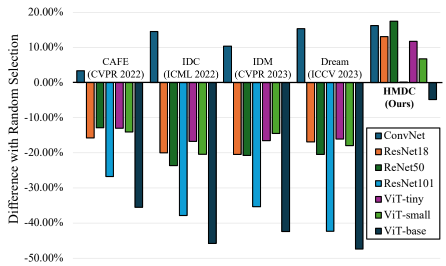
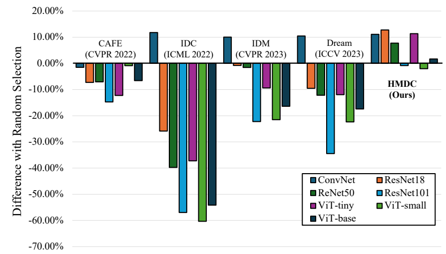
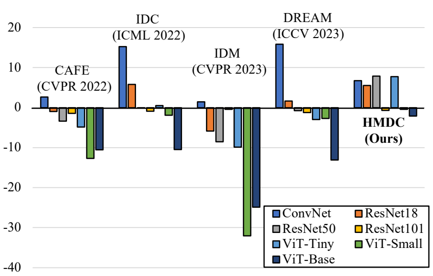



**(a)** The results obtained from CIFAR-10 [4] at IPC 1.

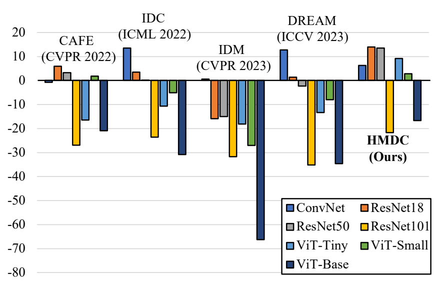



**(b)** The results obtained from CIFAR-10 at IPC 10.

**Fig. 1:** The graph depicts the performance gap between Random Data and condensation using each method. In most instances, images generated through these methods show a decrease in performance, except for ConvNet. Conversely, Heterogeneous Model Dataset Condensation (HMDC) demonstrates consistent performance across various models.




**(a)** The results obtained from CIFAR-100 at IPC 1.




**(b)** The results obtained from CIFAR-100 at IPC 10.

**Fig. 2:** The graph depicts the performance gap between Random Data and condensation using each method. In most instances, images generated through these methods show a decrease in performance, except for ConvNet. Conversely, Heterogeneous Model Dataset Condensation (HMDC) demonstrates consistent performance across various models.

# A Visualization of Comparison with Random

We have translated the tabulated data presented in Table 1 of the main text into a visual representation depicted in Figure 1. This graphical representation vividly illustrates the escalating impact of model dependency, a phenomenon particularly noticeable as the number of Images Per Class (IPC) decreases. Clearly, a majority of the generated images tend to exhibit bias towards Convolutional Neural Network (CNN) [6] and showcase suboptimal performance compared to random selection, posing a significant hindrance to their usability. However, it is noteworthy that Heterogeneous Model Dataset Condensation (HMDC) significantly alleviates this issue, thereby expanding the applicability and robustness of Dataset Condensation (DC). This visual insight contributes to a more nuanced understanding of the intricate dynamics of model dependency in the context of dataset condensation.

| IPC | Methods | ConvNet (0.3M) | ResNet18 (11M) | ResNet50 (22M) | ResNet101 (43M) | CNN Average | ViT-tiny (5.5M) | ViT-small (21M) | ViT-base (86M) | ViT Average | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Random | 22.51±0.30 | 39.39±3.43 | 39.95±8.40 | 51.17±1.04 | 38.26±3.29 | 32.49±3.59 | **58.33±1.42** | 41.05±5.99 | 43.96±3.66 | 40.70±3.45 |
| | CAFE | 8.76±0.09 | 14.61±1.80 | 17.03±0.97 | 19.75±0.60 | 15.04±0.86 | 8.38±0.22 | 19.35±2.40 | 7.76±1.16 | 11.83±1.26 | 13.66±1.03 |
| | IDM | 32.86±0.24 | 18.92±1.69 | 19.20±1.60 | 15.83±7.37 | 21.70±2.72 | 18.01±1.57 | 15.90±2.78 | 12.81±4.90 | 15.58±1.68 | 19.08±2.88 |
| | IDC | 37.48±0.29 | 18.81±2.44 | 16.85±0.96 | 11.15±0.58 | 21.07±1.07 | 10.82±1.81 | 10.82±1.81 | 14.93±1.74 | 12.72±1.01 | 12.72±1.01 |
| | DREAM | 37.85±0.35 | 22.50±2.89 | 19.47±3.71 | 8.82±1.24 | 22.16±2.04 | 14.53±1.26 | 10.93±1.32 | 11.93±4.23 | 12.46±2.27 | 18.00±2.14 |
| | HDMC | **38.74±0.37** | **52.45±2.78** | **57.40±4.40** | **51.32±0.88** | **49.98±2.11** | **39.25±6.05** | 53.50±7.22 | **49.43±10.9** | **47.39±2.51** | **48.87±4.65** |
| 10 | Random | 36.45±0.12 | 56.59±4.86 | 69.55±9.69 | **82.66±1.82** | 61.31±4.12 | 59.36±9.19 | **90.11±1.25** | 81.26±5.66 | 76.91±5.37 | 68.00±4.66 |
| | CAFE | 20.03±0.28 | 38.35±0.93 | 41.43±1.63 | 58.57±0.71 | 39.60±0.89 | 30.67±1.00 | 69.80±3.10 | 49.81±2.49 | 50.09±1.08 | 44.10±1.45 |
| | IDM | 46.49±0.17 | 55.82±1.16 | 68.01±2.41 | 60.42±6.39 | 57.68±2.53 | 49.96±2.75 | 68.61±9.74 | 64.87±13.67 | 61.15±5.53 | 59.17±5.18 |
| | IDC | **48.22±0.34** | 30.75±1.78 | 29.79±5.94 | 25.71±2.58 | 33.62±2.66 | 22.14±1.20 | 29.79±2.92 | 27.10±4.90 | 26.35±1.85 | 30.50±2.81 |
| | DREAM | 47.62±0.53 | 44.26±1.78 | 50.70±6.11 | 45.68±1.97 | 47.07±2.60 | 36.26±2.62 | 55.95±15.59 | 56.69±14.05 | 49.64±10.76 | 48.17±6.09 |
| | HDMC | 47.96±0.09 | **69.87±0.12** | **77.29±1.73** | 82.25±0.93 | **69.34±0.72** | **73.55±4.24** | 89.01±1.42 | **85.38±1.45** | **82.64±1.62** | **75.04±1.43** |

| IPC | Methods | ConvNet (0.3M) | ResNet18 (11M) | ResNet50 (22M) | ResNet101 (43M) | CNN Average | ViT-tiny (5.5M) | ViT-small (21M) | ViT-base (86M) | ViT Average | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Random | 45.47±0.39 | 73.55±1.02 | 80.96±6.48 | 91.33±0.57 | 72.83±2.11 | 71.68±5.66 | 96.19±0.10 | 94.68±1.88 | 87.52±2.54 | 79.13±2.30 |
| | CAFE | 27.79±0.09 | 49.67±1.10 | 55.35±2.41 | 68.71±1.07 | 50.38±1.17 | 58.60±4.30 | 84.06±0.39 | 84.19±0.48 | 75.62±2.23 | 61.20±1.41 |
| | IDM | 49.43±0.08 | 73.51±1.27 | **82.98±3.03** | 91.39±0.15 | 74.33±1.14 | 75.18±3.12 | 95.66±0.30 | 94.06±1.26 | 88.30±1.56 | 80.32±1.32 |
| | IDC | **52.90±0.23** | 63.69±0.69 | 72.60±0.49 | 67.15±12.6 | 64.09±3.50 | 53.53±3.87 | 77.39±15.5 | 74.41±12.0 | 68.44±5.97 | 65.95±6.48 |
| | DREAM | 52.63±0.34 | 76.00±3.05 | 85.04±0.69 | 90.11±0.89 | **75.95±1.24** | 69.25±5.03 | 79.75±1.92 | 92.02±2.57 | 84.81±2.10 | 93.17±0.84 |
| | HDMC | 52.06±0.58 | 75.78±0.60 | 84.06±0.84 | 88.91±0.36 | 75.20±0.59 | 74.76±4.32 | 90.43±1.08 | 89.62±3.02 | 84.94±1.63 | 79.37±1.54 |

**Table 1:** Experimental result achieved by employing condensed images from CIFAR-100 dataset for each condensation method across multiple models.

## B Additional Experiments on Other Datasets

We conducted experiments on CIFAR-100 [4] and TinyImageNet [5] to validate the effectiveness of Heterogeneous Model Dataset Condensation on datasets other than CIFAR-10.

Figure 2 illustrates the performance contrast with Random Image on CIFAR-100, in relation to Figure 1. As a comparison group, we used CAFE, IDM, IDC, and DREAM as in the main text. Once more, it is evident that HMDC exhibits commendable performance across the board, and distinctively, it avoids the model dependency challenges experienced by other methods. The results of these experiments are presented in Table 1.

We observe that CIFAR-100 exhibits strong performance across various models, excluding ConvNet, mirroring the patterns seen in CIFAR-10. Images generated at IPC 1 displayed notably low perceptual quality, yet remained valuable for training purposes. Although this may have influenced generalization performance to some extent, the overall performance distribution appears to be independent of the specific model employed.

We conducted an experiment to assess the performance of the proposed method on TinyImageNet. However, owing to the heightened computational requirements resulting from the increased resolution and number of classes, we opted for a partial training approach to discern trends. Despite the limited training, Dream was trained more in proportion and showing performance improvement, effectively illustrating the observed trends. Notably, Dream continued to exhibit model dependency on the larger dataset. In contrast, HMDC

| Methods | SimpleCNN | HuBERT |
|---|---|---|
| Dream | 10.00±0.00* | 12.08±1.19 |
| **HMDC (Ours)** | **17.08±1.91** | **12.50±2.50** |

**Table 2:** Table of results from experimenting with HMDC and Dream in audio modality. ∗ means failure to converge.

| Methods | SimpleRNN | BERT |
|---|---|---|
| Dream | 50.00±0.00* | 50.97±0.82 |
| **HMDC (Ours)** | 50.00±0.00* | **51.12±0.02** |

**Table 3:** Table of results from experimenting with HMDC and Dream in language modality. ∗ means failure to converge.

demonstrated the ability to mitigate performance disparities between models, showcasing a more consistent improvement across different model architectures.

## C Additional Modality Experiments

We extended our HMDC initially designed for the image domain to other audio and language domains. For each domain, ESC-10 [9] (10 classes) in audio and the IMDB Dataset [8] (2 classes) in text served as compression targets into 1 sample per class. We defined SimpleCNN (6 layers) and SimpleLSTM (6 layers) as correspondence for ConvNet, which is predominantly used in DC, respectively. Additionally, HuBERT [2] and BERT, structurally similar to ViT [1], were employed as secondary models to maintain heterogeneity. In each scenario, we applied the method of interpolating the temporal feature corresponding to the spatial feature. Tables 2 and 3 present the comparative results, demonstrating that our HMDC has the potential for application across various modalities.

## D Additional Implementation Details

We ran all of our experiments on the A5000 GPU. In Table 1 and 2 on the main paper, we re-implemented CAFE [10], IDM [12], IDC [3], and DREAM [7] based on their official implementations to enable model and seed flexibility. Each implementation prioritizes adherence to the paper's representation, with the official code considered secondarily. The computationally intensive experiments on TinyImageNet lasted a quarter of the iterations of the other experiments. We checked that each method exhibits comparable performance to the official code. For the evaluation process, we followed IDC's augmentation strategy, incorporating color adjustments, cropping, and CutMix [11]. Subsequently, we measured the performance on the test data after training for $1,000$ epochs at the specified learning rate for each model mentioned in the text, without any learning rate scheduling.

In the experiments detailed in Table 4 of the main text, when we applied Vision Transformer (ViT) [1] to DREAM, the image failed to converge, necessitating a reduction in the learning rate of the image from $5e^{-2}$ to $5e^{-5}$, as outlined in the original paper.

In Figure 5 on the main paper, we adhered to HMDC's training protocol and examined the Maximum Gradient Value both with and without the Gradient Balance Module at each of the 10 steps, conducting a total of 8 iterations.

We applied a learning rate of $1e^{-3}$ for the affine layer and layer-matching matrix when utilizing ViT-Small. Conversely, for all other cases, a learning rate of $1e^{-1}$ was employed. Additionally, to accommodate memory constraints, the batch sizes were set as 128, 16, 32, 16, 4, 128, 128, 64, and 4, respectively, starting after Random.

## References


1. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2020)
2. Hsu, W.N., et al.: Hubert: Self-supervised speech representation learning by masked prediction of hidden units. In: TASLP (2021)
3. Kim, J.H., Kim, J., Oh, S.J., Yun, S., Song, H., Jeong, J., Ha, J.W., Song, H.O.: Dataset condensation via efficient synthetic-data parameterization. In: International Conference on Machine Learning. pp. 11102–11118. PMLR (2022)
4. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
5. Le, Y., Yang, X.: Tiny imagenet visual recognition challenge. CS 231N **7**(7), 3 (2015)
6. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)
7. Liu, Y., Gu, J., Wang, K., Zhu, Z., Jiang, W., You, Y.: Dream: Efficient dataset distillation by representative matching. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 17314–17324 (October 2023)
8. Maas, A.L., et al.: Learning word vectors for sentiment analysis. In: Proceedings of ACL (2011)
9. Piczak, K.J.: ESC: Dataset for Environmental Sound Classification. In: Proceedings of ACM MM (2015)
10. Wang, K., Zhao, B., Peng, X., Zhu, Z., Yang, S., Wang, S., Huang, G., Bilen, H., Wang, X., You, Y.: Cafe: Learning to condense dataset by aligning features. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12196–12205 (2022)
11. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6023–6032 (2019)
12. Zhao, G., Li, G., Qin, Y., Yu, Y.: Improved distribution matching for dataset condensation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7856–7865 (2023)