A. Constructing Fibonacci Lattices

Fibonacci lattice constructs points that *"are evenly distributed with each of them representing almost the same area"* [30] in a unit square $[0,1)^2$ using the formula:

$$(x_i, y_i) = (i - \lfloor \frac{i}{\psi} \rfloor, \frac{i}{n}) \text{ for } 0 \leq i < n, \tag{12}$$

where $\psi = \lim_{n\to\infty}\left(\frac{F_{n+1}}{F_n}\right) = \frac{1+\sqrt{5}}{2}$. We can map this point distribution to a unit-length sphere $\mathbb{S}^2$ using cylindrical equal-area projection.

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} \to \begin{pmatrix} \theta_i = 2\pi x_i \\ \phi_i = \arccos(1-2y_i) \end{pmatrix} \to \begin{pmatrix} x_i' = \cos\theta_i \sin\phi_i \\ y_i' = \sin\theta_i \sin\phi_i \\ z_i' = \cos\phi_i \end{pmatrix}. \tag{13}$$

A.1. Generalizing the Spherical Fibonacci Lattice to higher dimensions

We provide the details to generalize the 3D spherical Fibonacci lattice [30] to higher dimensions [67].

Given a d-dimensional vector $\boldsymbol{u} = (u_1, u_2, \cdots, u_{d+1}) \in \mathbb{S}^d \subset \mathbb{R}^{d+1}$, we can also represent it in the hyperspherical coordinate system $(r, \varphi_1, \varphi_2, \cdots, \varphi_d)$, where $\varphi_1 \in [0, 2\pi]$, $\varphi_2, \cdots, \varphi_d \in [0, \pi]$, and specifically, $r = 1$ for $|\boldsymbol{u}| = 1$. The conversion to Cartesian coordinates is given as follows:

$$\begin{aligned} u_{d+1} &= \cos(\varphi_d) \\ u_d &= \sin(\varphi_d)\cos(\varphi_{d-1}) \\ &\vdots \\ u_2 &= \sin(\varphi_d)\sin(\varphi_{d-1})\cdots\sin(\varphi_2)\cos(\varphi_1) \\ u_1 &= \sin(\varphi_d)\sin(\varphi_{d-1})\cdots\sin(\varphi_2)\sin(\varphi_1). \end{aligned}$$

We examine the distribution over the angular coordinates $\Phi_{1,\cdots,d-1,d} \in [0, 2\pi] \times [0, \pi]^{d-1}$.

$$\begin{aligned} p(\Phi_{1:d}) &= p(\Phi_1, \Phi_2, \cdots, \Phi_d) &(14)\\ &= \rho(\Phi_1)\rho(\Phi_2|\Phi_1)\cdots\rho(\Phi_d|\Phi_1, \cdots, \Phi_{1:d-1}). &(15) \end{aligned}$$

The key observation is that the angles are independently distributed. To see this, if we fix $(\varphi_1, \cdots, \varphi_k)$, then $(\varphi_{k+1}, \cdots, \varphi_d)$ parameterizes a "subsphere" isoporphic to $\mathbb{S}^{d-k}$ with a rescaled radius $r' = \sin(\varphi_1)\cdots\sin(\varphi_k)$. In other words, $p(\Phi_{k+1:d}|\Phi_{1:k}) = p(\Phi_{k+1:d})$ for any $k \in [d]$. As such, Equation 15 can be simplified to:

$$p(\Phi_{1:d}) = \prod_{\alpha=1}^{d} p_\alpha(\Phi_\alpha). \tag{16}$$

The absolute value of the Jacobian determinant for the change of variables $(u_1, \cdots, u_{d+1}) \mapsto (r, \varphi_1, \cdots, \varphi_d)$ is

$$\left|\frac{\partial(u_1, \cdots, u_{d+1})}{\partial(r, \varphi_1, \cdots, \varphi_d)}\right| = r^d \prod_{k=2}^{d} \sin^{k-1}\varphi_k. \tag{17}$$

Therefore, Equation 15 reduces to

$$p(\Phi_{1:d}) \propto \prod_{k=2}^{d} \sin^{k-1}\varphi_k. \tag{18}$$

With Equation 16, we have $p(\Phi_k) \propto \sin^{k-1}\varphi_k$. Then we get the normalization constants $Z_k = \int_0^\pi \sin^{k-1}\varphi d\varphi = \frac{\sqrt{\pi}\cdot\Gamma(k/2)}{\Gamma((k+1)/2)}$ for $k = 2, \cdots, d$ and $Z_1 = \int_0^{2\pi} d\varphi_1 = 2\pi$. Finally, we arrive at

$$P(\Phi_k = \varphi_k) = \begin{cases} \frac{1}{2\pi}, & k = 1 \\ \frac{1}{\sqrt{\pi}}\frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}\sin^{k-1}\varphi_k, & k = 2, \cdots, d \end{cases} \tag{19}$$

Denote the cumulative distribution function with another variable Y

$$\begin{aligned} P(Y = y) = F_\Phi(\varphi) &= \int_0^{\varphi_k} p(\Phi = u)du &(20)\\ &= \begin{cases} \frac{1}{2\pi}y, & k = 1 \\ \dots \end{cases}. &(21) \end{aligned}$$

The Fibonacci-like spiral $\boldsymbol{Y}^{(n)} = \left(\boldsymbol{Y}_1^{(n)}, \cdots, \boldsymbol{Y}_d^{(n)}\right)$ is generated by the following formula:

$$\begin{aligned} \boldsymbol{Y}_d^{(n)} &= \frac{n}{N+1}, &(22)\\ \boldsymbol{Y}_{d-1}^{(n)} &= \{na_1\}, &(23)\\ &\vdots &(24)\\ \boldsymbol{Y}_1^{(n)} &= \{na_{d-1}\}, &(25)\\ & &(26) \end{aligned}$$

where $\{x\}$ refers to x's decimal part, *i.e.* $\{x\} = x - \lfloor x \rfloor$. $a_{1:d}$ satisfies $\frac{a_i}{a_j} \notin \mathbb{Q}, \forall i \neq j$.

The angles are given by taking the inverse:

$$\varphi_d^{(n)} = F^{-1}(Y_d^{(n)}) \tag{27}$$

$$\varphi[t+1] = \varphi[t] - \frac{F(\varphi[t]) - Y}{F'(\varphi[t])} \tag{28}$$

Table 10. **Generator matrix for the Leech lattice Λ_{24}**. The table is adapted from Table 4.12 in [18].

$$\frac{1}{\sqrt{8}}\begin{bmatrix}
8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
4 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 2 & 2 & 2 & 0 & 0 & 0 & 0 & 2 & 2 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 0 & 2 & 0 & 2 & 0 & 0 & 2 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 2 & 2 & 2 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\
2 & 2 & 0 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\
\hline
0 & 2 & 2 & 2 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 0 & 2 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 & 0 \\
-3 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}$$

B. Details of the Leech Lattice

Generator matrix. The generator matrix for the *unconstrained* Λ_{24} is given in Table 10.

C. Experimental Details

C.1. Architectures

ViT tokenizer. The architecture used to get the main results follows [91]. To conduct ablative studies in Table 8, we use a smaller architecture (ViT-small) for fast iteration. Both the hidden dimension and the number of layers are halved.

CNN-based VAR tokenizer. The architecture follows Infinity [36][10]. We use 7 scales: $[1^2, 2^2, 4^2, 6^2, 8^2, 12^2, 16^2]$, which amounts to 521 tokens in total.

Infinity-CC. The Infinity model used in the paper mostly follows the original Infinity paper [36]. We make two key changes: (1) The text condition is converted to class condition, which is representation by a single <SOS> token. The vocabulary size is augmented to $|\mathcal{C}| + K + 1$, where $K = 1000$. The additional 1 refers to the "no-class" index, which is used to randomly replace the original class index with a probability of 0.1, to enable classifier-free guidance (CFG) at the sampling phase. (2) We revert the shared AdaLN to an unshared version following VAR [75]. Table 11 summarizes the model configurations and parameter size in this paper.

[10] https://github.com/FoundationVision/BitVAE

C.2. Training specifications

ViT tokenizer. We train the image tokenizer with a batch size of 32 per GPU. We use AdamW optimizer with $(\beta_1, \beta_2) = (0.9, 0.99)$ with 1×10^{-4} weight decay. The base learning rate is 4×10^{-7} (or a total learning rate of 1×10^{-4}) and follows a half-period cosine annealing schedule. The model is trained for 1M steps, which amounts to 200 epochs over the entire ImageNet-1k training set. We use an ℓ_1 loss weight of 1, a perceptual loss weight of 0.1, and an adversarial loss weight of 0.1 throughout the experiments.

VAR tokenizer. We train the VAR tokenizer with a batch size of 8 per GPU. Two schedules are used: (1) the fast schedule trains the model for 500k iterations with 8 GPUs, which approximately sees the training data 25 epochs; (2)

Table 11. **Model configurations for Infinity-CC.**

layer	embed dim	# heads	# params (head)	# epochs
12	768	8	242M (151M)	50
16	1152	12	394M (226M)	200
24	1536	16	1B (300M)	350
32	2048	16	2.8B (402M)	400

Table 12. **VAR Tokenizer.**

	$\lvert\mathcal{C}\rvert$	rFID$_\downarrow$	LPIPS$_\downarrow$	SSIM$_\uparrow$	PSNR$_\uparrow$
(fast schedule)					
BSQ	16,384	1.82	0.1268	0.5626	19.989
$\mathbb{\Lambda}_{24}$-SQ†	16,384	1.36	0.1170	0.5957	20.639
BSQ	262,144	1.29	0.1106	0.6006	20.683
$\mathbb{\Lambda}_{24}$-SQ	196,560	1.08	0.1005	0.6280	21.315
$\mathbb{\Lambda}_{24}$-SQ (vf)	196,560	1.18	0.1088	0.6006	20.734
(standard schedule)					
BSQ	262,144	1.07	0.1064	0.6035	20.430
$\mathbb{\Lambda}_{24}$-SQ	196,560	**0.84**	**0.0954**	**0.6333**	**21.535**
$\mathbb{\Lambda}_{24}$-SQ (vf)	196,560	<u>0.92</u>	<u>0.1041</u>	<u>0.6118</u>	<u>21.188</u>

the standard schedule trains the model for 500k iterations with 32 GPUs, which is approximately 100 epochs.

D. More Results

D.1. VAR tokenization

First, we retrain a VAR tokenizer with a fast schedule (25 epochs). We use $\mathbb{\Lambda}_{24}$-SQ as the bottleneck with two codebook sizes: (1) the full codebook, whose bitrate is similar to BSQ ($d = 18$), and (2) a subset of 16,384 codes, whose bitrate is equivalent to BSQ ($d = 14$). From the upper half of Table 12, $\mathbb{\Lambda}_{24}$-SQ outperforms BSQ in all metrics in both cases. Next, we train a VAR tokenizer with the standard schedule (100 epochs) suggested in Infinity [36]. The full numbers are reported in the bottom half of Table 12, supplementing Figure 5.

D.2. VAR generation

Grid search of sampling parameters. We run a small-scale grid search of sampling hyperparameters for Infinity-CC with different prediction heads. We compare the gFID score on IN-1k by generating 10 samples per class (10k generated samples in total). From Figure 8, we conclude that the optimal top k varies significantly across different prediction head settings.

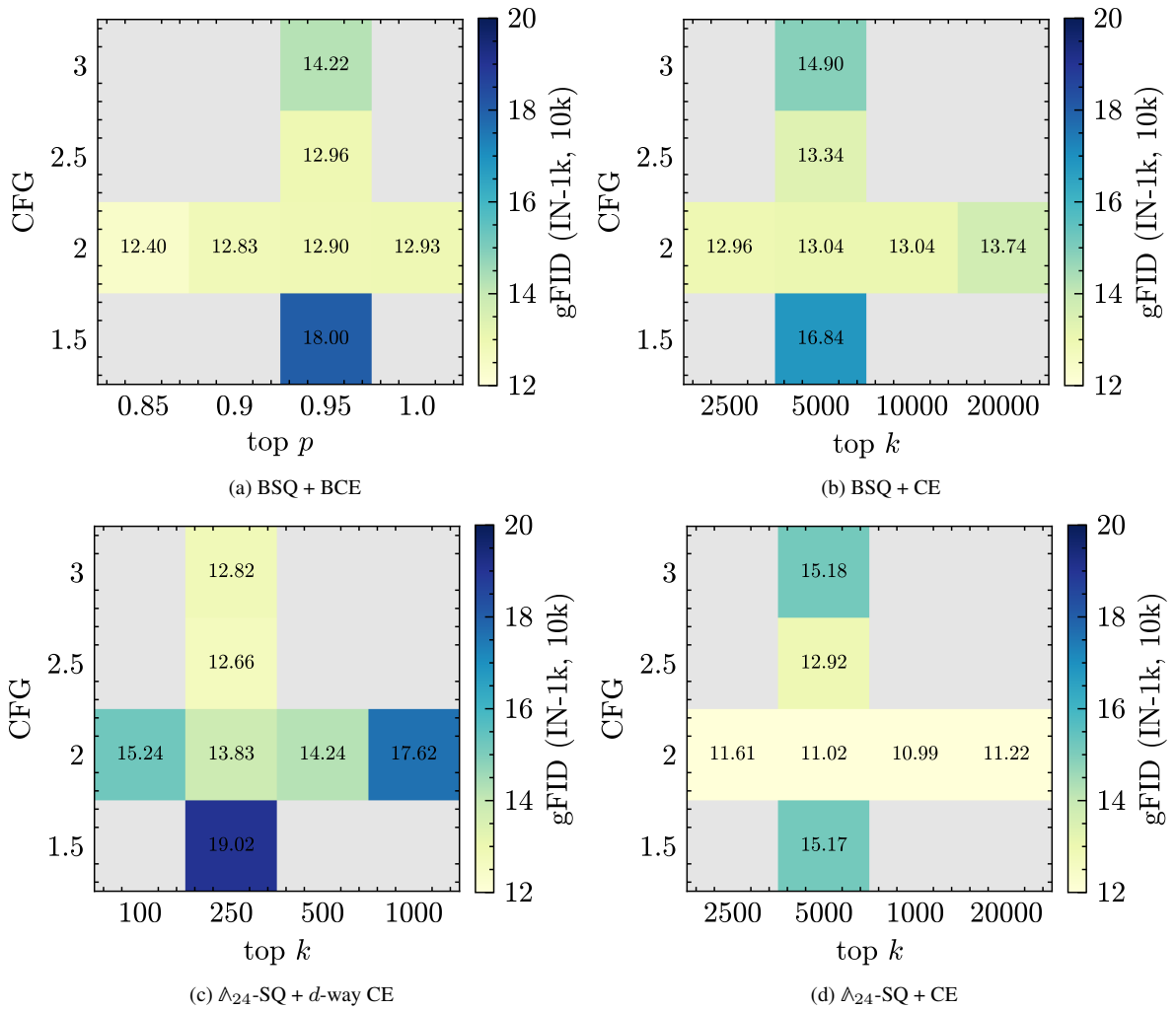
Advanced sampling techniques. In Section 4.3, we introduced advanced sampling techniques, including layerwise linearly scaling CFG and linearly scaling top-k. We show related ablation studies in Table 13. We use $\mathrm{lin}(x_0, s)$ to denote the linear scaling strategy, which starts from x_0 and changes by s per scale. We can see that both layerwise linear scaling CFG and top-k bring a noticeable improvement.

Table 13. **Advanced sampling techniques.** $\mathrm{lin}(x_0, \pm s)$ denotes the linear scaling strategy which starts from x_0 and increment/decrements by s per scale.

Tokenizer	rFID	CFG	top k	gFID
$\mathbb{\Lambda}_{24}$-SQ (25 ep)	1.08	2	5×10^3	8.78
$\mathbb{\Lambda}_{24}$-SQ (100 ep)	0.84	2	5×10^3	7.46
$\mathbb{\Lambda}_{24}$-SQ (100 ep)	0.84	$\mathrm{lin}(1, 0.33)$	5×10^3	6.81
$\mathbb{\Lambda}_{24}$-SQ (100 ep)	0.84	$\mathrm{lin}(1, 0.25)$	5×10^3	7.33
$\mathbb{\Lambda}_{24}$-SQ (100 ep)	0.84	$\mathrm{lin}(1, 0.33)$	$\mathrm{lin}(10^4, -10^3)$	6.68
$\mathbb{\Lambda}_{24}$-SQ (vf)	1.18	$\mathrm{lin}(1, 0.33)$	5×10^3	5.79
$\mathbb{\Lambda}_{24}$-SQ (vf)	1.18	$\mathrm{lin}(1, 0.33)$	$2,500$	5.41
$\mathbb{\Lambda}_{24}$-SQ (vf)	1.18	$\mathrm{lin}(1, 0.33)$	$\mathrm{lin}(2000, -100)$	5.30

It is also worth noting that, according to the bottom half of Table 13, the optimal k decreases when the tokenizer is trained with the VF loss. This is most likely because the probability density is more skewed, as is illustrated in Figure 4.

Qualitative Results. Figure 9 shows more generation results sampled by Infinity-CC + $\mathbb{\Lambda}_{24}$-SQ (2B). We cherry-pick the images and emphasize the quality *and* diversity.



(a) BSQ + BCE

(b) BSQ + CE

(c) $\mathbb{Z}_{24}$-SQ + d-way CE

(d) $\mathbb{Z}_{24}$-SQ + CE

Figure 8. **Hyperparameter grid search supplementing Table 9.** We use a fixed temperature $\tau = 1$ in (a-d) and top $p = 0.95$ in (b-d).

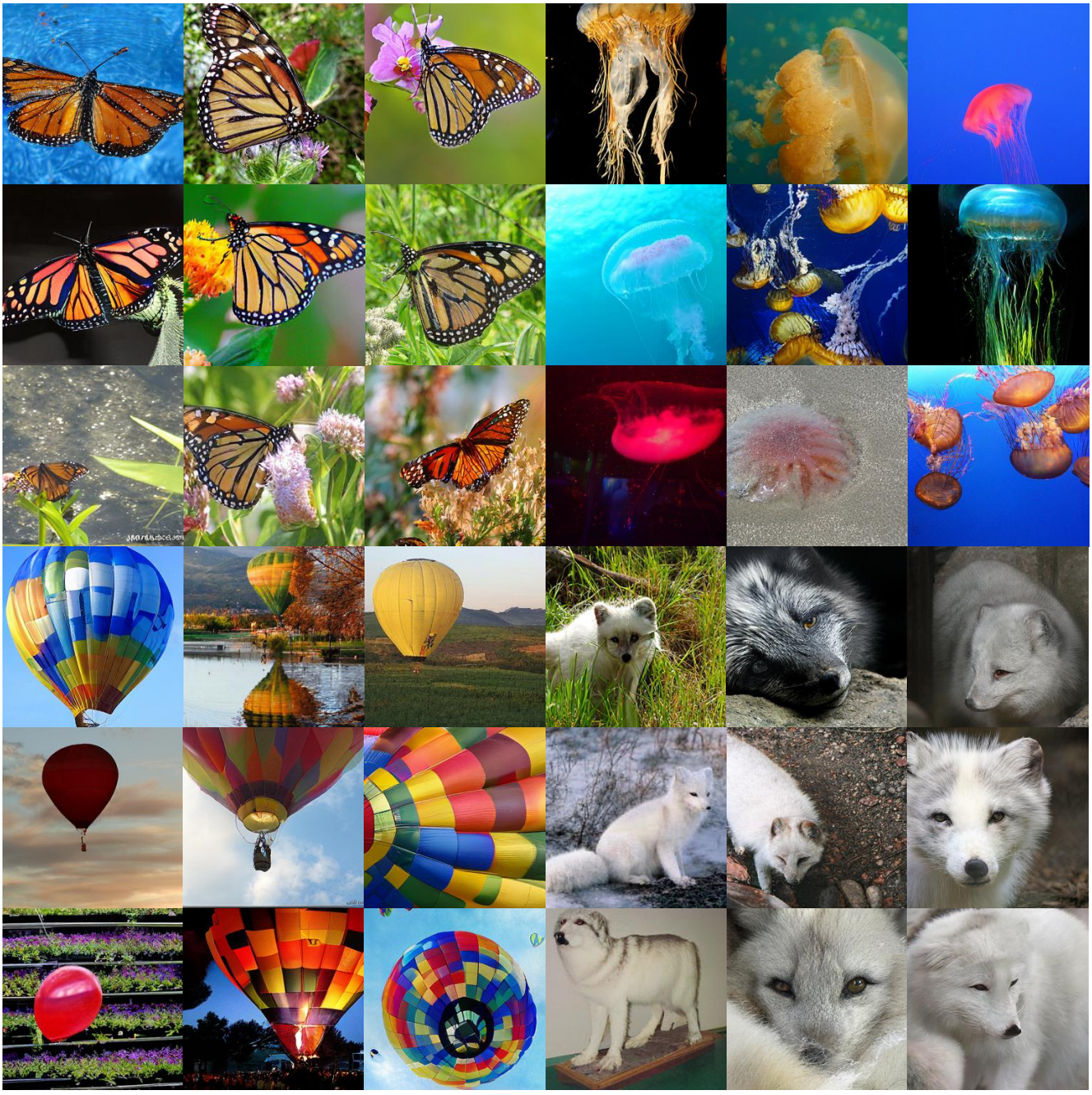


Figure 9. **More sampled generation results of Infinity-CC + $\mathbb{\Lambda}_{24}$-SQ (2B).** Classes are 323: monarch butterfly; 107: jellyfish; 417: balloon; 279: arctic fox.