JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

DINN360: Deformable Invertible Neural Network for Latitude-aware 360◦ Image Rescaling

Yichen Guo [†], Mai Xu[†], Lai Jiang[†‡], and Leonid Sigal[‡], Yunjin Chen

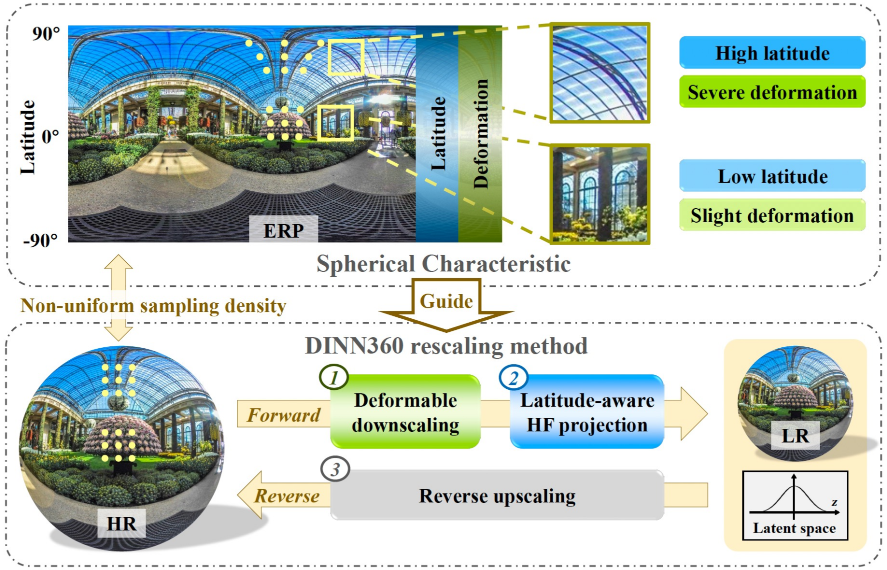
[†] Beihang University, Beijing, China; [‡] University of British Columbia, Vancouver, Canada

{yichenguo, MaiXu, cindydeng}@buaa.edu.cn

Paper tag: THU-PM-087

Quick intro: 360◦ image rescaling

- **Motivation and Contributions**



Spherical characteristics:
- Higher latitude → Severer deformation
- Lower latitude → More textures

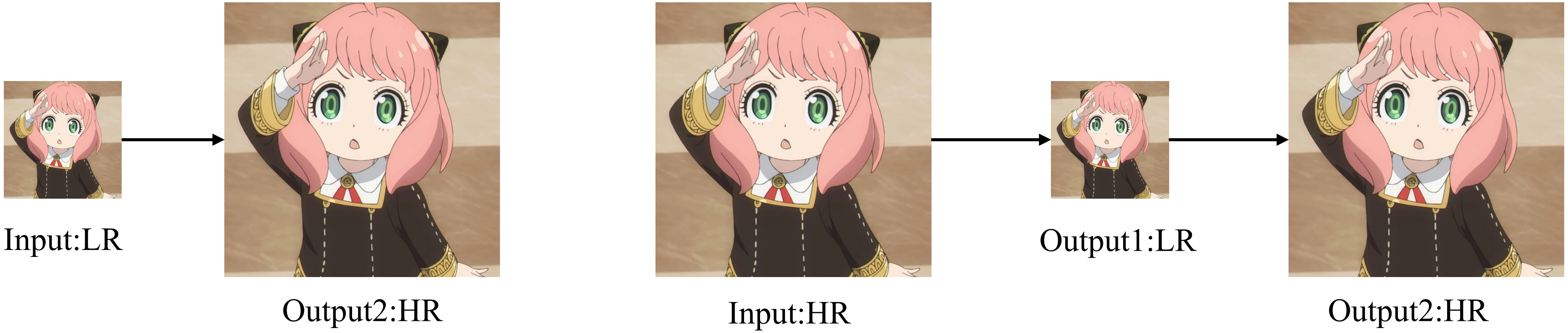
⇩ **Guide**

Model designing:
① Deformable downscaling
② Latitude-aware HF projection
③ Reverse upscaling

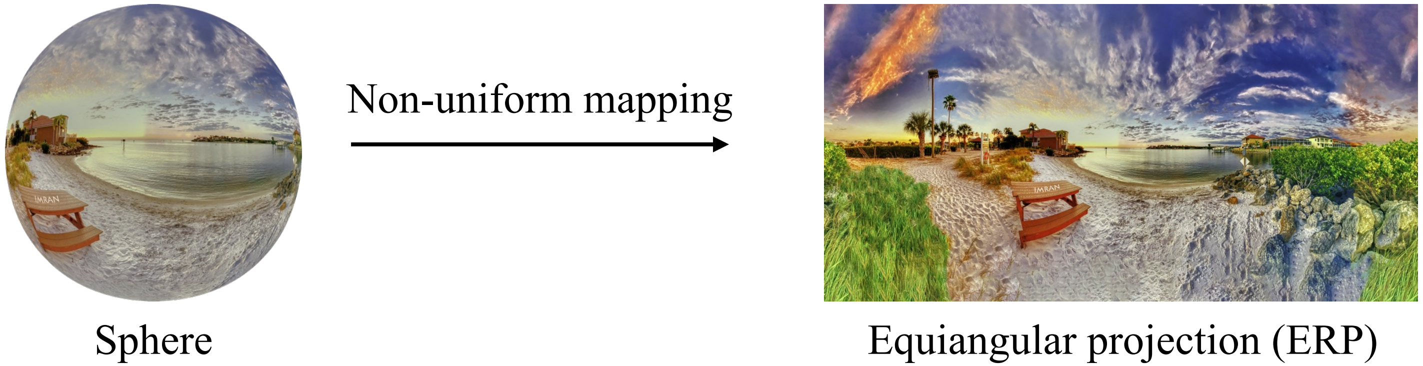
Contributions:
① We find how the low-level characteristics of 360◦ images change along with its latitude, benefiting the designs of our method.
② We propose a novel INN framework for 360◦ image rescaling, with the developed invertible deformable blocks to handle various spherical deformations.
③ We develop a latitude-aware conditional mechanism in our framework, to better preserve the high-frequency component of 360◦ images in a latitude-aware manner.

Background

- **Image rescaling:** HR → LR→ HR

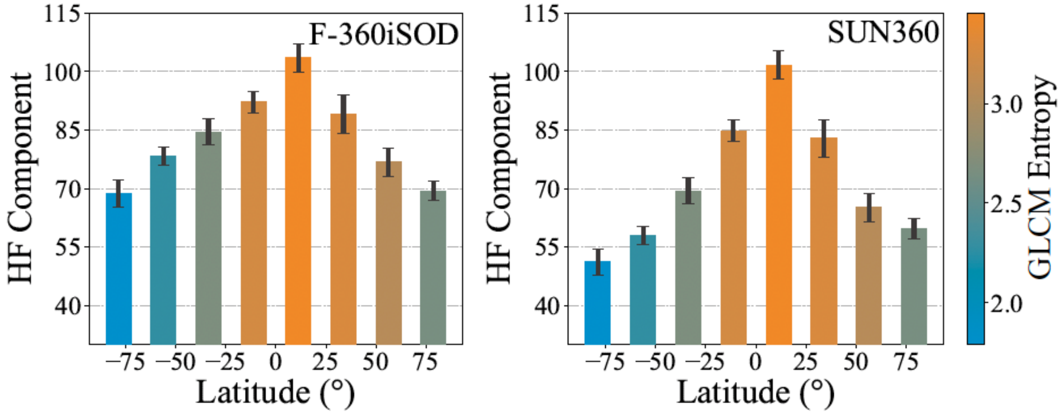


- **360∘ Images:** an omnidirectional view

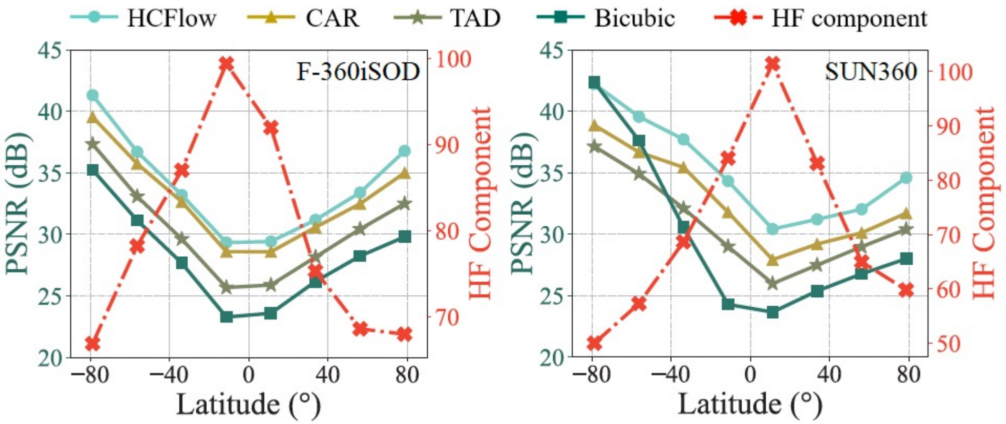


Analysis

- **Finding 1:** In 360◦ images, low-latitude regions tend to contain more textures, leading to larger HF components.

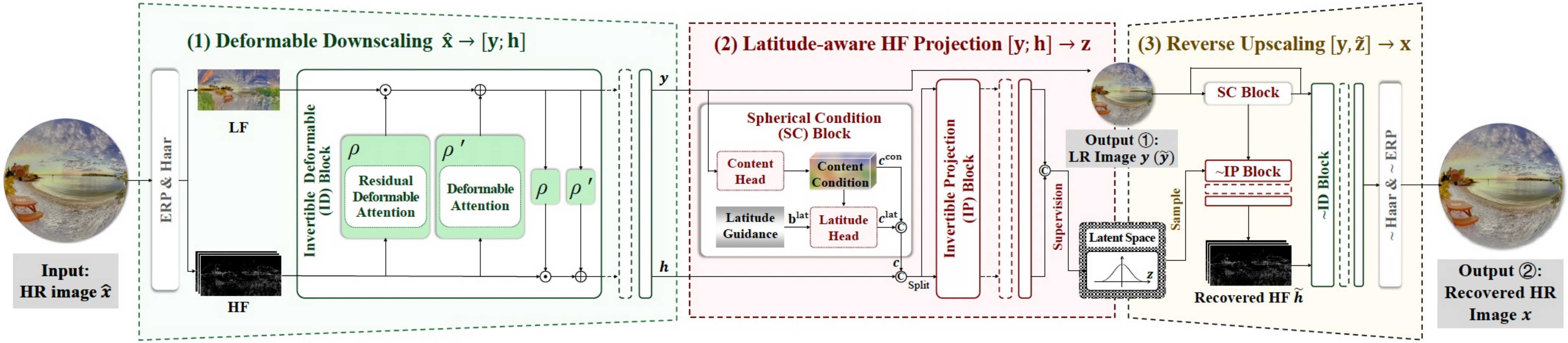


- **Finding 2:** In 360◦ images, the larger HF components at low-latitude regions result in worse rescaling performance for the existing 2D rescaling methods.



Method

- **Pipeline of DINN360**



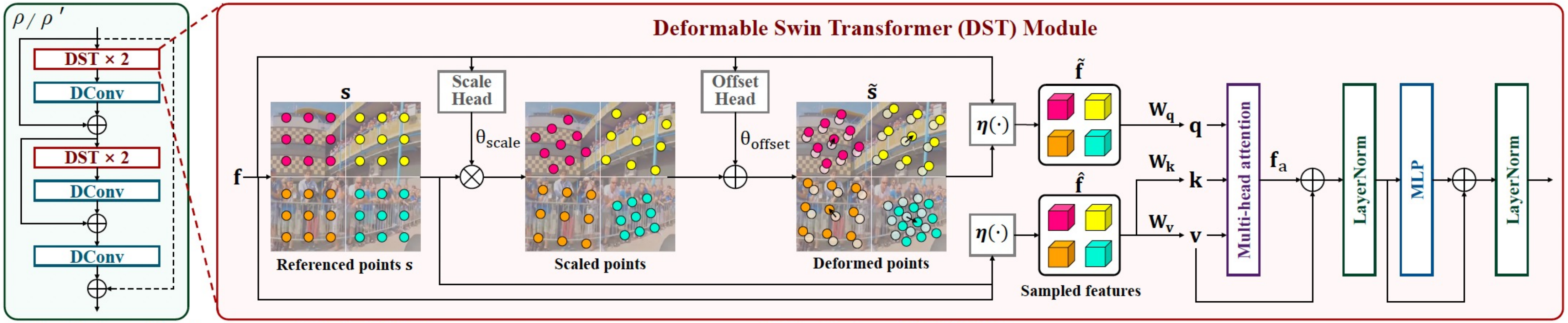
(1) Deformable downscaling $\hat{\mathbf{x}} \rightarrow [\mathbf{y}; \mathbf{h}]$: downscales the HR image $\hat{\mathbf{x}}$ into LR image $\mathbf{y}$ and HF component $\mathbf{h}$ by invertible deformable (ID) blocks.

(2) Latitude-aware HF projection $[\mathbf{y}; \mathbf{h}] \rightarrow \mathbf{z}$: projects the split HF component $\mathbf{h}$ into latent variable $\mathbf{z}$, which is conditional on LR image $\mathbf{y}$.

(3) Reverse upscaling $[\mathbf{y}; \tilde{\mathbf{z}}] \rightarrow \mathbf{x}$: recovers the HF component $\tilde{\mathbf{h}}$ and upscales the LR image into reconstructed HR image $\mathbf{x}$ by reversely passing stage (1) and stage (2).

Method

- **Deformable Swin Transformer (DST) Module**



(a) Architecture of functions ρ(·) and ρ′(·) in ID block

(b) Architecture of DST Module

(a) Affine functions in ID block: The functions are built in a deformable manner, upon the residual structure with deformable convolution (DConv) layers and the developed deformable swin transformer (DST) modules.

(b) Deformable swin transformer module: The referenced sampling points are scaled and shifted into deformed points by the scale head and offset head, which is used to produce the deformable token **q**.

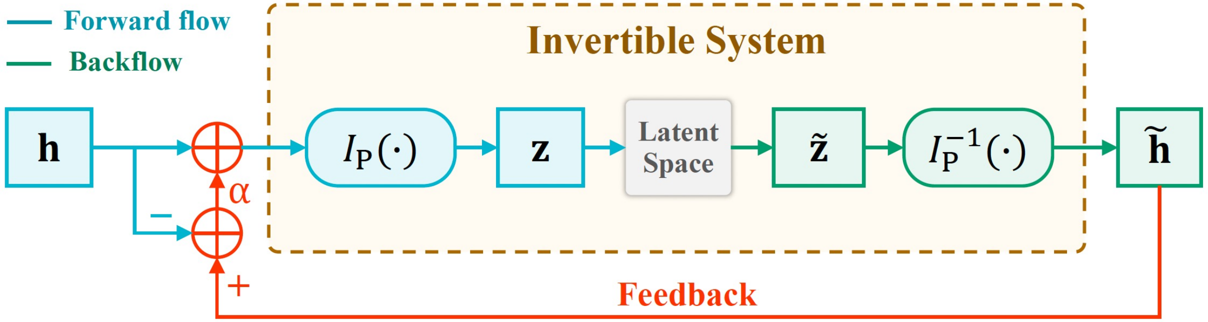
Method

- **Backflow training protocol**

Algorithm 1: Training process for $2\times$ rescaling.

Input: HR image $\hat{\mathbf{x}}$, LR image $\hat{\mathbf{y}}$ and distortion map $\hat{\mathbf{c}}^{lat}$.
Output: Trained $I_{\mathrm{D}}(\cdot)$ $I_{\mathrm{P}}(\cdot)$ and $G_{\mathrm{SC}}(\cdot)$.
Variables: Training variables Φ, latent variables $\mathbf{z}$, $\tilde{\mathbf{z}}$.
Parameters: λ_{H}, λ_{L}, $\lambda_{\mathbf{z}}$, α and learning rate lr.

Initialize Φ with Gaussian initialization.
while *Step* $<$ *max_steps* **do**
  $\mathbf{y}, \mathbf{h} = I_{\mathrm{D}}(\hat{\mathbf{x}})$.
  $\mathbf{c} = \left[\mathbf{c}^{\mathrm{con}}, \mathbf{c}^{\mathrm{lat}}\right] = G_{\mathrm{SC}}(\mathbf{y})$.
  if *Step* $<$ *backflow_steps* **then**
    $\tilde{\mathbf{h}} = I_{\mathrm{P}}^{-1}(\tilde{\mathbf{z}}, \mathbf{c})$.
    $\mathbf{h} = \alpha\tilde{\mathbf{h}} + (1-\alpha)\mathbf{h}$.
  end
  $\mathbf{z} = I_{\mathrm{D}}(\mathbf{h}, \mathbf{c})$.
  $\tilde{\mathbf{h}} = I_{\mathrm{P}}^{-1}(\tilde{\mathbf{z}}, \mathbf{c})$.
  $\mathbf{x} = I_{\mathrm{D}}^{-1}(\mathbf{y}, \tilde{\mathbf{h}})$.
  $\mathcal{L} = \lambda_{\mathrm{H}}\ell_1(\mathbf{x}, \hat{\mathbf{x}}) + \lambda_{\mathrm{L}}\ell_2(\mathbf{y}, \hat{\mathbf{y}}) + \lambda_{\mathbf{z}}\ell_2(\mathbf{z}, \tilde{\mathbf{z}})$.
  $\mathbf{\Phi} \leftarrow \mathbf{\Phi} - lr \cdot \nabla_{\mathbf{\Phi}}\mathcal{L}$.
end
return $\mathbf{\Phi}$.



For better training INN:

(1) Inspired by the proportional feedback in PID control;

(2) Regarding the forward and reverse model together as an invertible system;

(3) Minimize the gap between the generated and sampled latent variable.

Experiments

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- **Quantitative results**

Scale	Method	ODISR [4]	SUN360 [38]	F-360iSOD [42]	YouTube360 [27]
2×	Bicubic	29.46 ± 2.54 / 86.23 ± 5.05	30.06 ± 2.46 / 87.92 ± 4.85	30.68 ± 4.53 / 87.43 ± 7.23	34.93 ± 4.92 / 94.82 ± 4.47
	Bilinear	28.94 ± 2.45 / 83.15 ± 5.83	29.39 ± 2.40 / 85.09 ± 6.08	29.97 ± 4.23 / 84.61 ± 8.61	33.20 ± 4.21 / 9.255 ± 6.06
	Lanczos	28.58 ± 2.54 / 84.04 ± 5.57	29.16 ± 2.47 / 85.72 ± 5.48	29.86 ± 4.56 / 85.44 ± 8.11	34.26 ± 5.13 / 93.95 ± 5.24
	Bicubic & 360SR [28]	27.05 ± 2.36 / 80.46 ± 4.32	27.69 ± 2.20 / 81.55 ± 4.87	26.08 ± 4.17 / 78.08 ± 5.39	32.12 ± 3.77 / 89.84 ± 4.85
	Bicubic & 360SISR [27]	30.81 ± 2.90 / 87.44 ± 5.17	32.72 ± 2.79 / 90.53 ± 5.10	31.33 ± 4.81 / 89.63 ± 6.28	37.62 ± 5.21 / 96.23 ± 5.08
	TAD & TAU [16]	35.84 ± 3.28 / 96.12 ± 8.12	37.70 ± 2.68 / 97.17 ± 1.10	33.94 ± 5.11 / 93.87 ± 4.47	39.50 ± 4.08 / 98.22 ± 1.00
	CAR & EDSR [22, 32]	33.00 ± 3.51 / 91.31 ± 4.41	35.68 ± 3.37 / 93.91 ± 4.07	35.38 ± 5.48 / 93.05 ± 5.09	40.49 ± 5.24 / 97.75 ± 2.55
	IRN [39]	40.51 ± 3.52 / 98.63 ± 0.71	42.72 ± 2.73 / 99.11 ± 0.32	39.83 ± 5.74 / 97.83 ± 2.05	46.15 ± 4.02 / 99.50 ± 0.32
	HCFlow [21]	42.05 ± 3.79 / 99.02 ± 0.57	45.05 ± 3.00 / 99.49 ± 0.24	40.53 ± 5.78 / 97.92 ± 1.99	50.56 ± 3.07 / 99.71 ± 0.10
	DINN360	**42.64 ± 3.87 / 99.13 ± 0.52**	**45.72 ± 3.00 / 99.56 ± 0.21**	**40.77 ± 5.88 / 97.93 ± 2.21**	**50.75 ± 3.07 / 99.73 ± 0.10**
4×	Bicubic	25.39 ± 2.28 / 72.27 ± 7.45	25.38 ± 2.33 / 73.75 ± 8.83	26.16 ± 3.91 / 73.75 ± 12.38	28.29 ± 3.80 / 83.73 ± 10.58
	Bilinear	26.24 ± 2.27 / 72.96 ± 7.54	26.22 ± 2.29 / 74.72 ± 8.85	26.85 ± 3.78 / 74.30 ± 12.38	28.92 ± 3.53 / 83.94 ± 10.44
	Lanczos	24.97 ± 2.28 / 70.69 ± 7.64	24.99 ± 2.33 / 71.95 ± 9.05	25.77 ± 3.94 / 72.10 ± 12.84	27.97 ± 3.85 / 82.65 ± 11.02
	Bicubic & 360SR	25.42 ± 2.26 / 71.06 ± 6.89	25.42 ± 2.16 / 72.46 ± 8.64	25.19 ± 3.69 / 70.79 ± 9.83	28.43 ± 3.26 / 83.06 ± 9.36
	Bicubic & 360SISR	27.03 ± 2.45 / 76.15 ± 7.97	27.81 ± 2.44 / 80.45 ± 9.39	27.45 ± 4.35 / 78.79 ± 11.66	30.96 ± 3.87 / 89.36 ± 10.75
	TAD & TAU	28.98 ± 2.51 / 82.69 ± 5.91	29.70 ± 2.47 / 84.86 ± 6.21	28.71 ± 4.55 / 81.34 ± 10.40	33.24 ± 4.61 / 92.48 ± 6.08
	CAR & EDSR	29.61 ± 2.86 / 82.82 ± 6.76	31.32 ± 2.82 / 86.60 ± 7.49	31.33 ± 4.94 / 85.30 ± 9.10	34.85 ± 4.69 / 93.08 ± 6.45
	IRN	30.86 ± 3.06 / 87.47 ± 5.56	32.69 ± 2.92 / 90.41 ± 5.41	32.58 ± 5.19 / 88.95 ± 7.29	36.85 ± 4.78 / 95.86 ± 4.07
	HCFlow	31.48 ± 3.16 / 89.07 ± 5.02	33.62 ± 3.03 / 92.00 ± 4.78	32.40 ± 5.79 / 88.44 ± 8.85	40.31 ± 4.44 / 97.72 ± 2.13
	DINN360	**31.92 ± 3.26 / 89.90 ± 4.82**	**34.19 ± 3.12 / 92.77 ± 4.48**	**32.93 ± 5.90 / 89.34 ± 8.82**	**40.55 ± 4.29 / 97.89 ± 1.89**
8×	Bicubic	23.25 ± 2.19 / 64.10 ± 8.64	22.92 ± 2.21 / 65.18 ± 10.24	23.45 ± 3.48 / 64.12 ± 15.04	24.98 ± 3.06 / 74.70 ± 12.89
	Bilinear	24.16 ± 2.19 / 65.35 ± 8.65	23.81 ± 2.19 / 66.77 ± 10.20	24.25 ± 3.41 / 65.42 ± 14.89	25.78 ± 2.96 / 75.86 ± 12.65
	Lanczos	22.95 ± 2.19 / 63.15 ± 8.68	22.65 ± 2.21 / 63.98 ± 10.27	23.19 ± 3.49 / 63.05 ± 15.18	24.77 ± 3.08 / 73.78 ± 13.03
	Bicubic & 360SR	23.61 ± 2.06 / 64.15 ± 8.53	23.28 ± 2.17 / 65.11 ± 10.14	23.19 ± 3.17 / 63.30 ± 13.68	25.02 ± 2.85 / 78.19 ± 12.37
	Bicubic & 360SISR	24.63 ± 2.26 / 67.75 ± 8.99	24.56 ± 2.27 / 70.80 ± 10.66	24.53 ± 3.62 / 68.64 ± 14.76	26.28 ± 3.01/ 80.02 ± 12.73
	Bicubic & LAU-Net [4]	24.37 ± 2.22 / 66.64 ± 8.83	24.21 ± 2.26 / 69.37 ± 10.63	24.18 ± 3.57 / 66.94 ± 14.99	25.81 ± 2.94 / 77.33 ± 12.61
	TAD & TAU	26.36 ± 2.30 / 71.36 ± 7.86	26.50 ± 2.33 / 73.43 ± 9.36	25.94 ± 4.10 / 70.35 ± 14.15	28.36 ± 3.46 / 81.61 ± 11.04
	CAR & EDSR	25.97 ± 2.38 / 69.40 ± 8.82	26.40 ± 2.42 / 72.77 ± 10.75	26.87 ± 4.12 / 71.19 ± 14.36	27.98 ± 3.44 / 79.83 ± 11.27
	IRN	28.06 ± 2.72 / 77.41 ± 8.12	29.48 ± 2.74 / 82.02 ± 9.66	29.55 ± 4.89 / 80.03 ± 11.87	32.16 ± 4.24 / 89.01 ± 9.30
	HCFlow	28.25 ± 2.76 / 78.20 ± 8.00	29.77 ± 2.77 / 82.84 ± 9.42	29.83 ± 4.94 / 80.98 ± 11.47	34.19 ± 4.02 / 91.78 ± 7.14
	DINN360	**28.60 ± 2.86 / 79.17 ± 7.98**	**30.36 ± 2.87 / 84.02 ± 9.36**	**30.29 ± 5.13 / 82.07 ± 11.22**	**34.93 ± 4.17 / 92.58 ± 6.83**

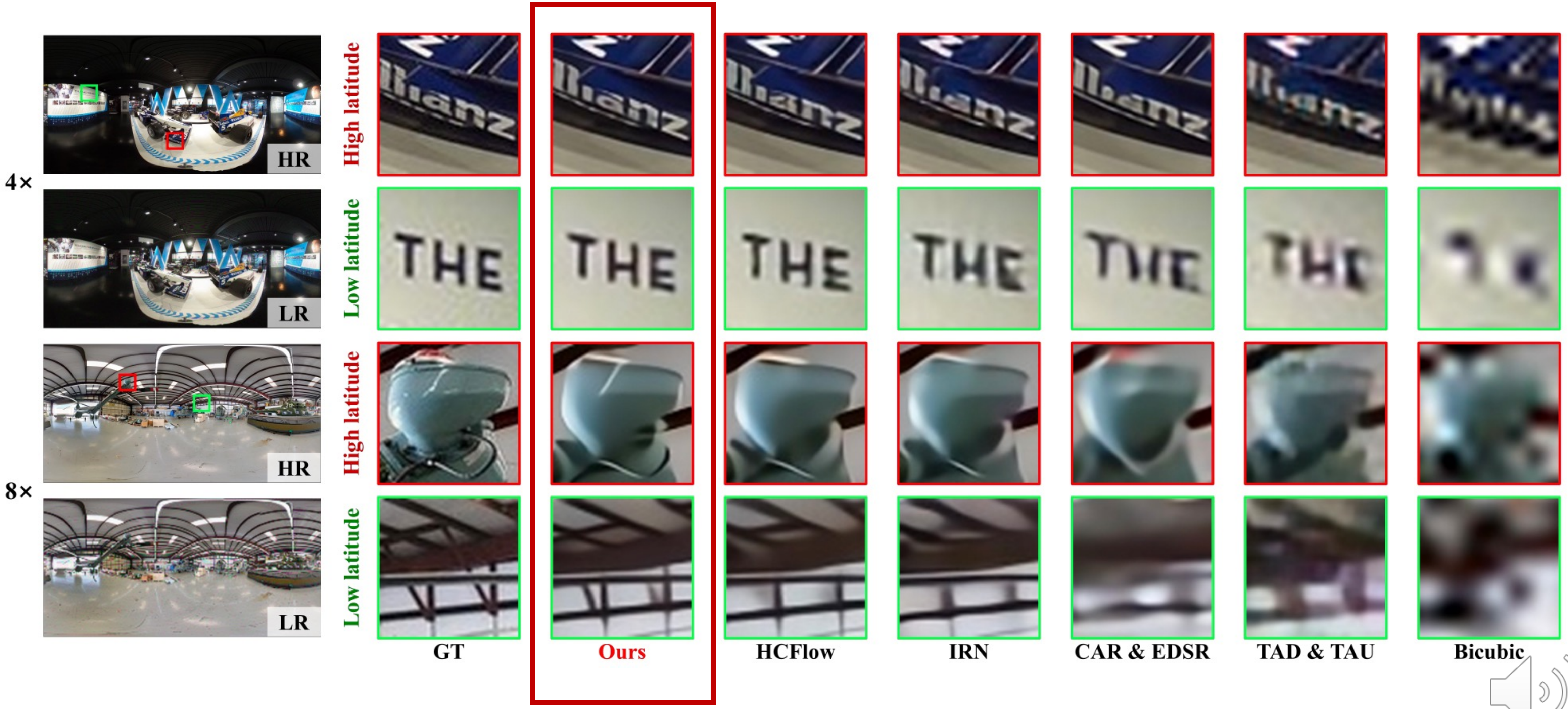
+0.19~0.67dB

+0.24~0.57dB

+0.35~0.74dB

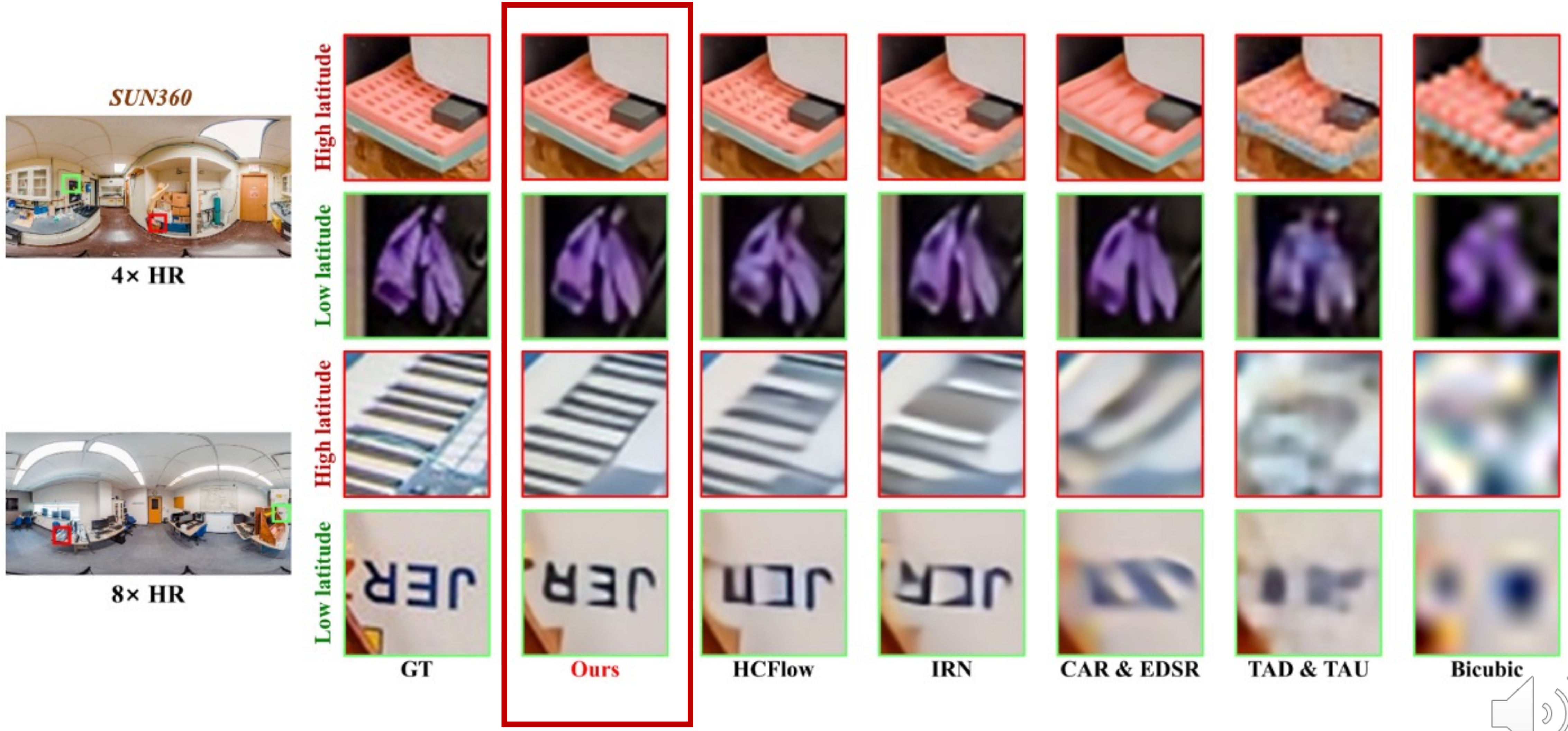
Experiments

- **Qualitative results**



Experiments

- **Qualitative results**



Experiments

- **Ablation results**

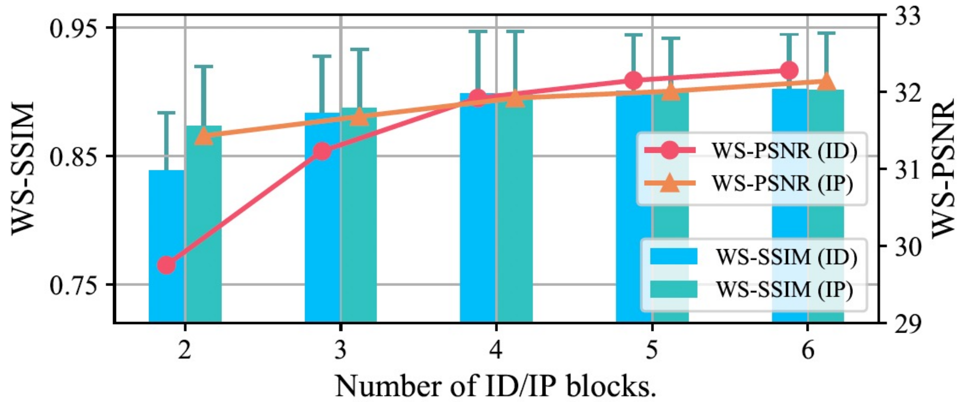


Figure 9. Ablation results on numbers of ID/IP blocks.

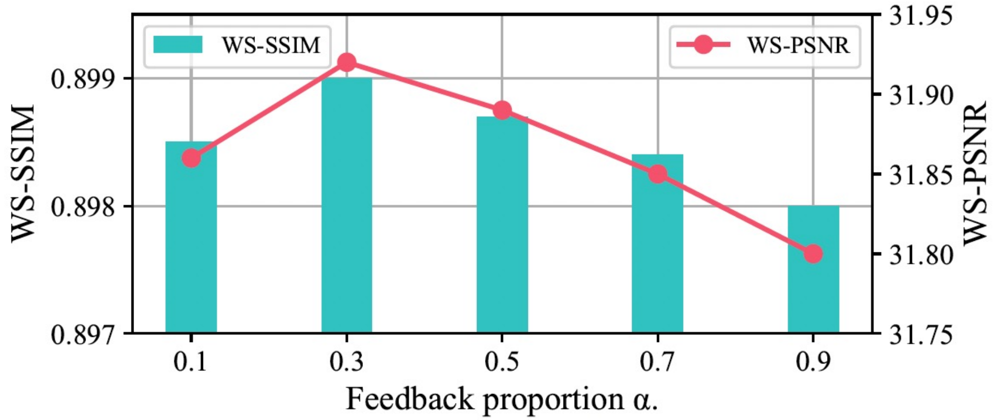


Figure 10. Ablation results on feedback ratio α in Eq. (13).

$$\mathbf{z} = I_{\mathrm{P}}\left(\mathbf{h} + \alpha(\widetilde{\mathbf{h}} - \mathbf{h})\right) = I_{\mathrm{P}}\left(\alpha I_{\mathrm{P}}^{-1}(\widetilde{\mathbf{z}}) + (1-\alpha)\mathbf{h}\right), \quad (13)$$

	Ablation settings	WS-PSNR	WS-SSIM
ID block	w/o DST module	31.79 ± 3.14	89.68 ± 4.17
	w/o deform	31.83 ± 3.63	89.75 ± 4.29
IP block	w/o latitude head	31.85 ± 3.21	89.74 ± 4.31
	w/o content head	31.76 ± 3.21	89.56 ± 4.08
backflow	w/o feedback	31.85 ± 3.32	89.83 ± 4.27
-	**DINN360**	**31.92 ± 3.26**	**89.90 ± 4.82**

Hyper-parameter values:

(1) Number of ID/IP blocks;

(2) Feedback ratio of backflow training protocol

Ablation studies:

(1) ID/IP block: deformable and latitude-aware

(2) Backflow: feedback protocol

Thanks

Wechat

Contact us

Multimedia Computing Towards Communications

(MC2) Lab

http://buaamc2.net/

Email: yichenguo@buaa.edu.cn