Constructive Assimilation: Boosting Contrastive Learning Performance through View Generation Strategies

Ligong Han[1], Seungwook Han[2], Shivchander Sudalairaj[2], Charlotte Loh[3], Rumen Dangovski[3], Fei Deng[1], Pulkit Agrawal[4], Dimitris Metaxas[1], Leonid Karlinsky[2], Tsui-Wei Weng[5], Akash Srivastava[2]

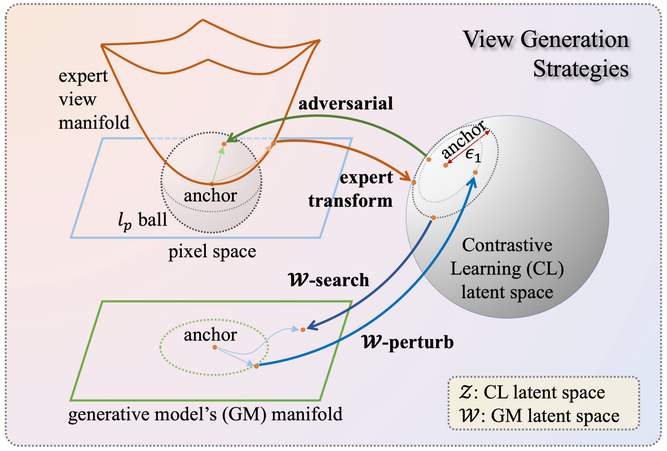
[1]Rutgers University [2]MIT-IBM Watson AI Lab [3]MIT EECS [4]MIT CSAIL [5]UCSD

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Motivation

- Transformations based on domain expertise (*expert transformations*), such as random-resized-crop and color-jitter, have proven *critical* to the success of contrastive learning techniques such as SimCLR.
- For imagery data, so far none of recent view generation methods has been able to outperform expert transformations.
- We tackle a different question: instead of replacing expert transformations with generated views, can we constructively *assimilate generated views* with expert transformations?

View Generation

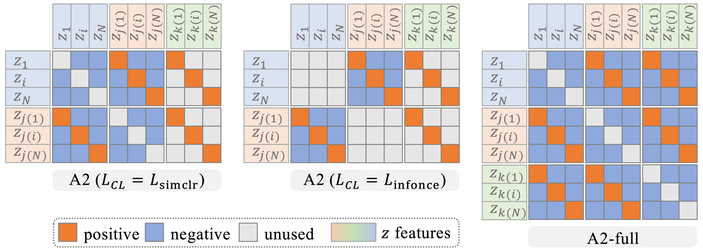


$$\{w_k^\star\}_{k=1}^n = \arg\min_{\{w_k\}} \left\{ \frac{1}{n} \sum_k \underbrace{\delta\left(\epsilon_1, \|f \circ g(w_k) - f(x_0)\|_2\right)}_{\text{boundary constraint}} + \underbrace{\lambda(\epsilon_2 - \bar{d}_n)^+}_{\text{uniformity}} \right\}$$

where $\delta(a,b)$ is L2 loss, $\bar{d}_n$ is the average Euclidean distances among generated views $\bar{d}_n = \frac{1}{n(n-1)} \sum_{j \neq k} \|f \circ g(w_j) - f \circ g(w_k)\|_2$, and $(\cdot)^+ := \max(0, \cdot)$ is a ReLU function.

View Generation

SimClR $$L_{\text{simclr}} = -\sum_{i \in \mathcal{I}} \log \frac{\exp(z_i \cdot z_{j(i)}/\tau)}{\sum_{a \in \mathcal{A}(i)} \exp(z_i \cdot z_a/\tau)}$$



Multiview Assimilation

$$L_{\text{multiview}} = L_{\text{CL}} - L_{\text{align}}, \quad \text{where} \quad L_{\text{CL}} = L_{\text{InfoNCE}} \quad \text{and} \quad L_{\text{align}} = \sum_{i \in \mathcal{I}} \frac{\alpha}{|k(i)|} \sum_{p \in k(i)} z_i^\top z_p / \tau$$

$$L_{\text{infonce}} = -\sum_{i \in \mathcal{I}_1} \log \frac{\exp(z_i \cdot z_{j(i)}/\tau)}{\sum_{a \in \mathcal{I}_2} \exp(z_i \cdot z_a/\tau)} - \sum_{i \in \mathcal{I}_2} \log \frac{\exp(z_i \cdot z_{j(i)}/\tau)}{\sum_{a \in \mathcal{I}_1} \exp(z_i \cdot z_a/\tau)}$$

$$L_{\text{A2-full}} = -\sum_{i \in \mathcal{I}} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i \cdot z_p/\tau)}{\sum_{a \in \mathcal{A}(i)} \exp(z_i \cdot z_a/\tau)}, \quad \text{where } P(i) = \{j(i)\} \cup \{k(i)\}$$

$$L_{\text{A2-simclr}} = L_{\text{simclr}} - \sum_{i \in \mathcal{I}} \frac{\alpha}{|k(i)|} \sum_{p \in k(i)} z_i \cdot z_p / \tau$$

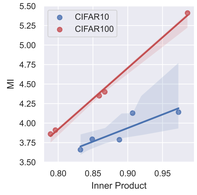
$$= -\sum_{i \in \mathcal{I}} \log \frac{\exp(z_i \cdot z_{j(i)}/\tau + \frac{\alpha}{|k(i)|} \sum_{p \in k(i)} (z_i \cdot z_p)/\tau)}{\sum_{a \in \mathcal{A}(i)} \exp(z_i \cdot z_a/\tau)}$$

Loss	CIFAR10 Acc@1	CIFAR10 5-NN	CIFAR100 Acc@1	CIFAR100 5-NN
A2-full	92.57	91.57	71.82	66.06
A2-SimCLR ($\alpha = 0.5$)	92.66	91.05	72.27	65.85
A2-InfoNCE ($\alpha = 0.5$)	**92.90**	90.95	**72.76**	66.46
A2-InfoNCE ($\alpha = 1$)	92.54	90.80	72.61	66.96

SCAN FOR CODE

Experiments

View Pairs	$I(f(X); f(\tilde{X}))$ CIFAR10	$I(f(X); f(\tilde{X}))$ CIFAR100	$\mathbb{E}[f(x)^\top f(\tilde{x})]$ CIFAR10	$\mathbb{E}[f(x)^\top f(\tilde{x})]$ CIFAR100
Original, Expert	4.14	5.41	0.973	0.986
Original, $\mathcal{W}$-*search*	4.13	4.40	0.907	0.867
$\mathcal{W}$-*search*, Expert	3.78	4.35	0.888	0.859
Original, $\mathcal{W}$-*perturb*	3.79	3.91	0.849	0.796
$\mathcal{W}$-*perturb*, Expert	3.66	3.86	0.832	0.789



View 1	View 2	View 3	Loss	CIFAR10	CIFAR100	TinyImageNet	Avg Rank
expert	expert	✗	SimCLR	92.04	70.41	47.48	4.67
expert	$\mathcal{W}$-*search*	✗	A1	91.86	71.69	**51.08**	2.67
expert	$\mathcal{W}$-*perturb*	✗	A1	91.09	70.83	50.18	4.67
expert	ViewMaker	✗	A1	82.91	41.87	26.40	8.00
ViewMaker	ViewMaker	✗	SimCLR	83.59	44.04	40.53	7.00
expert	expert	expert	A2	91.46	70.76	47.19	5.33
expert	expert	$\mathcal{W}$-*search*	A2	**92.90**	72.76	51.05	**1.67**
expert	expert	$\mathcal{W}$-*perturb*	A2	92.38	**72.95**	50.73	2.00
expert	expert	ViewMaker	A2	80.07	36.51	25.30	9.00