ICLR

Boost Self-Supervised Dataset Distillation via Parameterization, Predefined Augmentation, and Approximation

Sheng-Feng Yu[1,2], Jia-Jiun Yao[2], and Wei-Chen Chiu[2]

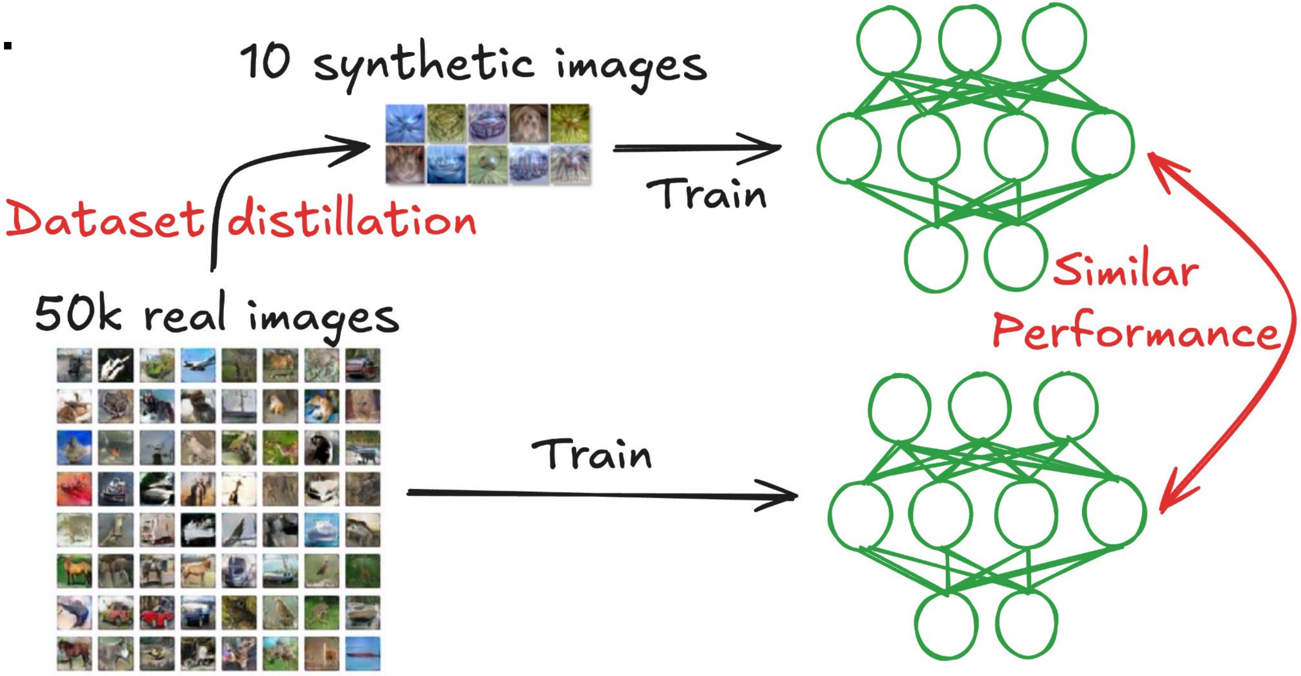
[1]Macronix International Co., Ltd. [2]National Yang Ming Chiao Tung University

MXIC
MACRONIX INTERNATIONAL CO., LTD.

NATIONAL YANG MING CHIAO TUNG UNIVERSITY

Dataset Distillation (DD)

Dataset Distillation (DD) aims to generate a compact, synthetic training dataset that retains the training effectiveness of the original dataset.



Settings of Dataset Distillation

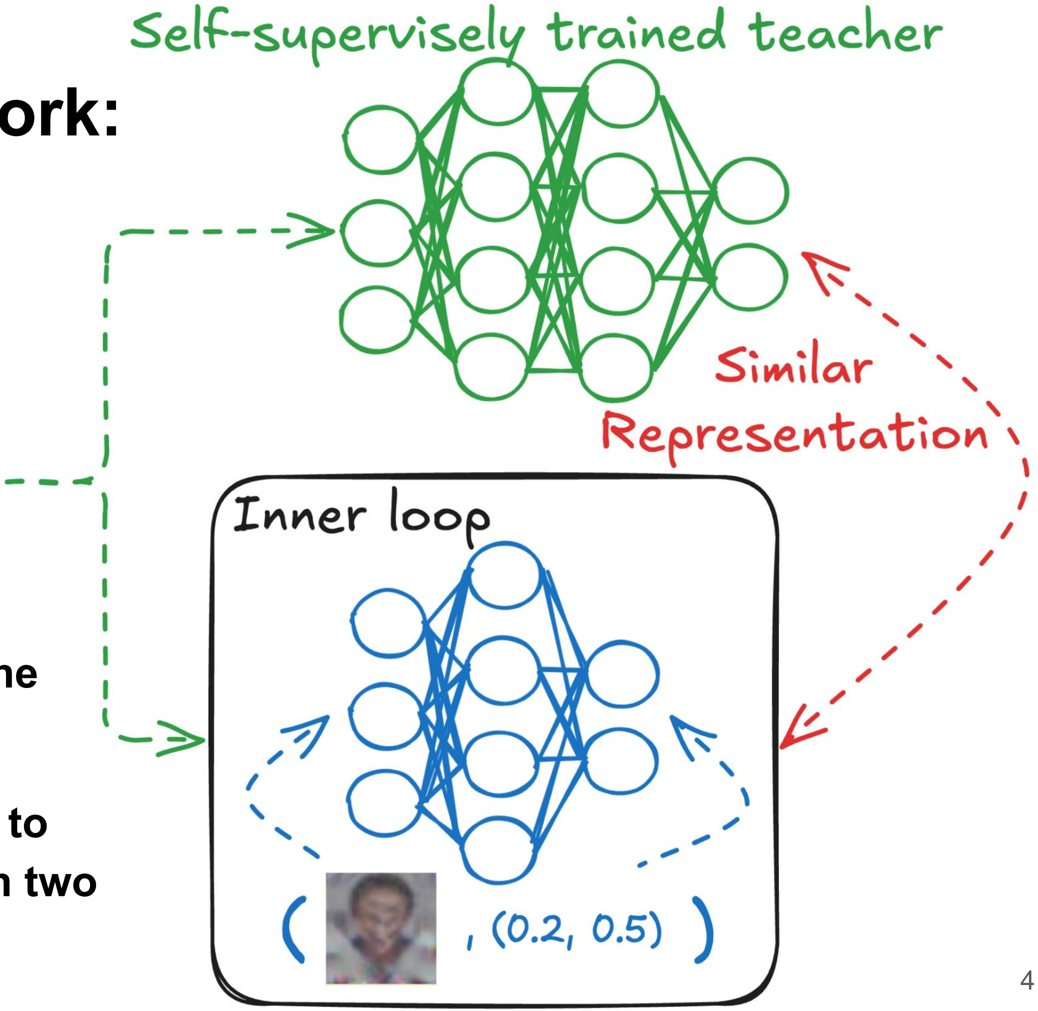
Most studies focus on supervised dataset distillation

- **Distillation requires extensive <u>human annotation</u>**
- **This usually leads to significant performance drops when evaluated on different models and tasks.**

Our work explores self-supervised dataset distillation

- **Solely relies on images and <u>self-supervisedly learned representations</u>**
- **This improves generalization across various models and tasks**

Basic Learning Framework: Bilevel Optimization



<u>Inner Loop</u>: Trains a network to fit on the synthetic dataset from scratch.

<u>Outer Loop</u>: Updates synthetic dataset to align the output of real images between two networks.

Motivations

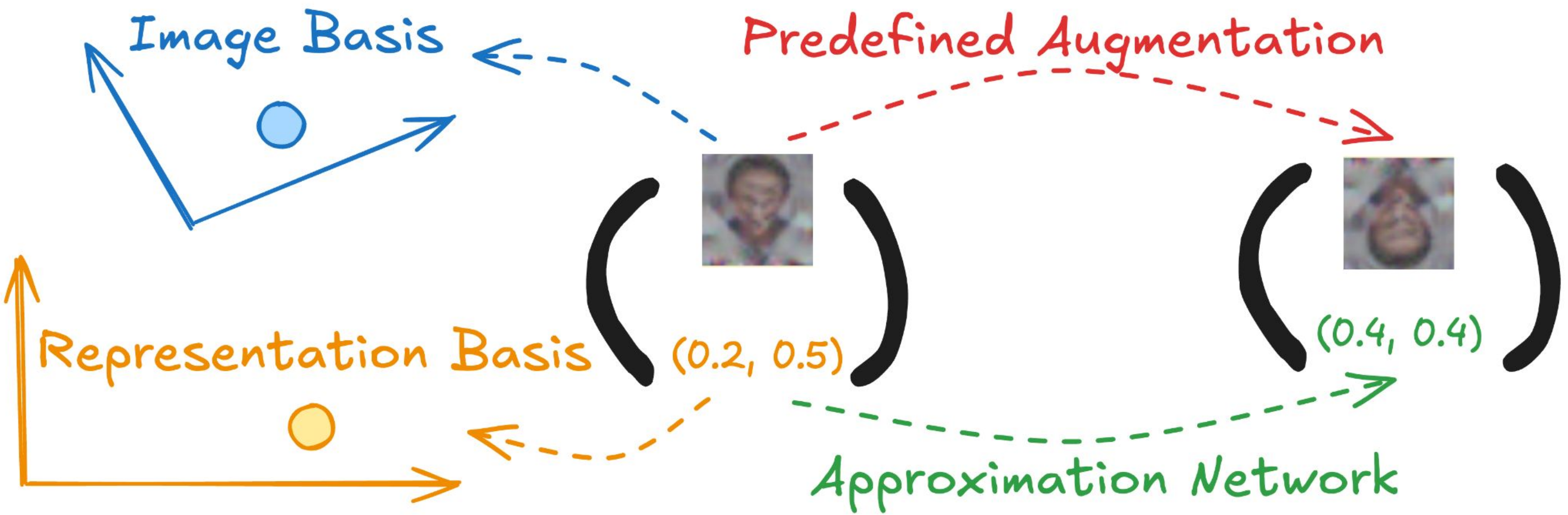
- **Raw images and representations might contain redundant informations, which cause inefficient data utilization.**

- **Random data augmentation - the key component in self-supervised learning - has been proven incompatible to bilevel optimization. [1]**

[1] Dong Bok Lee, et al. "Self-Supervised Dataset Distillation for Transfer Learning" ICLR2024

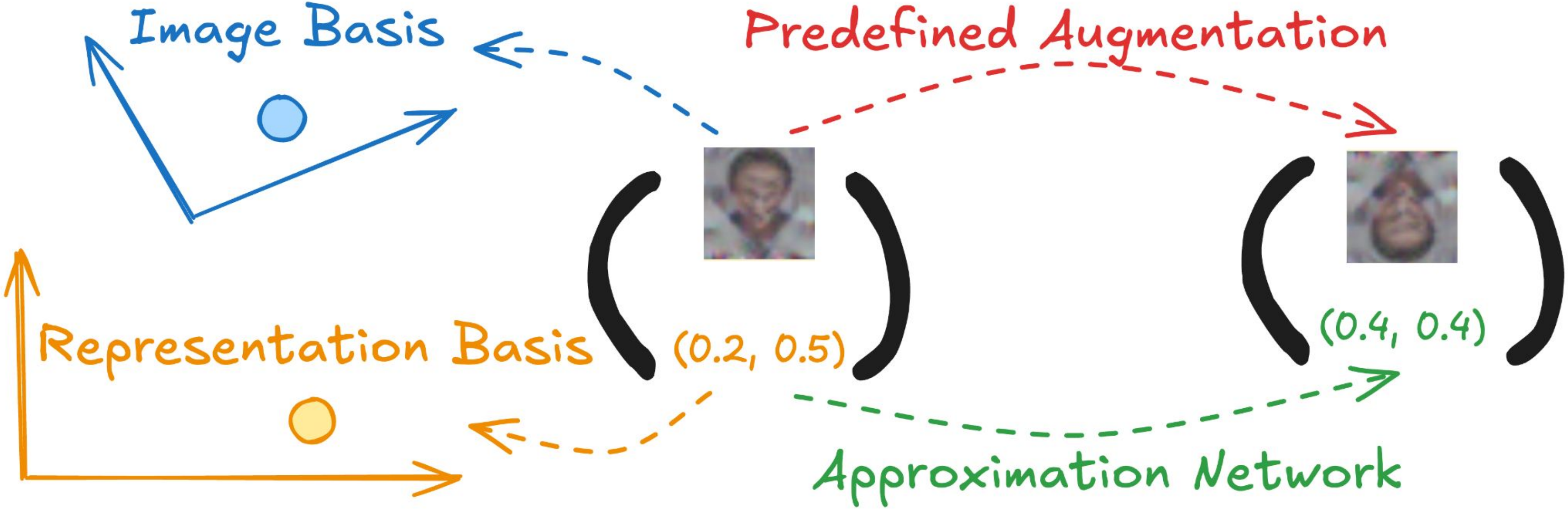
Our Contributions (1)

- **For better storage utilization, we build optimized orthogonal image basis and representation basis to store their information in a more compact manner.**



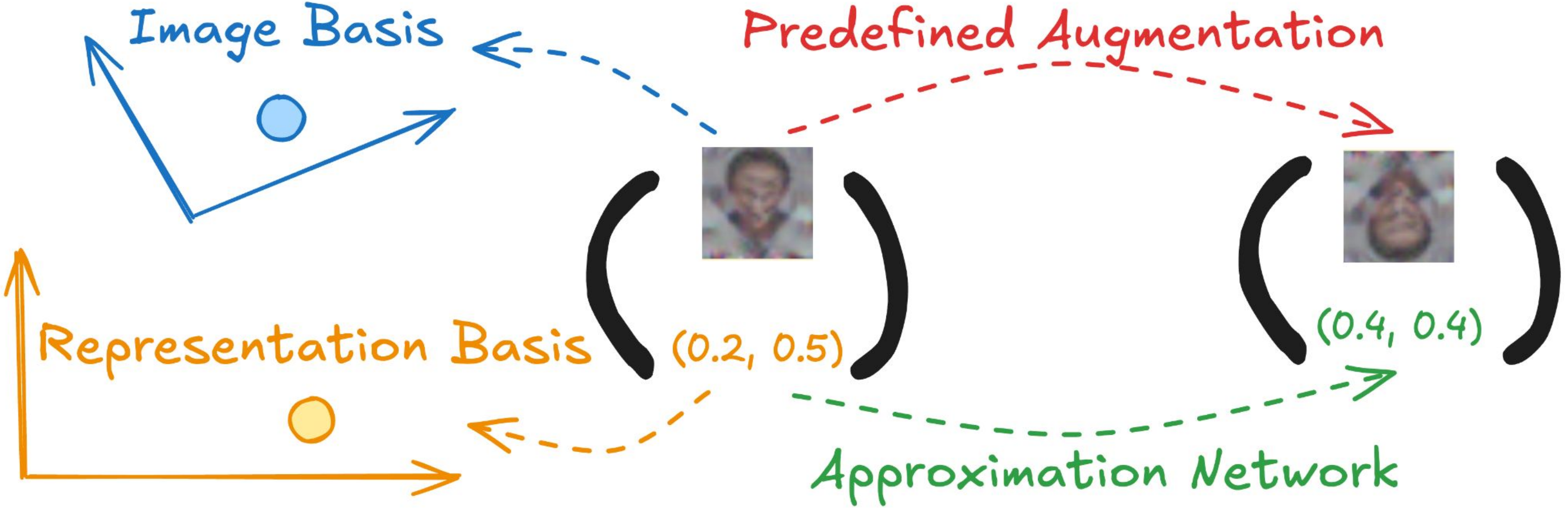
Our Contributions (2)

- For better storage utilization, we build optimized orthogonal image basis and representation basis to store their information in a more compact manner.
- Implementing predefined augmentations to mitigate instability in bilevel optimization.



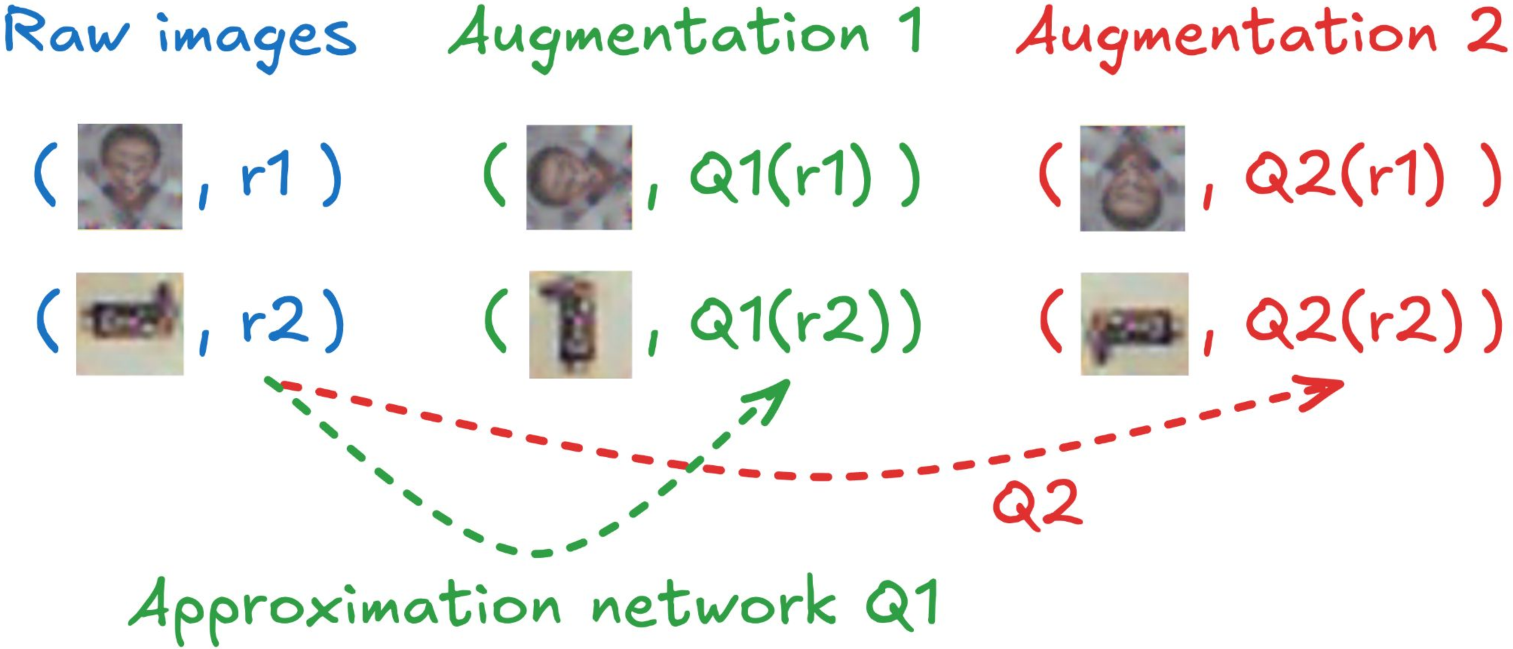
Our Contributions (3)

- For better storage utilization, we build optimized orthogonal image basis and representation basis to store their information in a more compact manner.
- Implementing predefined augmentations to mitigate instability in bilevel optimization.
- Developing approximation networks to predict the shift caused by predefined augmentation, which further enhance the data efficiency.



Approximation Networks

Instead of saving all augmented image and representation pairs, we only store the original image and representation pair along with approximation networks.



Cross-Architecture Performance

	ConvNet	VGG11	ResNet18	AlexNet	MobileNet	ViT
		Cross-architecture				
Random	43.66	23.76	19.26	28.82	11.99	20.70
DSA	39.38	19.97	20.11	31.57	9.58	20.03
IDM	38.71	14.24	19.05	33.71	8.18	17.41
DATM	38.73	26.04	<u>21.20</u>	29.31	10.17	20.11
KRR-ST	47.00	27.78	18.92	31.27	10.11	20.82
Ours	<u>52.41</u>	<u>35.35</u>	20.90	<u>36.88</u>	<u>24.14</u>	<u>23.33</u>

Downstream Tasks Transferability

	Downstream tasks			
	CIFAR100	CIFAR10	CUB2011	Stanford Dogs
Random	43.66	68.56	8.88	12.18
DSA	39.38	65.67	6.89	9.97
IDM	38.71	64.45	7.09	9.56
DATM	38.73	66.17	6.73	9.70
KRR-ST	47.00	72.14	10.43	13.42
Ours	<u>52.41</u>	<u>76.83</u>	<u>12.24</u>	<u>15.34</u>

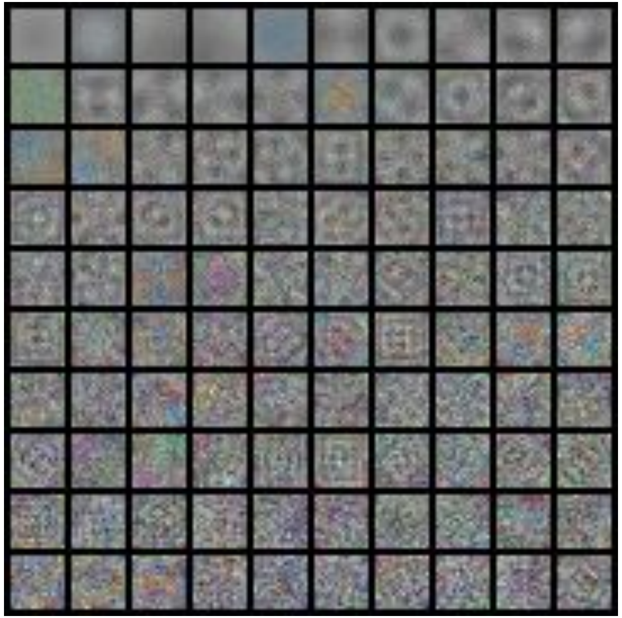
Distillation Result on Various Storage Budgets

	CIFAR100				
Memory Budget N	25	50	100	1000	5000
Random	41.61	41.80	43.66	49.82	52.76
DSA	-	-	39.38	36.16	36.25
IDM	-	-	38.71	37.21	42.19
DATM	-	-	38.73	44.98	46.23
KRR-ST	44.06	45.79	47.00	51.89	52.49
Ours	<u>51.41</u>	<u>52.08</u>	<u>52.41</u>	<u>53.54</u>	<u>55.53</u>

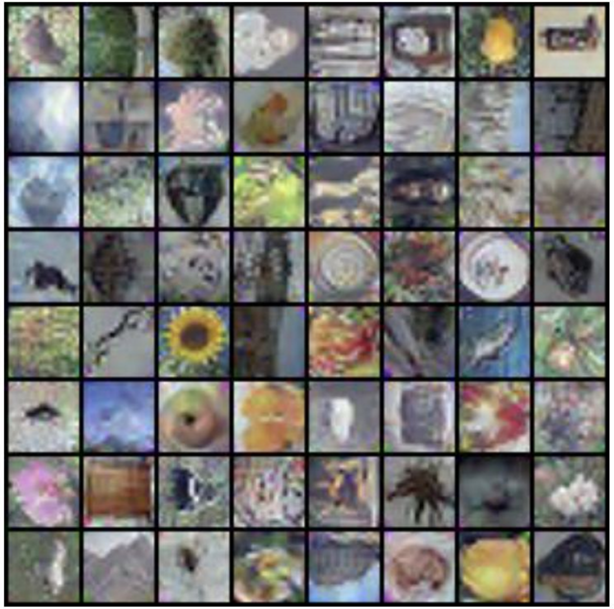
Ablation Study of the Proposed Components

	Accuracy
Baseline	47.00
+ Orthogonal Basis Parameterization	48.57
+ Predefined Augmentation and Approximation	52.41

Bases and Synthetic Images



CIFAR100 Bases



CIFAR100 Synthetic Images

Takeaways

- **Raw images and representations might contain redundant informations, which cause inefficient data utilization.**
  - **For better storage utilization, we build optimized orthogonal image basis and representation basis to store their information in a more compact manner.**
- **Random data augmentation - the key component in self-supervised learning - has been proven incompatible to bilevel optimization.**
  - **Implementing predefined augmentations to mitigate instability in bilevel optimization.**
  - **Developing approximation networks to predict the shift caused by the predefined augmentation, which further enhance the data efficiency.**
- **We demonstrate a better performance and cross-architecture generalizability.**

Thank you for your listening