Supplementary Material for D^{4}M: Dataset Distillation via Disentangled Diffusion Model

Duo Su[1,5,6,†] Junjie Hou[2,5,6,†] Weizhi Gao[3] Yingjie Tian[4,5,6,7,∗] Bowen Tang[8]

[1]School of Computer Science and Technology, UCAS [2]Sino-Danish College, UCAS
[3]Department of Computer Science, NCSU [4]School of Economics and Management, UCAS
[5]Research Center on Fictitious Economy and Data Science, CAS
[6]Key Laboratory of Big Data Mining and Knowledge Management, CAS
[7]MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation, UCAS
[8]Institute of Computing Technology, CAS
https://junjie31.github.io/D4M/

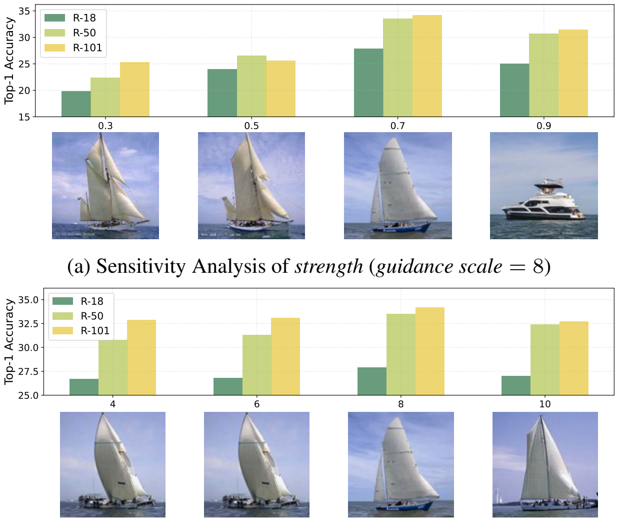
1. Experimental Settings

In our experimental framework, we primarily concentrate on the parameters of the synthesis and the Training-Time Matching (TTM) processes. For the synthesis phase, Stable Diffusion (V1-5) serves as the core mechanism in Latent Diffusion Model implementation. Based on the insights of Sec. 2.1, we calibrate the *strength* and *guidance scale* parameters at 0.7 and 8, respectively. During the prototype learning, the Mini-Batch k-Means algorithm is employed, with an in-depth ablation study of cluster number variations presented in Sec. 2.2. Furthermore, in scenarios where the IPC is less than 100, we adjust the cluster numbers to match the IPC. Within the TTM process, the comprehensive parameter settings of student networks are provided in Tab. 3.

2. Hyper-parameter Analysis

2.1. Sensitivity Analysis

There are two hyper-parameters in the diffusion model with text prompts, *i.e.* *strength* ($0 < s < 1$) and *guidance scale* ($g > 1$). Conceptually, the *strength* quantifies the extent of noise infusion into the latent features (prototypes). The diffusion model predominantly disregards these features in scenarios where *strength* equals 1. Furthermore, an elevated *guidance scale* fosters the generation of images that more precisely align with the text prompt. Based on the hyper-parameter tuning results in Fig. 1a and Fig. 1b, we suggest setting *strength* $= 0.7$ and *guidance scale* $= 8$.



(a) Sensitivity Analysis of *strength* (*guidance scale* $= 8$)

(b) Sensitivity Analysis of *guidance scale* (*strength* $= 0.7$)

Figure 1. **Sensitivity analysis of *strength* and *guidance scale*.** Quantitative results are evaluated on ResNet. Furthermore, qualitative results are presented to illustrate the variations corresponding to parameter adjustments.

2.2. Number of Prototypes

To ensure the feature diversity of the distilled dataset, multiple prototypes are learned for each category in our experiments. We select 10 or 50 prototypes to generate distilled ImageNet-1K datasets (IPC-100/200) respectively, *i.e.* synthesizing multiple images per prototype. These datasets are then trained across three distinct ResNet architectures, with the corresponding outcomes detailed in Tab. 1.

Given the marginal disparity observed between the ex-

[†] Equal contribution. ∗ Corresponding authors.

perimental results of the two groups, we conducted an independent sample t-test. The alternative hypothesis is that the true difference in means is not equal to 0. According to the p-value, at a significance threshold of 0.05, the performance variations of each group are not statistically significant, which means that the distilled datasets are not sensitive to the number of prototypes.

In addition, the t-SNE visualization results of D^{4}M on ImageNet-1K are displayed in Fig. 2. Except for a few outliers, the features extracted from the D^{4}M distilled ImageNet-1K dataset are compact and discriminative for both different and similar categories.

IPC	Prototypes	R18	R50	R101	p-value
100	10	59.0	64.4	65.9	0.7391
	50	59.3	65.4	66.5	
200	10	62.4	67.6	68.2	
	50	62.6	67.8	68.1	

Table 1. **Ablation study on the cluster numbers.** We utilize 10 or 50 prototypes to synthesize images for IPC-100/200 and evaluate them with ResNet.

3. Quantitative Analysis

In the main text, we delve into the enhancement of input-output image space consistency constraints for addressing cross-architecture generalization challenges. This section presents a direct comparative analysis of the image quality yielded by D^{4}M against the benchmark, as detailed in Tab. 2.

Dataset	Method	IS↑	FID↓	KID↓
ImageNet-1K	SRe^{2}L	28.872	59.119	0.047
	D^{4}M	**49.381**	**9.419**	**0.003**
Tiny-ImageNet	SRe^{2}L	6.243	74.814	0.055
	D^{4}M	**25.866**	**34.702**	**0.020**

Table 2. Quantitative results of distilled image. Comparing the quality of distilled images using IPC-50 on ImageNet-1K and Tiny-ImageNet, D^{4}M consistently outperforms SRe^{2}L across IS, FID, and KID metrics. This demonstrates that the distilled images produced by D^{4}M exhibit higher image quality.

Firstly, we employ the Inception Score (IS) to assess the clarity $p(y \mid x)$ of the synthetic images and the feature diversity $p(y)$ of the generative model G. The IS quantifies the KL divergence between the probability distribution and the conditional probability distribution of the features, as extracted by the Inception V3 model:

$$\mathrm{IS} = \exp\left(\mathbb{E}_{x\sim p_G} D_{KL}(p(y \mid x) \| p(y))\right). \tag{1}$$

Moreover, to demonstrate that the D^{4}M enhances the consistency between synthetic and real images, we compute the Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) metrics for these datasets. Empirical evaluations demonstrate that D^{4}M is capable of generating a variety of high-resolution images while maintaining consistency between the input and output image spaces.

4. More Visualizations

We randomly select the visualizations to enhance the understanding of our methods and easier to reference. The distilled CIFAR-10 and CIFAR-100 are shown in Fig. 3 and Fig. 4. Furthermore, the distilled ImageNet-1K is shown in Fig. 5-Fig. 14. (**more pages after this paragraph**)

(a) ImageNet-1K and Tiny-ImageNet

Settings	Values
network	ResNet
input size	224
batch size	1024
epoch	300
augmentation	RandomResizedCrop
min scale	0.08
max scale	1
temperature	20
optimizer	AdamW
learning rate	0.001
weight decay	0.01
learning rate schedule	cosine decay

(b) CIFAR-10 and CIFAR-100

Settings	Values
network	ConvNet
input size	32
batch size	100
epoch	500
augmentation	RandomResizedCrop
min scale	0.08
max scale	1
temperature	20
optimizer	AdamW
learning rate	0.001
weight decay	0.01
learning rate schedule	cosine decay

Table 3. Parameter settings of the student networks.

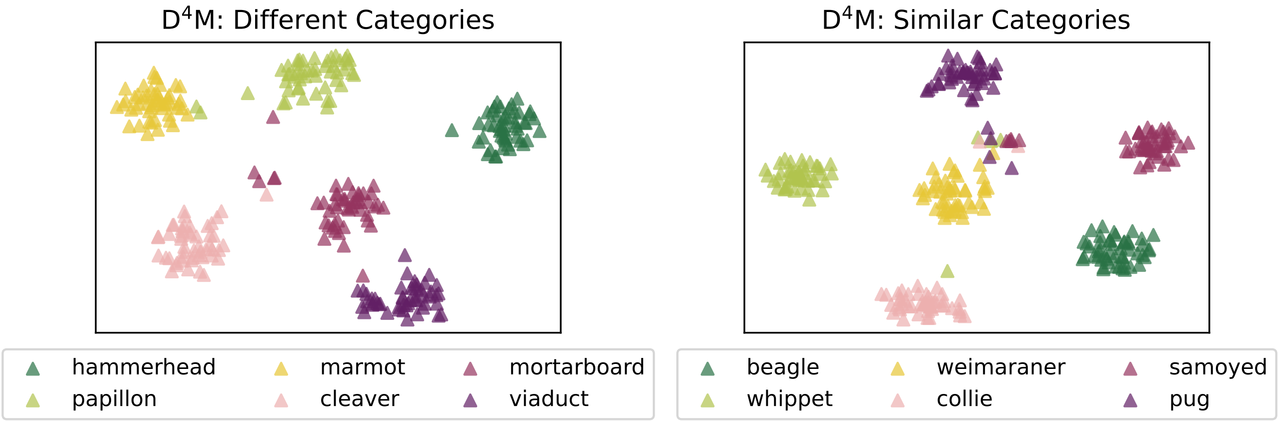


Figure 2. **T-SNE visualizations on ImageNet-1K.** The features are extracted by ResNet-18.

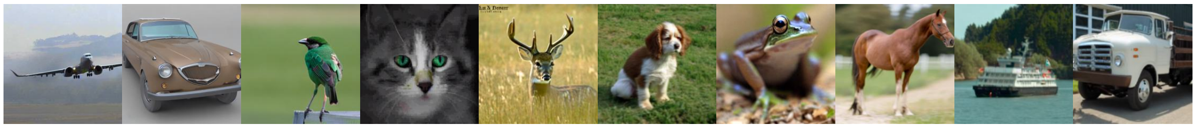


Figure 3. More visualizations selected from the distilled CIFAR-10 (Class 0-9)

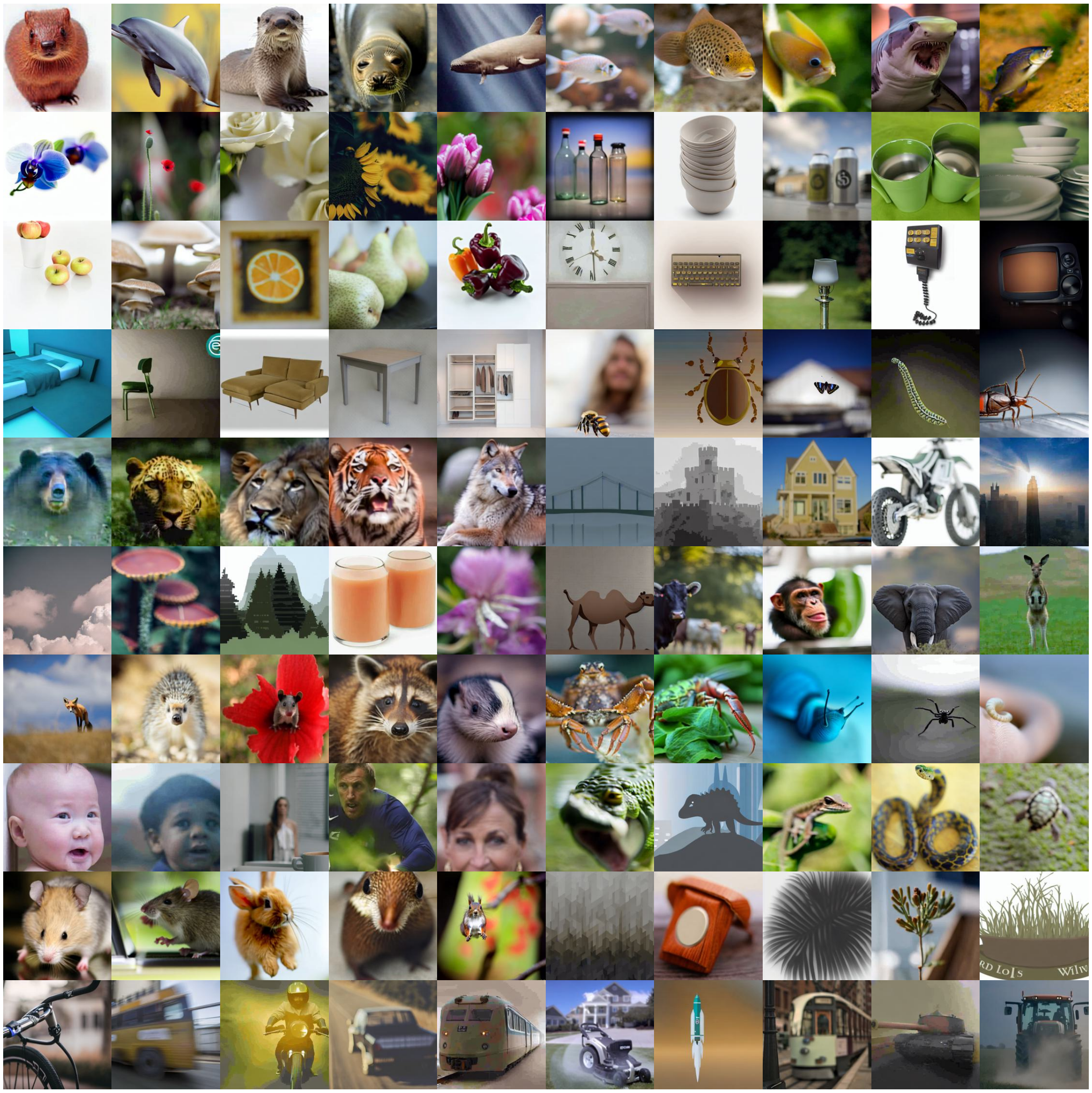


Figure 4. More visualizations selected from the distilled CIFAR-100 (Class 0-99)

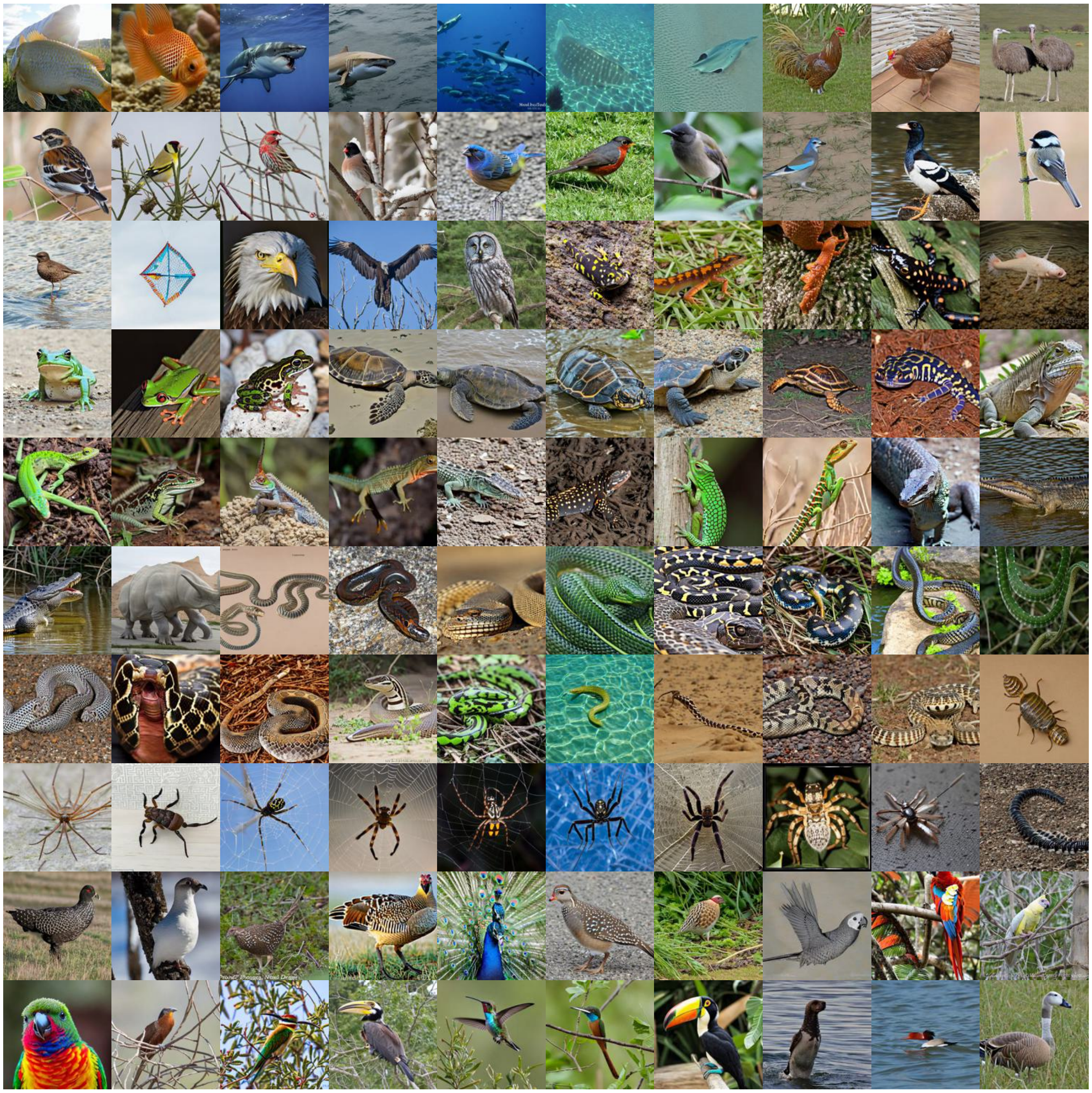


Figure 5. More visualizations selected from the distilled ImageNet-1K (Class 0-99)

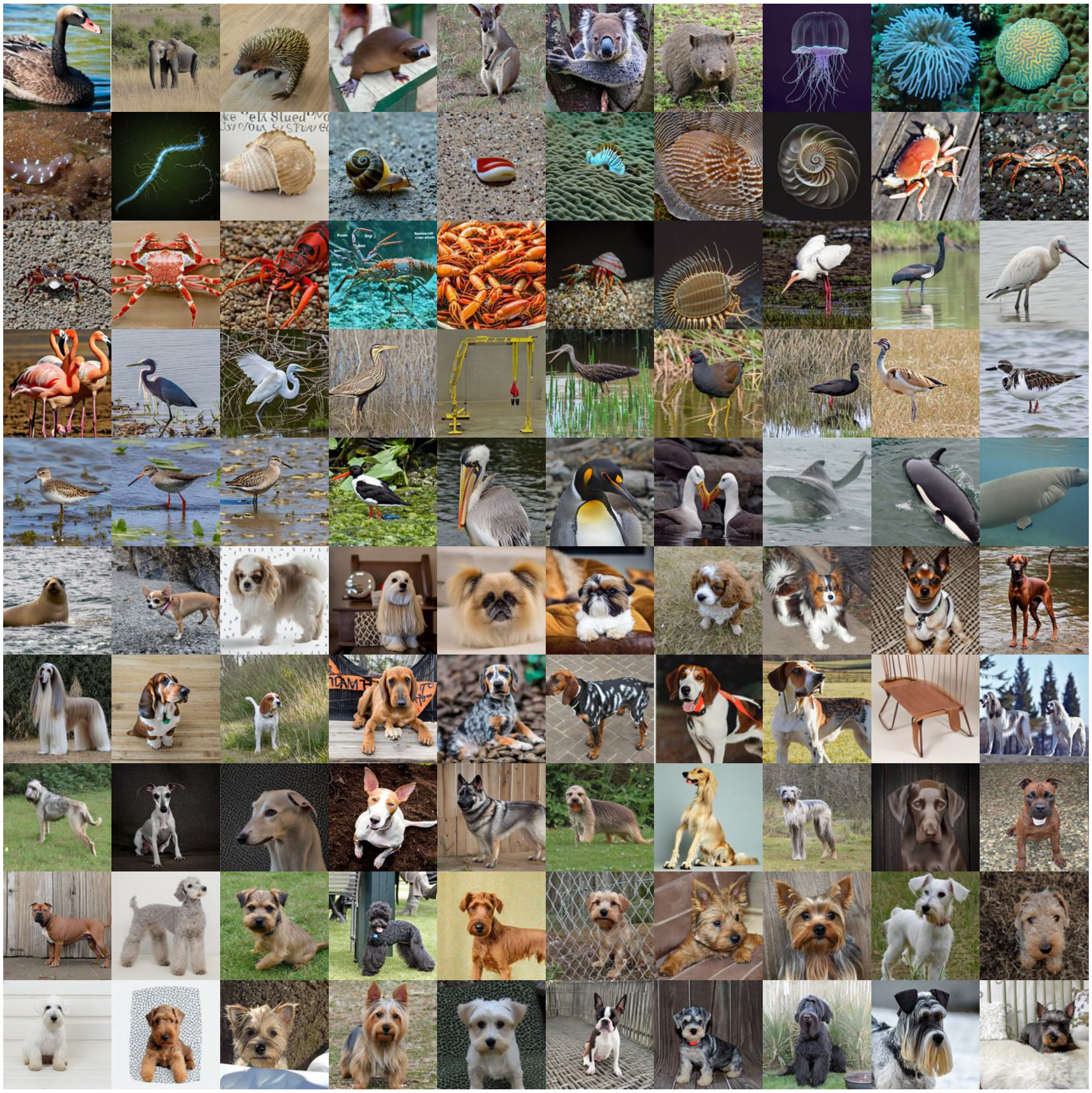


Figure 6. More visualizations selected from the distilled ImageNet-1K (Class 100-199)

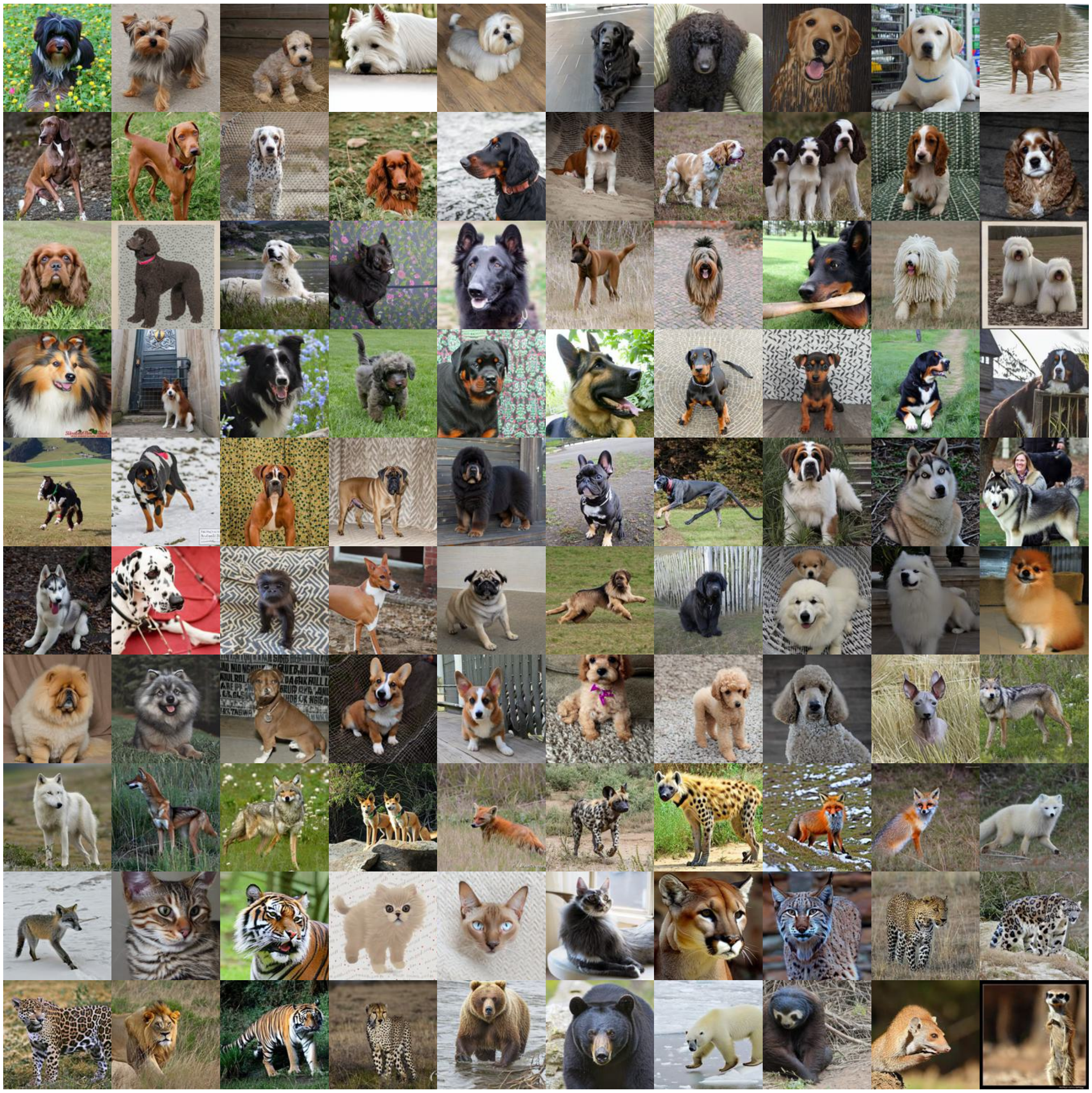


Figure 7. More visualizations selected from the distilled ImageNet-1K (Class 200-299)

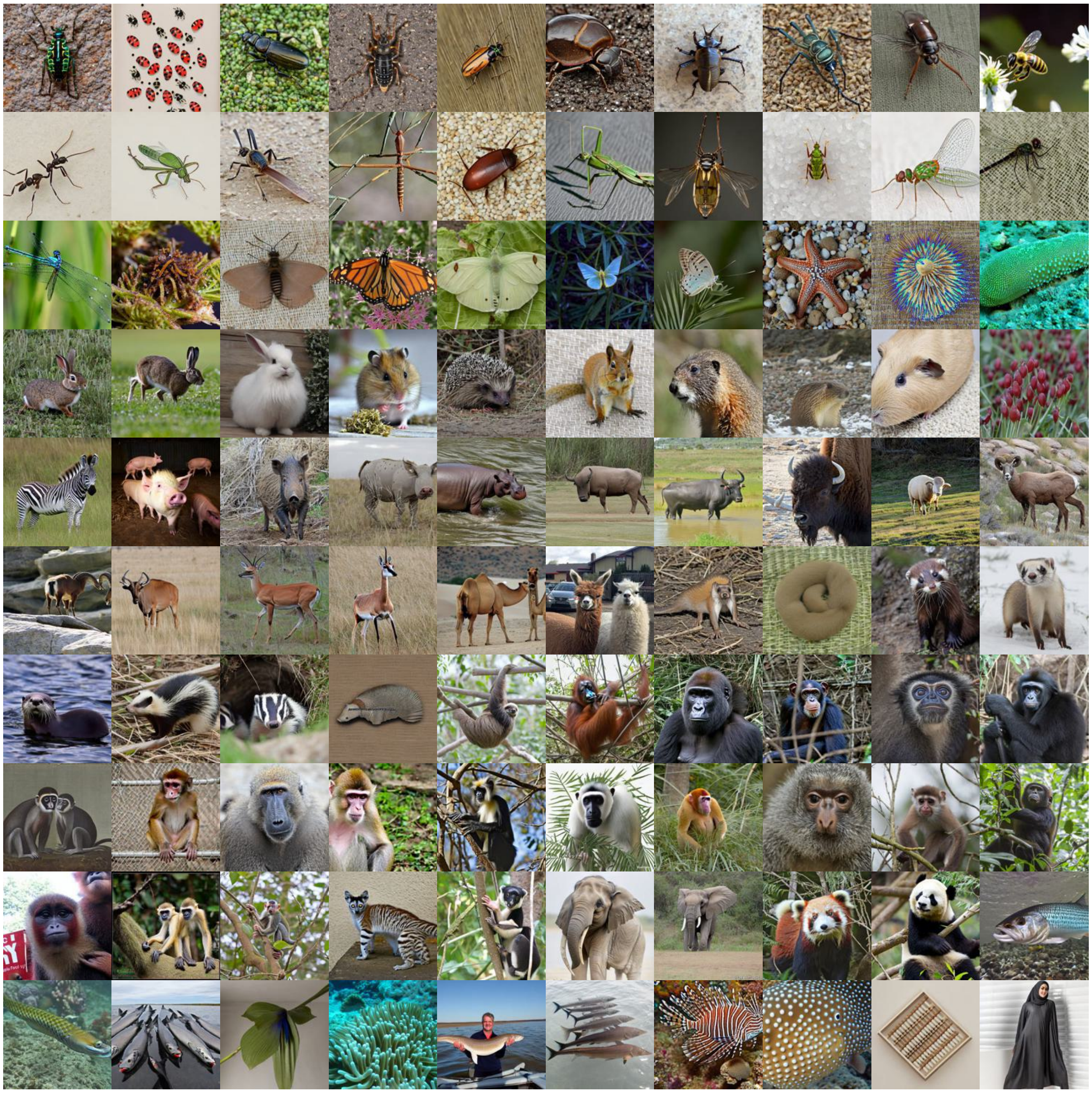


Figure 8. More visualizations selected from the distilled ImageNet-1K (Class 300-399)

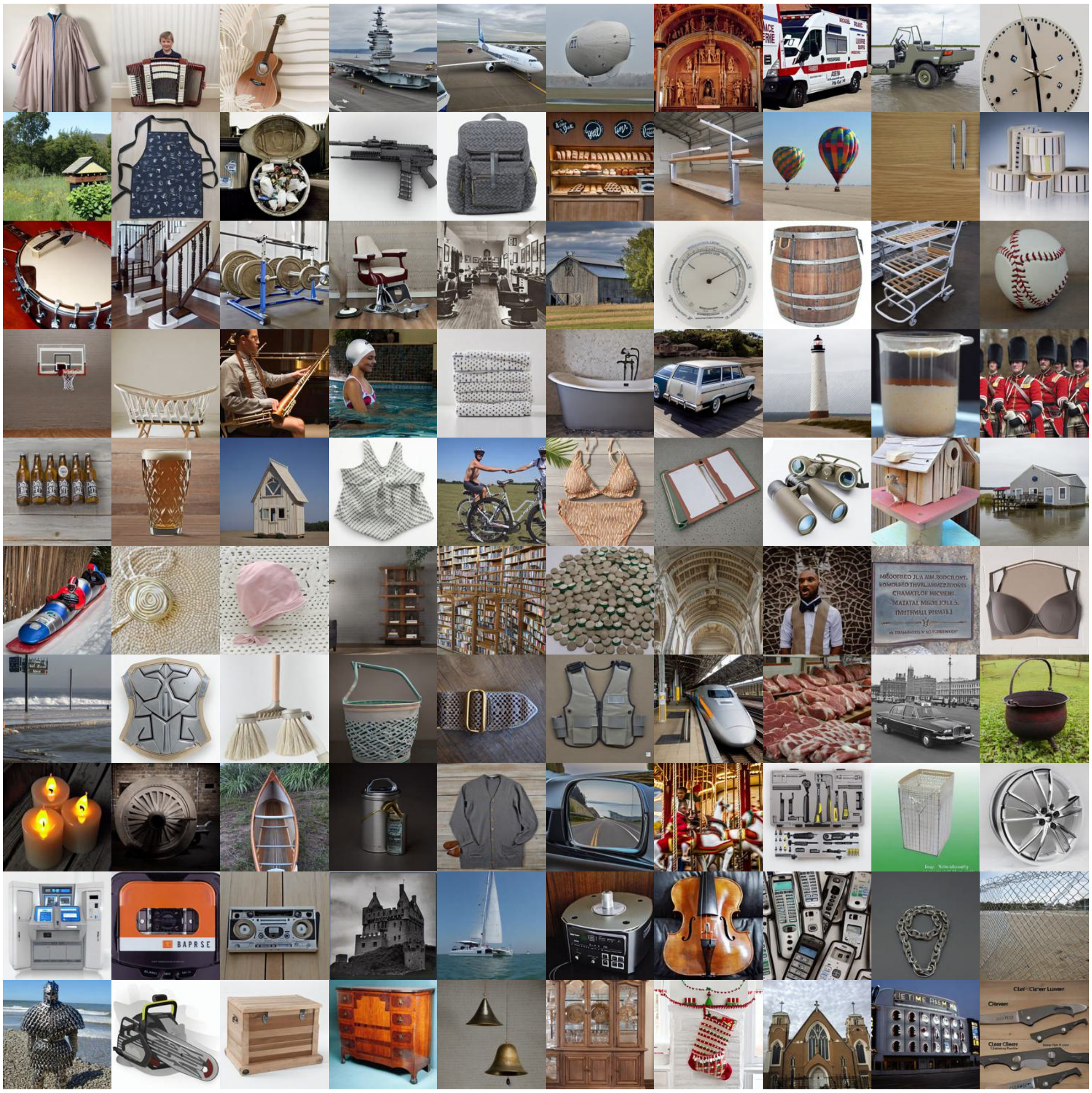


Figure 9. More visualizations selected from the distilled ImageNet-1K (Class 400-499)

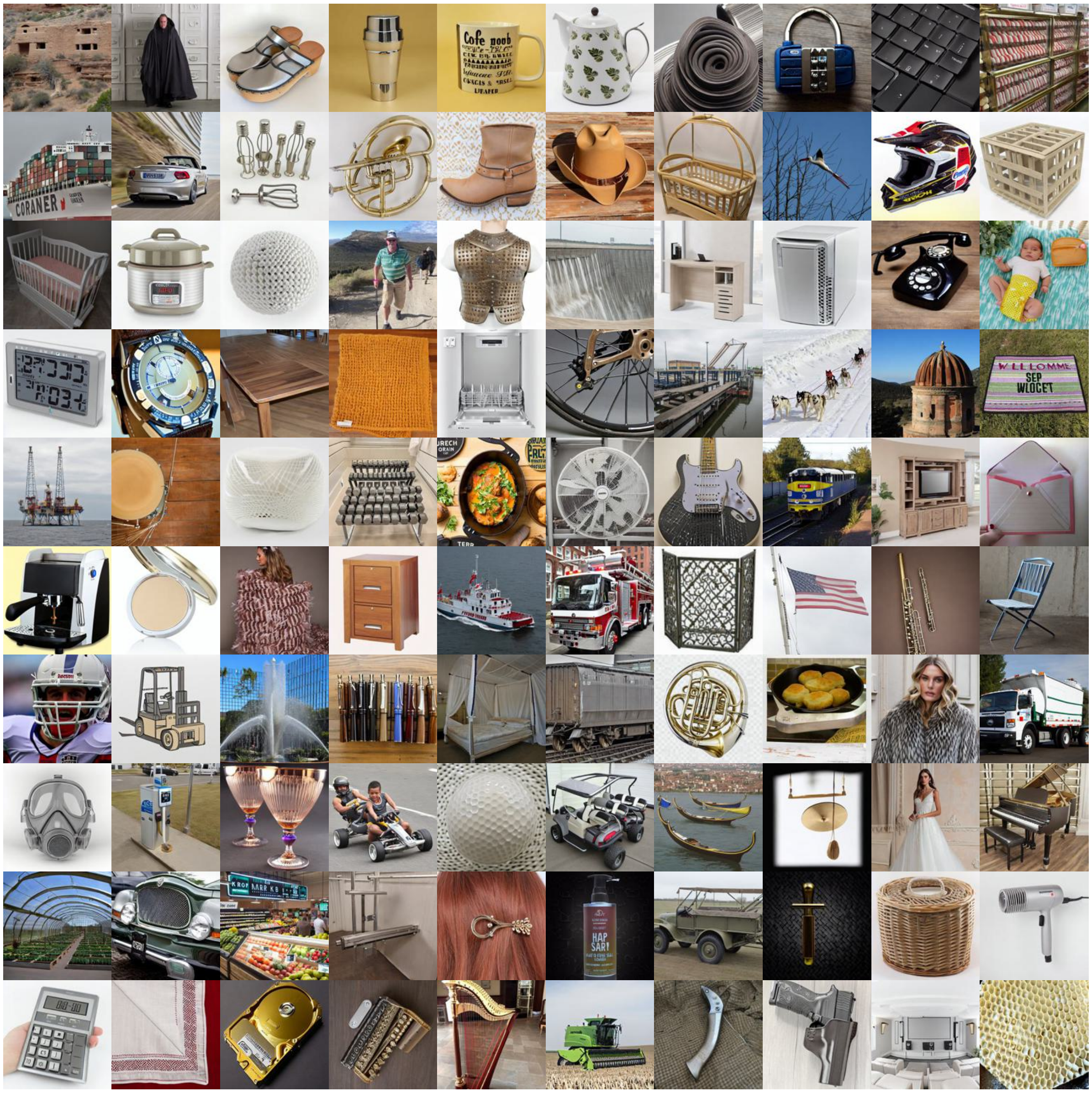


Figure 10. More visualizations selected from the distilled ImageNet-1K (Class 500-599)

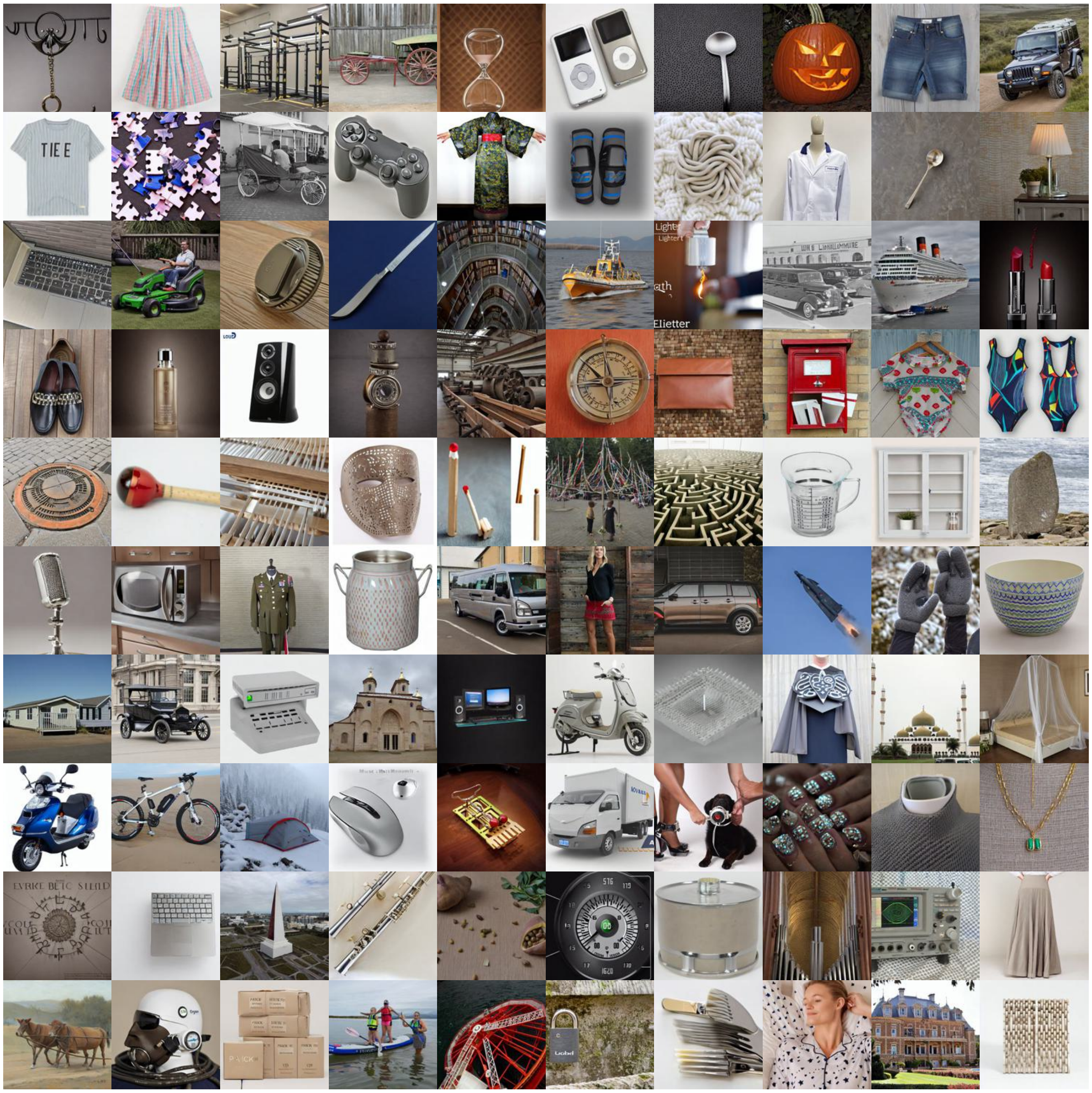


Figure 11. More visualizations selected from the distilled ImageNet-1K (Class 600-699)

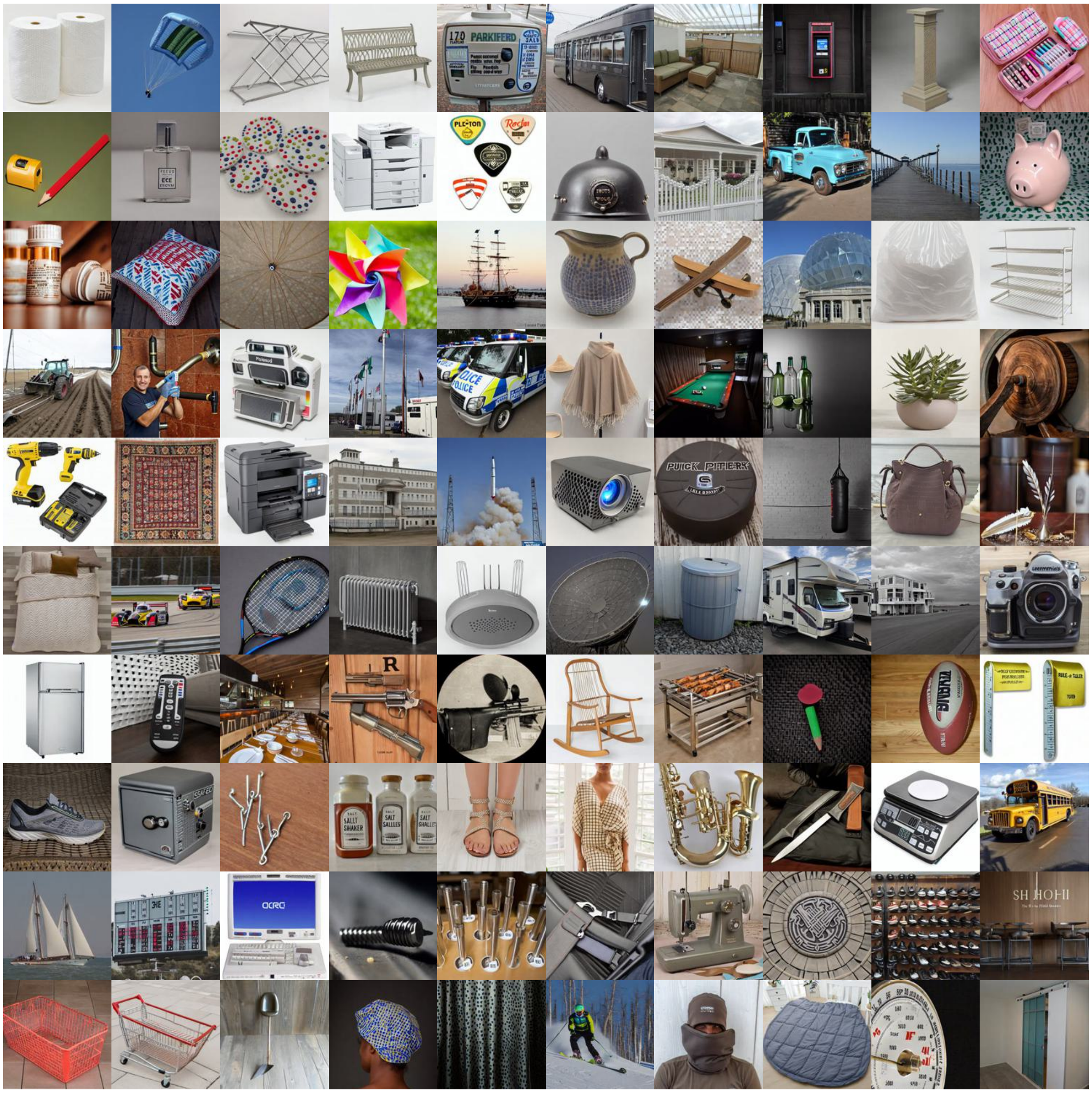


Figure 12. More visualizations selected from the distilled ImageNet-1K (Class 700-799)

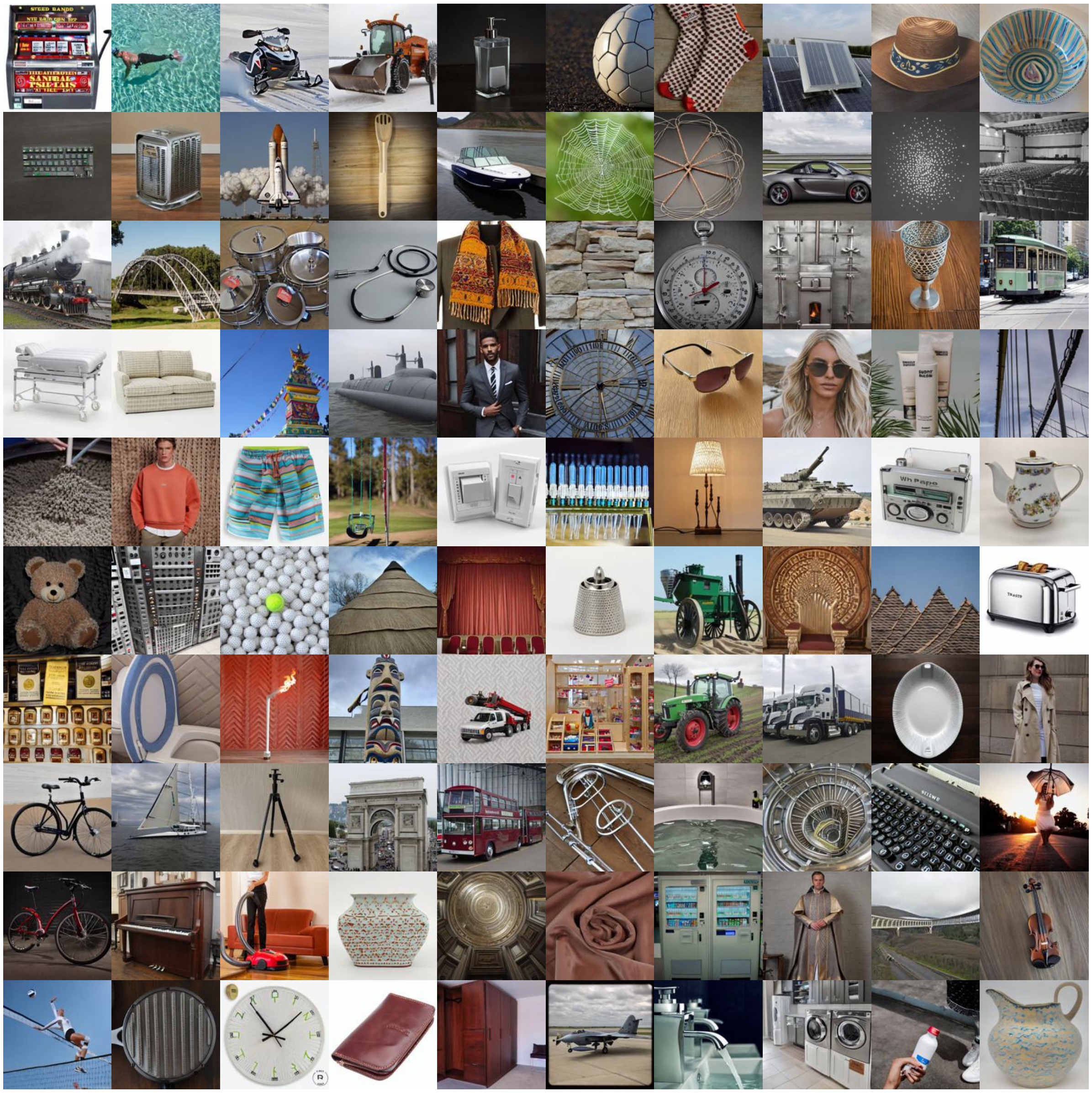


Figure 13. More visualizations selected from the distilled ImageNet-1K (Class 800-899)

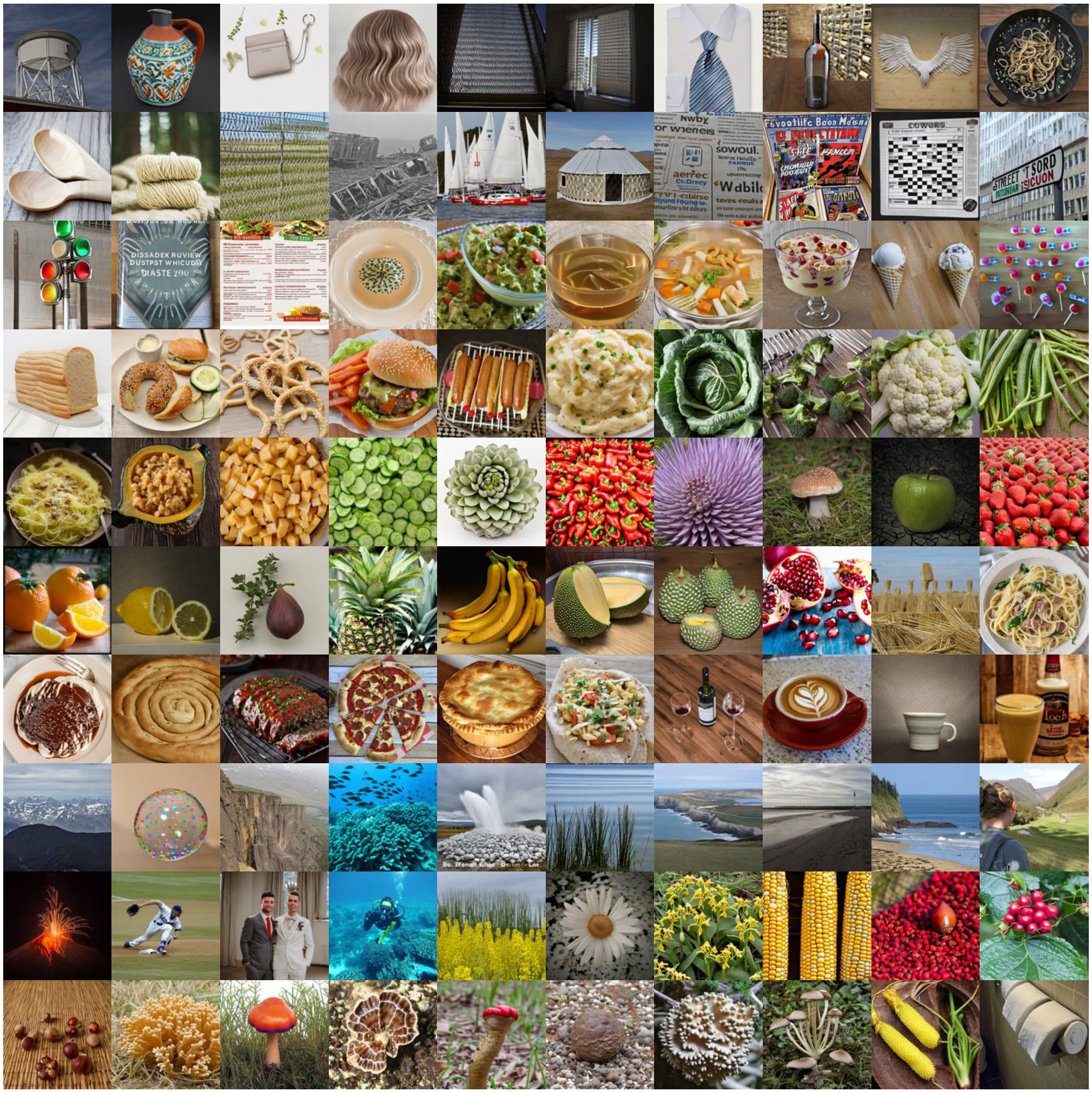


Figure 14. More visualizations selected from the distilled ImageNet-1K (Class 900-999)