Curriculum Coarse-to-Fine Selection for High-IPC Dataset Distillation

Supplementary Material

A. Implement Details

In this section, we introduce more implementation details of CCFS. In the main results, we choose CDA as the base distillation method. We utilize its official code to synthesize $\mathcal{D}_{\text{distill}}$ for CIFAR-10/100 and Tiny-ImageNet. We also leverage the pre-generated soft label approach for the final synthetic data as CDA. Here, we don't elaborate details of the dataset distillation. We provide implementation details of the subsequent curriculum selection and the final evaluation below.

A.1. CIFAR-10/100

Hyper-parameter Setting. In curriculum selection, we set the default number of curriculum phases to 3 and evenly distribute the samples to be selected among them. In each curriculum, we train a modified ResNet-18 model from scratch on the current synthetic dataset as the filter, using equal training epochs as those in the final evaluation. We use pre-calculated Forgetting scores and apply our coarse-to-fine selection strategy on the training set, excluding previously selected samples. For evaluation, we train the identical ResNet-18 on the final synthetic dataset and follow the same training settings as the filter. The hyperparameter settings are shown in Table 1.

Table 1. Hyperparameter settings on CIFAR-10/100.

config	value
difficulty score	Forgetting
number of curricula	3
optimizer	SGD
base learning rate	0.1
momentum	0.9
weight decay	5e-4
learning rate schedule	cosine decay
augmentation	RandomResizedCrop

For the hyperparameter—training epochs, we set the same training epochs for both the filter and the final evaluation model. The number of training epochs varies based on the target IPC. We assign more training epochs to smaller IPC settings, following the settings in other dataset distillation methods. Table 2 shows the specific settings of the training epochs.

For the hyperparameter—batch size, We configure it based on the current size of the synthetic dataset considering its progressive growth across curriculum phases. As the size of the synthetic dataset grows, we appropriately increase the batch size in the filter training. For evaluation, we similarly set the evaluation model's training batch size based on the size of the final synthetic dataset. Table 3 presents the detailed settings for the batch size.

Table 2. Training epochs configuration on CIFAR-10/100.

Compression Ratio	5%	10%	20%	30%
Training Epochs	500	500	250	200

Table 3. Batch size configuration for both the filter and the evaluation training according to the size of the current on CIFAR-10/100.

Compression Ratio	$\leq 5\%$	$5\% - 20\%$	$> 20\%$
Batch Size	32	64	128

A.2. Tiny-ImageNet

Hyper-parameter Setting. In curriculum selection, we set the default number of curriculum phases to 3 and evenly distribute the samples to be selected among them. In each curriculum, we train a modified ResNet-18 model from scratch on the current synthetic dataset as the filter, using equal training epochs as those in the final evaluation. We use pre-calculated Forgetting scores and apply our coarse-to-fine selection strategy on the training set, excluding previously selected samples. For evaluation, we train the identical ResNet-18 on the final synthetic dataset and follow the same training settings as the filter. We uniformly set the training epochs to 100 and the batch size to 64 for both the filter training across curriculum phases and the final evaluation. The hyperparameter settings are shown in Table 4.

Table 4. Parameter setting on CIFAR-10/100.

config	value
difficulty score	Forgetting
number of curricula	3
optimizer	SGD
base learning rate	0.2
momentum	0.9
weight decay	1e-4
learning rate schedule	cosine decay
augmentation	RandomResizedCrop
training epochs	100
batch size	64

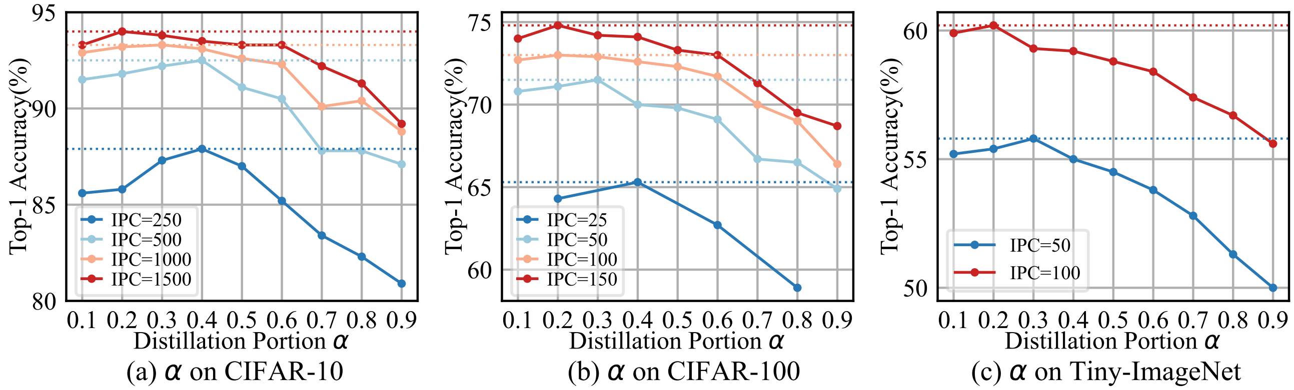

Figure 1. **Impact of different distillation portion α on CIFAR-10/100 and Tiny-ImageNet.** We recommend a small distillation portion α in high-IPC settings.

Dataset	CIFAR-10				CIFAR-100			
IPC	250	500	1000	1500	25	50	100	150
Ratio	5%	10%	20%	30%	5%	10%	20%	30%
SelMatch	82.8	85.9	90.4	91.3	50.9	54.5	62.4	67.4
CCFS w/ MTT	**83.2**	**86.3**	**91.0**	**92.1**	**51.6**	**56.0**	**65.2**	**69.2**

Table 5. **Results of CCFS with MTT as the base dataset distillation method.** CCFS with MTT still outperforms SelMatch across all high-IPC settings, showcasing its excellent scalability.

B. Distillation Portion

The portion α of $\mathcal{D}_{\text{distill}}$ in the final synthetic dataset is another key hyperparameter. In the main table, we report the results of the best distillation portion α in each setting. Here, we provide results of other α settings. As shown in Figure 1, in high-IPC settings, the optimal distillation portion α is typically between 0.2 and 0.4. We recommend a small distillation portion α in high-IPC settings.

C. CCFS with MTT

In the main results, we use CDA to get $\mathcal{D}_{\text{distill}}$. However, our curriculum selection framework is independent of the base dataset distillation method and can be applied to other dataset distillation methods. To verify the scalability of CCFS, we also provide results using MTT as the dataset distillation method. We compare them with SelMatch, which is also based on the MTT approach. We follow the same experimental setup as SelMatch to evaluate the synthetic datasets on ResNet-18. The results in Table 5 demonstrate that CCFS with MTT still outperforms SelMatch across all high-IPC settings, showcasing its excellent scalability.

D. More Experimental Results

In the ablation study, we present the results of other combinations in the selection strategy on CIFAR-100 with IPC=50 and demonstrate that the simplest-misclassified strategy is the optimal combination. Here, we provide experimental results of more datasets and more IPC settings to further validate the effectiveness of the simplest-misclassified combination. As shown in Table 6, 7 and 8, the simplest-misclassified combination consistently outperforms others in all settings. This further validates the effectiveness of our coarse-to-fine selection strategy.

Table 6. CIFAR-10

IPC	classified			misclassified		
	random	hard	simple	random	hard	simple
250	84.2	86.0	85.4	87.0	86.4	**87.9**
500	89.8	90.5	90.8	91.8	91.6	**92.5**
1000	91.5	91.8	91.9	92.6	92.2	**93.2**
1500	92.4	93.0	92.9	93.2	92.9	**93.8**

Table 7. CIFAR-100

IPC	classified			misclassified		
	random	hard	simple	random	hard	simple
25	59.2	52.5	60.5	62.9	51.6	**65.3**
50	66.8	63.5	66.8	70.1	65.0	**71.5**
100	70.7	69.1	70.4	72.0	71.0	**73.0**
150	72.1	71.6	71.2	73.3	72.7	**74.8**

Table 8. Tiny-ImageNet

IPC	classified			misclassified		
	random	hard	simple	random	hard	simple
50	52.1	48.4	52.5	52.9	46.5	**55.8**
100	58.1	56.4	57.7	58.2	54.9	**60.2**

E. Visualization

We present more visualizations of the synthetic datasets, including CIFAR-10 with IPC=250 (ratio=5%) and 1500 (ratio=30%) in Figure 2 and 3, resp., CIFAR-100 with IPC=25 (ratio=5%) and 150 (ratio=30%) in Figure 4 and 5, resp., and Tiny-ImageNet with IPC=50 (ratio=10%) and 100 (ratio=20%) in Figure 6 and 7, respectively. In each visualization, we show partial images from 10 classes in the dataset (corresponding to 10 columns). The first six rows denote the selected real images $\mathcal{D}_{\text{real}}$, while the last four rows correspond to the distilled images $\mathcal{D}_{\text{distill}}$. For $\mathcal{D}_{\text{real}}$, we display two samples of median difficulty per class selected at each curriculum phase. The visualizations demonstrate the progressive difficulty of selected samples across curriculum phases and show that higher IPC settings tend to select more challenging samples than lower IPC settings within the same phase.

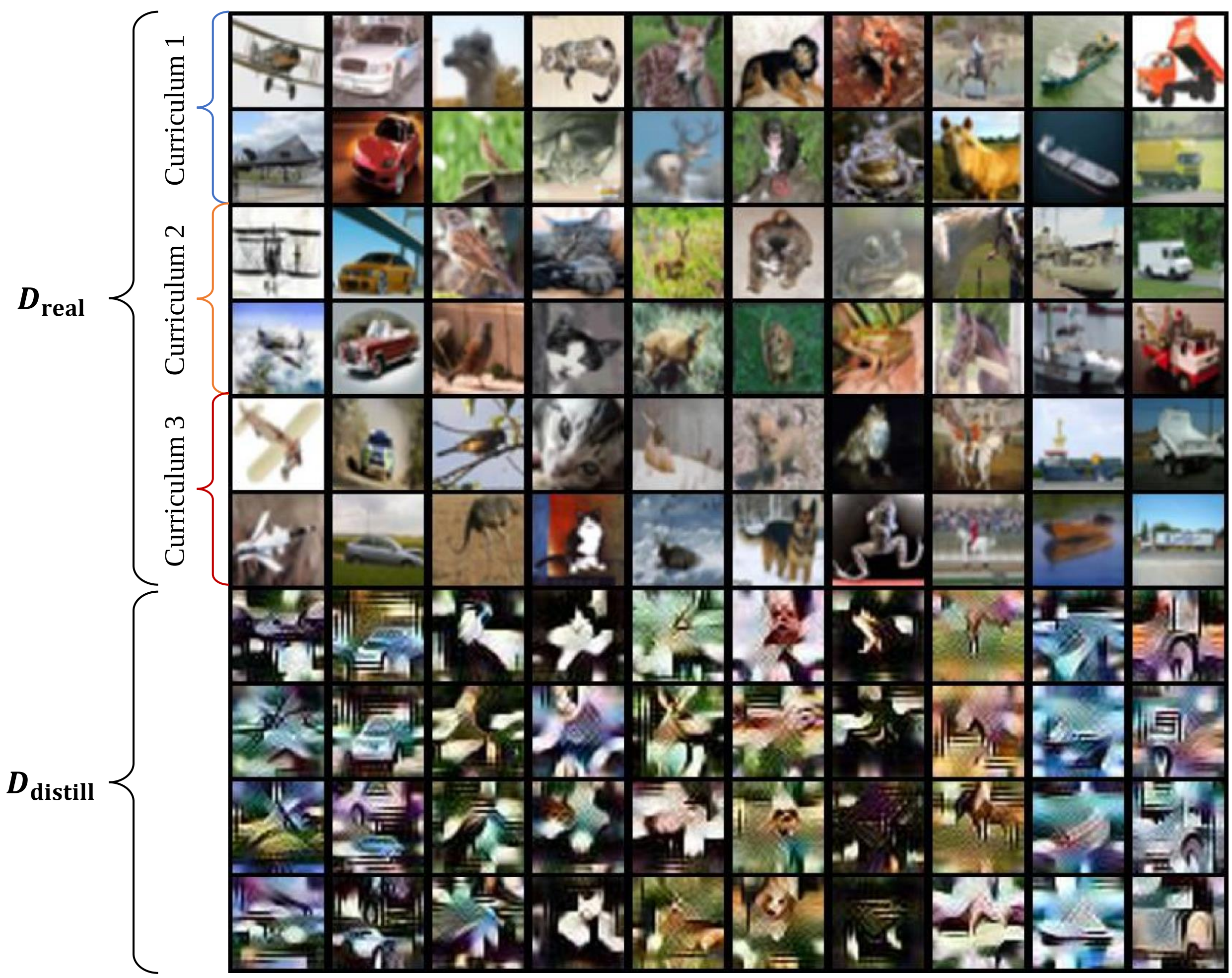

Figure 2. Visualization of the synthetic dataset (CIFAR-10, IPC=250)

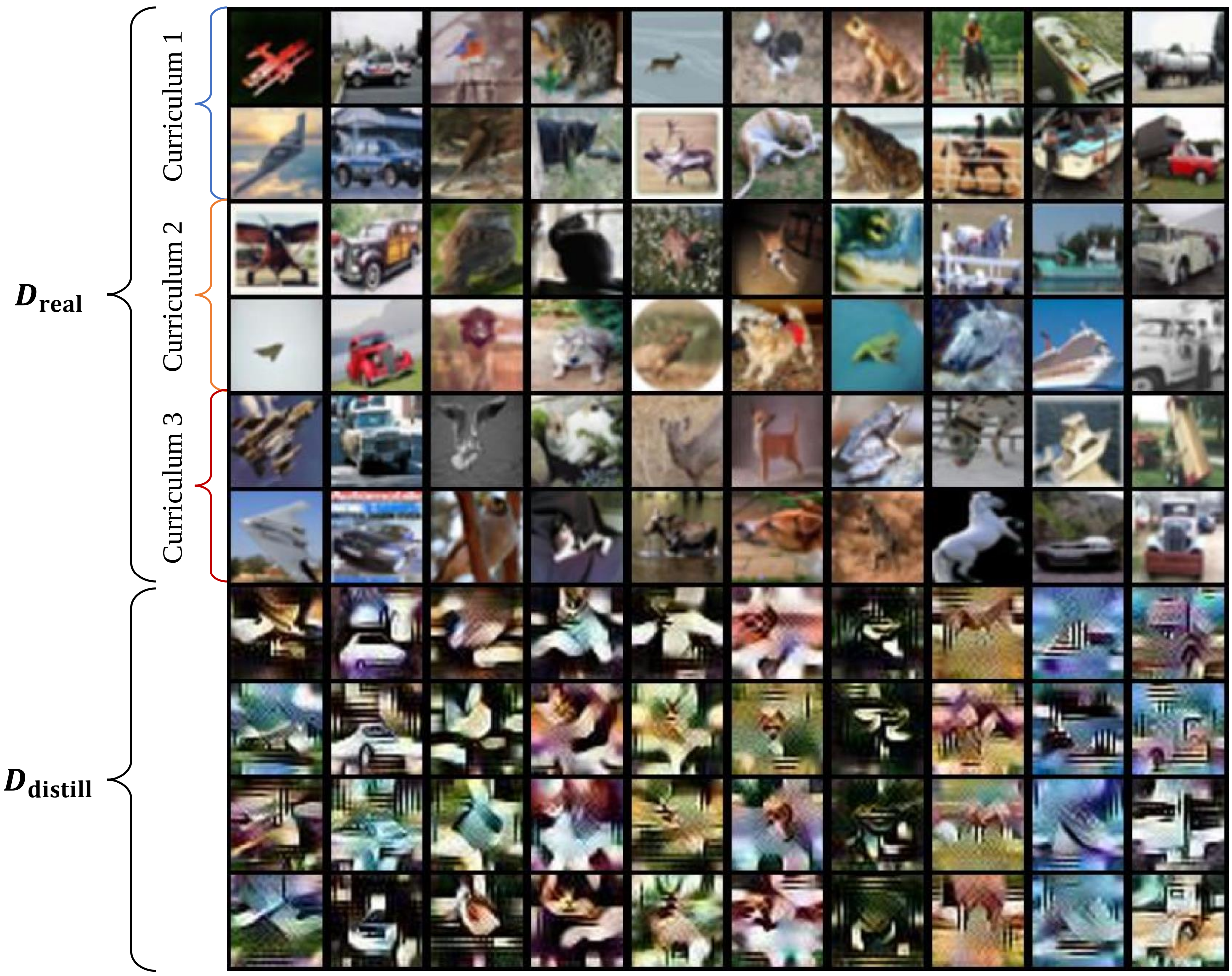

Figure 3. Visualization of the synthetic dataset (CIFAR-10, IPC=1500)

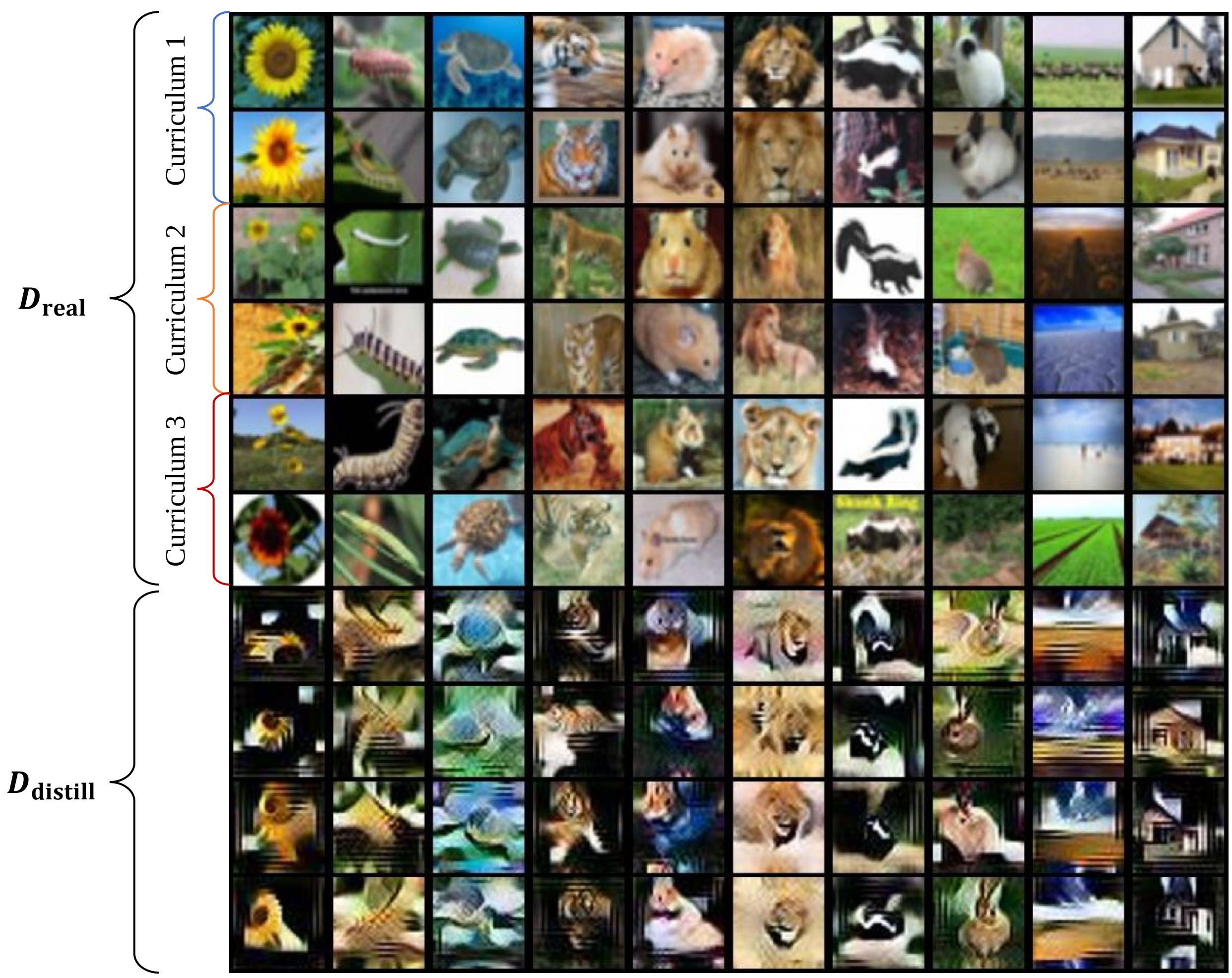

Figure 4. Visualization of the synthetic dataset (CIFAR-100, IPC=25)

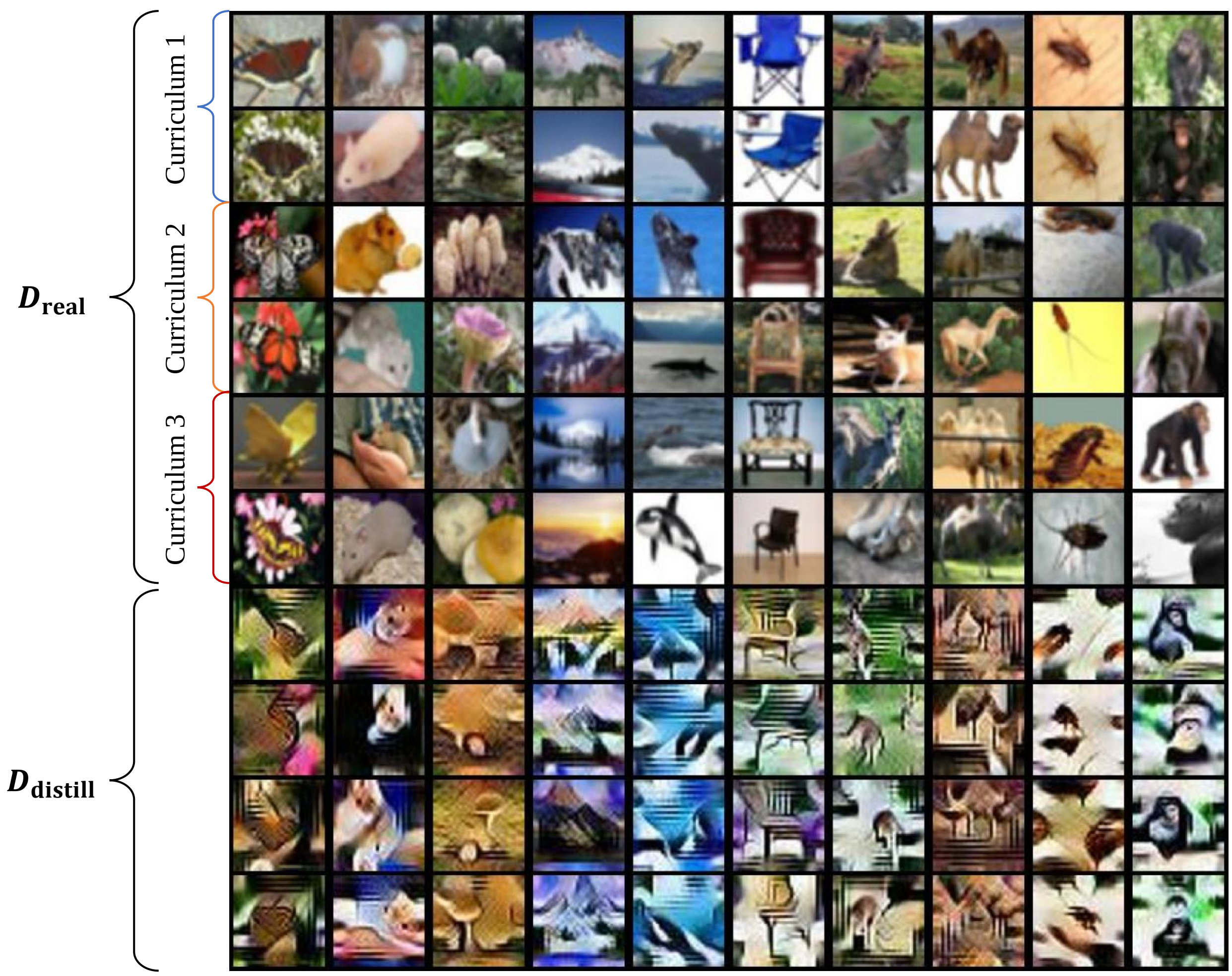

Figure 5. Visualization of the synthetic dataset (CIFAR-100, IPC=150)

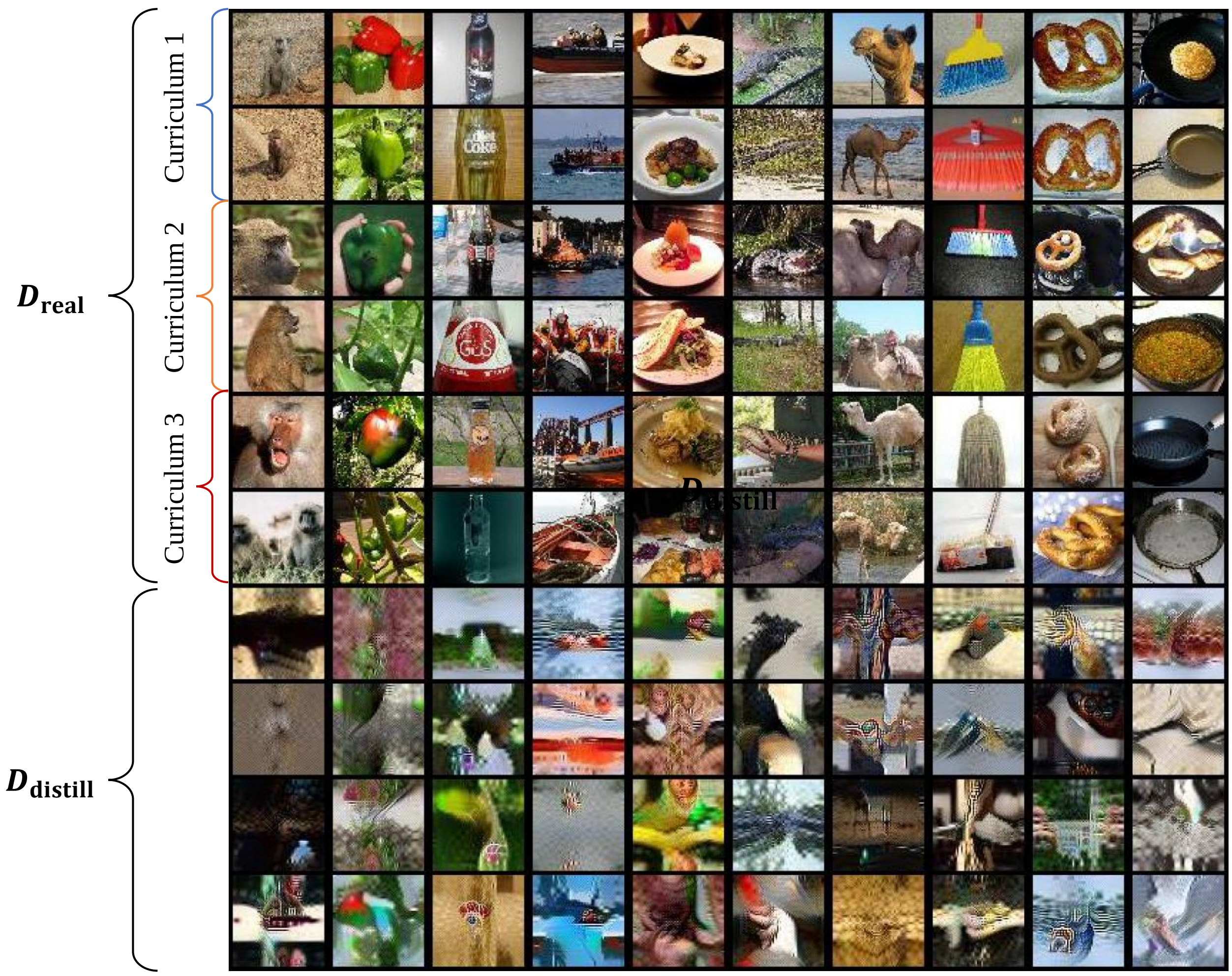

Figure 6. Visualization of the synthetic dataset (Tiny-ImageNet, IPC=50)

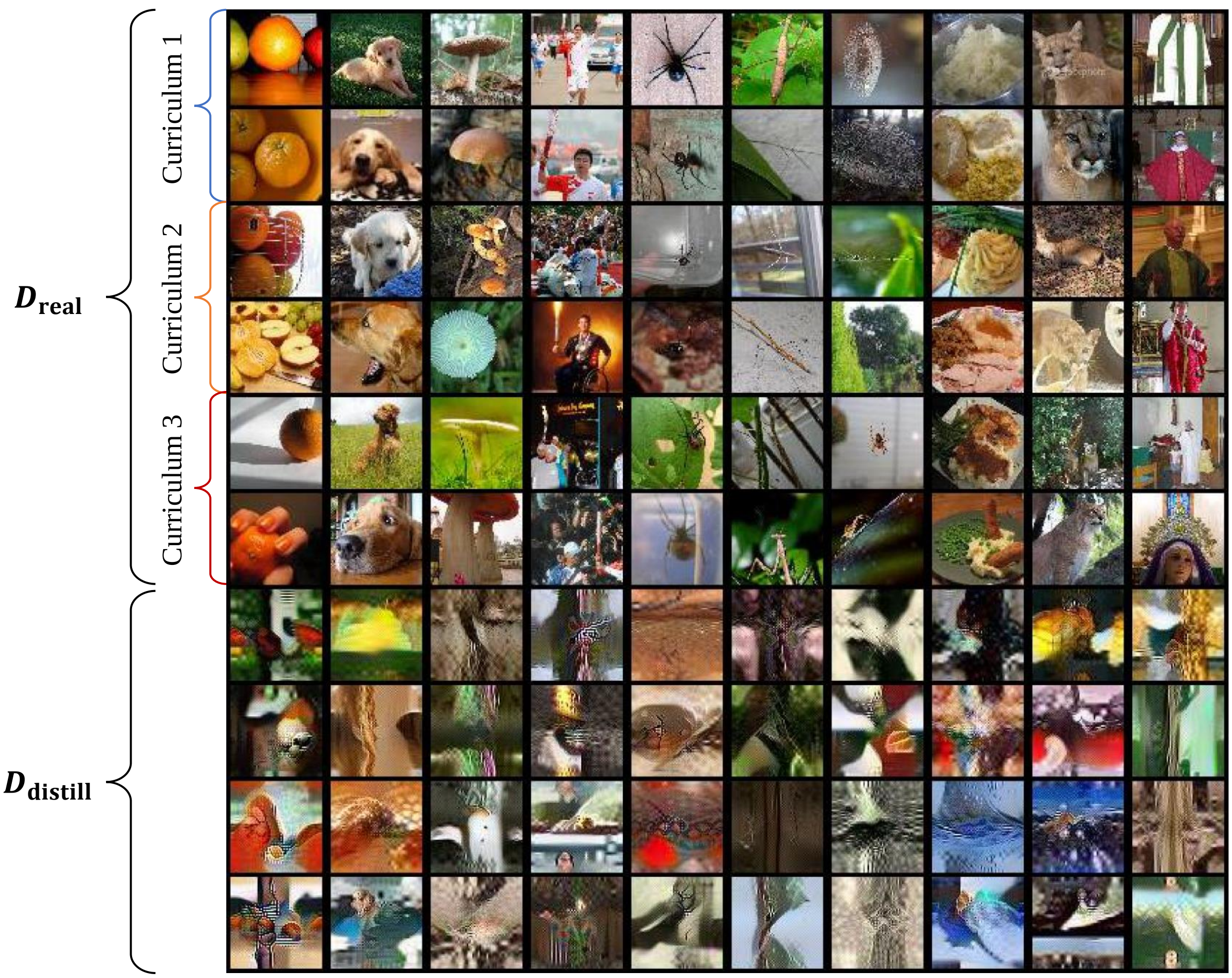

Figure 7. Visualization of the synthetic dataset (Tiny-ImageNet, IPC=100)