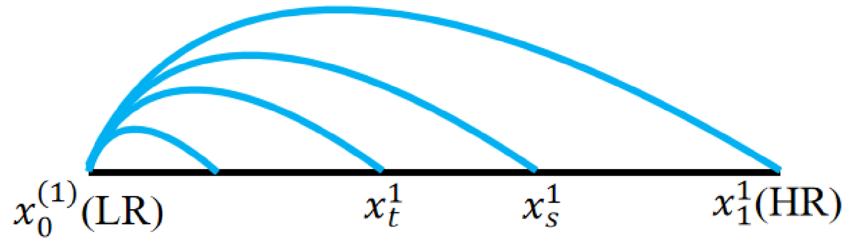
# One-Step Flow for Image Super-Resolution with Tunable Fidelity-Realism Trade-offs


Yuanzhi Zhu[*†1,2] Ruiqing Wang[*1] Shilin Lu[3] Junnan Li[2] Hanshu Yan[2] Kai Zhang[‡1]

[1]Nanjing University [2]Rhymes.AI [3]Nanyang Technological University

yuanzhi-zhu.github.io/
wrqcodedoge.github.io/


## TL;DR

**OFTSR** is a one-step flow-based SR model combining noise-augmented conditional flow with a novel distillation algorithm, enabling a seamless trade-off between high-fidelity and realistic generation in 1 NFE.

## Motivation & Highlights

Diffusion and flow SR models are slow and struggle to balance realism with fidelity, making a single-step, user-controllable method highly desirable.

**Highlights**:

- **Fidelity-Realism Trade-off**: Continuous control between fidelity and perceptual realism.
- **Single-Step Distillation**: preserving the teacher flow trajectory.
- **User-Controlled Output**: Generation quality tuned via a scalar $t$.
- **Teacher-Agnostic**: Compatible w/ any pretrained SR diffusion model.
- **Scalable**: Extends efficiently to large-scale diffusion models.

## Distillation Loss

To preserve the fidelity-realism trade-off, student $\mathbf{v}_\phi$ is trained so that its outputs lie on the same ODE trajectory as teacher $\mathbf{v}_\theta$ for any $t < s$:

$$\mathbf{x}_s = \mathbf{x}_t + (s-t)\,\mathbf{v}_\theta(\mathbf{x}_{t,\mathrm{LR}}, t) \tag{1}$$

Intermediate states are produced by the one-step student via:

$$\mathbf{x}_t = \mathbf{x}_0 + t\,\mathbf{v}_\phi(\mathbf{x}_{0,\mathrm{LR}}, t) \tag{2}$$

Substituting Equation (2) into Equation (1) yields:

$$s(\mathbf{v}_\phi(\mathbf{x}_{0,\mathrm{LR}}, s) - \mathbf{v}_\phi(\mathbf{x}_{0,\mathrm{LR}}, t)) = (s-t)(\mathbf{v}_\theta(\mathbf{x}_{t,\mathrm{LR}}, t) - \mathbf{v}_\phi(\mathbf{x}_{0,\mathrm{LR}}, t)). \tag{3}$$

Rearranging and simplifying leads to the final distillation objective:

$$\boxed{\mathcal{L}_{\mathrm{distill}}(\phi) = \mathbb{E}\left[\left\|\mathbf{v}_\phi(\mathbf{x}_{0,\mathrm{LR}}, s) - \mathrm{SG}\left[\mathbf{v}_\phi(\mathbf{x}_{0,\mathrm{LR}}, t) + \frac{\mathrm{d}t}{s}(\mathbf{v}_\theta(\mathbf{x}_{t,\mathrm{LR}}, t) - \mathbf{v}_\phi(\mathbf{x}_{0,\mathrm{LR}}, t))\right]\right\|_2^2\right]}$$

**Solver**

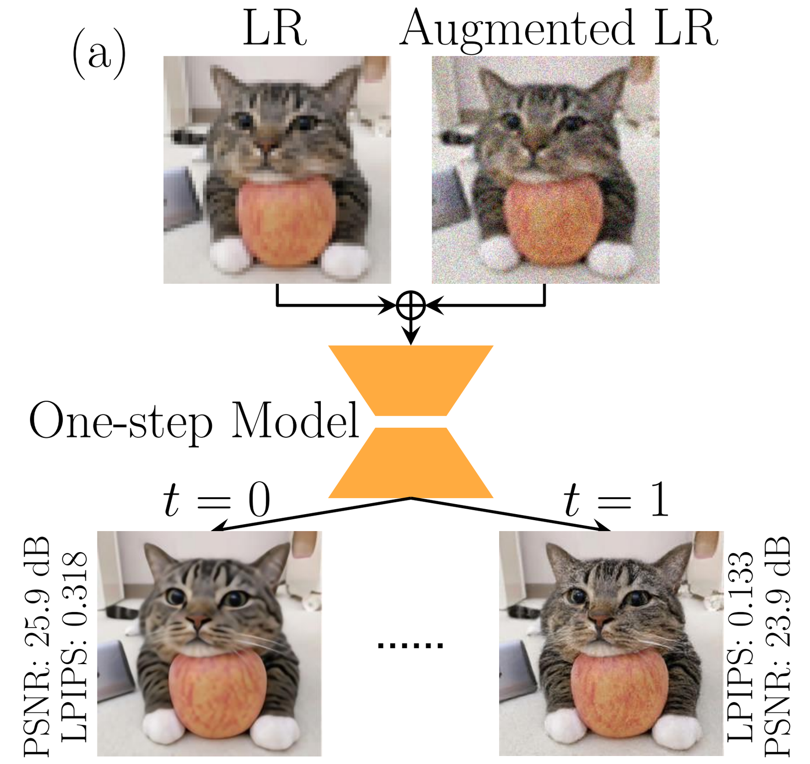
RK2 variants for $\mathbf{v}_\theta$ evaluation:

- **Midpoint** — main experiments
- **Heun's** — ablation
- **Ralston's** — ablation

**Key Properties**

→ No divide-by-zero ($s > t > 0$)
→ Stop-gradient SG[·] for stability
→ Related to MeanFlow & AlignYourFlow

## Noise-augmented Conditional Flow



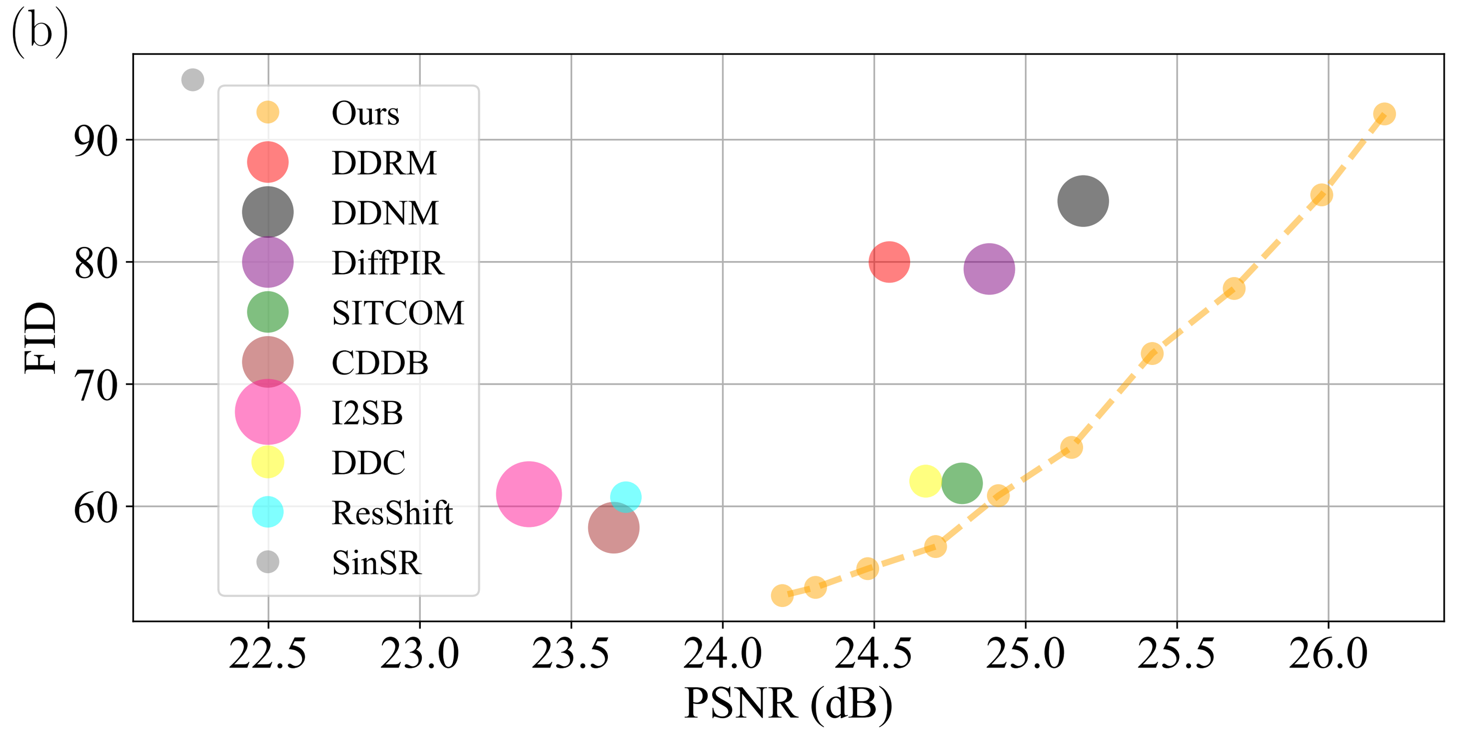




*Figure:* (a) Our model takes a LR image concatenated with its noise-augmented version as input, generating HR outputs with tunable realism-fidelity trade-off via interpolation parameter $t$. PSNR and LPIPS values are reported on each output. (b) Comparison of diffusion and flow-based SR methods on ImageNet 256×256, where bubble radius indicates the number of NFEs.

## OFTSR Distillation Pipeline

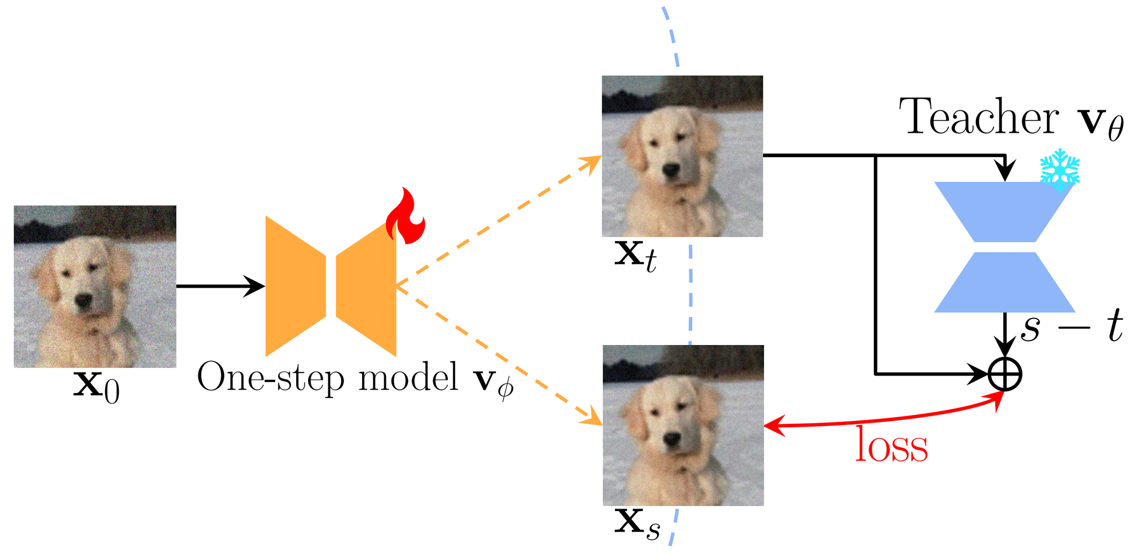



*Figure:* We leverage the teacher to align the one-step intermediate outputs, $\mathbf{x}_t$ and $\mathbf{x}_s$, along teacher's PF-ODE trajectory. For simplicity, LR conditioning is omitted here.

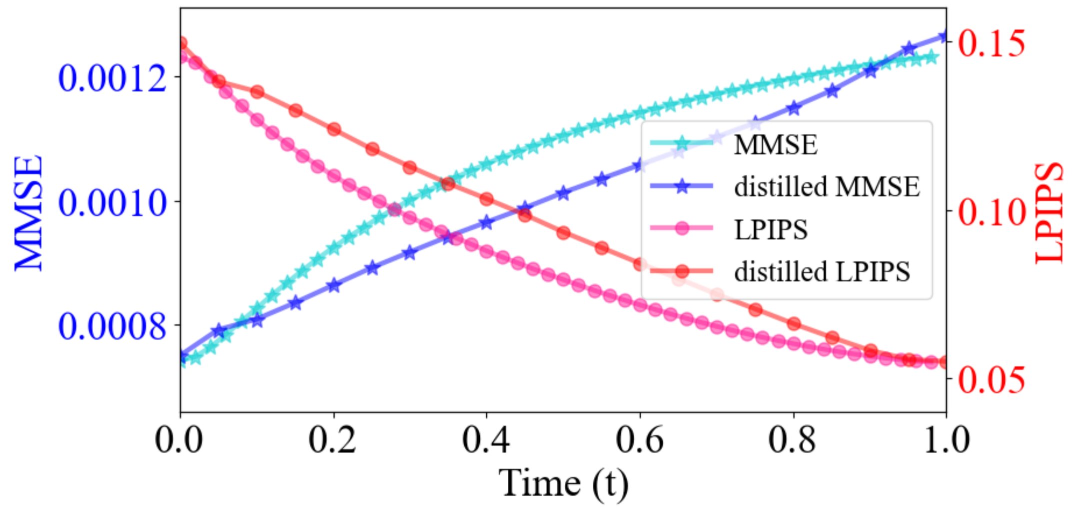



*Figure:* Continuous trade-off between MMSE (fidelity) and LPIPS (realism) controlled by parameter $t$. Performance for both pretrained teacher and distilled model are displayed.

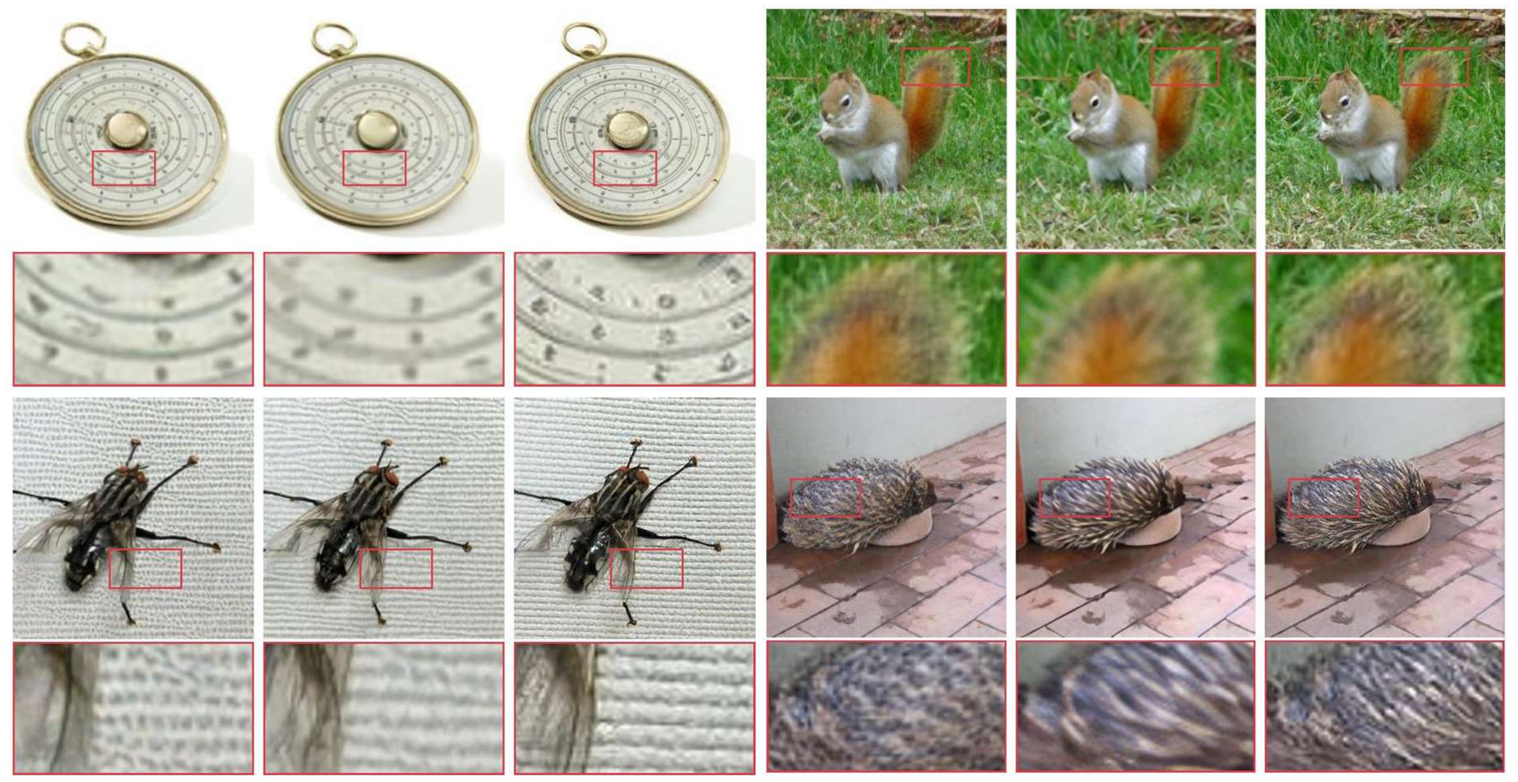
## Fidelity-Realism Trade-offs in One-step

Realism ⟷ Fidelity

| GT | $t=1$ | $t=0.8$ | $t=0.6$ | $t=0.4$ | $t=0.2$ | $t=0$ | LR |
|---|---|---|---|---|---|---|---|
| LPIPS / PSNR | 0.055 / 27.66 | 0.090 / 28.92 | 0.120 / 29.56 | 0.142 / 29.88 | 0.157 / 30.02 | 0.160 / 30.03 | 0.438 / 27.48 |

## Experimental Results

**1. Distilling OFTSR teacher**



I2SB CDDB Ours DDC ResShift Ours

**2. Distilling DiT4SR teacher**

LR CTMSR OSEDiff AddSR TSDSR OFTSR

## Future work

We plan to explore other Image Restoration tasks with OFTSR.