



南京大學  
NANJING UNIVERSITY

vivo

# UltraHR-100K: Enhancing UHR Image Synthesis with A Large-Scale High-Quality Dataset

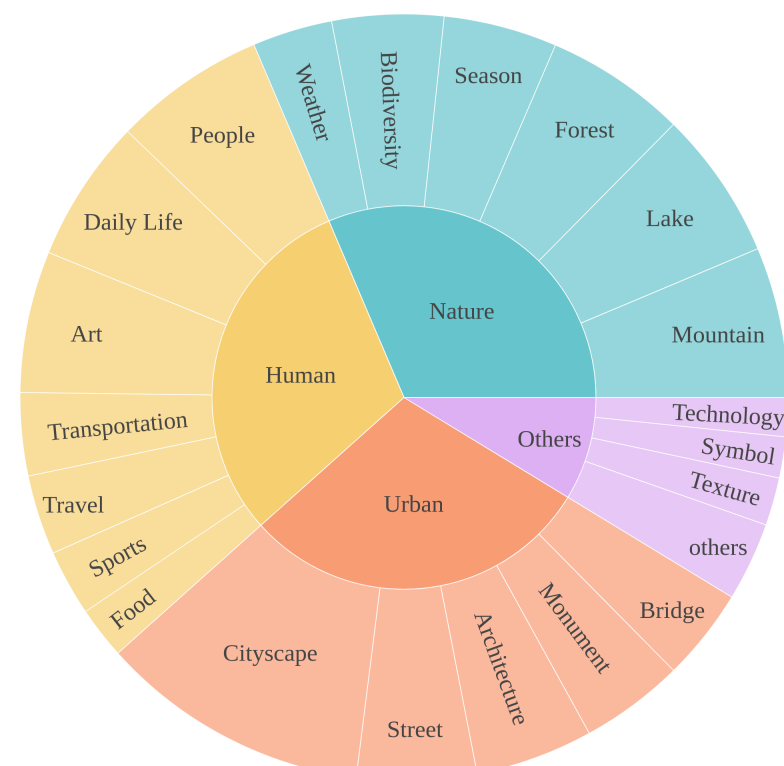
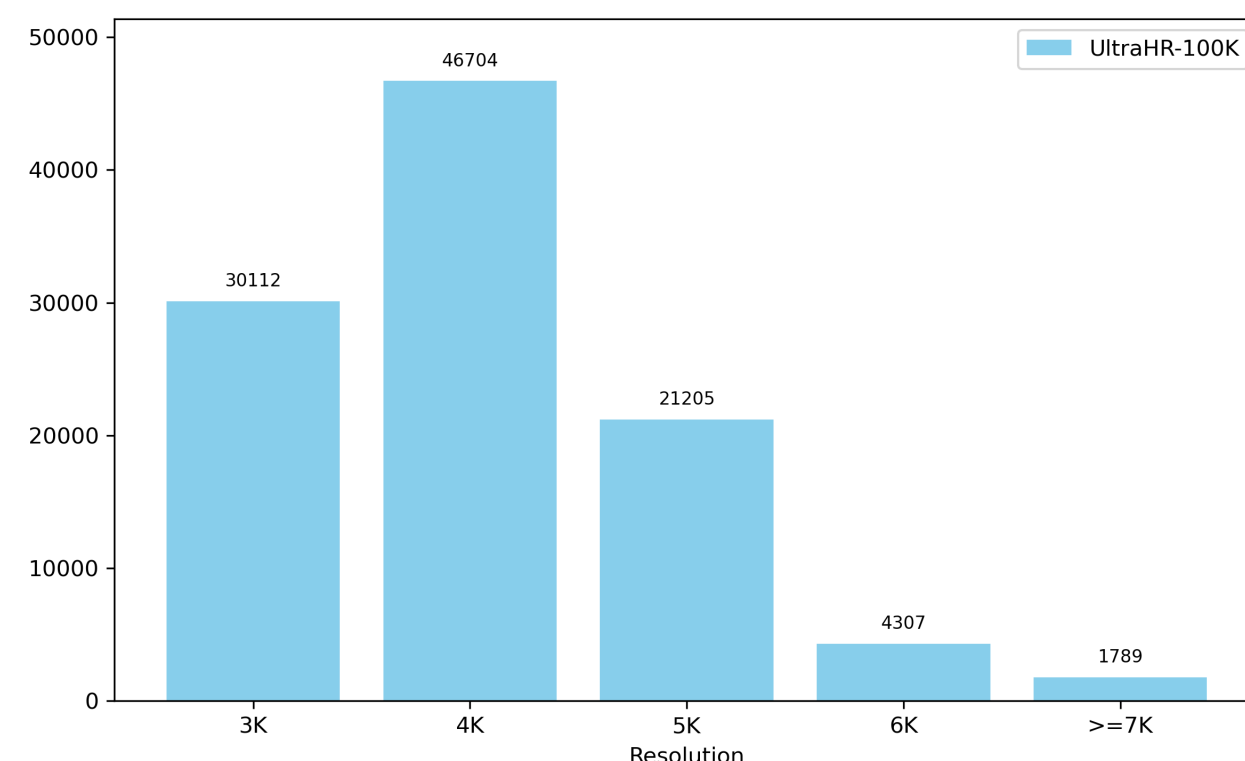
Chen Zhao<sup>1\*</sup>, En Ci<sup>1\*</sup>, Yunzhe Xu<sup>2</sup>, Tiehan Fan<sup>1</sup>, Shanyan Guan<sup>2</sup>, Yanhao Ge<sup>2</sup>, Jian Yang<sup>1</sup>, Ying Tai<sup>1†</sup>

\* Equal Contribution  
† Corresponding Author

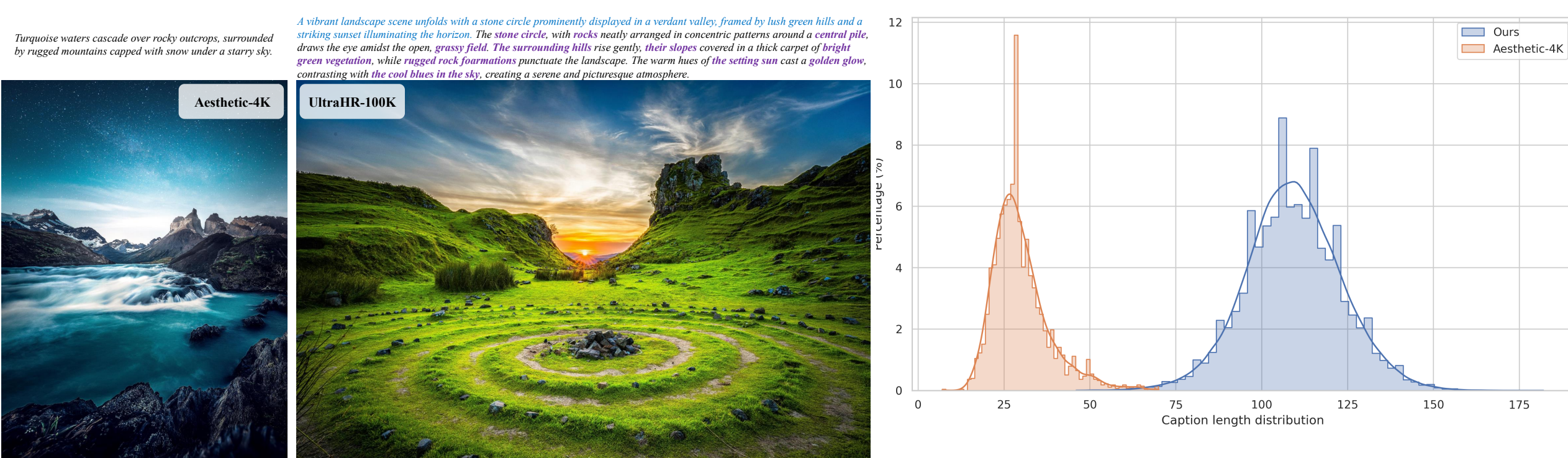
<sup>1</sup>State Key Laboratory of Novel Software Technology, Nanjing University <sup>2</sup>vivo



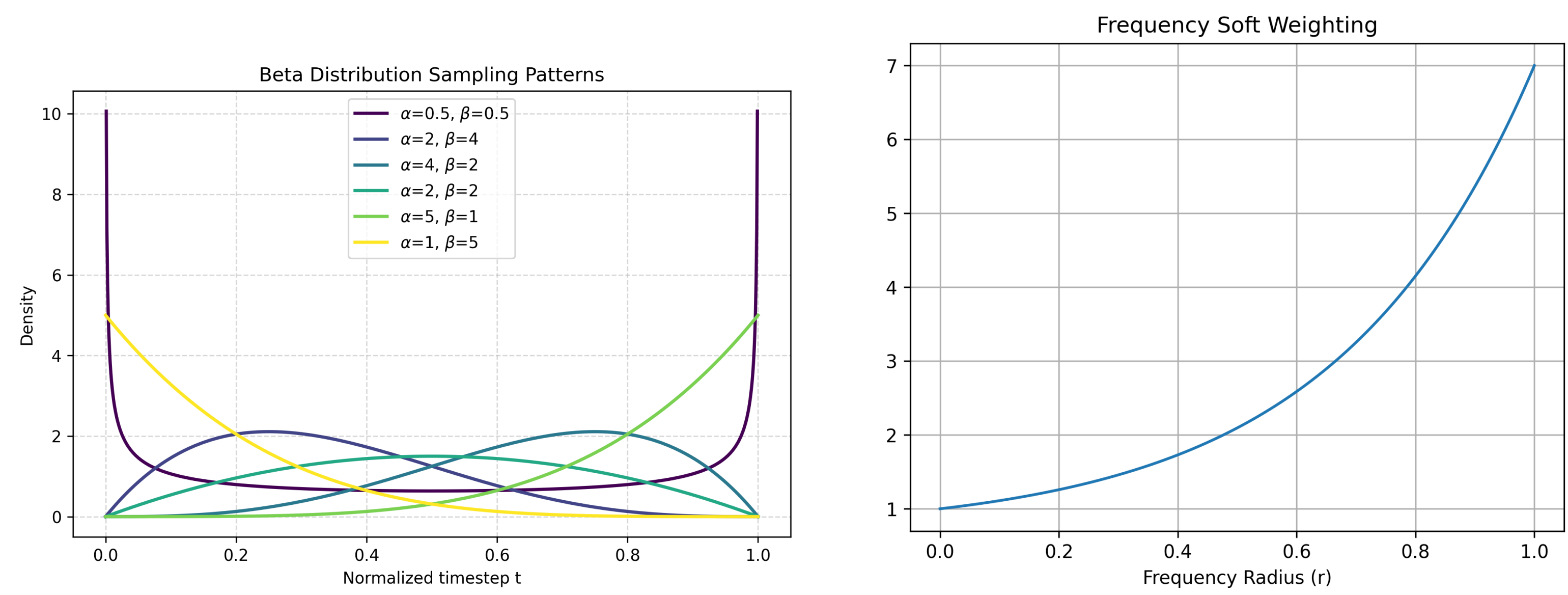
## UltraHR-100K



## Contribution



## Methodology



$$t \sim \text{Beta}(\alpha, \beta).$$

$$\pi_{\text{beta}}(t; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} t^{\alpha-1} (1-t)^{\beta-1}$$

$$\mathcal{L}_{\text{freq}} = \mathbb{E} \left[ |w(\mathbf{r}) \cdot \hat{\mathbf{x}} - w(\mathbf{r}) \cdot \hat{\mathbf{y}}|^2 \right],$$

$$w(\mathbf{r}) = 1 + \lambda \cdot \frac{\exp(\gamma \mathbf{r}) - 1}{\exp(\gamma) - 1}, \quad \mathbf{r} \in [0, 1],$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{diff}} + \lambda_{\text{freq}} \cdot \mathcal{L}_{\text{freq}}$$

## Quantitative Comparisons

Method	FID ↓	FID <sub>patch</sub> ↓	IS ↑	IS <sub>patch</sub> ↑	CLIP ↑	FG-CLIP ↑
FLUX [7] + BSRGAN [62]	37.651	43.143	11.773	5.389	31.45	28.02
SD3.5 [6] + BSRGAN [62]	31.870	25.598	12.780	5.456	31.75	28.66
I-Max(FLUX) [63]	37.667	37.835	11.991	4.391	31.49	27.78
HiFlow(FLUX) [13]	35.892	38.327	11.767	4.620	31.52	27.75
Pixart- $\sigma$ [17]	33.171	32.198	12.212	5.390	31.78	28.65
SANA [15]	37.070	38.795	11.778	<b>5.649</b>	31.70	28.60
Diffusion4K [16]	39.857	38.515	10.832	3.235	31.41	26.48
<b>Ours(UltraHR-100K)</b>	33.995	20.932	12.502	5.020	<b>31.85</b>	28.65
<b>Ours(UltraHR-100K+FAPT)</b>	<b>31.748</b>	<b>15.795</b>	<b>12.995</b>	5.104	31.82	<b>28.68</b>

## Qualitative Comparisons



## User & Ablation Study

Method	Overall Quality	Detail Quality	Text-Image Alignment	Preference
Pixart- $\sigma$ [17]	14%	10%	16%	18%
SANA [15]	4%	8%	8%	6%
Diffusion4K [16]	12%	4%	6%	6%
<b>Ours</b>	<b>70%</b>	<b>78%</b>	<b>72%</b>	<b>70%</b>

Model	DOTS	SWFR	Dataset	FID ↓	FID <sub>patch</sub> ↓	CLIP ↑
LoRA	×	×	Full	35.07	35.02	31.80
A	×	×	Full	33.99	20.93	<b>31.85</b>
B	✓	×	Full	32.57	19.95	31.79
C	✓	✓	Part	32.75	18.42	31.81
D	✓	✓	Full	<b>31.74</b>	<b>15.79</b>	31.82