - Calculate valid masks V_L and V_R
    - Final converted features $\mathrm{F}_{L\to R}, \mathrm{F}_{R\to L}$

$$\mathbf{F}_{\mathrm{R}\to\mathrm{L}} = \mathbf{V}_{\mathrm{L}} \odot \mathbf{F'}_{\mathrm{R}\to\mathrm{L}} + (\mathbf{1} - \mathbf{V}_{\mathrm{L}}) \odot \mathbf{F}_{\mathrm{L}},$$

$$\mathbf{F}_{\mathrm{L}\to\mathrm{R}} = \mathbf{V}_{\mathrm{R}} \odot \mathbf{F'}_{\mathrm{L}\to\mathrm{R}} + (\mathbf{1} - \mathbf{V}_{\mathrm{R}}) \odot \mathbf{F}_{\mathrm{R}},$$

iPASSR

- Reconstruction
  - Similar to the feature extraction
    - Residual dense block (RDB) as the basic block
    - Combination of RDBs , Channel Attention (CA), and sup-pixel layer to generate super-resolved image

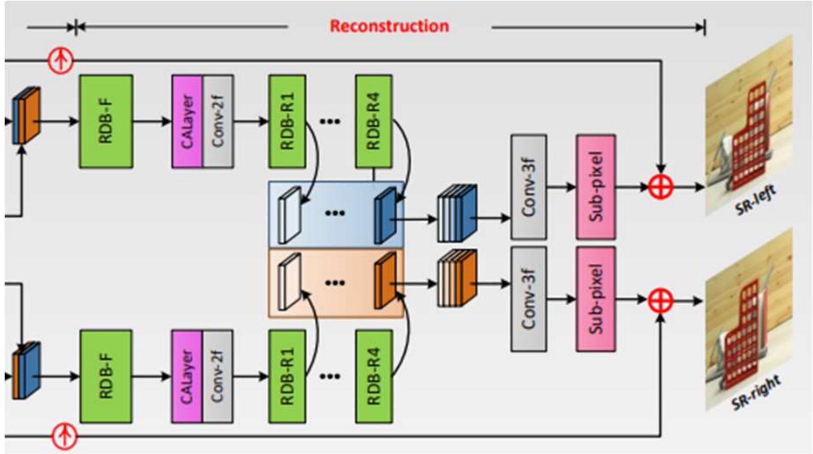
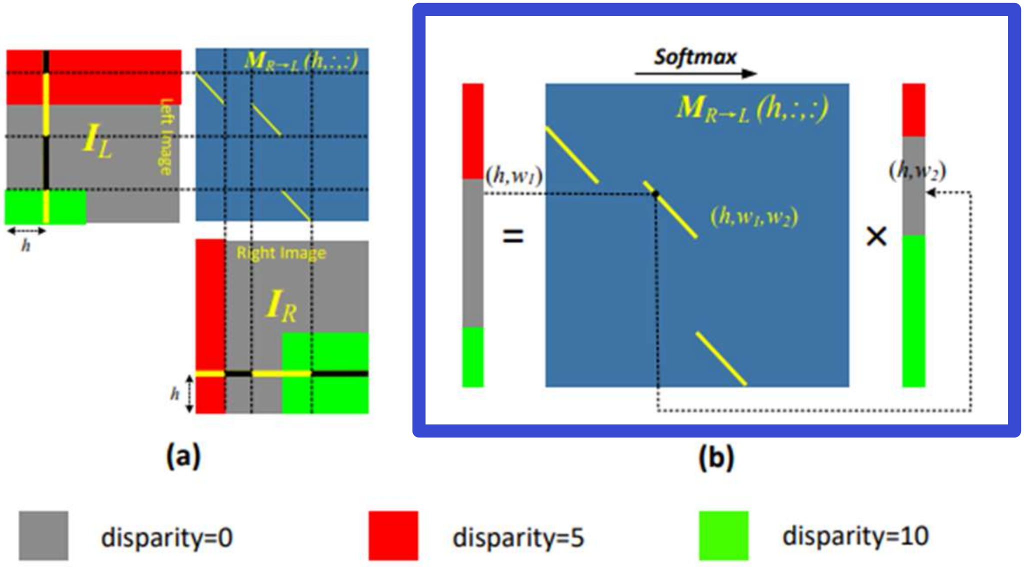


Figure 12: Reconstruction

iPASSR

- Inline Occlusion Handling Scheme
  - Occlusion derived
    - By checking the stereo consistency using the attention maps
  - Toy example
    - How occlusion implicitly encoded in the parallax attention maps



$$I_{R\to L} = M_{R\to L} \otimes I_R$$

Figure 12: Reconstruction

iPASSR

- Inline Occlusion Handling Scheme
  - $P_L(h, w_1)$ represent the possibility
    - $I_L(h, w_1)$ converted to I_R and re-converted to $I_L(h, w_1)$

$$\mathbf{P}_{\mathrm{L}}(h, w_1) = \sum_{w_2=1}^{W} \mathbf{M}_{\mathrm{R}\rightarrow\mathrm{L}}(h, w_1, w_2) \cdot \mathbf{M}_{\mathrm{L}\rightarrow\mathrm{R}}(h, w_2, w_1).$$

  - Valid masks

$$V_L = \tanh(\tau P'_L), \qquad for\ left\ valid\ mask$$

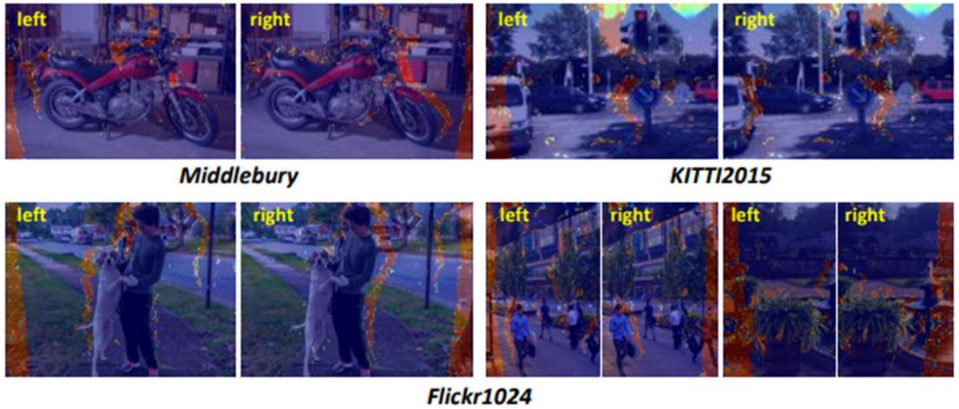


Figure 13: An illustration of valid masks

iPASSR

- Total Losses
  - SR, residual photometric, residual cycle , smoothness , and residual stereo consistency losses

$$\mathcal{L} = \mathcal{L}_{\text{SR}} + \lambda(\mathcal{L}^{\text{res}}_{\text{photo}} + \mathcal{L}^{\text{res}}_{\text{cycle}} + \mathcal{L}_{\text{smooth}} + \mathcal{L}^{\text{res}}_{\text{cons}})$$

  - SR losses
    - L1 distance between the SR and GT stereo images

$$\mathcal{L}_{\text{SR}} = \| \mathbf{I}^{\text{SR}}_{\text{L}} - \mathbf{I}^{\text{HR}}_{\text{L}} \|_1 + \| \mathbf{I}^{\text{SR}}_{\text{R}} - \mathbf{I}^{\text{HR}}_{\text{R}} \|_1$$

  - Residual photometric & cycle losses
    - Illuminance intensity vary significantly
      - Exposure difference and non-Lambertain surfaces
    - Used residual images to improve the robustness

$$\mathrm{X_L} = \left|\mathrm{I_L^{HR}} - \mathrm{I_L^{lR}} \uparrow\right|_{\downarrow}, \qquad \mathrm{X_R} = \left|\mathrm{I_R^{HR}} - \mathrm{I_R^{lR}} \uparrow\right|_{\downarrow}$$

    - $\mathrm{X_L}$ *and* $\mathrm{X_R}$ represent the absolute values of the left and right residual images

iPASSR

- Residual photometric and cycle losses
  - Benefits
    - More consistent and illuminance-robust stereo correspondence
    - Pay more attention to texture-rich regions

$$
\begin{aligned}
\mathcal{L}_{\text{photo}}^{\text{res}} = & \parallel V_{\mathrm{L}} \odot (\mathbf{X}_{\mathrm{L}} - \mathbf{M}_{\mathrm{R}\rightarrow\mathrm{L}} \otimes \mathbf{X}_{\mathrm{R}}) \parallel_1 \\
& + \parallel V_{\mathrm{R}} \odot (\mathbf{X}_{\mathrm{R}} - \mathbf{M}_{\mathrm{L}\rightarrow\mathrm{R}} \otimes \mathbf{X}_{\mathrm{L}}) \parallel_1,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L}_{\text{cycle}}^{\text{res}} = & \parallel V_{\mathrm{L}} \odot (\mathbf{X}_{\mathrm{L}} - \mathbf{M}_{\mathrm{R}\rightarrow\mathrm{L}} \otimes \mathbf{M}_{\mathrm{L}\rightarrow\mathrm{R}} \otimes \mathbf{X}_{\mathrm{L}}) \parallel_1 \\
& + \parallel V_{\mathrm{R}} \odot (\mathbf{X}_{\mathrm{R}} - \mathbf{M}_{\mathrm{L}\rightarrow\mathrm{R}} \otimes \mathbf{M}_{\mathrm{R}\rightarrow\mathrm{L}} \otimes \mathbf{X}_{\mathrm{R}}) \parallel_1
\end{aligned}
$$

iPASSR

- Smoothness loss
  - Encourage smoothness in correspondence space

$$\mathcal{L}_{\text{smooth}} = \sum_{\mathbf{M}} \sum_{i,j,k} (\| \mathbf{M}(i,j,k) - \mathbf{M}(i+1,j,k) \|_1 + \| \mathbf{M}(i,j,k) - \mathbf{M}(i,j+1,k+1) \|_1),$$

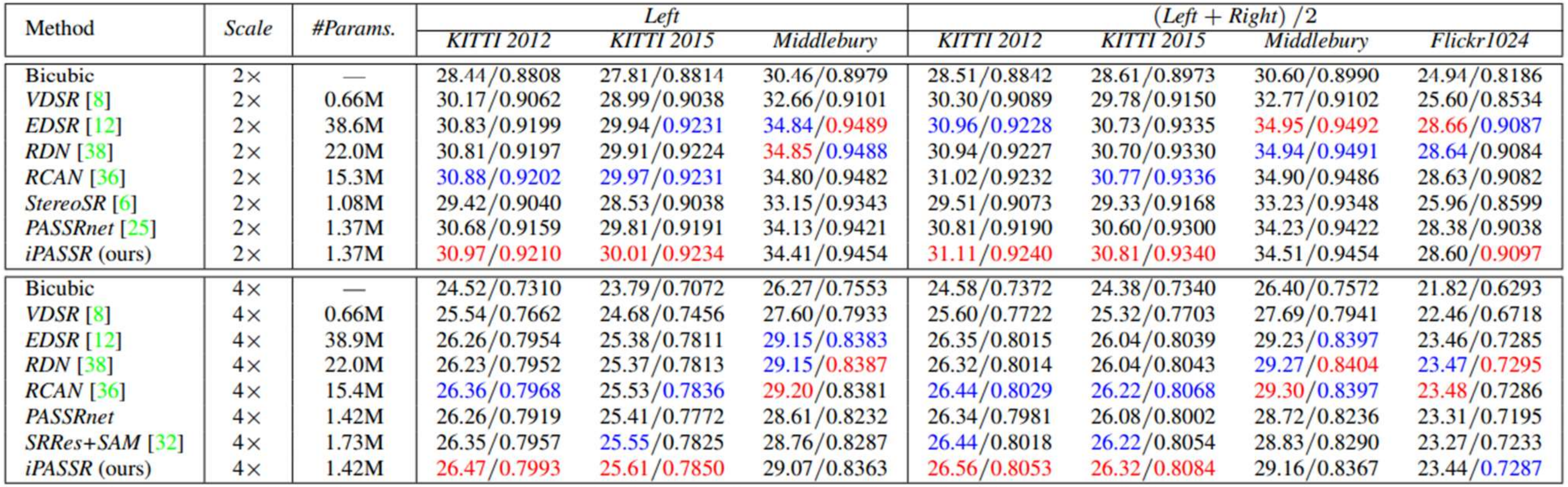
- Residual stereo consistency loss
  - LR residuals between super-resolved images and ground truth images

$$\mathrm{Y_L} = \left|\mathrm{I_L^{HR}} - \mathrm{I_L^{SR}}\right|_{\downarrow}, \qquad \mathrm{X_R} = \left|\mathrm{I_R^{HR}} - \mathrm{I_R^{SR}}\right|_{\downarrow}$$

$$\mathcal{L}_{\text{cons}}^{\text{res}} = \| V_{\mathrm{L}} \odot (\mathbf{Y}_{\mathrm{L}} - \mathbf{M}_{\mathrm{R}\rightarrow\mathrm{L}} \otimes \mathbf{Y}_{\mathrm{R}}) \|_1 + \| V_{\mathrm{R}} \odot (\mathbf{Y}_{\mathrm{R}} - \mathbf{M}_{\mathrm{L}\rightarrow\mathrm{R}} \otimes \mathbf{Y}_{\mathrm{L}}) \|_1 \ .$$

Results

- Qualitative results



Method	*Scale*	*#Params.*	*Left*			*(Left + Right) /2*			
			KITTI 2012	*KITTI 2015*	*Middlebury*	*KITTI 2012*	*KITTI 2015*	*Middlebury*	*Flickr1024*
Bicubic	2×	—	28.44/0.8808	27.81/0.8814	30.46/0.8979	28.51/0.8842	28.61/0.8973	30.60/0.8990	24.94/0.8186
VDSR [8]	2×	0.66M	30.17/0.9062	28.99/0.9038	32.66/0.9101	30.30/0.9089	29.78/0.9150	32.77/0.9102	25.60/0.8534
EDSR [12]	2×	38.6M	30.83/0.9199	29.94/0.9231	34.84/0.9489	30.96/0.9228	30.73/0.9335	34.95/0.9492	28.66/0.9087
RDN [38]	2×	22.0M	30.81/0.9197	29.91/0.9224	34.85/0.9488	30.94/0.9227	30.70/0.9330	34.94/0.9491	28.64/0.9084
RCAN [36]	2×	15.3M	30.88/0.9202	29.97/0.9231	34.80/0.9482	31.02/0.9232	30.77/0.9336	34.90/0.9486	28.63/0.9082
StereoSR [6]	2×	1.08M	29.42/0.9040	28.53/0.9038	33.15/0.9343	29.51/0.9073	29.33/0.9168	33.23/0.9348	25.96/0.8599
PASSRnet [25]	2×	1.37M	30.68/0.9159	29.81/0.9191	34.13/0.9421	30.81/0.9190	30.60/0.9300	34.23/0.9422	28.38/0.9038
iPASSR (ours)	2×	1.37M	30.97/0.9210	30.01/0.9234	34.41/0.9454	31.11/0.9240	30.81/0.9340	34.51/0.9454	28.60/0.9097
Bicubic	4×	—	24.52/0.7310	23.79/0.7072	26.27/0.7553	24.58/0.7372	24.38/0.7340	26.40/0.7572	21.82/0.6293
VDSR [8]	4×	0.66M	25.54/0.7662	24.68/0.7456	27.60/0.7933	25.60/0.7722	25.32/0.7703	27.69/0.7941	22.46/0.6718
EDSR [12]	4×	38.9M	26.26/0.7954	25.38/0.7811	29.15/0.8383	26.35/0.8015	26.04/0.8039	29.23/0.8397	23.46/0.7285
RDN [38]	4×	22.0M	26.23/0.7952	25.37/0.7813	29.15/0.8387	26.32/0.8014	26.04/0.8043	29.27/0.8404	23.47/0.7295
RCAN [36]	4×	15.4M	26.36/0.7968	25.53/0.7836	29.20/0.8381	26.44/0.8029	26.22/0.8068	29.30/0.8397	23.48/0.7286
PASSRnet	4×	1.42M	26.26/0.7919	25.41/0.7772	28.61/0.8232	26.34/0.7981	26.08/0.8002	28.72/0.8236	23.31/0.7195
SRRes+SAM [32]	4×	1.73M	26.35/0.7957	25.55/0.7825	28.76/0.8287	26.44/0.8018	26.22/0.8054	28.83/0.8290	23.27/0.7233
iPASSR (ours)	4×	1.42M	26.47/0.7993	25.61/0.7850	29.07/0.8363	26.56/0.8053	26.32/0.8084	29.16/0.8367	23.44/0.7287

Figure 14: Quantitative results achieved by different methods

Results

- Visual results

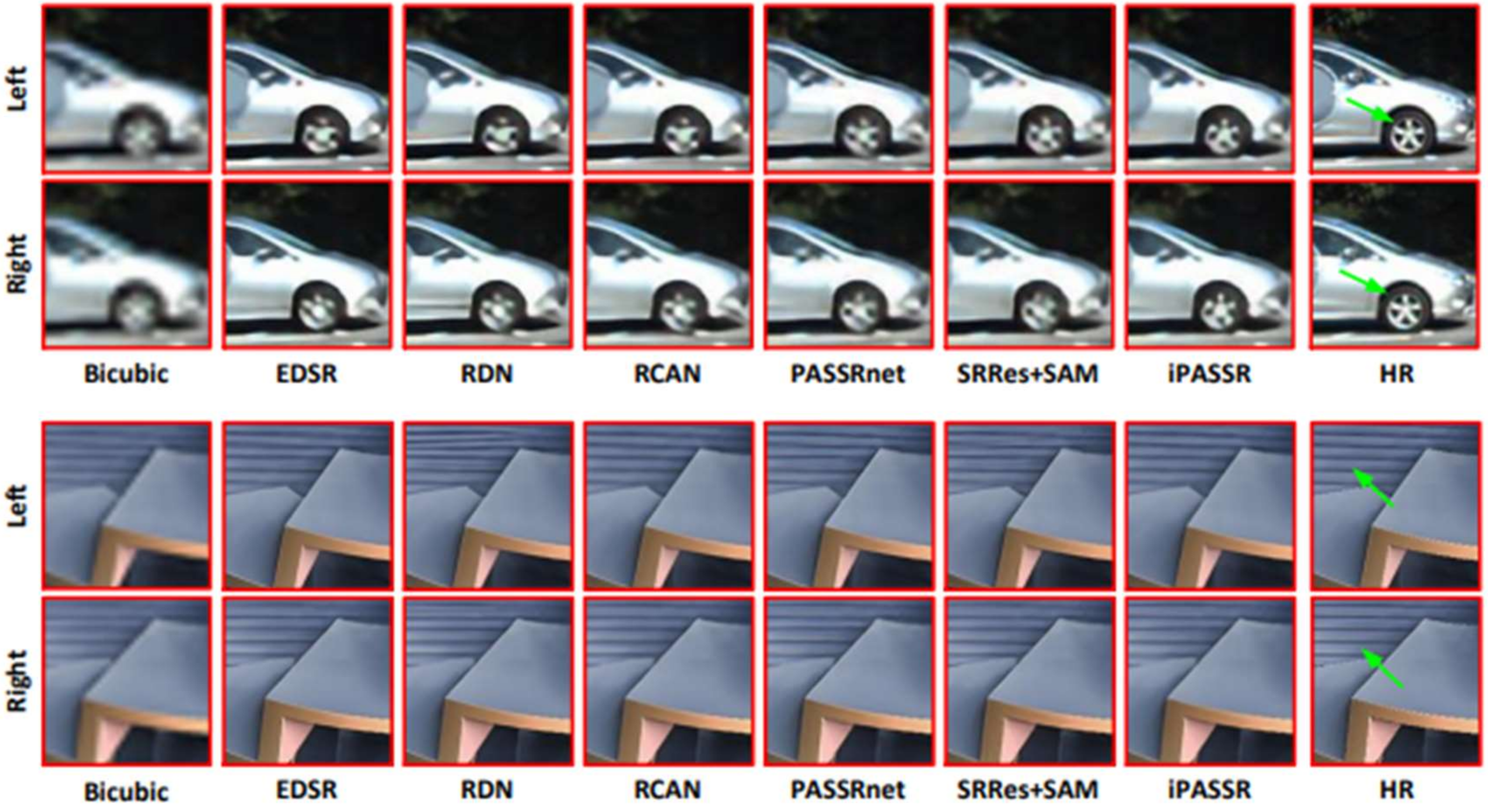


Figure 15: Visual results (4 X) achieved by different methods

Results

- Visual results

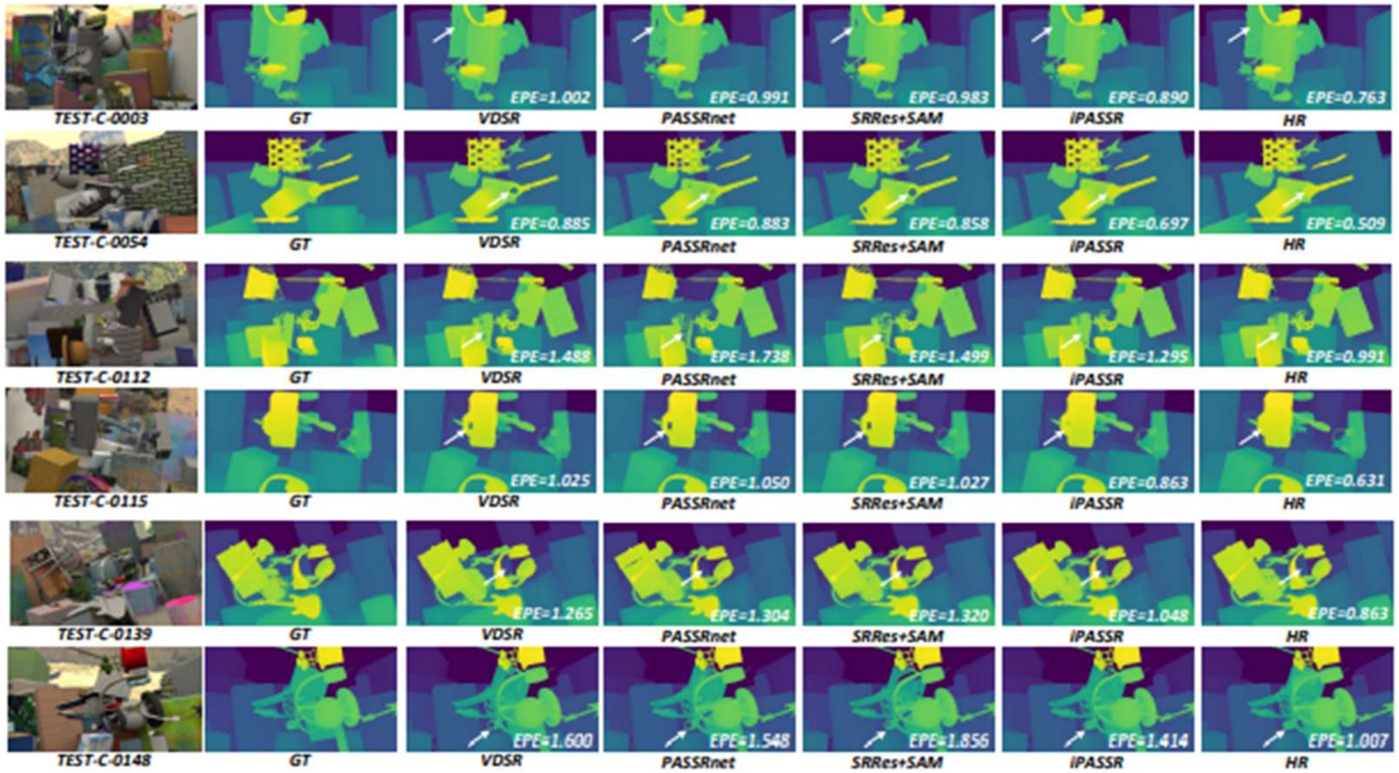


Figure 16: Qualitative results achieved by GwcNet using 4xSR stereo images generated by different SR methods

NAFSSR[4]

- Contributions
  - 1st Place in NTIRE 2022 Stereo Image Super-resolution Challenge
  - NAFSSR
    - SOTA performance with fewer parameter
    - Faster inference
    - Representation through a simple stereo crosse attention module
- Overview

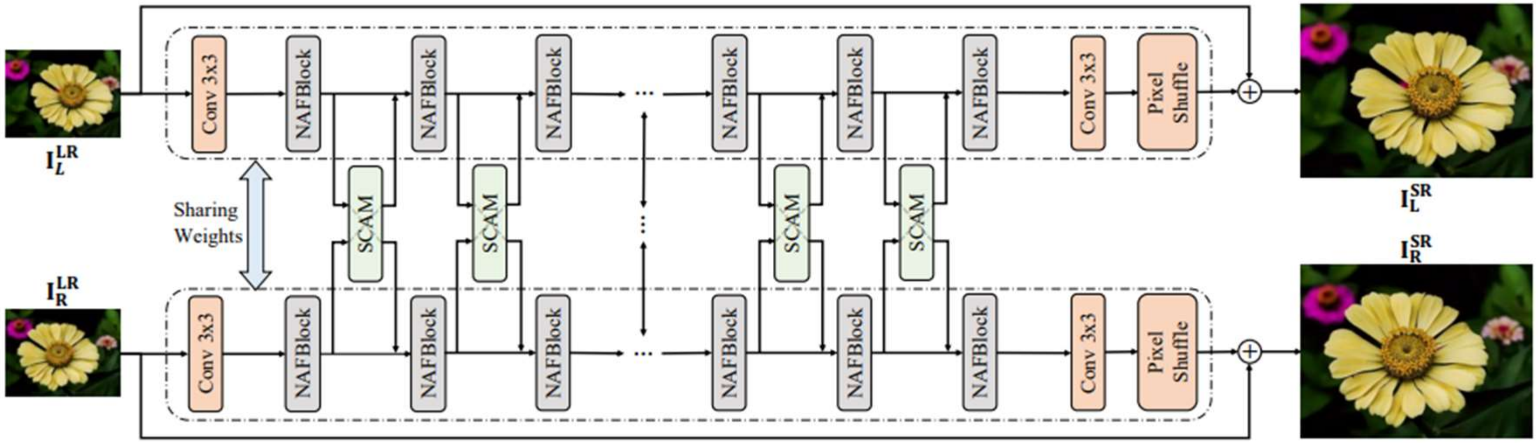


Figure 17: Overall architecture of NAFSSR

NAFSSR

- NAFBlock

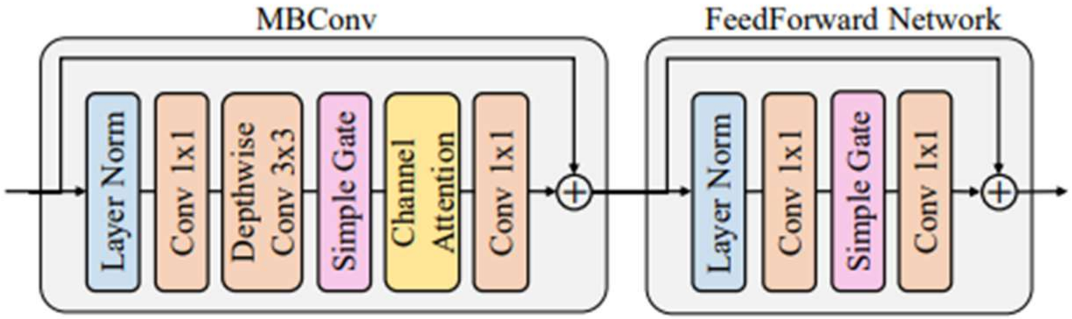


Figure 18: Architecture of NAFBlock

1. Mobile convolution module (MBConv)
   - Based on point-wise and depth-wise convolution with channel attention
2. Feed-forward network (FFN)
   - Implemented by point-wise convolution

- Simple gate mechanism
  - Makes block nonlinear activation free
  - Replaced nonlinear activation (ReLU, GELU)

$$\mathrm{SimpleGate(X)} = X_1 \odot X_2$$

NAFSSR

- Stereo Cross Attention Module (SCAM)
  - Scaled dot-Product Attention

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{softmax}\left(\mathbf{Q}\mathbf{K}^T/\sqrt{C}\right)\mathbf{V}$$

    - Query matrix projected by source intra-view feature
    - Key, Value matrices projected by target intra-view feature
  - Highly symmetric under epipolar constraint
    - Same Q and K to represent each intra-view features
    - Calculate correlation of cross-view features on horizontal line

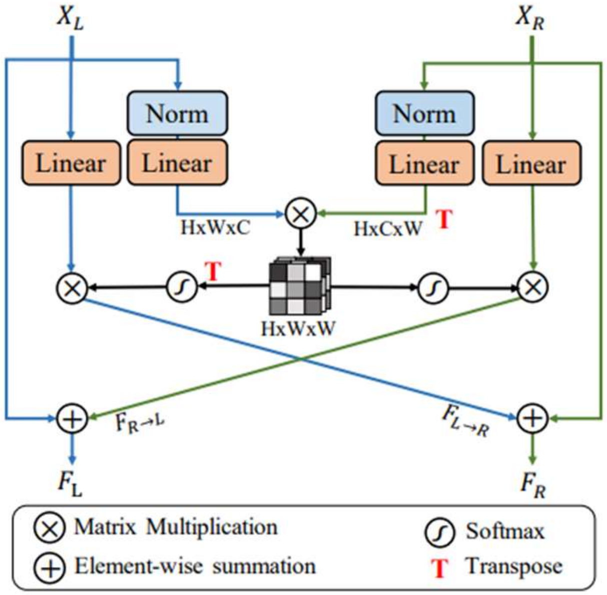


Figure 19: Stereo Cross Attention Module

$$\mathbf{F}_{\mathbf{R}\to\mathbf{L}} = \text{Attention}(\mathbf{W}_1^{\mathbf{L}}\bar{\mathbf{X}}_{\mathbf{L}}, \mathbf{W}_1^{\mathbf{R}}\bar{\mathbf{X}}_{\mathbf{R}}, \mathbf{W}_2^{\mathbf{R}}\mathbf{X}_{\mathbf{R}}),$$
$$\mathbf{F}_{\mathbf{L}\to\mathbf{R}} = \text{Attention}(\mathbf{W}_1^{\mathbf{R}}\bar{\mathbf{X}}_{\mathbf{R}}, \mathbf{W}_1^{\mathbf{L}}\bar{\mathbf{X}}_{\mathbf{L}}, \mathbf{W}_2^{\mathbf{L}}\mathbf{X}_{\mathbf{L}}),$$

  - Fusion

$$\mathbf{F}_{\mathbf{L}} = \gamma_L \mathbf{F}_{\mathbf{R}\to\mathbf{L}} + \mathbf{X}_{\mathbf{L}},$$
$$\mathbf{F}_{\mathbf{R}} = \gamma_R \mathbf{F}_{\mathbf{L}\to\mathbf{R}} + \mathbf{X}_{\mathbf{R}},$$

NAFSSR

- Training Strategies
  - Super-Resolution
    - Train models with small patches cropped from full-resolution images
  - Data augmentation
    - Horizontally and vertically flipped
    - Channel shuffle
- Loss
  - Pixel-wise L1 distance

$$\mathcal{L} = \left\|\mathbf{I}_{\mathrm{L}}^{\mathrm{SR}} - \mathbf{I}_{\mathrm{L}}^{\mathrm{HR}}\right\|_1 + \left\|\mathbf{I}_{\mathrm{R}}^{\mathrm{SR}} - \mathbf{I}_{\mathrm{R}}^{\mathrm{HR}}\right\|_1$$

NAFSSR

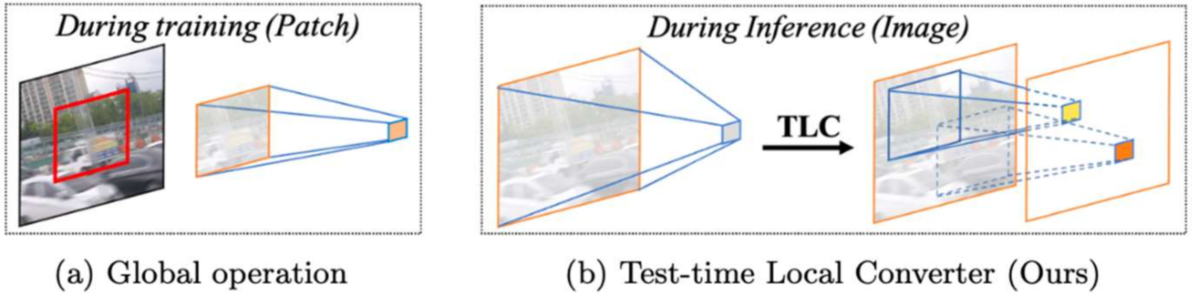


Figure 20: Test-Time Local Converter

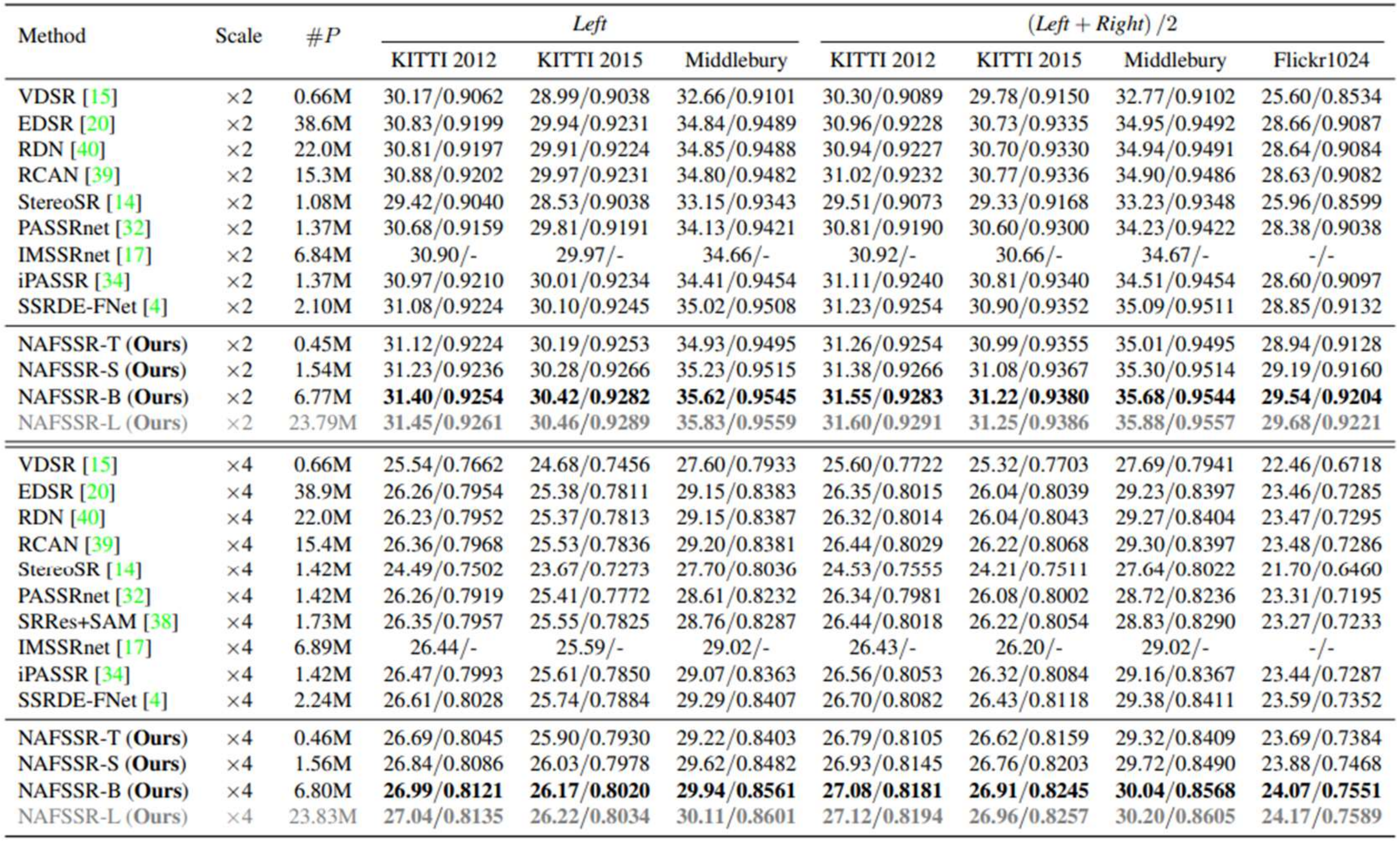
- Train-test Inconsistency
  - Train: patch-based features
  - Inference: image-based features
  - For stereo super-resolution task
    - Regional range of inputs for training and inference varies greatly (patch only 4.5% of LR images)
  - Channel attention
    - Aggregate global spatial information
    - Redistributes the pooled information to input features

$$\mathrm{CA}(\mathbf{X}) = \mathbf{X} * \mathbf{W}\,\mathrm{pool}(\mathbf{X}),$$

  - Apply TLSC[5]
    - Converts global average pooling to local average pooling during inference
    - Extract representations based on local spatial region of features as in training phase

NAFSSR

- Quantitative results



Method	Scale	#P	*Left*			(*Left* + *Right*) /2			
			KITTI 2012	KITTI 2015	Middlebury	KITTI 2012	KITTI 2015	Middlebury	Flickr1024
VDSR [15]	×2	0.66M	30.17/0.9062	28.99/0.9038	32.66/0.9101	30.30/0.9089	29.78/0.9150	32.77/0.9102	25.60/0.8534
EDSR [20]	×2	38.6M	30.83/0.9199	29.94/0.9231	34.84/0.9489	30.96/0.9228	30.73/0.9335	34.95/0.9492	28.66/0.9087
RDN [40]	×2	22.0M	30.81/0.9197	29.91/0.9224	34.85/0.9488	30.94/0.9227	30.70/0.9330	34.94/0.9491	28.64/0.9084
RCAN [39]	×2	15.3M	30.88/0.9202	29.97/0.9231	34.80/0.9482	31.02/0.9232	30.77/0.9336	34.90/0.9486	28.63/0.9082
StereoSR [14]	×2	1.08M	29.42/0.9040	28.53/0.9038	33.15/0.9343	29.51/0.9073	29.33/0.9168	33.23/0.9348	25.96/0.8599
PASSRnet [32]	×2	1.37M	30.68/0.9159	29.81/0.9191	34.13/0.9421	30.81/0.9190	30.60/0.9300	34.23/0.9422	28.38/0.9038
IMSSRnet [17]	×2	6.84M	30.90/-	29.97/-	34.66/-	30.92/-	30.66/-	34.67/-	-/-
iPASSR [34]	×2	1.37M	30.97/0.9210	30.01/0.9234	34.41/0.9454	31.11/0.9240	30.81/0.9340	34.51/0.9454	28.60/0.9097
SSRDE-FNet [4]	×2	2.10M	31.08/0.9224	30.10/0.9245	35.02/0.9508	31.23/0.9254	30.90/0.9352	35.09/0.9511	28.85/0.9132
NAFSSR-T (**Ours**)	×2	0.45M	31.12/0.9224	30.19/0.9253	34.93/0.9495	31.26/0.9254	30.99/0.9355	35.01/0.9495	28.94/0.9128
NAFSSR-S (**Ours**)	×2	1.54M	31.23/0.9236	30.28/0.9266	35.23/0.9515	31.38/0.9266	31.08/0.9367	35.30/0.9514	29.19/0.9160
NAFSSR-B (**Ours**)	×2	6.77M	**31.40/0.9254**	**30.42/0.9282**	**35.62/0.9545**	**31.55/0.9283**	**31.22/0.9380**	**35.68/0.9544**	**29.54/0.9204**
NAFSSR-L (Ours)	×2	23.79M	31.45/0.9261	30.46/0.9289	35.83/0.9559	31.60/0.9291	31.25/0.9386	35.88/0.9557	29.68/0.9221
VDSR [15]	×4	0.66M	25.54/0.7662	24.68/0.7456	27.60/0.7933	25.60/0.7722	25.32/0.7703	27.69/0.7941	22.46/0.6718
EDSR [20]	×4	38.9M	26.26/0.7954	25.38/0.7811	29.15/0.8383	26.35/0.8015	26.04/0.8039	29.23/0.8397	23.46/0.7285
RDN [40]	×4	22.0M	26.23/0.7952	25.37/0.7813	29.15/0.8387	26.32/0.8014	26.04/0.8043	29.27/0.8404	23.47/0.7295
RCAN [39]	×4	15.4M	26.36/0.7968	25.53/0.7836	29.20/0.8381	26.44/0.8029	26.22/0.8068	29.30/0.8397	23.48/0.7286
StereoSR [14]	×4	1.42M	24.49/0.7502	23.67/0.7273	27.70/0.8036	24.53/0.7555	24.21/0.7511	27.64/0.8022	21.70/0.6460
PASSRnet [32]	×4	1.42M	26.26/0.7919	25.41/0.7772	28.61/0.8232	26.34/0.7981	26.08/0.8002	28.72/0.8236	23.31/0.7195
SRRes+SAM [38]	×4	1.73M	26.35/0.7957	25.55/0.7825	28.76/0.8287	26.44/0.8018	26.22/0.8054	28.83/0.8290	23.27/0.7233
IMSSRnet [17]	×4	6.89M	26.44/-	25.59/-	29.02/-	26.43/-	26.20/-	29.02/-	-/-
iPASSR [34]	×4	1.42M	26.47/0.7993	25.61/0.7850	29.07/0.8363	26.56/0.8053	26.32/0.8084	29.16/0.8367	23.44/0.7287
SSRDE-FNet [4]	×4	2.24M	26.61/0.8028	25.74/0.7884	29.29/0.8407	26.70/0.8082	26.43/0.8118	29.38/0.8411	23.59/0.7352
NAFSSR-T (**Ours**)	×4	0.46M	26.69/0.8045	25.90/0.7930	29.22/0.8403	26.79/0.8105	26.62/0.8159	29.32/0.8409	23.69/0.7384
NAFSSR-S (**Ours**)	×4	1.56M	26.84/0.8086	26.03/0.7978	29.62/0.8482	26.93/0.8145	26.76/0.8203	29.72/0.8490	23.88/0.7468
NAFSSR-B (**Ours**)	×4	6.80M	**26.99/0.8121**	**26.17/0.8020**	**29.94/0.8561**	**27.08/0.8181**	**26.91/0.8245**	**30.04/0.8568**	**24.07/0.7551**
NAFSSR-L (Ours)	×4	23.83M	27.04/0.8135	26.22/0.8034	30.11/0.8601	27.12/0.8194	26.96/0.8257	30.20/0.8605	24.17/0.7589

Figure 21: Quantitative results achieved by different methods

Results

- Visual results

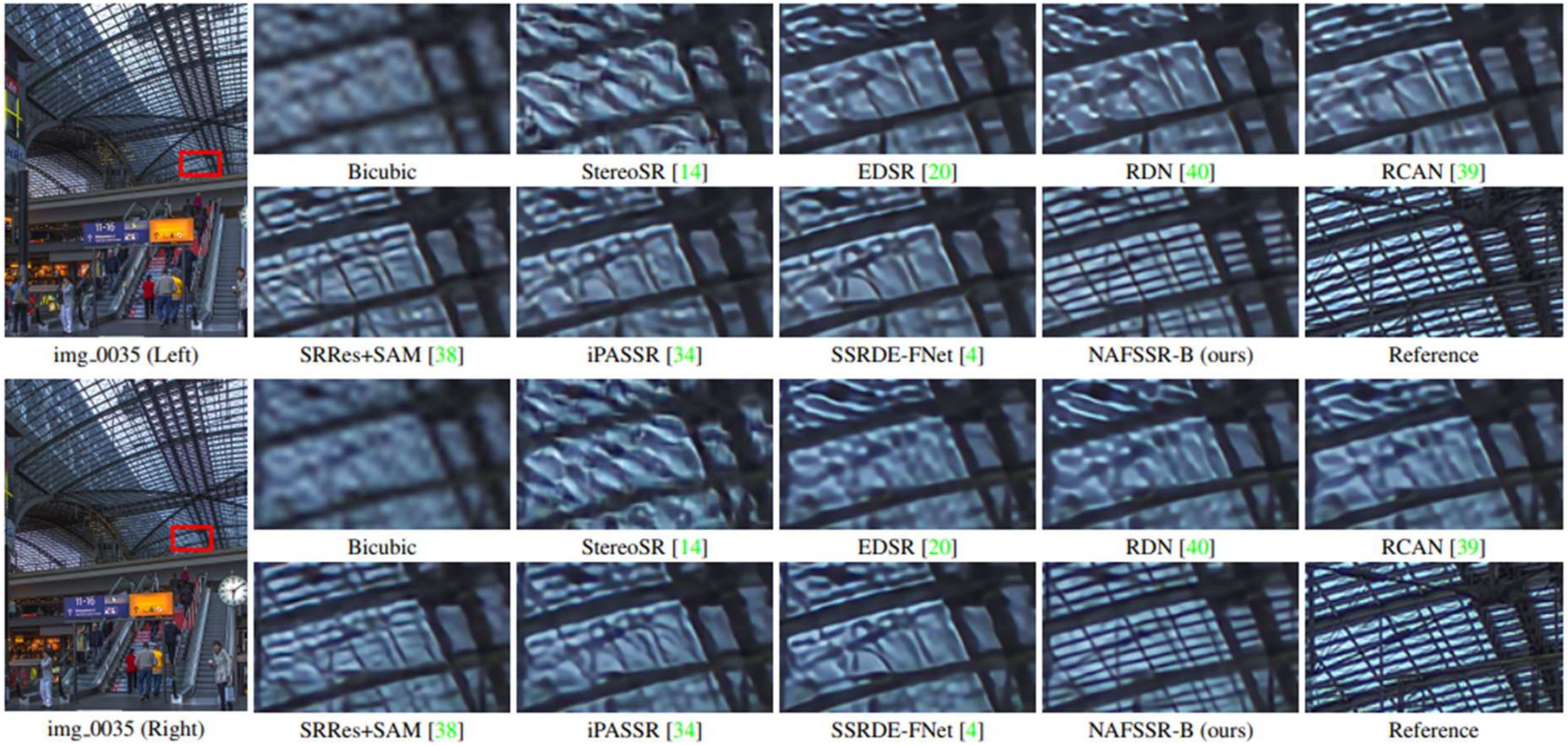


Figure 23: Visual results achieved by different methods

Results

- Visual results

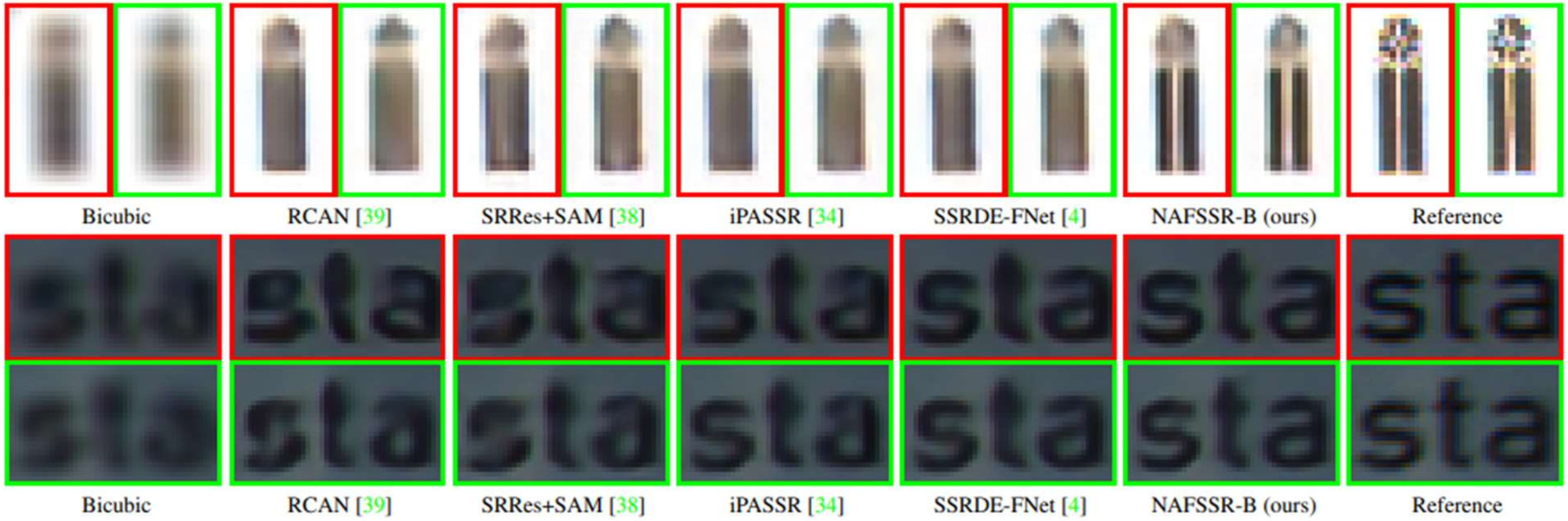


Figure 23: Visual results achieved by different methods

Conclusion

- Stereo Super Resolution task
  - Super resolution task + stereo matching task
  - The cross-view information important for performance
- iPASSR
  - A bi-directional parallax attention module (biPAM)
  - An inline occlusion handling scheme
  - Residual losses to achieve robustness to illuminance changes
- NAFSSR
  - NAFBlcok for intra-view feature extraction
  - Stereo cross attention (SCAM) for cross-view feature
  - Solve the train-test inconsistency

감사합니다