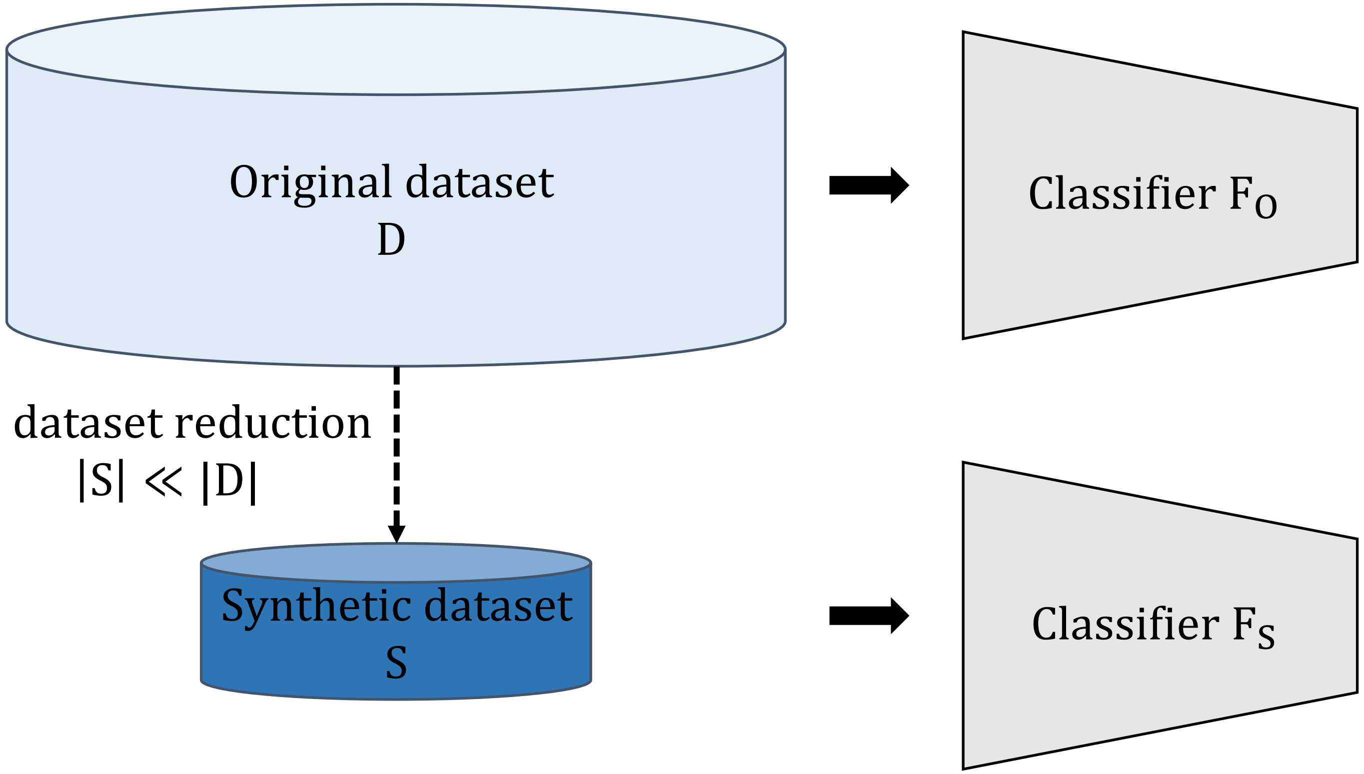
ICML
International Conference On Machine Learning

Dataset Condensation with Contrastive Signals

Saehyung Lee Sanghyuk Chun Sangwon Jung Sangdoo Yun Sungroh Yoon

SEOUL NATIONAL UNIVERSITY
NAVER AI LAB

Introduction



Minimize the performance gap between F_O *and* F_S

Introduction

(a) Trailer Truck

(b) Police Van

Task-irrelevant features: {Wheels, Head lights, Roads, Trees, ...}

Task-relevant features: {Logo, Police sign, Trailers, ...}

Source: Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and FeiFei, L. ImageNet: A Large-Scale Hierarchical Image Database. In CVPR09, 2009.

Class-Wise Gradient Matching (DC)

$$\mathcal{L} = \sum_{c=0}^{C-1} D\left(\nabla_{\theta_t} L(\mathcal{X}^c; \theta_t), \nabla_{\theta_t} L(\mathcal{S}^c; \theta_t)\right),$$

$\mathcal{X}^c$: a set of real images for class c

$\mathcal{S}^c$: a set of synthetic images for class c

θ_t: the model parameter at time t

C: the number of classes

$D(\cdot, \cdot)$: a distance function

$L(\cdot;)$: the cross-entropy loss function

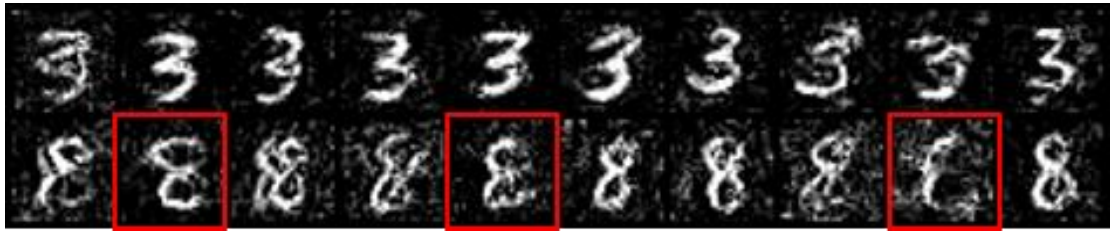
Our Gradient Matching (DCC)

$$\mathcal{L} = \sum_{c=0}^{C-1} D\left(\nabla_{\theta_t} L(\mathcal{X}^c; \theta_t), \nabla_{\theta_t} L(\mathcal{S}^c; \theta_t)\right)$$

$$\mathcal{L} = D\left(\frac{\boxed{\sum_{c=0}^{C-1}} \nabla_{\theta_t} L(\mathcal{X}^c;\theta_t)}{\sum_{c=0}^{C-1} \|\mathcal{X}^c\|}, \frac{\boxed{\sum_{c=0}^{C-1}} \nabla_{\theta_t} L(\mathcal{S}^c;\theta_t)}{\sum_{c=0}^{C-1} \|\mathcal{S}^c\|}\right)$$

Comparison

3 vs. 8 class-wise gradient matching



3 vs. 8 our gradient matching

Flipped 3 vs. 8 our gradient matching

Comparison

Dataset	Img/cls	Random	Selection-based C-score	GraNd	EL2N	Synthesis-based DC	DSA	Ours DCC	DSAC	Ω
SVHN	1	14.6±1.6	-	19.6±0.5	19.1±0.6	34.6±2.0	36.0±2.0	34.3±1.6	**47.5±2.6**	92.1±0.2
	10	35.1±4.1	-	37.5±1.6	32.5±1.2	76.2±0.6	78.9±0.5	76.2±0.8	**80.5±0.6**	
	50	70.9±0.9	-	69.1±0.7	68.7±0.7	82.7±0.3	84.4±0.4	83.3±0.2	**87.2±0.3**	
CIFAR-10	1	14.4±2.0	21.7±0.6	21.8±0.5	20.9±0.6	28.2±0.7	28.7±0.7	32.9±0.8	**34.0±0.7**	81.6±0.3
	10	26.0±1.2	31.6±0.4	32.3±0.4	32.3±0.4	44.7±0.6	52.1±0.6	49.4±0.5	**54.5±0.5**	
	50	43.4±1.0	39.8±0.4	41.2±0.3	40.7±0.3	54.8±0.5	60.6±0.4	61.6±0.4	**64.2±0.4**	
CIFAR-100	1	4.2±0.3	8.0±0.3	8.8±0.3	8.8±0.3	12.8±0.3	13.9±0.4	13.3±0.3	**14.6±0.3**	52.5±0.3
	10	14.6±0.5	18.1±0.2	17.8±0.2	17.3±0.2	26.6±0.3	32.4±0.3	30.6±0.4	**33.5±0.3**	
	50	29.7±0.4	30.4±0.3	27.6±0.2	27.7±0.2	32.1±0.3	38.6±0.3	**40.0±0.3**	39.3±0.4	

Conclusion

- We show that DC primarily focuses on the class-wise gradient while overlooking **contrastive signals**.
- To address this issue, we propose the **Dataset Condensation with Contrastive signals (DCC)** method.
- In our experiments, we demonstrate that the proposed DCC **outperforms** DC in fine-grained classification tasks and general benchmark datasets

Thank you!

https://github.com/Saehyung-Lee/DCC