ResShift: Efficient Diffusion Model for Image Super-resolution by Residual Shifting

Zongsheng Yue, Jianyi Wang, Chen Change Loy

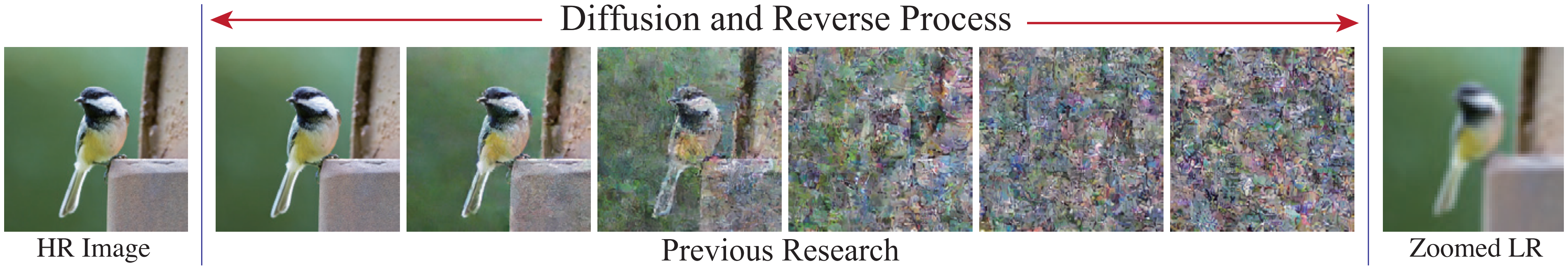
MMLab@NTU, S-Lab for Advanced Intelligence

https://www.mmlab-ntu.com/

Motivation

➢ Current diffusion model (High resolution image ⇄ Random Gaussian Noise)

- Efficiency limitation due to the too long transition path.
- Fidelity issue due to the randomness in sampling.



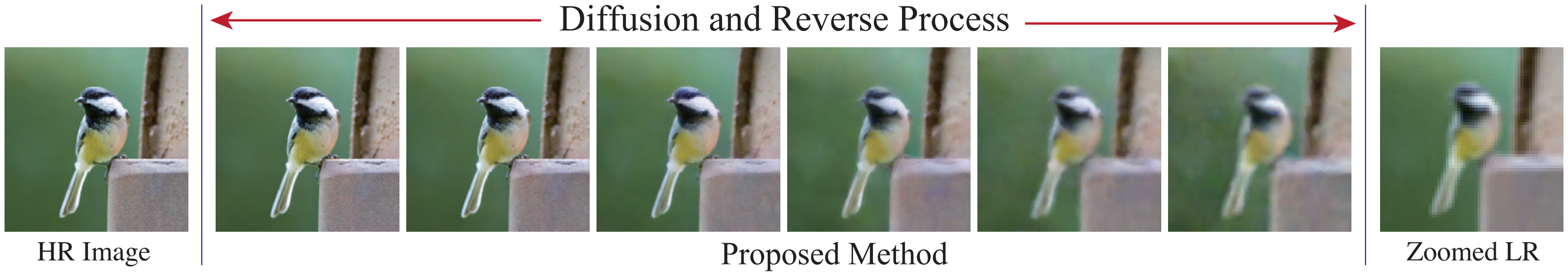
Motivation

➢ Current diffusion model (High resolution image ⇄ Random Gaussian Noise)

- Efficiency limitation due to the too long transition path.
- Fidelity issue due to the randomness in sampling.

➢ Proposed diffusion model (High resolution image ⇄ ***Low resolution image***)

- Improving efficiency by shifting the residuals between LR-HR image pairs.
- Achieving a better fidelity-realism trade-off.

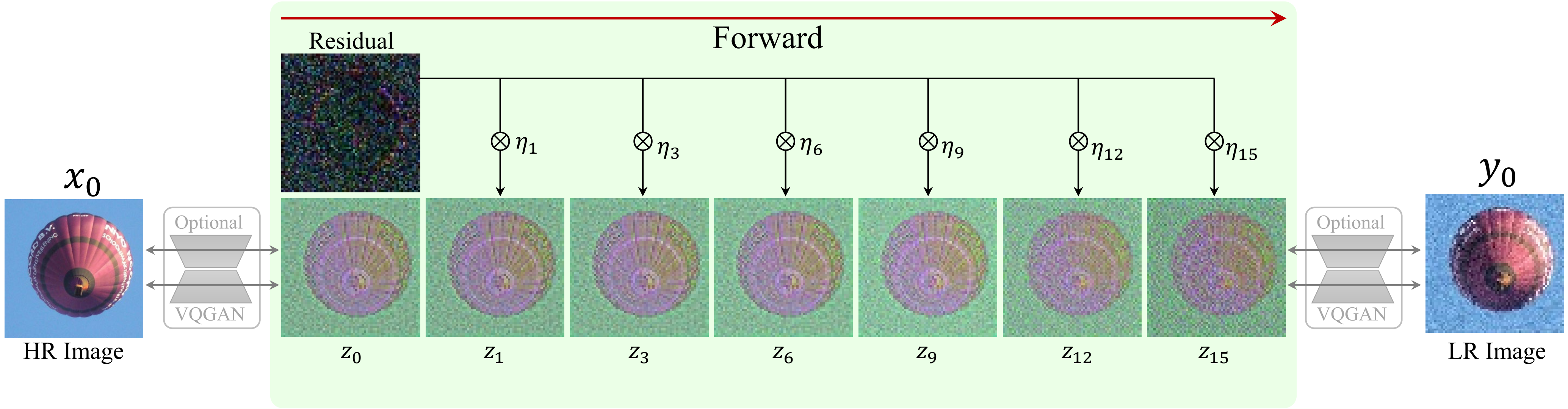


Model Design (Forward Process)

Shifting sequence: $\{\eta_i\}_{i=1}^{T}$, $\eta_1 \to 0$, $\eta_T \to 1$

$$q(x_t|x_0, y_0) = N(x_t; x_0 + \eta_t e_0, \kappa^2 \eta_t I), \;\; e_0 = y_0 - x_0$$

where κ is a hyper-parameter controlling the *noise variance*.



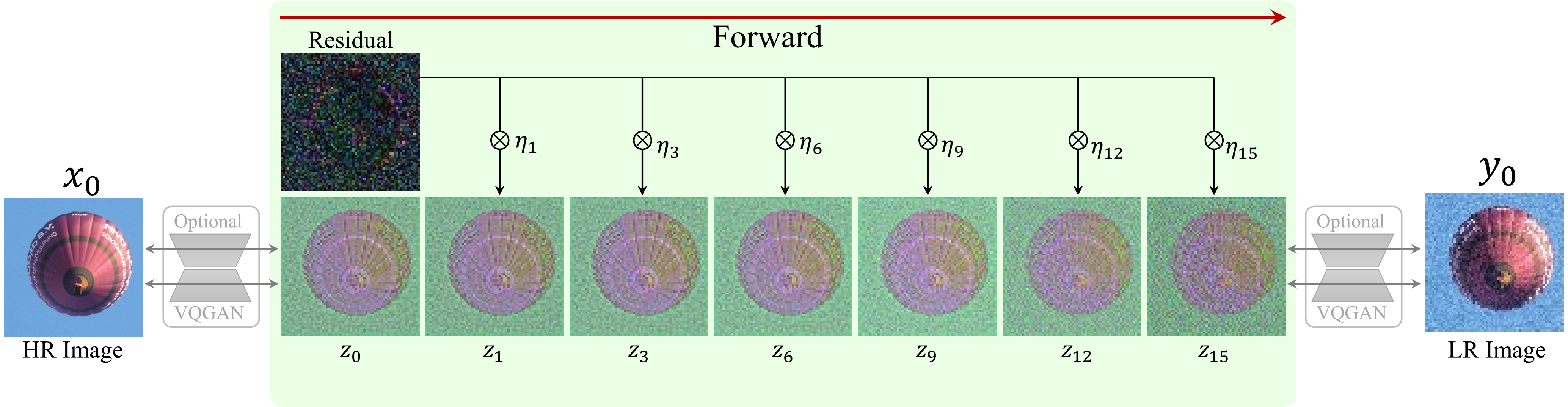
Model Design (Forward Process)

Extension to latent space:

$$q(x_t|x_0, y_0) = N(x_t; x_0 + \eta_t e_0, \kappa^2 \eta_t I), \;\; e_0 = y_0 - x_0$$

$$q(z_t|z_0, g_0) = N(z_t; z_0 + \eta_t e_0, \kappa^2 \eta_t I), \;\; e_0 = g_0 - z_0$$

$$z_0 = \text{Encoder}(x_0), \quad g_0 = \text{Encoder}(y_0)$$



Model Design (Reverse Optimization)

True posterior:

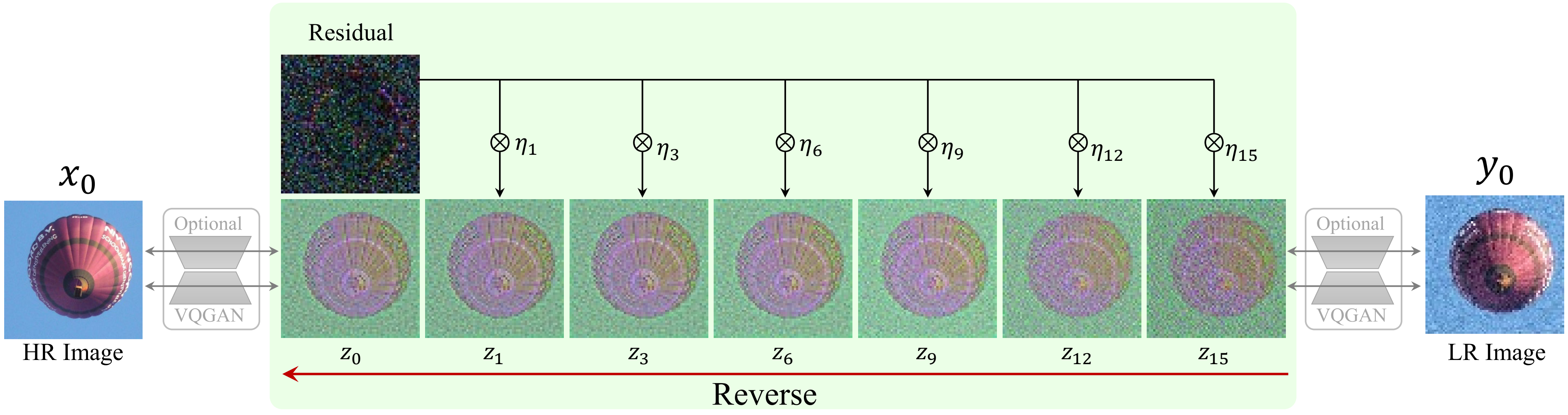
$$q(x_{t-1}|x_t, x_0, y_0) = N\left(x_{t-1}\middle|\frac{\eta_{t-1}}{\eta_t}x_t + \frac{\alpha_t}{\eta_t}x_0, \kappa^2\frac{\eta_{t-1}}{\eta_t}\alpha_t I\right)$$

Approximation:

$$p_\theta(x_{t-1}|x_t, y_0) = N\left(x_{t-1}\middle|\mu_\theta(x_t, y_0, t), \kappa^2\frac{\eta_{t-1}}{\eta_t}\alpha_t I\right)$$

$$\mu_\theta(x_t, y_0, t) = \frac{\eta_{t-1}}{\eta_t}x_t + \frac{\alpha_t}{\eta_t}f_\theta(x_t, y_0, t)$$

Predicting x_0

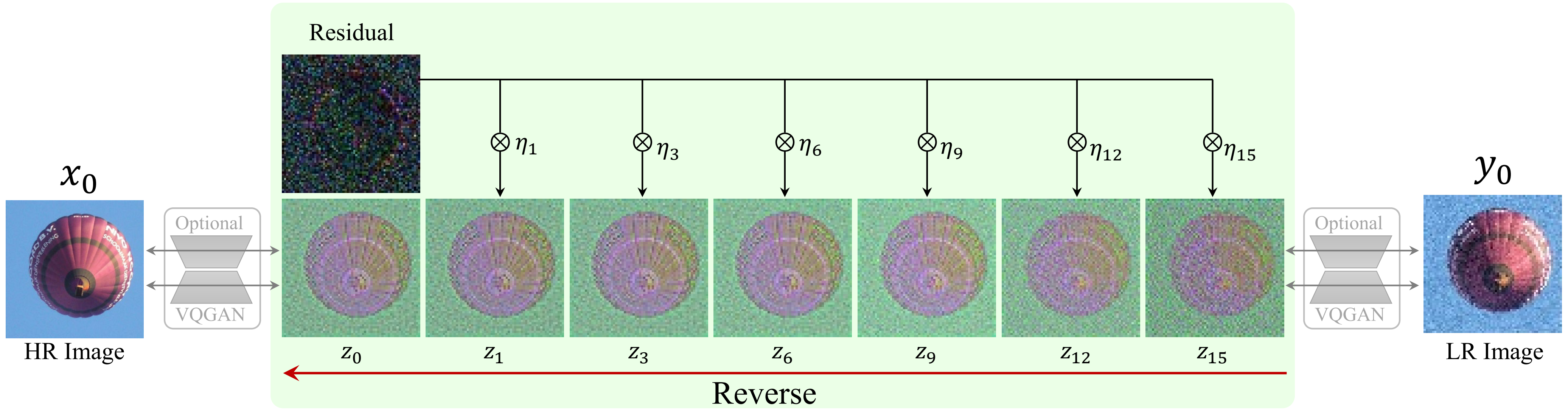


Model Design (Reverse Optimization)

Objective function:

$$\min_{\theta} \sum_{t} D_{KL}[q(x_{t-1}|x_t, x_0, y_0)||p_\theta(x_{t-1}|x_t, y_0)]$$

$$\min_{\theta} \sum_{t} w_t \|f_\theta(x_t, y_0, t) - x_0\|_2^2$$



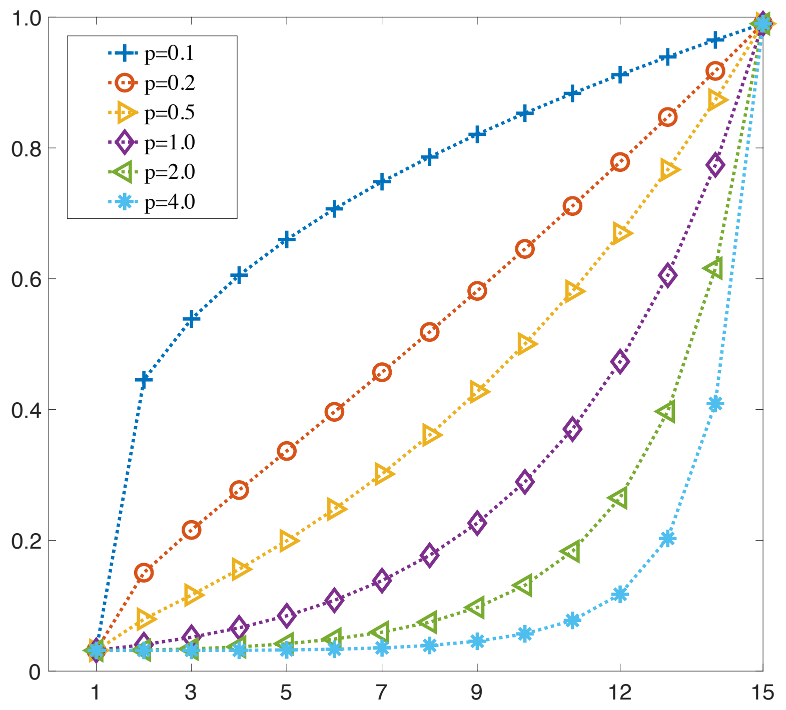
Model Design (Noise Schedule)

Shifting sequence: $\{\eta_i\}_{i=1}^{T}$, $\eta_1 \to 0$, $\eta_T \to 1$

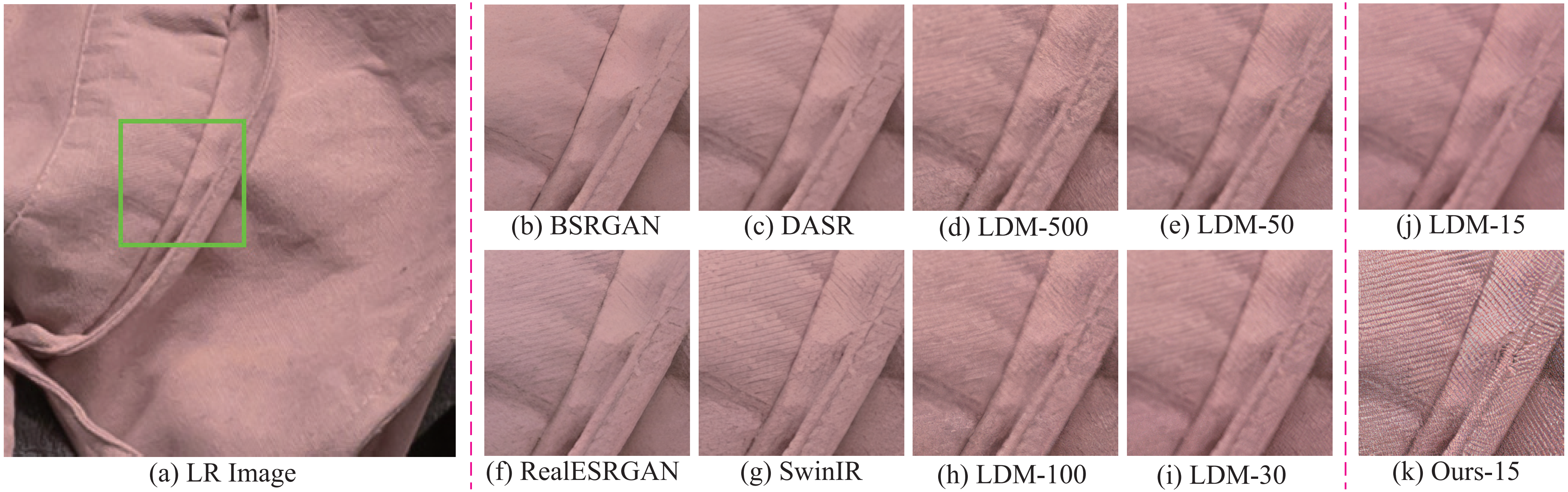
$$\sqrt{\eta_t} = \sqrt{\eta_1} \times {b_0}^{\beta_t}, \quad t = 2, \cdots, T-1$$

$$\beta_t = \left(\frac{t-1}{T-1}\right)^{p} \times (T-1)$$

$$b_0 = \exp\left[\frac{1}{2(T-1)} \log \frac{\eta_T}{\eta_1}\right]$$



Efficiency Evaluation



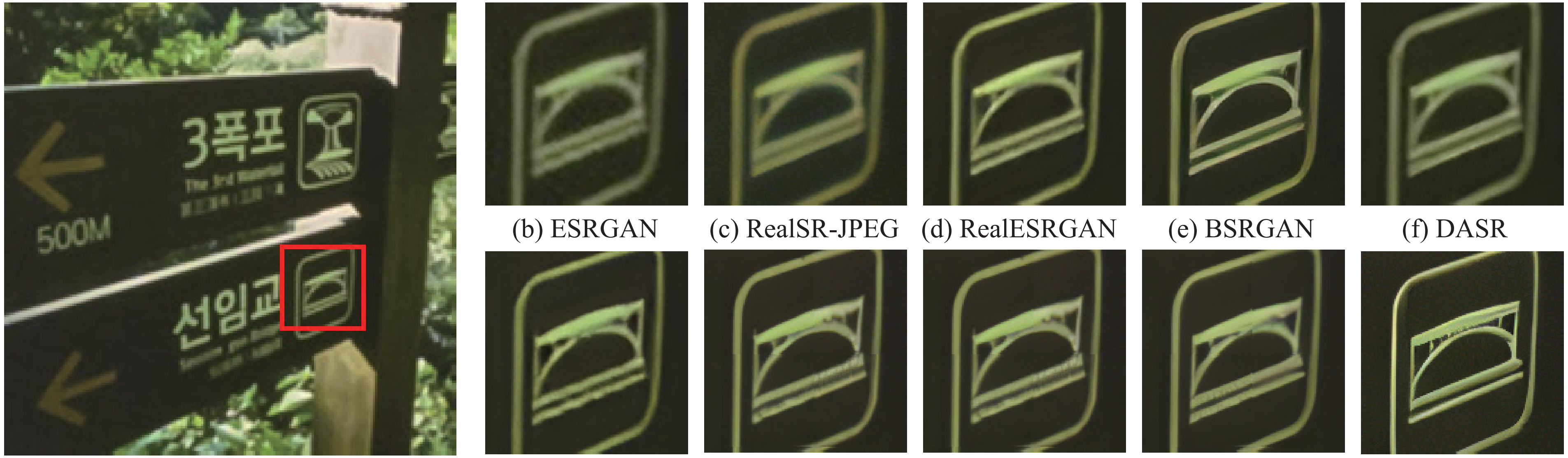
Metrics	Methods						
	LDM-15	LDM-30	LDM-100	LDM-200	LDM-500	LDM-1000	*ResShift*
LPIPS↓	0.269	0.248	0.244	0.245	0.246	0.248	0.231
CLIPIQA↑	0.512	0.572	0.620	0.630	0.634	0.636	0.592
Runtime (s)	0.102	0.184	0.413	0.853	2.094	4.171	0.105
# Parameters (M)	113.60						118.59

Evaluation on Real-world Dataset

(a) LR Image (b) ESRGAN (c) RealSR-JPEG- (d) RealESRGAN (e) BSRGAN (f) DASR (g) SwinIR (h) LDM-100 (i) LDM-30 (j) LDM-15 (k) *ResShift*-15

(a) LR Image (b) ESRGAN (c) RealSR-JPEG (d) RealESRGAN (e) BSRGAN (f) DASR (g) SwinIR (h) LDM-100 (i) LDM-30 (j) LDM-15 (k) *ResShift*-15

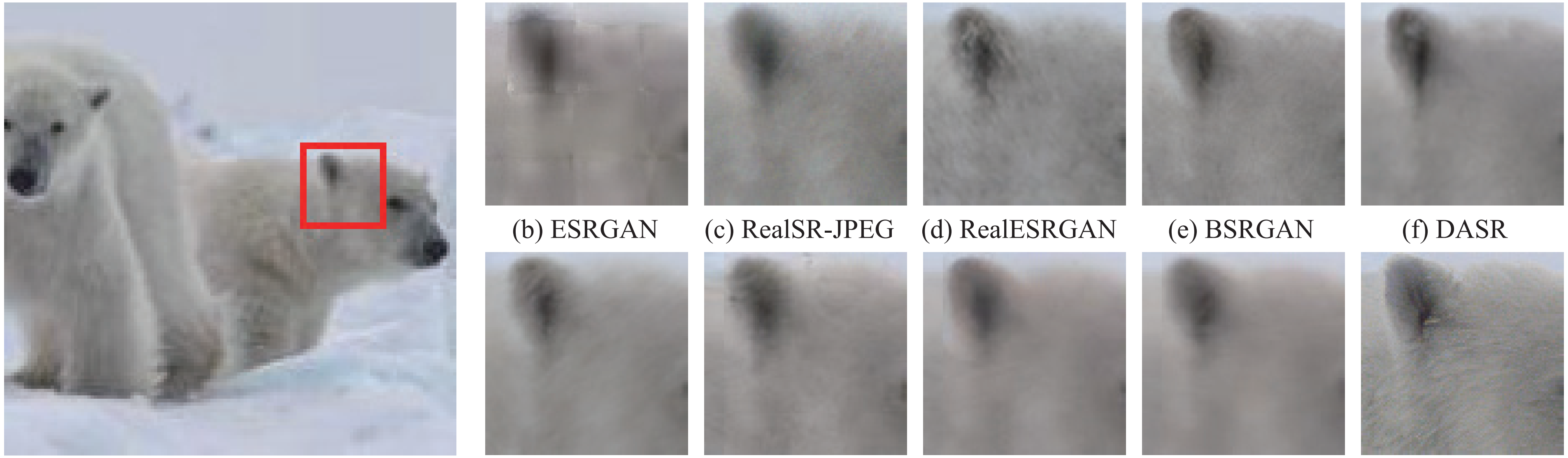
Evaluation on Real-world Dataset



(a) LR Image

(b) ESRGAN (c) RealSR-JPEG (d) RealESRGAN (e) BSRGAN (f) DASR

(g) SwinIR (h) LDM-100 (i) LDM-30 (j) LDM-15 (k) *ResShift*-15



(a) LR Image

(b) ESRGAN (c) RealSR-JPEG (d) RealESRGAN (e) BSRGAN (f) DASR

(g) SwinIR (h) LDM-100 (i) LDM-30 (j) LDM-15 (k) *ResShift*-15

Paper & Code & Demos

Thanks!

https://github.com/zsyOAOA