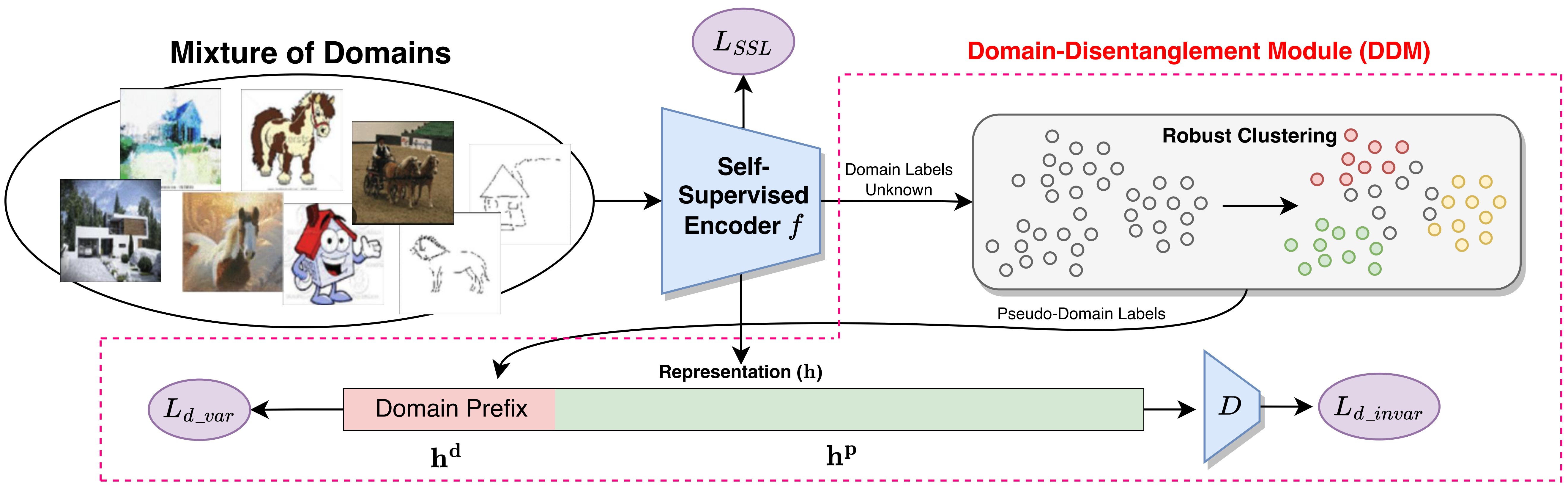
BMVC 2023

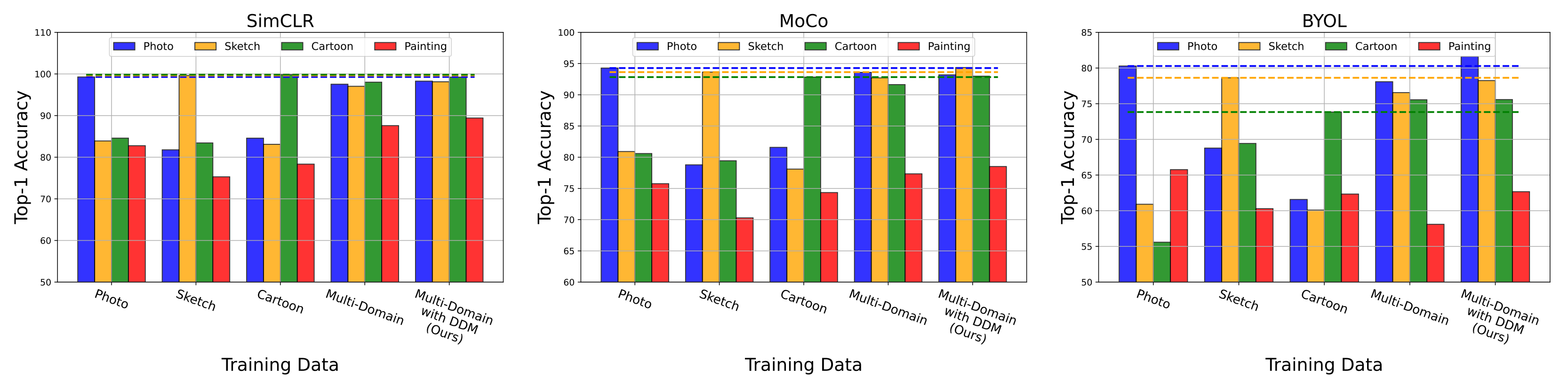
Adapting Self-Supervised Representations to Multi-Domain Setups

Neha Kalibhat, Sam Sharpe, Jeremy Goodsitt, Bayan Bruss, Soheil Feizi



Self-Supervised Models under Multi-Domain Regimes

Current state-of-the-art self-supervised approaches, are effective when trained on individual domains but show limited generalization on unseen domains. We observe that these models poorly generalize even when trained on a mixture of domains, making them unsuitable to be deployed under diverse real-world setups.



A Closer Look at Representations

The natural clustering of classes among SimCLR representations disappears when we move from CIFAR-10 to Colored-CIFAR. There is almost no overlap between the most activating features of each class between the red and green domains. The domain information (color) and instance information (actual content of the image) are somewhat interleaved in these representations, causing different sets of features to be strongly activated for the same class based on the domain.

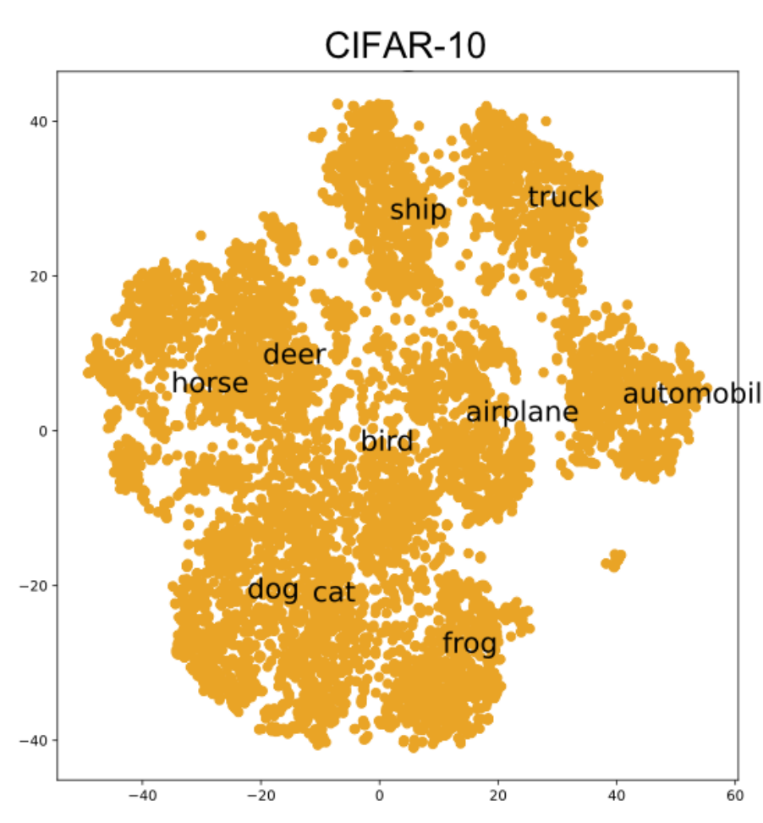


Figure 1. Accuracy: 90.18

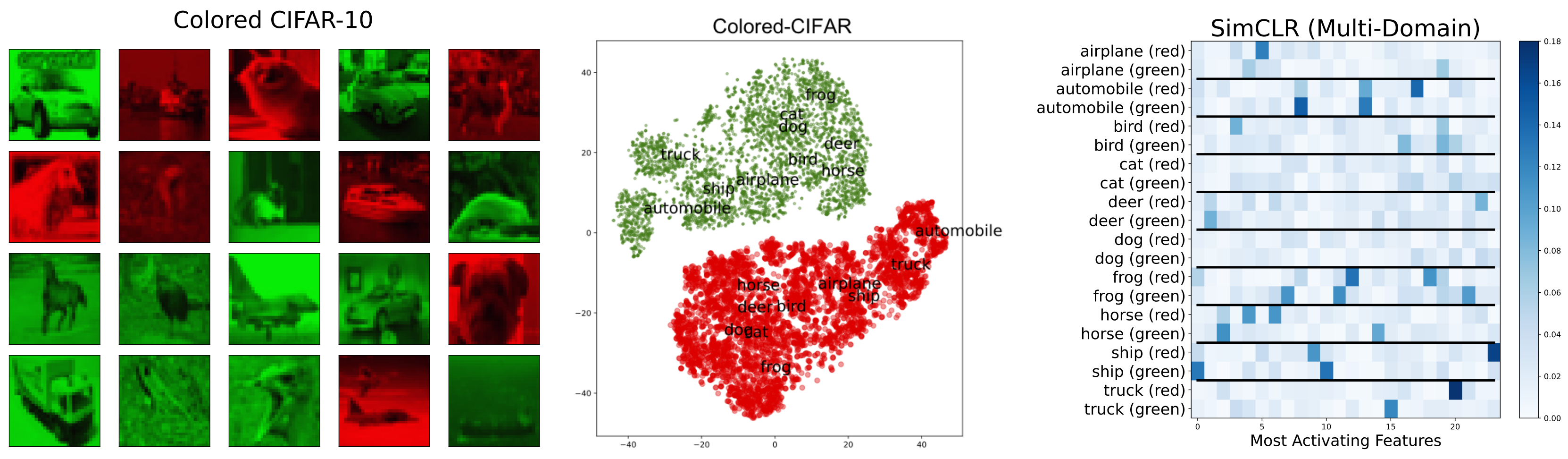


Figure 2. Accuracy: 78.52

Domain Disentanglement Module for Self-Supervised Representations

Let $\mathbf{h}_i \in \mathbb{R}^r$ denote the i^{th} representation with domain y_i. We call $\mathbf{h}_i^d$ ($\mathbf{h}_{i,0..k}$) as the *domain-variant* portion and $\mathbf{h}_i^p$ ($\mathbf{h}_{i,k..r}$) as the *domain-invariant* portion. We train the domain prefix of the i^{th} sample according to the following contrastive optimization, $L_{i_{d\_var}} = \log \frac{\sum_{j=1}^{2N} \mathbb{1}_{j \neq i} \mathbb{1}_{y_i = y_j} sim(\mathbf{h}_i^d, \mathbf{h}_j^d)}{\sum_{j=1}^{2N} \mathbb{1}_{y_i \neq y_j} sim(\mathbf{h}_i^d, \mathbf{h}_j^d)}$.

It should not be possible to predict the the domain label y_i from the representation $\mathbf{h}_i^p$ therefore, we pass each $\mathbf{h}_i^p$ through a domain discriminator $D(.)$ and minimize the Wasserstein distance, $L_{i_{d\_invar}} = D(\mathbf{h}_i^p, y_i) - D(\mathbf{h}_i^p, y_{rand})$, where $y_{rand} \sim \mathbb{P}(y)$. The final optimization for the encoder ($f(.)$) and the discriminator ($D(.)$) is,

$$\max_f \sum_{i=1}^{2N} \left[\lambda L_{i_{ssl}} + L_{i_{d\_var}} + L_{i_{d\_invar}}\right]$$

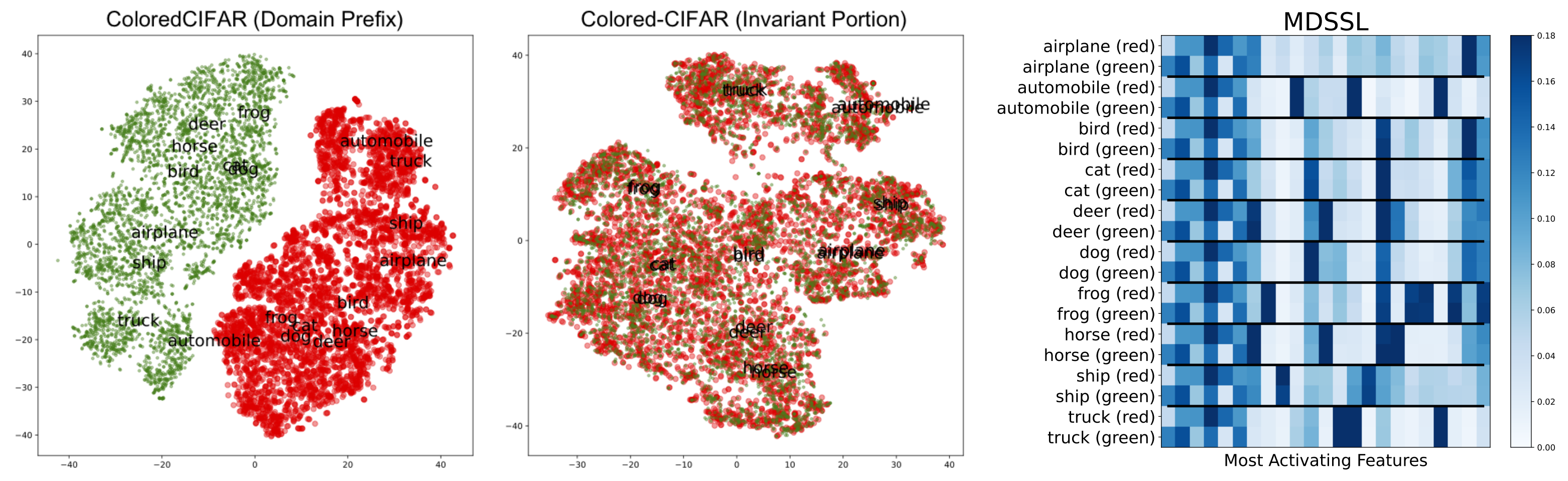


Figure 3. Accuracy: 87.06

Self-Supervised Baselines Trained with DDM

Model	Top-1 Accuracy (Baseline / with DDM) Photo	Sketch	Cartoon	Painting (Unseen)	Average
SimCLR	97.54 / **98.28**	**98.12** / 97.04	98.03 / **99.24**	87.59 / **89.42**	95.32 / **96.00**
MoCo	**93.59** / 93.19	92.71 / **94.36**	91.63 / **92.98**	77.34 / **78.51**	88.81 / **89.76**
BYOL	78.08 / **81.61**	76.55 / **78.24**	75.55 / **75.58**	58.10 / **62.67**	72.07 / **74.53**
DINO	93.67 / **95.25**	94.33 / **96.42**	79.44 / **81.77**	72.12 / **74.43**	85.89 / **86.97**
SimSiam	83.68 / **84.71**	80.97 / **85.44**	**93.75** / 92.59	57.98 / **64.09**	79.09 / **81.71**
Barlow Twins	**85.09** / 83.94	85.44 / **88.07**	92.0 / **92.83**	59.01 / **62.67**	80.39 / **81.89**

Table 1. SSL baselines trained on PACS (Photo, Sketch and Cartoon) with DDM

Model	Top-1 Accuracy (Baseline / with DDM) Painting	Real	Sketch	Clipart (Unseen)	Infograph (Unseen)	Quickdraw (Unseen)	Average
SimCLR	74.49 / **75.99**	79.31 / **82.02**	85.86 / **86.26**	68.60 / **70.48**	34.75 / **39.25**	22.98 / **24.38**	60.99 / **63.06**
MoCo	70.20 / **73.08**	**89.79** / 86.37	86.66 / **88.15**	65.10 / **68.91**	34.56 / **34.75**	19.89 / **22.12**	61.03 / **62.23**
BYOL	56.87 / **59.82**	77.60 / **79.67**	71.43 / **75.21**	50.67 / **55.86**	27.4 / **30.68**	19.33 / **22.85**	50.55 / **54.02**
DINO	**79.53** / 79.11	86.46 / **86.88**	75.8 / **76.50**	66.32 / **73.76**	30.83 / **32.12**	27.71 / **29.08**	61.11 / **62.90**
SimSiam	77.55 / **78.78**	82.02 / **85.88**	86.52 / **88.38**	67.43 / **71.53**	27.03 / **30.56**	22.29 / **25.67**	60.47 / **63.47**
Barlow Twins	56.78 / **61.18**	79.06 / **80.16**	71.56 / **73.90**	60.40 / **64.33**	26.11 / **28.82**	18.67 / **21.70**	52.09 / **55.01**

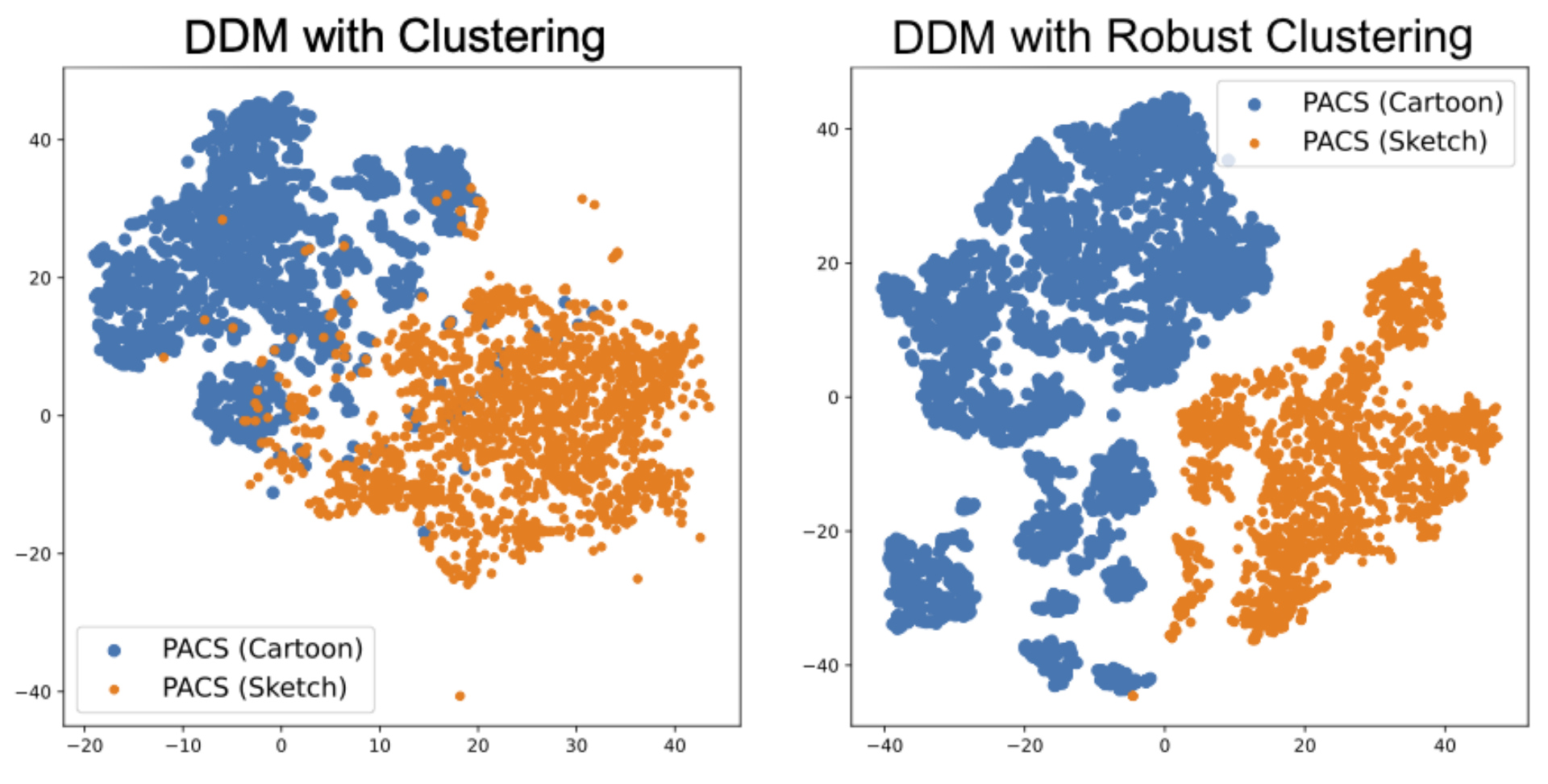
Table 2. SSL baselines trained on DomainNet (Painting, Real and Sketch) with DDM

DDM with Robust Clustering

When domain labels are not available, we discover M clusters (K-Means after warmup) with centroids $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M$ and assign pseudo-domain-labels to each sample, provided the sample is not an outlier,

$$\max\left\{\frac{\|\mathbf{h}_i - \mathbf{c}_m\|^2}{\|\mathbf{h}_i - \mathbf{c}_n\|^2} : 1 \leq m \leq M, 1 \leq n \leq M\right\} > 1 + \epsilon$$

We repeat clustering at regular intervals of training decaying ϵ from 1 to 0 as more and more samples become inliers to the discovered clusters.



Model	Top-1 Accuracy (Baseline / with DDM and robust clustering) CIFAR-10	STL-10	CIFAR-100	Tiny-ImageNet (Unseen)	Average
SimCLR	89.43 / **90.03**	79.77 / **81.01**	63.33 / **64.90**	49.58 / **51.22**	70.53 / **71.79**
MoCo	**90.80** / 90.69	80.02 / **81.60**	61.57 / **64.28**	37.16 / **39.55**	67.38 / **69.03**
BYOL	88.31 / **89.68**	75.07 / **75.72**	64.82 / **65.56**	50.04 / **51.10**	69.56 / **70.52**
DINO	90.61 / **92.96**	**84.7** / 82.35	62.63 / **63.57**	49.52 / **52.46**	71.87 / **72.84**
SimSiam	87.02 / **87.38**	72.15 / **73.78**	**62.08** / 61.90	33.11 / **34.78**	63.59 / **64.46**
Barlow Twins	88.31 / **89.01**	75.59 / **76.11**	65.03 / **66.89**	40.27 / **41.31**	67.30 / **68.33**

Table 3. SSL baselines trained on a mixture of CIFAR-10, STL-10 and CIFAR-100 using DDM and robust clustering

University of Maryland, College Park
BMVC 2023
nehamk@umd.edu