

## A ALGORITHM

---

### Algorithm 1 Multisize Dataset Condensation

---

**Input:** Full dataset  $\mathcal{B}$ , model  $\Theta$ , MLS selection period  $\Delta t$ , learning rate of the synthetic dataset  $\lambda$ , learning rate of the model  $\eta$ , outer loop iterations  $T$ , inner loop epochs  $E$ , and class loop iterations  $C$ .

**Output:** Synthetic dataset  $\mathcal{S}$

- 1: Initialize synthetic dataset  $\mathcal{S}$
- 2: Initialize the most learnable subset (MLS)  $\mathcal{S}_{\text{MLS}}$
- 3: **for**  $t = 1$  **to**  $T$  **do** ▷ Outer loop
- 4:     Randomly initialize model weight  $\theta_t$
- 5:     **for**  $e = 1$  **to**  $E$  **do** ▷ Inner loop
- 6:         **for**  $c = 1$  **to**  $C$  **do** ▷ Class loop
- 7:             Sample class-wise mini-batches  $B_c \sim \mathcal{B}, S_c \sim \mathcal{S}$
- 8:             Update  $\mathcal{S}_c$  with **subset loss** according to Eq. 10
- 9:         **end for**
- 10:          $\theta_{t,e+1} \leftarrow \theta_{t,e} - \eta \nabla_{\theta} \ell(\theta_{t,e}; B)$  ▷ Update model with real image mini-batch  $B \sim \mathcal{B}$
- 11:     **end for**
- 12:     **if**  $t \% \Delta t$  is 0 **then** ▷ Every  $\Delta t$  iterations
- 13:         Select  $\mathcal{S}_{\text{MLS}}$  according to Eq. 9
- 14:     **end if**
- 15: **end for**
- 16: **return** Synthetic dataset  $\mathcal{S}$

---

Algo. 1 provides the algorithm of the proposed MDC method.

## B EXPERIMENT

### B.1 EXPERIMENT SETTINGS

#### Datasets:

- SVHN [Netzer et al. (2011)] contains street digits of shape  $32 \times 32 \times 3$ . The dataset contains 10 classes including digits from 0 to 9. The training set has 73257 images, and the test set has 26032 images.
- CIFAR-10 [Krizhevsky et al. (2009)] contains images of shape  $32 \times 32 \times 3$  and has 10 classes in total: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The training set has 5,000 images per class and the test set has 1,000 images per class, containing in total 50,000 training images and 10,000 testing images.
- CIFAR-100 [Krizhevsky et al. (2009)] contains images of shape  $32 \times 32 \times 3$  and has 100 classes in total. Each class contains 500 images for training and 100 images for testing, leading to a total of 50,000 training images and 10,000 testing images.
- ImageNet-10 [Deng et al. (2009)] is a subset of ImageNet-1K [Deng et al. (2009)] containing images with an average  $469 \times 387 \times 3$  pixels but reshaped to resolution of  $224 \times 224 \times 3$ . It contains 1,280 training images per class on average and a total of 50,000 images for testing (validation set). Following [Kim et al. (2022b)], the ImageNet-10 contains 10 classes: 1) poke bonnet, 2) green mamba, 3) langur, 4) Doberman pinscher, 5) gyromitra, 6) gazelle hound, 7) vacuum cleaner, 8) window screen, 9) cocktail shaker, and 10) garden spider.

**Augmentation:** Following IDC [Kim et al. (2022b)], we perform augmentation during training networks in condensation and evaluation, and we use coloring, cropping, flipping, scaling, rotating, and mixup. When updating network parameters, image augmentations are different for each image in a batch; when updating synthetic images, the same augmentations are utilized for the synthetic images and corresponding real images in a batch.

- Color which adjusts the brightness, saturation, and contrast of images.

- Crop which pads the image and then randomly crops back to the original size.
- Flip which flips the images horizontally with a probability of 0.5.
- Scale which randomly scales the images by a factor according to a ratio.
- Rotate which rotates the image by a random angle according to a ratio.
- Cutout which randomly removes square parts of the image, replacing the removed parts with black squares.
- Mixup which randomly selects a square region within the image and replaces this region with the corresponding section from another randomly chosen image. It happens at a probability of 0.5.

**Multi-formation Settings.** For all results we use IDC [Kim et al. \(2022b\)](#) as the “basic condensation method” otherwise stated. Following its setup, we use a multi-formation factor of 2 for SVHN, CIFAR-10, CIFAR-100 datasets and a factor of 3 for ImageNet-10.

**Reason for Using Large Batch Size for  $IPC > 32$ .** For CIFAR-10, CIFAR-100 and SVHN, we use the default batch size (128) when  $IPC < 32$  and a larger batch size (256) when  $32 \leq IPC \leq 64$ . The reason is that our method is based on IDC [Kim et al. \(2022b\)](#) which uses a multi-formation factor of  $f = 2$  for CIFAR-10, CIFAR-100, and SVHN datasets. The multi-formation function splits a synthetic image into  $f^2 = 2^2 = 4$  images during the condensation process. To ensure all samples in a subset can be sampled during condensation, we increase the subsets when the number of images exceeds the default batch size, which is 128. With a multi-formation factor  $f = 2$ , the maximum IPC of each sampling process is  $IPC = 32$  (i.e.,  $IPC \times 2^2 \leq 128$ ). For ImageNet-10  $IPC_{20}$ , a multi-formation factor of  $f = 3$  is used. Hence, we use a batch size of 256 (i.e.,  $128 \leq 20 \times 3^2 \leq 256$ ).

**MTT Settings.** The reported numbers of MTT [Cazenavette et al. \(2022\)](#) are obtained without ZCA normalization to keep all methods using the standard normalization technique.

## B.2 FEATURE DISTANCE CALCULATION

	1	2	3	4	5	6	7	8	9
$t = 1$	3012	1678	1249	1013	896	807	738	701	675
$t = 50$	2596	1294	891	661	514	429	373	332	298
Diff.	416	384	358	352	382	378	365	369	377
Rate of change	8.32	7.68	7.16	7.04	7.64	7.56	7.30	7.38	7.54

Table 8: Feature distance of subsets summed over inner loop training. CIFAR-10,  $IPC_{10}$ .

Tab. 8 presents the feature distance computed at a specific outer loop  $t$  without imposing the subset loss. The table conveys two pieces of information. First, the feature loss of a smaller subset is always greater than that of a larger subset. That is a reason why we need to find the rate of change. Otherwise,  $\mathcal{S}_{[1]}$  will always be selected. Second, the feature distance of the smallest subset changes the most, and this contributes to why we select  $\mathcal{S}_{[1]}$  as the subset initialization.

## B.3 THE INFLUENCE OF BASIC CONDENSATION METHOD

	1	2	3	4	5	6	7	8	9	10	Avg.
IDC	49.55	53.75	56.39	59.33	58.13	60.62	62.06	63.59	65.25	66.79	59.55
DREAM	49.70	55.12	55.84	57.59	58.72	62.92	63.61	64.71	66.26	67.44	60.19

Table 9: Comparison between the proposed method applied to IDC [Kim et al. \(2022b\)](#) and to DREAM [Liu et al. \(2023b\)](#) on CIFAR-10  $IPC_{10}$ .

Tab. 9 shows our method works on other basic condensation methods such as DREAM [Liu et al. \(2023b\)](#).

## B.4 PRIMARY RESULTS WITH STANDARD DEVIATION

Dataset		1	2	3	4	5	6	7	8	9	10	Avg. Diff.	
SVHN	A	68.50 <sup>†</sup> <sub>±0.9</sub>	75.27 <sub>±0.3</sub>	79.55 <sub>±0.4</sub>	81.85 <sub>±0.2</sub>	83.33 <sub>±0.1</sub>	84.53 <sub>±0.3</sub>	85.66 <sub>±0.3</sub>	86.05 <sub>±0.1</sub>	86.25 <sub>±0.2</sub>	87.50 <sup>†</sup> <sub>±0.3</sub>	81.85	-
	B	68.50 <sup>†</sup> <sub>±0.9</sub>	71.65 <sub>±0.1</sub>	71.27 <sub>±0.9</sub>	71.92 <sub>±0.3</sub>	73.28 <sub>±0.3</sub>	70.74 <sub>±0.4</sub>	71.83 <sub>±0.4</sub>	71.08 <sub>±0.8</sub>	71.97 <sub>±1.0</sub>	71.55 <sub>±0.7</sub>	71.38	-
	C	35.48 <sub>±0.4</sub>	51.55 <sub>±0.6</sub>	60.42 <sub>±1.0</sub>	67.97 <sub>±0.5</sub>	74.38 <sub>±0.5</sub>	77.65 <sub>±0.7</sub>	81.70 <sub>±0.2</sub>	83.86 <sub>±0.5</sub>	85.96 <sub>±0.4</sub>	87.50 <sup>†</sup> <sub>±0.3</sub>	70.65	0
	Ours	63.26 <sub>±1.0</sub>	67.91 <sub>±0.7</sub>	72.15 <sub>±1.0</sub>	74.09 <sub>±0.3</sub>	77.54 <sub>±0.4</sub>	78.17 <sub>±0.3</sub>	80.92 <sub>±0.3</sub>	82.82 <sub>±0.5</sub>	84.27 <sub>±0.3</sub>	86.38 <sub>±0.2</sub>	76.75	+6.10
CIFAR-10	A	50.80 <sub>±0.3</sub>	54.85 <sub>±0.4</sub>	59.79 <sub>±0.2</sub>	61.84 <sub>±0.1</sub>	62.49 <sub>±0.3</sub>	64.59 <sub>±0.1</sub>	65.53 <sub>±0.2</sub>	66.33 <sub>±0.1</sub>	66.82 <sub>±0.3</sub>	67.50 <sup>†</sup> <sub>±0.5</sub>	62.05	-
	B	50.80 <sub>±0.3</sub>	53.17 <sub>±0.4</sub>	55.09 <sub>±0.4</sub>	56.17 <sub>±0.3</sub>	55.80 <sub>±0.3</sub>	56.98 <sub>±0.3</sub>	57.60 <sub>±0.3</sub>	57.78 <sub>±0.1</sub>	58.22 <sub>±0.4</sub>	58.38 <sub>±0.1</sub>	56.00	-
	C	27.49 <sub>±0.8</sub>	38.50 <sub>±0.5</sub>	45.29 <sub>±0.1</sub>	50.85 <sub>±0.5</sub>	53.60 <sub>±0.3</sub>	57.98 <sub>±0.2</sub>	60.99 <sub>±0.5</sub>	63.60 <sub>±0.2</sub>	65.71 <sub>±0.1</sub>	67.50 <sup>†</sup> <sub>±0.5</sub>	53.15	0
	Ours	49.55 <sub>±0.7</sub>	53.75 <sub>±0.3</sub>	56.39 <sub>±0.3</sub>	59.33 <sub>±0.2</sub>	58.13 <sub>±0.4</sub>	60.62 <sub>±0.2</sub>	62.06 <sub>±0.3</sub>	63.59 <sub>±0.1</sub>	65.25 <sub>±0.3</sub>	66.79 <sub>±0.2</sub>	59.55	+6.40
CIFAR-100	A	28.90 <sup>†</sup> <sub>±0.2</sub>	34.28 <sub>±0.2</sub>	37.35 <sub>±0.2</sub>	39.13 <sub>±0.1</sub>	41.15 <sub>±0.4</sub>	42.65 <sub>±0.4</sub>	43.62 <sub>±0.3</sub>	44.48 <sub>±0.2</sub>	45.07 <sub>±0.1</sub>	45.40 <sub>±0.4</sub>	40.20	-
	B	28.90 <sup>†</sup> <sub>±0.2</sub>	30.63 <sub>±0.1</sub>	31.64 <sub>±0.0</sub>	31.76 <sub>±0.2</sub>	32.61 <sub>±0.2</sub>	32.85 <sub>±0.2</sub>	33.03 <sub>±0.3</sub>	33.04 <sub>±0.2</sub>	33.32 <sub>±0.2</sub>	33.39 <sub>±0.2</sub>	32.12	-
	C	14.38 <sub>±0.2</sub>	21.76 <sub>±0.2</sub>	28.01 <sub>±0.2</sub>	32.21 <sub>±0.3</sub>	35.27 <sub>±0.3</sub>	39.09 <sub>±0.2</sub>	40.92 <sub>±0.1</sub>	42.69 <sub>±0.2</sub>	44.28 <sub>±0.2</sub>	45.40 <sub>±0.4</sub>	34.40	0
	Ours	27.58 <sub>±0.2</sub>	31.83 <sub>±0.0</sub>	33.59 <sub>±0.2</sub>	35.42 <sub>±0.1</sub>	36.93 <sub>±0.1</sub>	38.95 <sub>±0.4</sub>	40.70 <sub>±0.1</sub>	42.05 <sub>±0.1</sub>	43.86 <sub>±0.1</sub>	44.34 <sub>±0.2</sub>	37.53	+3.13

(a) Results of SVHN, CIFAR-10, CIFAR-100 targeting IPC<sub>10</sub>.

Dataset		1	2	3	4	5	6	7	8	9	10
SVHN	A	68.50 <sup>†</sup> <sub>±0.9</sub>	75.27 <sub>±0.3</sub>	79.55 <sub>±0.4</sub>	81.85 <sub>±0.2</sub>	83.33 <sub>±0.1</sub>	84.53 <sub>±0.3</sub>	85.66 <sub>±0.3</sub>	86.05 <sub>±0.1</sub>	86.25 <sub>±0.2</sub>	87.50 <sup>†</sup> <sub>±0.3</sub>
	C	34.90 <sub>±0.9</sub>	46.52 <sub>±0.4</sub>	52.23 <sub>±0.9</sub>	56.30 <sub>±0.4</sub>	62.25 <sub>±0.5</sub>	65.34 <sub>±0.5</sub>	68.84 <sub>±0.3</sub>	69.57 <sub>±1.7</sub>	71.95 <sub>±0.5</sub>	74.69 <sub>±0.2</sub>
	Ours	58.77 <sub>±1.5</sub>	67.72 <sub>±0.3</sub>	69.33 <sub>±0.5</sub>	72.26 <sub>±0.4</sub>	75.02 <sub>±0.3</sub>	73.71 <sub>±0.7</sub>	74.50 <sub>±0.5</sub>	74.63 <sub>±0.6</sub>	76.21 <sub>±0.4</sub>	76.87 <sub>±0.7</sub>
CIFAR-10	A	50.80 <sub>±0.3</sub>	54.85 <sub>±0.4</sub>	59.79 <sub>±0.2</sub>	61.84 <sub>±0.1</sub>	62.49 <sub>±0.3</sub>	64.59 <sub>±0.1</sub>	65.53 <sub>±0.2</sub>	66.33 <sub>±0.1</sub>	66.82 <sub>±0.3</sub>	67.50 <sup>†</sup> <sub>±0.5</sub>
	C	27.87 <sub>±0.4</sub>	35.69 <sub>±0.5</sub>	41.93 <sub>±0.2</sub>	45.29 <sub>±0.2</sub>	47.54 <sub>±0.4</sub>	51.96 <sub>±0.4</sub>	53.51 <sub>±0.3</sub>	55.59 <sub>±0.1</sub>	56.62 <sub>±0.2</sub>	58.26 <sub>±0.1</sub>
	Ours	47.83 <sub>±0.6</sub>	52.18 <sub>±0.2</sub>	56.29 <sub>±0.1</sub>	58.52 <sub>±0.2</sub>	58.75 <sub>±0.4</sub>	60.67 <sub>±0.3</sub>	61.90 <sub>±0.1</sub>	62.74 <sub>±0.2</sub>	62.32 <sub>±0.2</sub>	62.64 <sub>±0.2</sub>
CIFAR-100	A	28.90 <sup>†</sup> <sub>±0.2</sub>	34.28 <sub>±0.2</sub>	37.35 <sub>±0.2</sub>	39.13 <sub>±0.1</sub>	41.15 <sub>±0.4</sub>	42.65 <sub>±0.4</sub>	43.62 <sub>±0.3</sub>	44.48 <sub>±0.2</sub>	45.07 <sub>±0.1</sub>	45.40 <sub>±0.4</sub>
	C	12.66 <sub>±0.1</sub>	18.35 <sub>±0.1</sub>	23.76 <sub>±0.4</sub>	26.92 <sub>±0.4</sub>	29.12 <sub>±0.2</sub>	32.23 <sub>±0.1</sub>	34.21 <sub>±0.4</sub>	35.71 <sub>±0.3</sub>	37.18 <sub>±0.3</sub>	38.25 <sub>±0.3</sub>
	Ours	26.34 <sub>±0.2</sub>	29.71 <sub>±0.3</sub>	31.74 <sub>±0.4</sub>	32.95 <sub>±0.4</sub>	34.49 <sub>±0.3</sub>	36.36 <sub>±0.2</sub>	38.49 <sub>±0.4</sub>	39.59 <sub>±0.1</sub>	40.43 <sub>±0.4</sub>	41.35 <sub>±0.2</sub>

(b) Results of SVHN, CIFAR-10, CIFAR-100 targeting IPC<sub>50</sub>.

Table 10: Comparisons between the proposed method and three different baselines built with the IDC [Kim et al. \(2022b\)](#). <sup>†</sup> represents the numbers reported in the original paper. Results from sub-table (b) are divided into two parts due to limited space.

In Tab. [10](#), we list the primary results with standard deviation for synthetic datasets with IPC<sub>10</sub> and IPC<sub>50</sub>, including SVHN, CIFAR-10, and CIFAR-100 datasets. The standard deviation is computed from three randomly initialized networks since the same subset is selected for each run. Even by taking into account these standard deviations, our method shows a consistent improvement in the average accuracy.

## B.5 ACCURACY OF SUBSETS DURING CONDENSATION PROCESS

	100	200	300	400	500	600	700	800	900	1000
1	49.40	50.20	50.10	51.00	50.00	50.60	49.80	48.70	48.90	49.74
2	47.30	47.60	50.30	50.10	48.20	49.50	49.50	53.10	53.60	53.41
3	48.40	48.60	57.80	57.90	55.30	57.20	55.80	55.70	56.00	57.46
4	52.30	51.80	57.40	57.60	55.00	57.60	56.60	55.60	55.90	56.93
5	52.80	53.60	56.30	56.50	56.20	57.10	56.30	55.80	56.20	56.59
6	57.00	57.90	59.60	58.50	59.20	60.10	60.00	59.00	58.50	59.58
7	58.10	59.50	61.10	60.00	61.50	61.50	61.30	60.70	60.50	61.24
8	60.50	61.90	62.30	62.20	62.50	63.30	63.10	63.10	62.50	63.13
9	62.00	63.10	64.10	64.20	63.90	64.40	65.30	64.20	64.30	64.72
10	63.40	65.10	64.60	64.50	65.90	64.90	66.50	65.80	66.40	66.03
Avg.	55.12	55.93	58.36	58.25	57.77	58.62	58.42	58.17	58.28	58.88
	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
1	49.19	49.58	49.36	49.38	49.41	49.68	49.68	49.14	49.41	49.55
2	53.40	53.05	54.14	53.78	53.93	53.11	53.48	53.79	53.86	53.75
3	57.00	57.19	57.03	57.51	56.61	56.74	56.58	56.54	56.20	56.39
4	57.12	57.38	56.81	57.17	57.17	56.75	56.88	56.51	58.88	59.33
5	56.88	56.49	56.84	56.80	56.72	56.33	56.33	56.64	58.05	58.13
6	59.58	59.37	59.29	59.62	59.62	59.38	59.59	59.22	60.55	60.62
7	61.35	61.51	61.41	61.20	61.66	61.48	61.32	61.13	61.72	62.06
8	62.78	63.93	63.38	63.55	63.20	63.64	63.41	63.18	63.41	63.59
9	64.75	64.68	65.12	65.36	65.14	65.32	64.88	65.52	65.40	65.25
10	66.15	66.56	66.22	66.22	66.80	66.36	66.64	66.88	66.91	66.79
Avg.	58.82	58.97	58.96	59.06	59.03	58.88	58.88	58.86	59.44	59.55

Table 11: Accuracy of subsets evaluated at different outer loops. CIFAR-10, IPC<sub>10</sub>.

Tab. 11 provides the accuracy of subsets evaluated at different outer loops for Fig. 4. The table illustrates that training with our method for 100 outer loop achieves a higher average accuracy (i.e., 55.12%) than condensing with the previous method for 2,000 outer loops (i.e., 53.15%).

## C VISUALIZATION

### C.1 CLASS-WISE MLS SELECTION

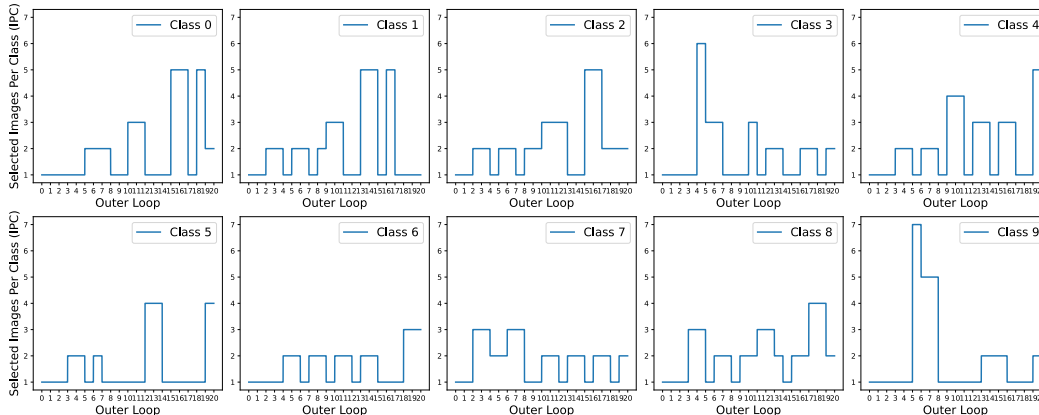


Figure 7: Visualization of selected subsets using a class-wise approach.

**Visualization of Class-wise MLS Selection.** Fig. 6 presents the choice of MLS of each class at every selection round. Compared to the non-class-wise manner (Fig. 3b), the class-wise manner selection tends to select relatively larger subsets.

### C.2 VISUALIZATION OF CONDENSED IMAGES

Fig. 8, 9 show the effectiveness of the proposed method. Note that the two figures using a multi-formation factor of 1 are for the purpose of better visualization. All the experimental results reported are using the same setting as the primary result listed in Tab. 1. Fig. 10, 11, and 12 present the visualizations of MDC using a factor of 2 Kim et al. (2022b).

Fig. 13 uses a factor of 3 for ImageNet. Through comparing the images (class: gazelle hound) highlighted by orange, red and green boxes in Fig. 13, we observe the similar pattern shown in Fig. 6 that our MDC has large distortion.

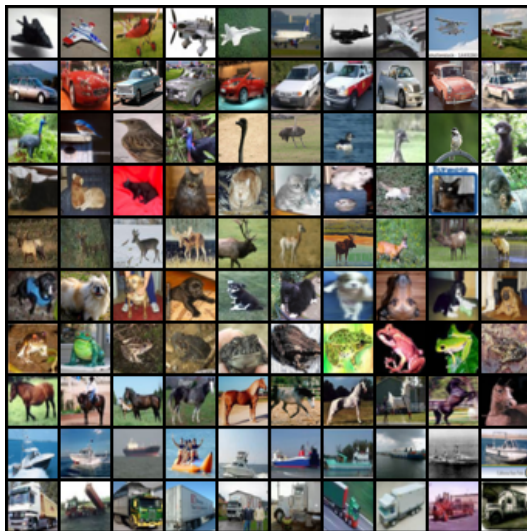


Figure 8: Visualization of the initialization of CIFAR-10,  $IPC_{10}$ .



Figure 9: Visualization of the proposed condensation method. (a) and (b) are IDC Kim et al. (2022b) condensed to  $IPC_1$  and  $IPC_{10}$ , respectively. (c) is the proposed method, MDC. CIFAR-10.

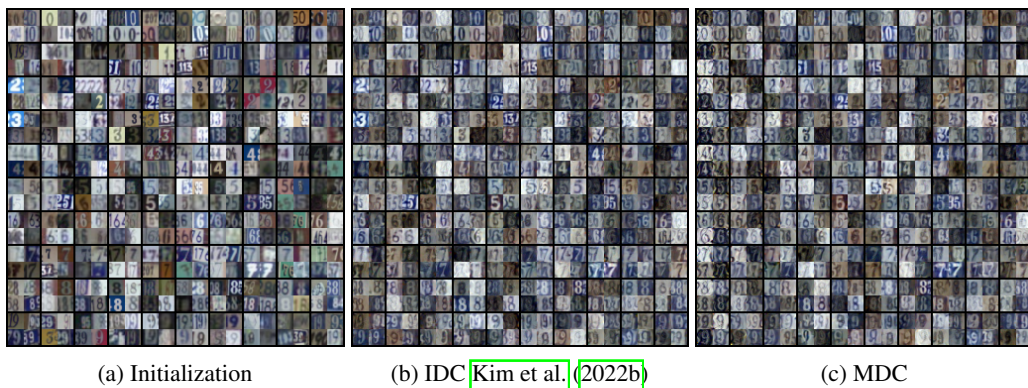


Figure 10: Visualization of the proposed condensation method. The number of the multi-formation factor is 2, which means each condensed image is merged with four original images. SVHN,  $IPC_{10}$ .

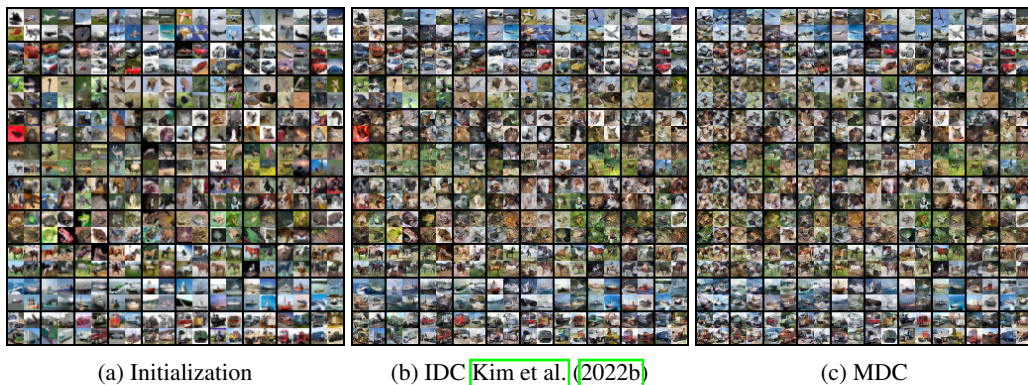


Figure 11: Visualization of the proposed condensation method. The number of the multi-formation factor is 2, which means each condensed image is merged with four original images. CIFAR-10,  $IPC_{10}$ .

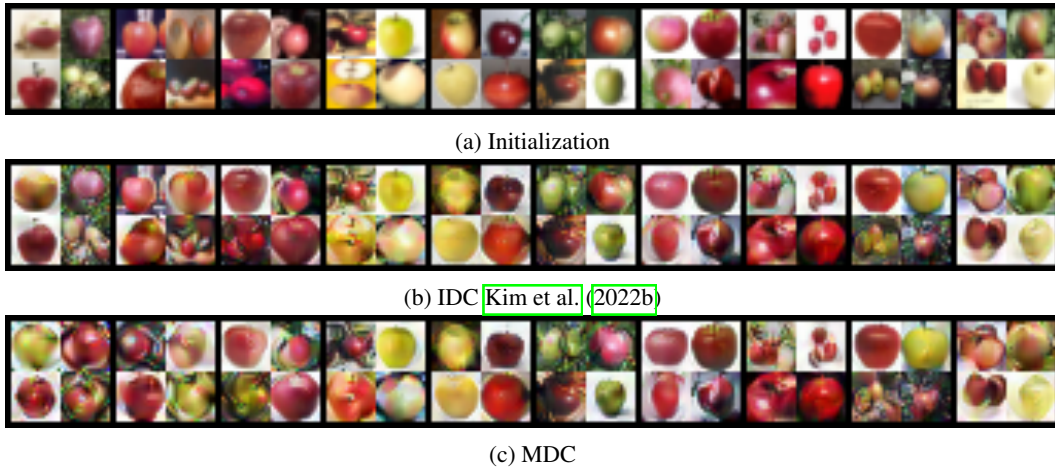


Figure 12: Visualization of the proposed condensation method. CIFAR-100, IPC<sub>10</sub>, class: apple.

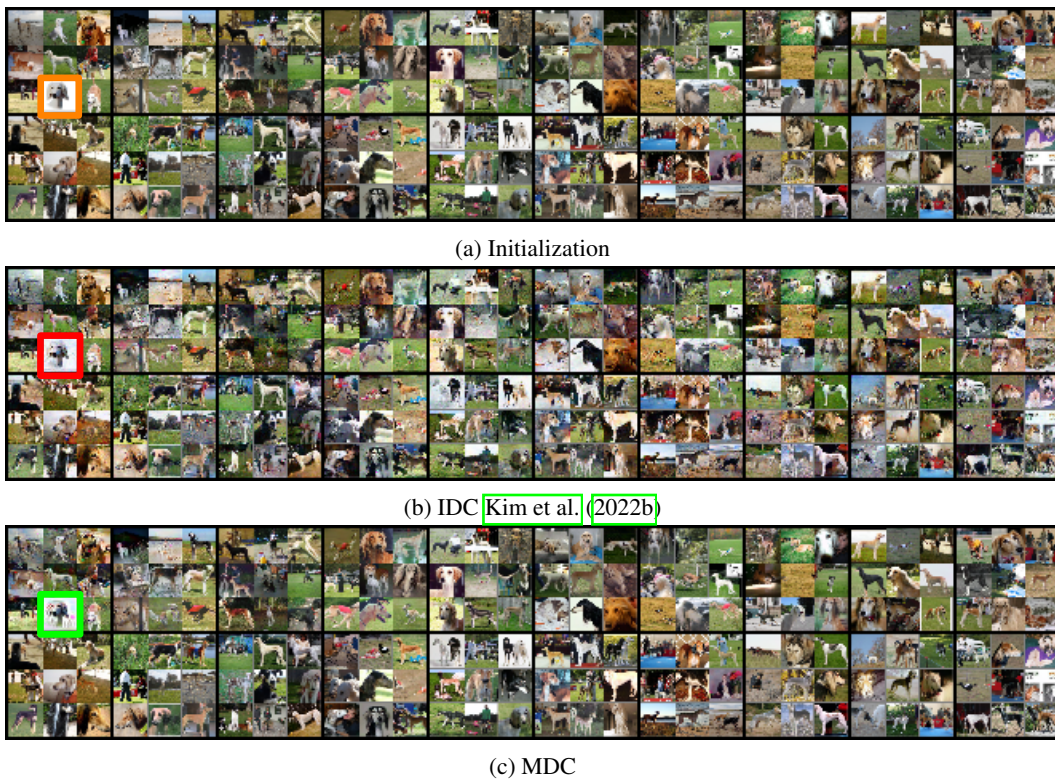


Figure 13: Visualization of on ImageNet targeting IPC<sub>20</sub>. The number of the multi-formation factor is 3, which means each condensed image is merged with nine original images. class: gazelle hound.