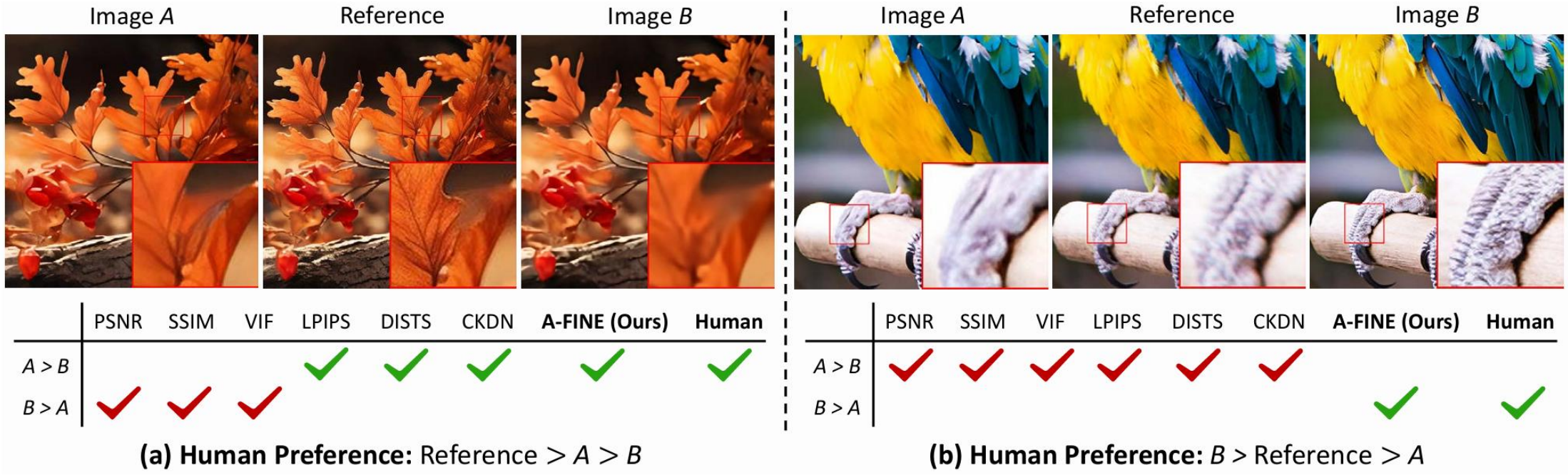
Toward Generalized Image Quality Assessment: Relaxing the Perfect Reference Quality Assumption

Du Chen[1,3*], Tianhe Wu[2,3*], Kede Ma[2], and Lei Zhang[1,3]
[1]Department of Computing, The Hong Kong Polytechnic University
[2]Department of Computer Science, City University of Hong Kong
[3]OPPO Research Institute

https://tianhewu.github.io/A-FINE-page.github.io/

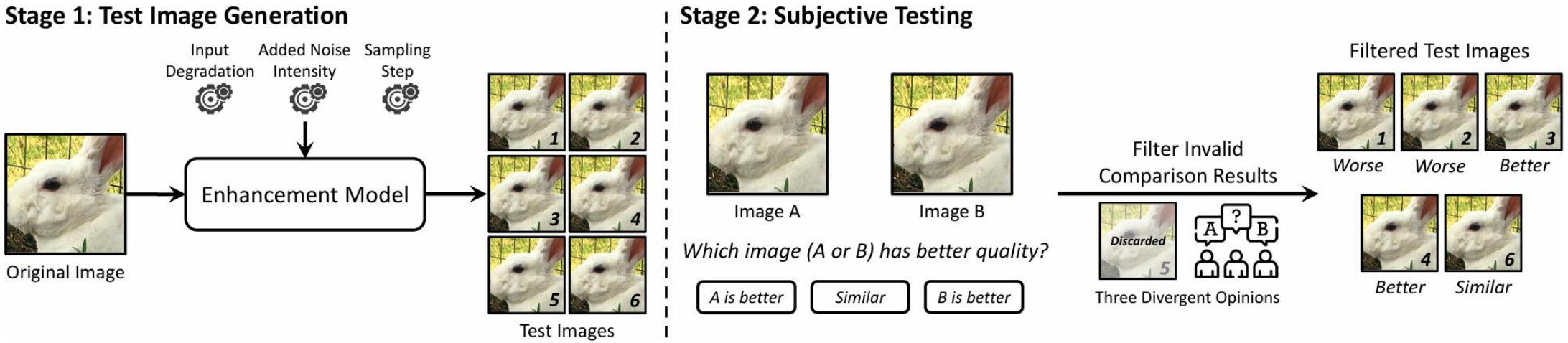
Motivation

- Full-reference image quality assessment hypothesis
  - Reference images are of the highest quality
  - Perceptual distance as the quality score for the test image
- **Existing generative SR or restoration models can generate images with higher quality than the reference**



DiffIQA Database

- Training: Image enhancement network based on diffusion model
- Enhanced 179,208 images
- 76,515 worse, 24,654 similar, 76,150 better (compared to reference) image pairs

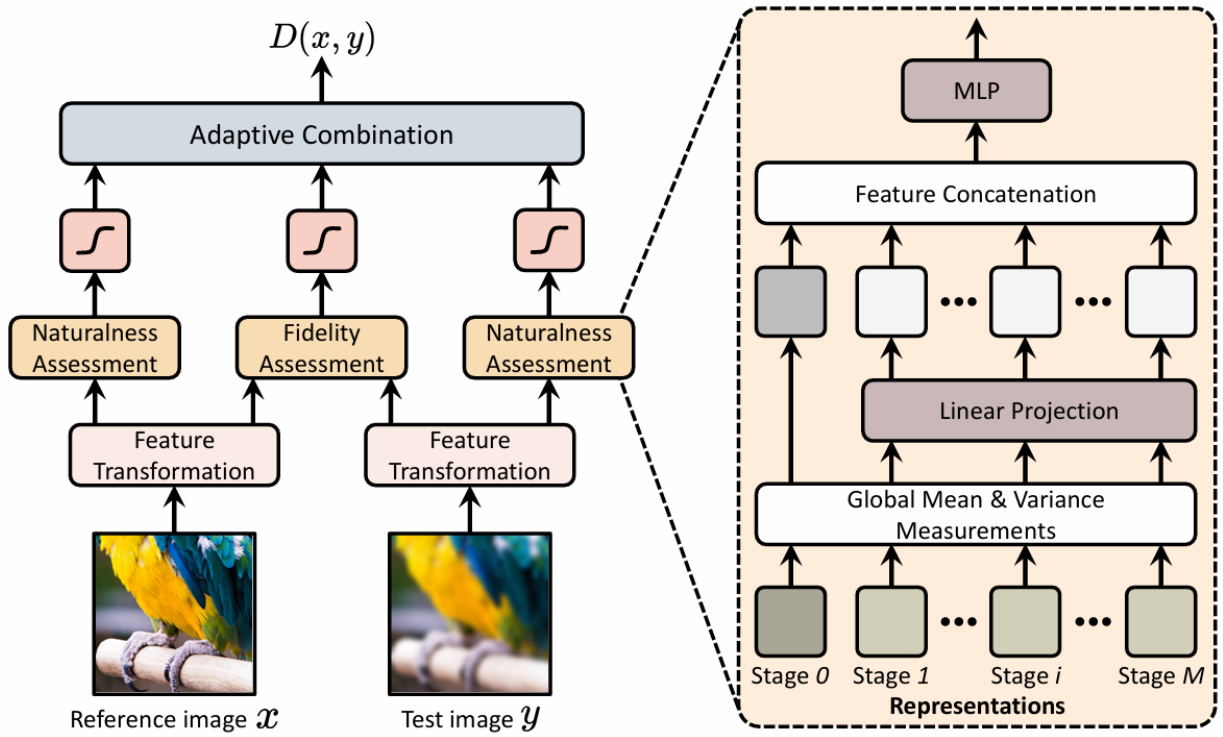


SRIQA-Bench

- Regression-based SR networks
  - SwinIR and RRDB
- Generation-based SR networks
  - Real-ESRGAN, BSR GAN, HGGT, SUPIR, SeeSR, StableSR, SinSR and OSEDiff

Dataset	# of Ref. Images	# of Test Images	Distortion / Enhancement Type	Image Resolution	# of Human Annotations	Perfect Reference Quality Assumption
LIVE [30]	29	779	Simulated	480×720 to 768×512	25k	Necessary
CSIQ [15]	30	866	Simulated	512×512	25k	Necessary
TID2013 [26]	25	3k	Simulated	512×384	500k	Necessary
KADID-10K [20]	81	10.1k	Simulated	512×384	303.8k	Necessary
PIPAL [10]	250	29k	Simulated / GAN-based	288×288	1.1m	Necessary
BAPPS [56]	187.7k	375.4k	Simulated / DNN-based	64×64	484.3k	Necessary
DiffIQA (Ours)	29.9k	177.3k	Diffusion-based	512×512	537.6k	Not Necessary
SRIQA-Bench (Ours)	100	1.1k	DNN- / GAN- / Diffusion-based	512×512	55k	Not Necessary

Computation of A-FINE



- Combine fidelity and naturalness

$$D(x, y) = F(x, y) + \lambda(x, y)N(y)$$

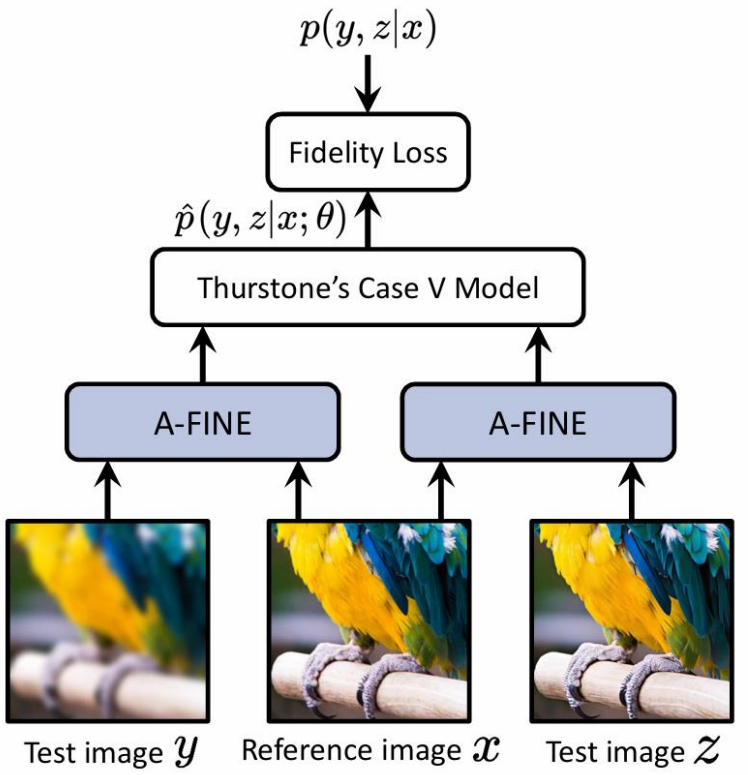
$$\lambda(x, y) = \exp\left(k(N(x) - N(y))\right)$$

- Texture and structure similarity

$$L(x_j^{(i)}, y_j^{(i)}) = \frac{2\mu_{x_j}^{(i)}\mu_{y_j}^{(i)} + c_1}{\left(\mu_{x_j}^{(i)}\right)^2 + \left(\mu_{y_j}^{(i)}\right)^2 + c_1}$$

$$S(x_j^{(i)}, y_j^{(i)}) = \frac{2\sigma_{x_j y_j}^{(i)} + c_2}{\left(\sigma_{x_j}^{(i)}\right)^2 + \left(\sigma_{y_j}^{(i)}\right)^2 + c_2}$$

Training Procedure of A-FINE



- **Learning-to-rank**

$$p(y, z|x) = \begin{cases} 1 & \text{if } Q(y|x) > Q(z|x) \\ 0.5 & \text{if } Q(y|x) = Q(z|x) \\ 0 & \text{otherwise,} \end{cases}$$

$$\hat{p}(y, z|x; \theta) = \Phi\left(\frac{D(x, y; \theta) - D(x, z; \theta)}{\sqrt{2}}\right)$$

$$\ell(y, z|x; \theta) = 1 - \sqrt{p(y, z|x)\hat{p}(y, z|x; \theta)} - \sqrt{(1 - p(y, z|x))(1 - \hat{p}(y, z|x; \theta))}.$$

Main Results

Scenario	Method	Training Dataset	TID2013	KADID	PIPAL	Average	DiffIQA Ref < Test	DiffIQA Ref > Test	DiffIQA Average	All Average
Standard	PSNR	N.A.	75.8	74.9	70.7	72.2	18.2	92.1	45.6	58.9
	SSIM [43]	N.A.	68.9	74.0	72.1	72.4	20.1	93.0	47.1	60.0
	MS-SSIM [42]	N.A.	83.4	81.8	72.5	75.9	20.1	93.0	47.1	61.5
	FSIM [53]	N.A.	86.0	83.4	76.2	79.0	20.2	93.1	47.2	63.1
	VSI [54]	N.A.	87.3	84.8	76.2	79.5	19.7	93.1	46.9	63.2
	LPIPS [56]	BAPPS	78.7	77.0	74.3	75.4	23.7	94.7	50.0	62.7
	LPIPS-FT	Combined	72.5	78.2	71.7	73.6	35.4	91.6	55.6	64.6
	DISTS [5]	KADID	78.4	81.4	75.3	77.2	21.4	94.8	48.6	62.9
	DISTS-FT	Combined	78.4	81.9	72.1	75.3	38.2	89.5	56.7	66.0
	AHIQ [14]	PIPAL	74.6	76.4	79.3	78.1	34.1	88.1	54.1	66.1
	AHIQ-FT	Combined	81.0	79.7	74.9	76.7	78.4	73.8	76.7	76.7
	TOPIQ [1]	KADID	**90.4**	**94.3**	80.5	**85.1**	22.1	**95.1**	49.1	67.1
	TOPIQ-FT	Combined	78.9	85.0	79.0	80.6	78.6	74.2	77.0	78.8
Generalized	VIF [29]	N.A.	78.5	75.2	72.4	73.7	20.0	92.8	46.9	60.3
	PCQI [35]	N.A.	66.6	65.4	56.7	59.9	17.3	90.3	44.3	52.1
	SFSN [61]	N.A.	75.6	70.5	69.8	70.5	19.5	89.6	45.4	58.0
	CKDN [60]	PIPAL	76.9	70.9	79.8	77.1	33.3	82.4	51.4	64.3
	CKDN-FT	Combined	75.0	80.1	68.1	72.0	**79.4**	71.0	76.4	74.2
	A-FINE (Ours)	Combined	88.1	88.3	**81.0**	83.6	78.5	82.3	**79.9**	**81.8**

- **There is a balance between two assessment scenarios**

Other Results

SRIQA-Bench

Method	Regression-based	Generation-based	All
PSNR	80.7	41.7	34.7
SSIM [43]	83.0	45.3	37.4
MS-SSIM [42]	83.0	45.6	37.6
FSIM [53]	85.3	49.5	41.0
VSI [54]	81.3	50.1	41.2
LPIPS [56]	82.0	63.9	65.8
LPIPS-FT	84.7	63.8	72.2
DISTS [5]	83.3	66.6	72.4
DISTS-FT	**86.0**	63.9	71.4
AHIQ [14]	83.7	70.0	68.4
AHIQ-FT	71.0	71.5	69.6
TOPIQ [1]	83.7	63.9	67.0
TOPIQ-FT	78.3	73.0	77.7
VIF [29]	85.3	47.1	38.9
PCQI [35]	79.0	39.8	32.2
SFSN [61]	80.3	48.4	39.8
CKDN [60]	45.0	60.1	47.4
CKDN-FT	76.7	64.3	59.1
A-FINE (Ours)	83.3	**78.9**	**82.4**

Backbone

Backbone	Standard	DiffIQA	SRIQA-Bench		
			Reg.	Gen.	All
TOPIQ-FT	80.6	77.0	78.3	73.0	77.7
VGG16	77.6	77.0	79.0	75.0	79.8
ResNet50 (ImageNet)	74.8	69.6	84.7	70.7	77.2
ResNet50 (CLIP)	76.1	71.1	**85.2**	70.3	75.6
ViT-B/32 (ImageNet)	81.0	77.7	81.3	75.5	80.4
ViT-B/32 (CLIP)	**83.6**	**79.9**	83.3	**78.9**	**82.4**

Training Data

Training Dataset	Standard	DiffIQA	SRIQA-Bench		
			Reg.	Gen.	All
Standard	**84.1**	65.6	**86.7**	71.8	78.7
DiffIQA	70.6	79.6	78.3	72.9	76.0
Combined	83.6	**79.9**	83.3	**78.9**	**82.4**

Training Strategy

Training Strategy	Standard	DiffIQA	SRIQA-Bench		
			Reg.	Gen.	All
Single-Phase	79.7	79.6	79.3	75.1	77.9
Three-Phase	**83.6**	**79.9**	**83.3**	**78.9**	**82.4**

Thanks

Tianhe Wu

2025.5.24

wth22@mails.tsinghua.edu.cn