

Performance on CIFAR-10, CIFAR-100, Imagenet-100

Table 1: Performance comparison for the trace maximization formulation (Ours) and VICReg on small-scale datasets. Configuration: ResNet-18 backbone, 3-layer projector [2048-2048-2048], 1000 epochs for CIFAR, 400 epochs for ImageNet-100. The same hyperparameters are used across methods, they are favourable for VICReg and were not tuned for the trace formulation. Mean and standard deviation for top-1 / top-5 validation set accuracy are reported across 5 runs for CIFAR-10 and 4 runs for CIFAR-100. Only 1 run is reported for ImageNet-100 as it takes longer time to train, we will report results across more runs by the end of the discussion period.

	CIFAR-10		CIFAR-100		ImageNet-100	
	top-1	top-5	top-1	top-5	top-1	top-5
VICReg	91.15 \pm 0.16	99.64 \pm 0.05	66.76 \pm 0.37	89.39 \pm 0.3	79.28	94.64
Ours	91.19 \pm 0.13	99.67 \pm 0.04	67.35 \pm 0.38	89.91 \pm 0.11	78.36	94.3

Incoherence

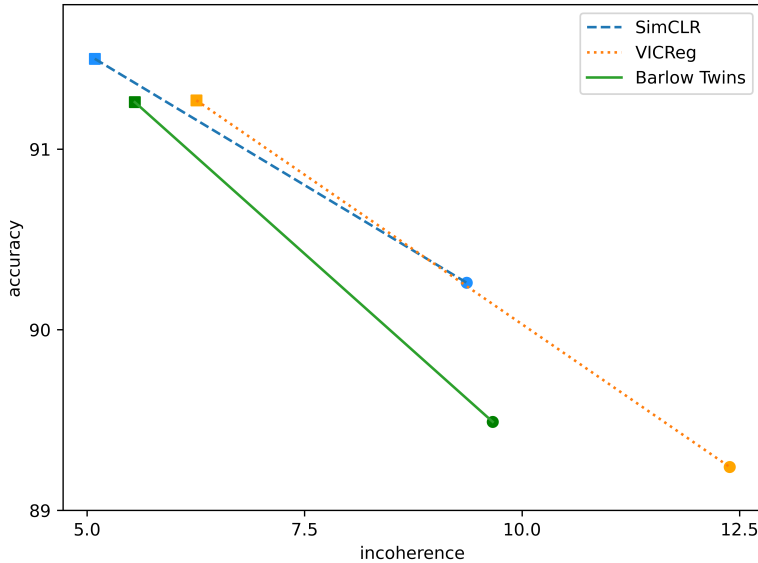


Figure 1: Incoherence across methods. We train ResNet18 backbone with a 3-layer projection head against three common SSL objectives: SimCLR, VICReg and BarlowTwins. Each line represents a method in the coordinate plane with incoherence as x-axis and accuracy on downstream task as y-axis. Top *square* marker denotes the incoherence of the backbone outputs (representations) and lower *circle* — incoherence of the projection head outputs (embeddings). While all methods enjoy better performance on the downstream task when representations are used rather than embeddings, i.e. squares are higher than circles, this correlates with incoherence, i.e. squares are to the left of circles.