Dataset Condensation via Efficient Synthetic-Data Parameterization

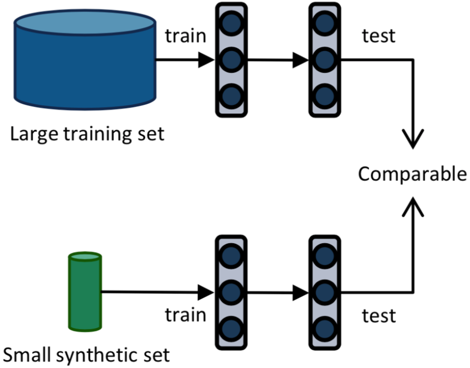
Jang-Hyun Kim[1], Jinuk Kim[1], Seong Joon Oh[2,3], Sangdoo Yun[2], Hwanjun Song[2], Joonhyun Jeong[2], Jung-Woo Ha[2], Hyun Oh Song[1]

[1]Seoul National University [2]NAVER AI [3]University of Tübingen

ICML 2022

About dataset condensation

Dataset condensation seeks to **synthesize** a small dataset ($\mathcal{S}$) that has comparable training performance to the original training set ($\mathcal{T}$).

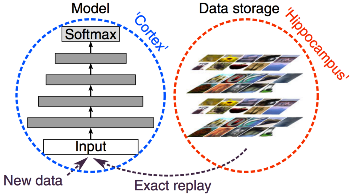


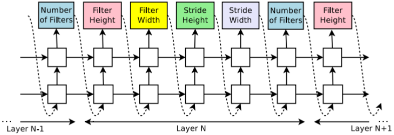
Zhao et al., Dataset Condensation with Gradient Matching, ICLR, 2021

About dataset condensation

The synthesized datasets have a variety of applications:

- increasing the efficiency of replay exemplars in **continual learning**.
- accelerating **neural architecture search**.





Ven et al., Brain-inspired replay for continual learning with artificial neural networks, Nature, 2020

Zoph et al., Neural Architecture Search with Reinforcement Learning, ICLR, 2017

Shortcomings of existing methods

1. **Inefficiency**
   - Pixel-level optimization of the synthetic data:

$$\mathcal{S}^* = \underset{\mathcal{S}\in\mathbb{R}^{n\times m}}{\operatorname{argmin}}\ D(\mathcal{S}, \mathcal{T}).$$

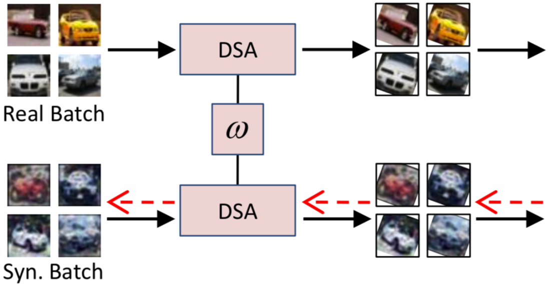


Figure: Gradient matching with differentiable siamese augmentation (DSA).

Zhao et al., Dataset Condensation with Differentiable Siamese Augmentation, ICML, 2021

Shortcomings of existing methods

1. Inefficiency
   - Pixel-level optimization of the synthetic data.
2. **Small-scale** experiments
   - Only tackles small-scale benchmarks such as CIFAR-10.
   - Test with a 3-layer convolutional network which has low upper-bound performance.

Shortcomings of existing methods

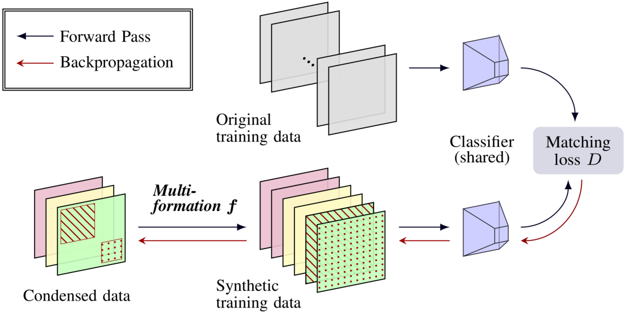
1. Inefficiency
   - Pixel-level optimization of the synthetic data.
2. Small-scale experiments
   - Only tackles small-scale benchmarks such as CIFAR-10.
   - Test with a 3-layer convolutional network which has low upper-bound performance.
3. **Lack of theoretical** interpretation
   - Lacks theoretical understanding of why certain objectives work better.

Our contributions

1. **Efficiency**
   - We propose a novel framework with efficient synthetic-data parameterization (*multi-formation*).

$$\mathcal{S}^* = \underset{\mathcal{S}\in\mathbb{R}^{n\times m}}{\operatorname{argmin}}\, D(f(\mathcal{S}), \mathcal{T}) \qquad \text{(Optimization)}$$

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \ell(\theta; f(\mathcal{S}^*)). \qquad \text{(Evaluation)}$$



Our contributions

1. Efficiency
   - We propose a novel framework with efficient synthetic-data parameterization.

2. **Large-scale** experiments
   - Condense high-resolution data (ImageNet) and set benchmarks.
   - Multiple test models including ResNet and EfficientNet.

Our contributions: Improved optimization techniques

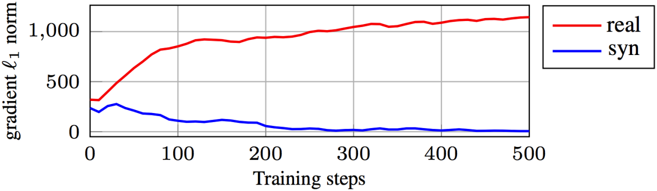


Figure: Gradient norm during the condensation stage.

Existing gradient matching methods suffer from **gradient norm divergence**.

- We propose effective optimization techniques for regularizing the optimization.

Our contributions

1. Efficiency
   - We propose an efficient synthetic-data parameterization strategy for condensation.
2. Large-scale experiments
   - Condense high-resolution data (ImageNet) and set benchmarks.
   - Multiple test models including ResNet and EfficientNet.
3. **Theoretical** interpretation
   - Our method achieves the lower bound of the original objective function.

$$\min_{\mathcal{S}\in\mathbb{R}^{n\times m}} D(f(\mathcal{S}),\mathcal{T}) \leqslant \min_{\mathcal{S}\in\mathbb{R}^{n\times m}} D(\mathcal{S},\mathcal{T}).$$

Results: classification (CIFAR-10)

IDC denotes our full methods and *IDC-I* denotes our method without multi-formation.

- IDC and IDC-I achieve the state-of-the-art performance over various compression ratio (pixel/class) and test models.

Pixel/Class	Test Model	Random	Herding	DSA	DM	IDC-I	IDC
10×32×32 (0.2%)	ConvNet-3	37.2	41.7	52.1†	53.8	58.3 (+4.5)	**67.5** (+13.7)
	ResNet-10	34.1	35.9	32.9	42.3	50.2 (+7.9)	**63.5** (+21.2)
	DenseNet-121	36.5	36.7	34.5	39.0	49.5 (+10.5)	**61.6** (+22.6)
50×32×32 (1%)	ConvNet-3	56.5	59.8	60.6†	65.6	69.5 (+3.9)	**74.5** (+8.9)
	ResNet-10	51.2	56.5	49.7	58.6	65.7 (+7.1)	**72.4** (+13.8)
	DenseNet-121	55.8	59.0	49.1	57.4	63.1 (+4.1)	**71.8** (+12.8)

Table: CIFAR-10 top-1 accuracy.

Results: classification (ImageNet)

Our method outperforms on high-resolution data (ImageNet subset).

- The gains range from 3%p to 26%p!

Class	Pixel/Class	Test Model	Random	Herding	DSA	DM	IDC-I	IDC
10	10×224×224 (0.8%)	ResNetAP-10	46.9	50.4	52.7	52.3	61.4 (+8.7)	**72.8** (+10.5)
		ResNet-18	43.3	47.0	44.1	41.7	56.2 (+9.2)	**73.6** (+26.6)
		EfficientNet-B0	46.3	50.2	48.3	45.0	58.7 (+8.5)	**74.7** (+24.5)
100	10×224×224 (0.8%)	ResNetAP-10	20.7	22.6	21.8	22.3	29.2 (+6.6)	**46.7** (+24.1)
		ResNet-18	15.7	15.9	13.5	15.8	23.3 (+7.4)	**40.1** (+24.2)
		EfficientNet-B0	22.4	24.5	19.9	20.7	27.7 (+3.2)	**36.3** (+11.8)

Table: ImageNet subset top-1 accuracy.

Results: classification (Speech Commands)

We also verify the effectiveness of our approach on speech data.

Spectrogram/ Class	Random	Herding	DSA	DM	IDC-I	IDC
10×64×64 (1%)	42.6	56.2	65.0	69.1	73.3 (**+4.2**)	**82.9** (**+13.8**)
20×64×64 (2%)	57.0	72.9	74.0	77.2	83.0 (**+5.8**)	**86.6** (**+9.4**)

Table: Google Speech Commands top-1 accuracy.

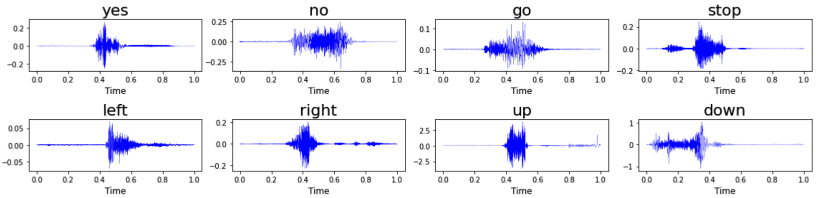


Figure: Condensed speech samples.

Results: continual learning

Our condensed data shows the best performance on exemplar-based continual learning setting.

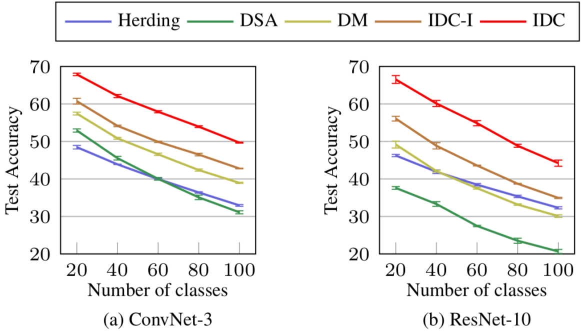


Figure: Class incremental learning on CIFAR-100.

Qualitative results

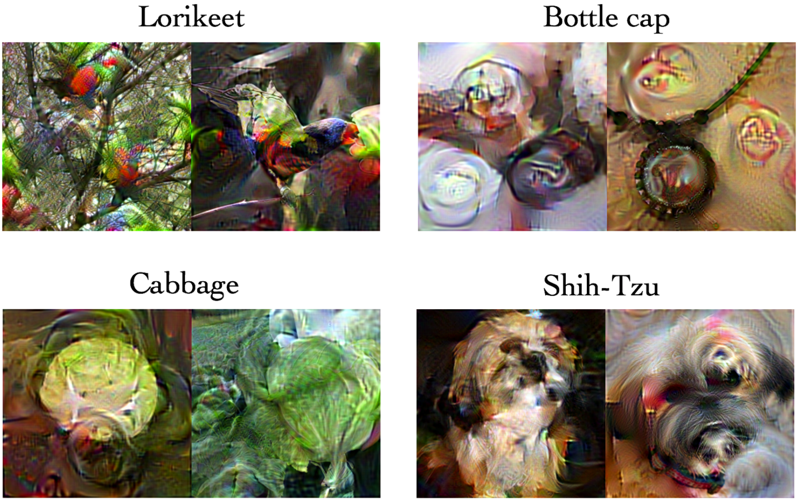


Figure: Representative synthetic data samples from ImageNet.

Conclusion

- We propose a novel condensation algorithm with **synthetic-data parameterization** for efficient condensation.
- Our method achieves the **best condensation performance** on vision and audio domains with various test models and compression ratios.
- Our method also shows promising results on exemplar-based continual learning.
- https://github.com/snu-mllab/Efficient-Dataset-Condensation