UniRes: Universal Image Restoration for Complex Degradations

Mo Zhou[1,2], Keren Ye[1], Mauricio Delbracio[1], Peyman Milanfar[1], Vishal M. Patel[2], Hossein Talebi[1]
[1]Google, [2]Johns Hopkins University

ICCV OCT 19-23, 2025 HONOLULU HAWAII
arXiv:

Motivation & Problem

Real-world degraded images usually contain a complex mixture of degradations, stemming from capture conditions, device limitations, and post processing. For instance, low-resolution, motion blur, defocus blur, and real noise in different extents may co-exist on the same image.

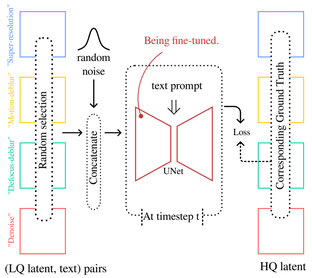
Key Challenges:

* Arbitrary complex degradations: real-world images present an unpredictable combination of degradation types and intensities.
* Data scarcity: curating (low-quality, high-quality) image pairs for complex degradations is extremely difficult.
* Generalization failure: Models trained on single degradation types often fail when faced with complex, unseen mixtures.

Our Contributions:

* UniRes: universal and flexible diffusion-based framework that restores images with complex degradations in an end-to-end manner.
* DiversePhotos: a challenging benchmark dataset curated from real-world photos, each containing at least two distinct degradation types, to better evaluate model robustness.

Our Approach :: Training

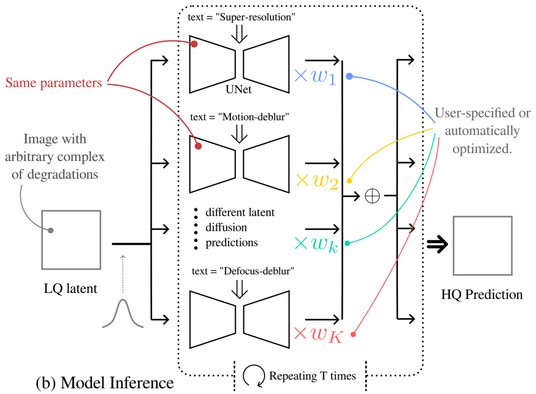


(a) Model Training

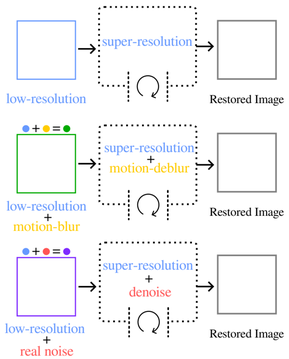
Task Description	Condition Inputs: Image	Condition Inputs: Text	Output	Training Data
Super resolution	LQ	"Super-resolution"	HQ	DF2K [3, 36], LSDIR [32]
Motion deblur	LQ	"Motion-deblur"	HQ	GoPro [42], OID-Motion
Defocus deblur	LQ	"Defocus-deblur"	HQ	DPDD [2]
Denoise	LQ	"Denoise"	HQ	SIDD [1]

Multi-task training detail of the proposed method. (To enable classifier free guidance, we also randomly drop the text during training, which at the same time trains the model for blind restoration.)

Our Approach :: Inference



(b) Model Inference



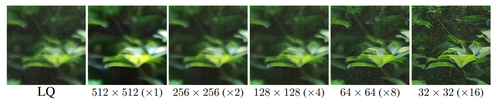
Flexibility of Combination at Inference

$$\tilde{\epsilon}_\theta(z_t, z_{\mathrm{LQ}}; w) \triangleq \sum_{k=1}^{K} w_k \cdot \epsilon_\theta(z_t, z_{\mathrm{LQ}}, s_k)$$

where $w_k \in \mathbb{R}$, $\sum_{k=1}^{K} w_k = 1$, and s_k is the text prompt corresponding to the k-th prediction task.

Task Types during Inference:

1. Super resolution
2. Motion deblur
3. Defocus deblur
4. Denoise
5. Blind restoration (text="")
6. DownLQ super-resolution



Fidelity vs Quality: "DownLQ" is our trade-off mechanism, where the input LQ image is first downscaled by a factor, and then bicubic-upscaled back to its original resolution for super resolution. Larger factors elicit the model to generate more.

Weight Selection Methods:

1. Grid search
2. Best combination weights found in grid search
3. Predict from features

Method	Combination Weights	ClipIQA	MUSIQ	ManIQA
UniRes	Grid search	0.6519	68.22	0.5021
UniRes	Most frequent 8 sets of combination weights	0.6613	68.02	0.5101
UniRes	Most frequent 6 sets of combination weights	0.6633	67.92	0.5096
UniRes	Most frequent 4 sets of combination weights	0.6655	67.68	0.5095
UniRes	Most frequent 2 sets of combination weights	0.6581	66.89	0.5052
UniRes	Most frequent 1 set of combination weights	0.6590	66.44	0.5042
UniRes	Average optimal combination weights	0.5941	62.10	0.4266
UniRes	Random Forest (skip search)	0.5873	61.91	0.4257

Results with different weight selection methods on DiversePhotos×1 dataset.

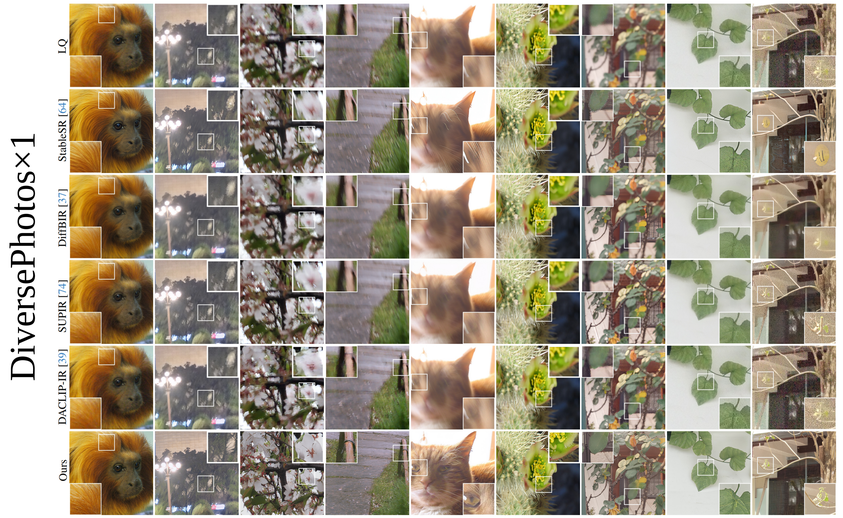
Grid Search

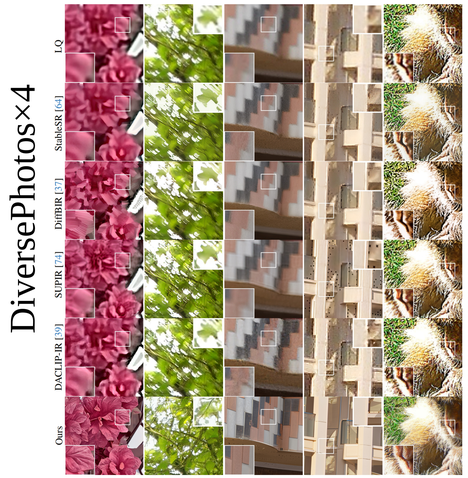
Then, the optimization is done by grid search within a predefined range $[\gamma, \delta]^K$, where $\gamma \leqslant \delta$ and $\gamma, \delta \in \mathbb{R}$:

$$w^* = \arg\max_{w \in \Omega} Q(g(x, w))$$

$$\text{where} \quad \Omega \triangleq \left\{ w \in [\gamma, \delta]^K \middle| \sum_{k=1}^{K} w_k = 1 \right\}$$

In this paper, we empirically adopt MUSIQ [27] as the $Q(\cdot)$ function. The range $[\gamma, \delta]$ as well as the grid search interval are hyper-parameters provided by the user.





Model	ClipIQA [63]	MUSIQ [27]	ManIQA [72]
DiversePhotos×1 (160 images, size 512 × 512)			
StableSR [64]	0.6227	61.39	0.3992
DiffBIR [37]	0.6453	59.97	0.4922
SUPIR [74]	0.5060	51.68	0.3745
DACLIP-IR [39]	0.3497	46.16	0.2567
UniRes	0.6519	68.22	0.5021
Real60 [74] (60 images, size 512 × 512)			
StableSR [64]	0.7593	72.06	0.4997
DiffBIR [37]	0.7851	70.10	0.5772
SUPIR [74]	0.8217	71.61	0.6716
DACLIP-IR [39]	0.3802	54.57	0.2765
UniRes	0.7894	75.15	0.6080

Real-world restoration with 1x upscaling.

Model	PSNR	SSIM	LPIPS↓	ClipIQA	MUSIQ	ManIQA
DiversePhotos×4 (160 images, size 128 × 128)						
StableSR [64]	-	-	-	0.5177	41.53	0.2983
DiffBIR [37]	-	-	-	0.6190	54.09	0.4551
SUPIR [74]	-	-	-	0.5308	39.11	0.3403
DACLIP-IR [39]	-	-	-	0.2924	32.99	0.2339
UniRes	-	-	-	0.6050	62.40	0.4656

Real-world restoration with 4x upscaling. (low-resolution dominates in this case.)