Semi-supervised learning made simple with self-supervised clustering

CVPR 2023 - TUE-AM-303







Fini Enrico*, Astolfi Pietro*, Alahari Karteek, Alameda-Pineda Xavier, Mairal Julien, Nabi Moon, Ricci Elisa

University of Trento, INRIA, SAP AI research

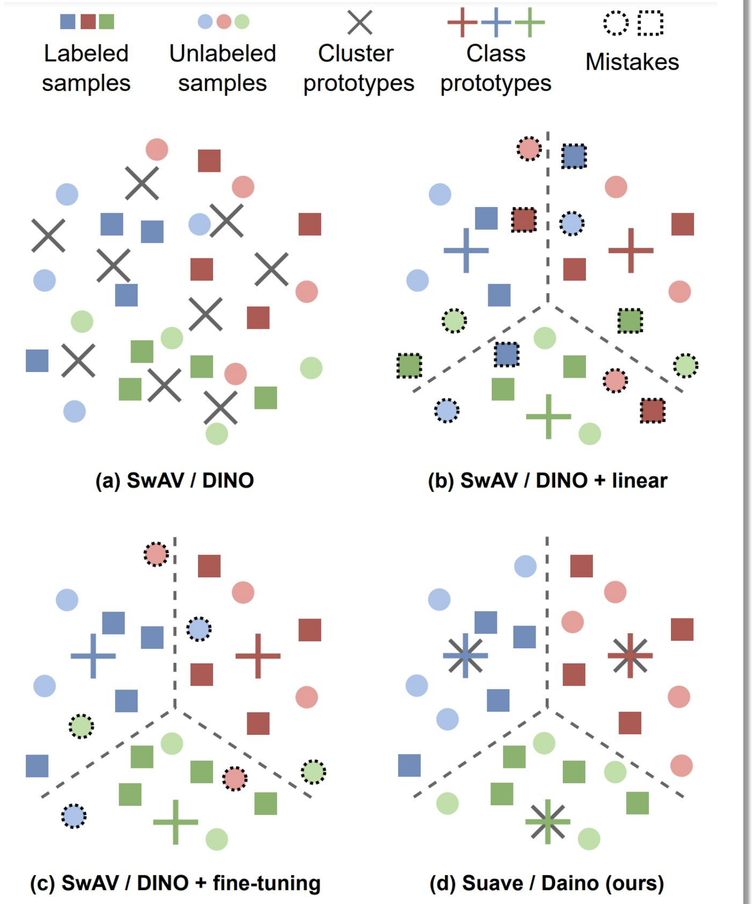
Semi-supervised learning made simple with self-supervised clustering

Enrico Fini*[1] Pietro Astolfi*[1,2] Karteek Alahari[2] Xavier Alameda-Pineda[2] Julien Mairal[2] Moin Nabi[3] Elisa Ricci[1,4]

[1] University of Trento [2] Inria [3] SAP AI Research [4] Fondazione Bruno Kessler

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Self-supervised → Semi-supervised



Simple multi-tasking: supervised learning + self-supervised clustering

[Background] Clustering-based self-supervision (e.g. **SwAV** & **DINO**): contrast predicted cluster assignments of correlated views of an image. **Cluster prototypes** are learned without supervision and might not be aligned with the actual classes.

[Framework] Our semi-supervised learning framework: use a small labeled set to encourage prototypes to align with the underlying classes. In practice, we **multi-task** a **supervised cross-entropy loss** with a **clustering-based self-supervised loss**.

[Methods] Two clustering-based semi-supervised models fit into our framework:

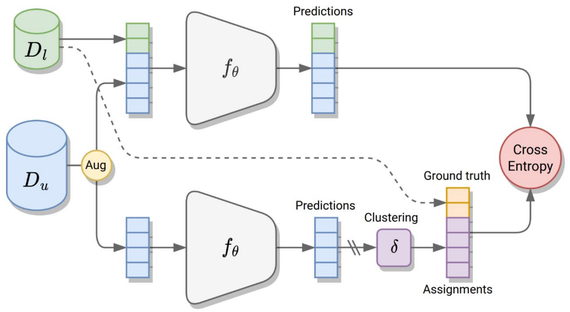
- **Suave** (inspired by SwAV), uses Sinkhorn-Knopp to equipartition the batch
- **Daino** (inspired by DINO), uses a momentum encoder, centering and sharpening

$$\ell(\mathbf{s}, \bar{\mathbf{y}}), \quad \bar{\mathbf{y}} = \begin{cases} \mathbf{y}, & \mathbf{x} \in \mathcal{D}_l \\ \hat{\mathbf{y}}, & \mathbf{x} \in \mathcal{D}_u \end{cases}$$

$$\hat{\mathbf{y}} = \delta(\hat{\mathbf{p}}, \mathbf{c}, \epsilon)$$

Prediction Context Temperature

Architecture overview & pseudo-code



Pseudo-code for Suave and Daino.

```
# x_l, x_u: labeled and unlabeled samples
# label: one-hot ground truth for x_l
def training_step(x_l, x_u, label):

    # correlated views
    x1_u, x2_u = aug(x_u), aug(x_u)

    # forward pass
    p_l = f(x_l)
    p1_u, p2_u = f(x1_u), f(x2_u)

    # compute assignments via clustering
    assign = delta(p2_u.detach())

    # concatenate predictions and targets
    pred = cat(p_l, p1_u)
    target = cat(label, assign)

    # compute loss
    loss = cross_entropy(pred, target)
    return loss
```

Results

Suave and Daino match or outperform SOTA, despite being simpler.

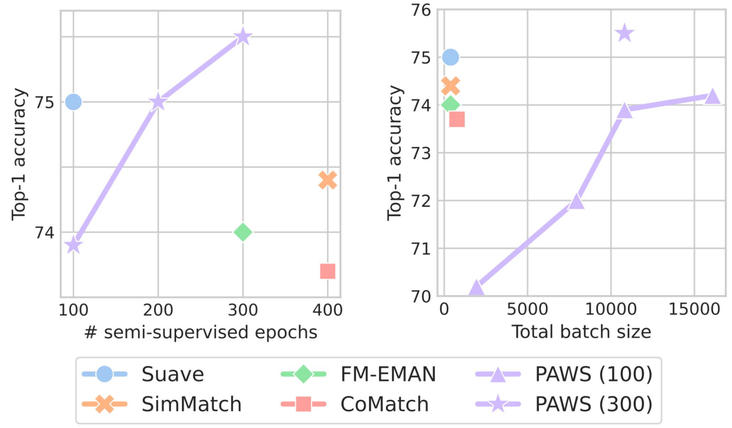
Table 3. Comparison with the state-of-the-art on ImageNet-1k. All the models reported use ResNet-50. In the first and the second column, we indicate within brackets whether a model is initialized from a self-supervised checkpoint and the number of epochs

Method	Epochs	Batch size		Acc@1	
		Unlab.	Lab.	10%	1%
with similar batch size and number of epochs					
S^4L-Rotation [72]	200†	256	256	61.4	-
FM-DA (MoCo v2) [44]	(800) 400	640	160	72.2	59.9
PAWS [6]	100	256	1680	70.2	-
CoMatch (MoCo v2) [44]	(800) 400	640	160	73.7	67.1
FM-EMAN (MoCo-EMAN) [13]	(800) 300	320	64	74.0	63.0
SimMatch [76]	400	320*	64*	74.4	**67.2**
DebiasPL (MoCo-EMAN) [62]	(800) 50	640	128	-	65.3
	(800) 200	640	128	-	66.5
Suave	(100) 100	256	128	73.6	63.8
	(200) 100	256	128	74.3	65.0
	(800) 100	256	128	**75.0**	66.2

Table 1. Comparison with the state-of-the-art on CIFAR100.

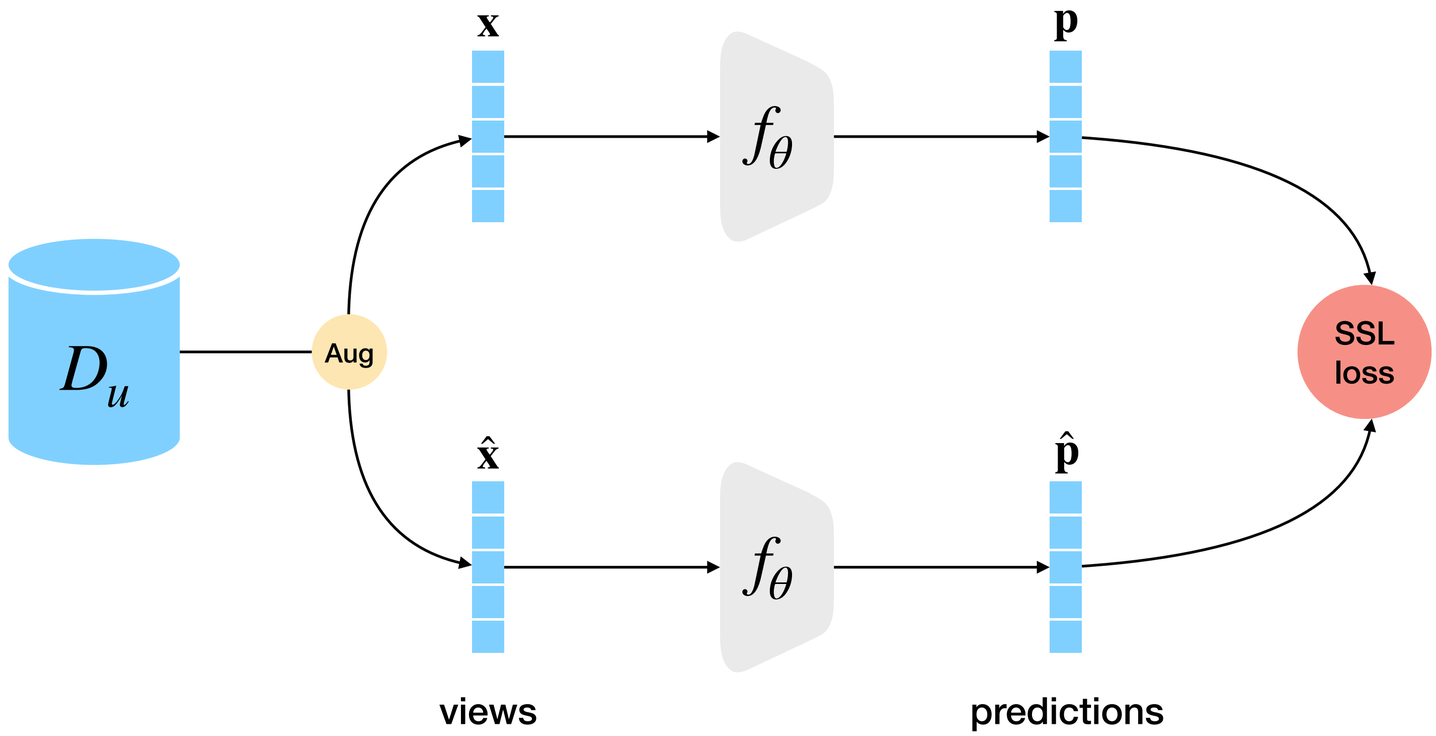
Method	Acc@1		
	400	2500	10000
Π-Model [42]	-	42.8	62.1
Mean Teacher [57]	-	46.1	64.2
MixMatch [12]	32.4	60.2	72.2
UDA [64]	53.6	72.3	77.5
ReMixMatch [11]	55.7	72.6	77.0
FixMatch [55]	50.1	71.4	76.8
Dash [66]	55.2	72.8	78.0
CoMatch [44]	60.0	73.0	78.2
Meta Pseudo Labels [50]	55.8	72.3	77.5
FlexMatch [73]	60.1	73.5	78.1
FixMatch+DM [46]	59.8	74.1	79.6
NP-Match [60]	61.1	74.0	78.8
ConMatch [40]	61.1	74.6	-
SimMatch [76]	62.2	74.9	79.4
CCSSL [67]	61.2	75.7	80.1
Daino	61.1	75.2	79.2
Suave	**64.6**	**77.0**	**81.6**

Suave and Daino are efficient and work well with small batch sizes



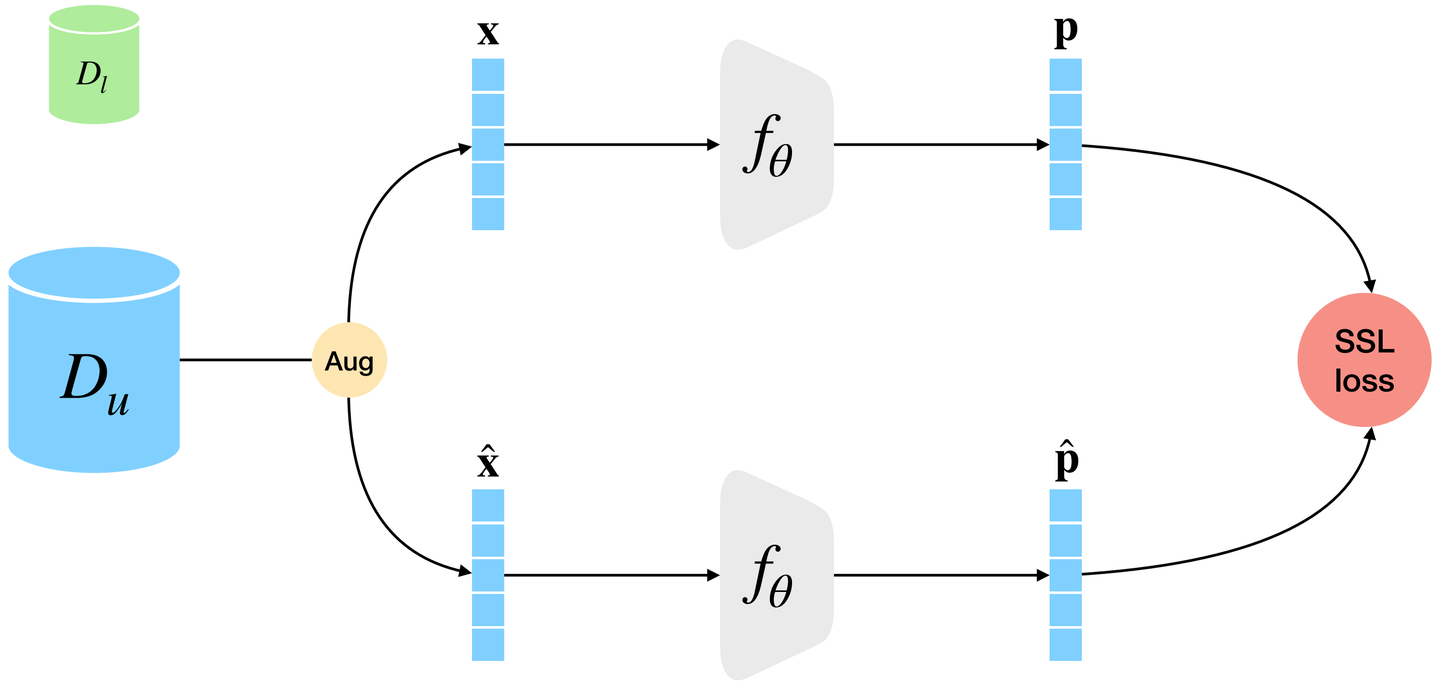
Unsupervised learning

With self-supervised models



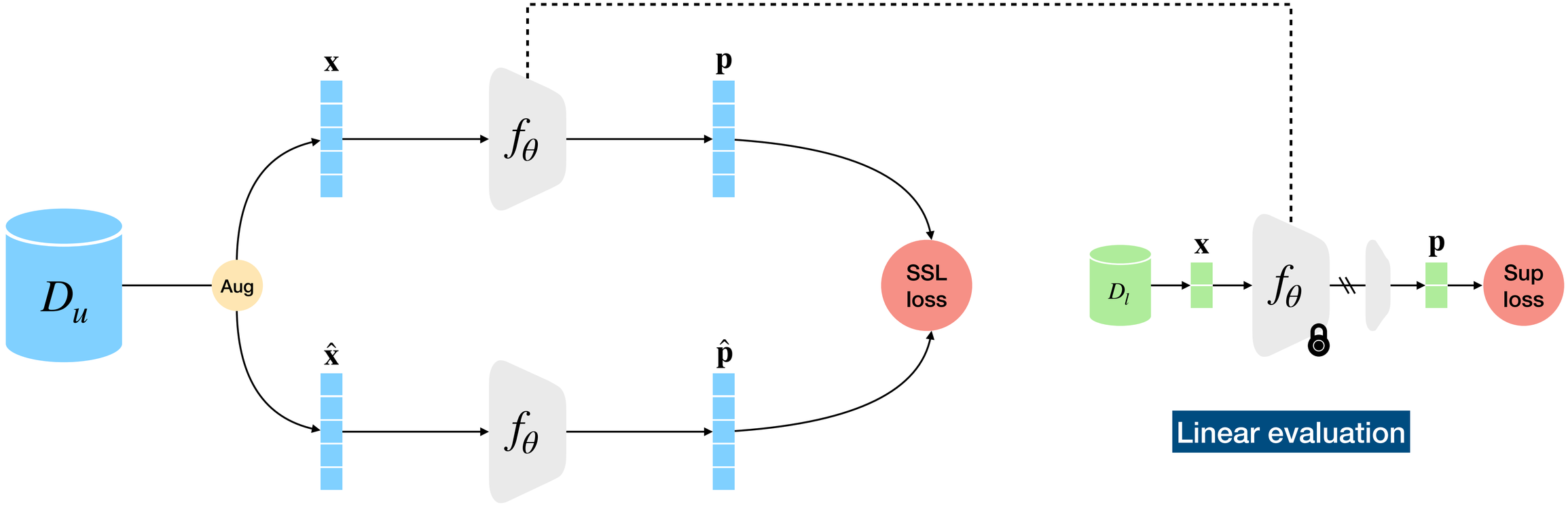
Semi-supervised learning

With pre-trained self-supervised models



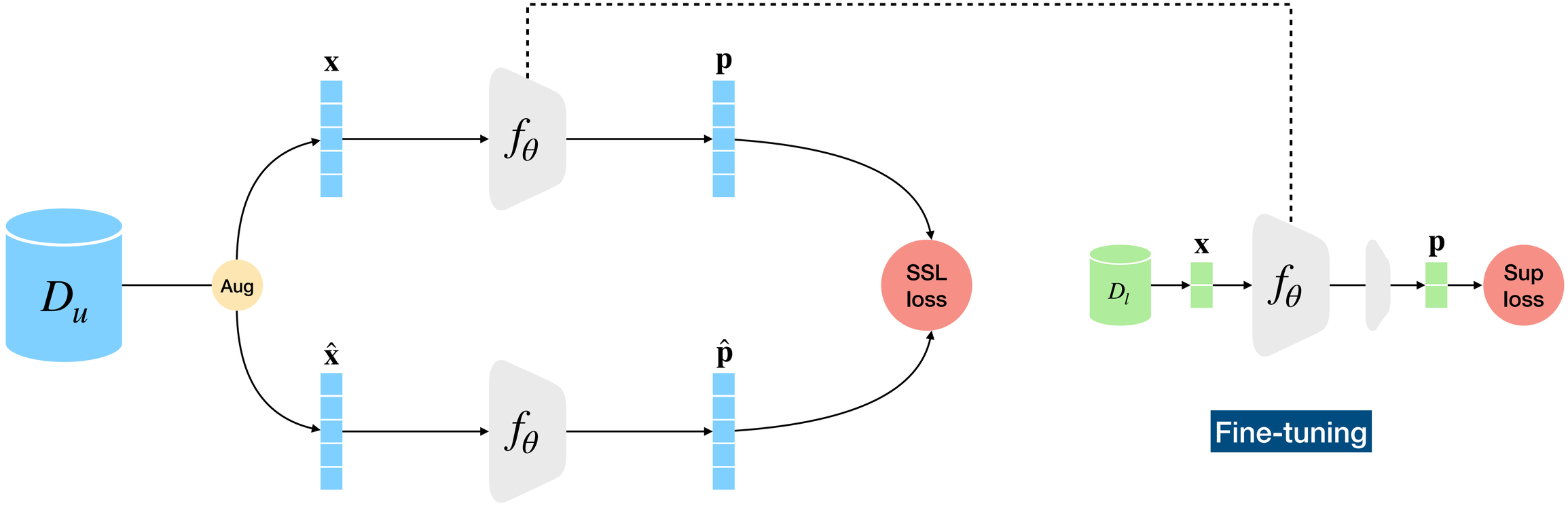
Semi-supervised learning

With pre-trained self-supervised models



Semi-supervised learning

With pre-trained self-supervised models

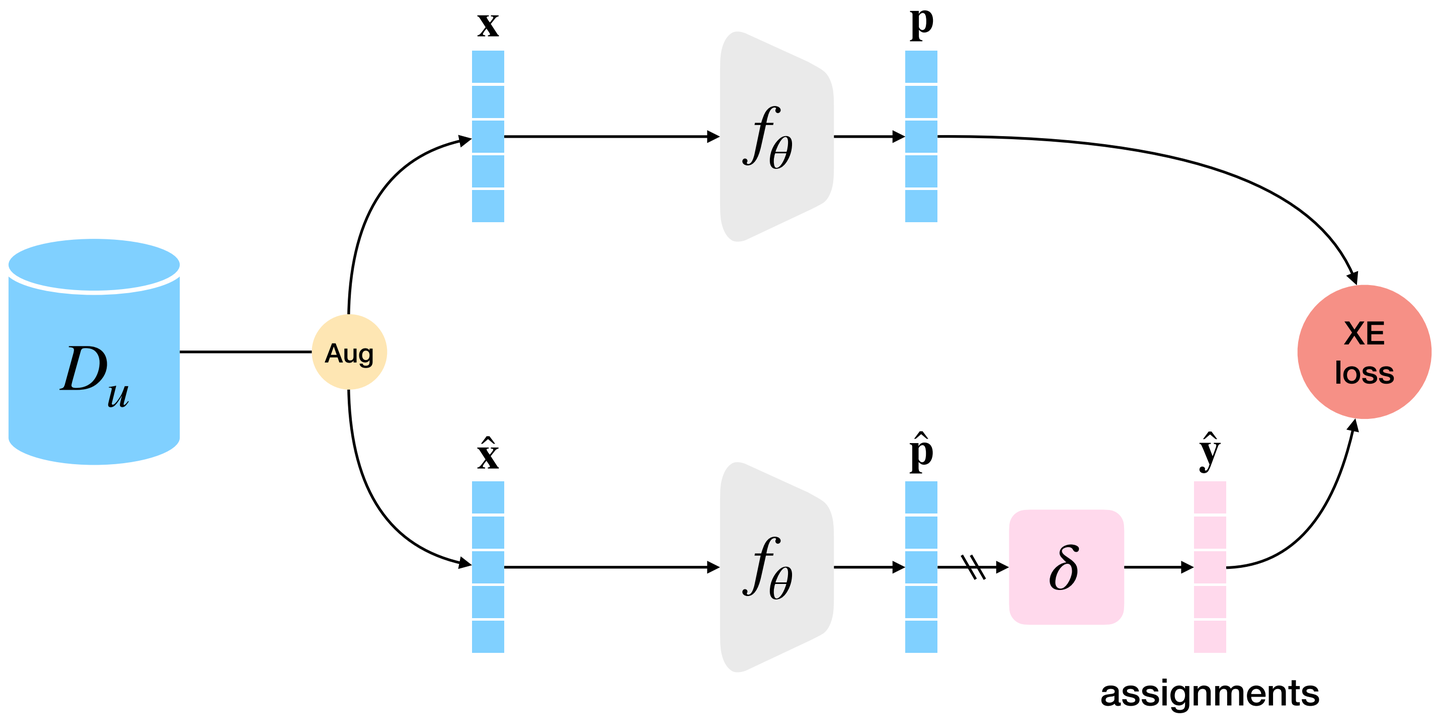


Self-supervised → Semi-supervised

S4L - Zhai et al (2019); PAWS - Assran et al (2021)

Clustering-based

Self-supervised



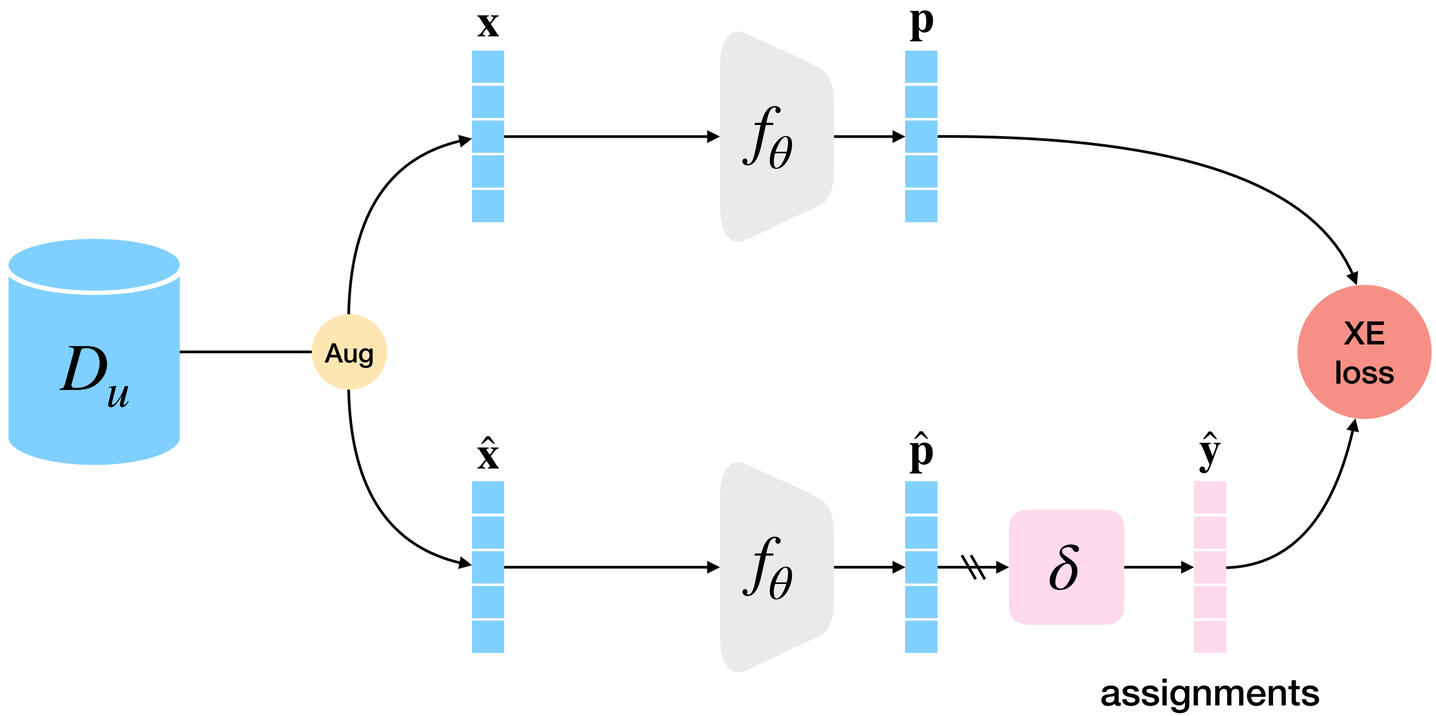
Clustering

$$\hat{\mathbf{y}} = \delta(\hat{\mathbf{p}}, \mathbf{c}, \epsilon)$$

context temperature

Clustering-based

Self-supervised



Clustering

$$\hat{\mathbf{y}} = \delta(\hat{\mathbf{p}}, \mathbf{c}, \epsilon)$$

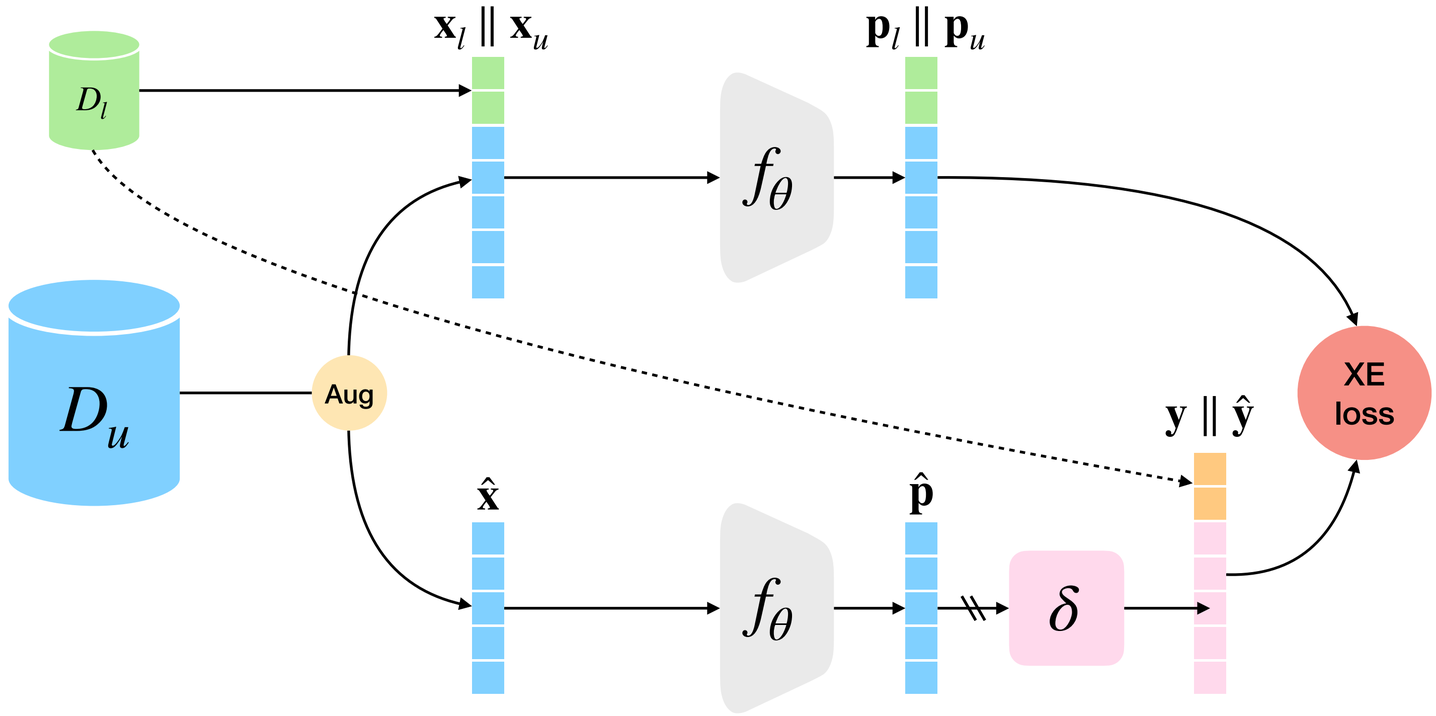
context temperature

XE loss

$$\ell(\mathbf{p}, \hat{\mathbf{y}}) = -\sum_{k=1}^{K} \hat{\mathbf{y}}_k \log\left(\sigma(\mathbf{p}/\tau)_k\right)$$

Clustering-based

Semi-supervised

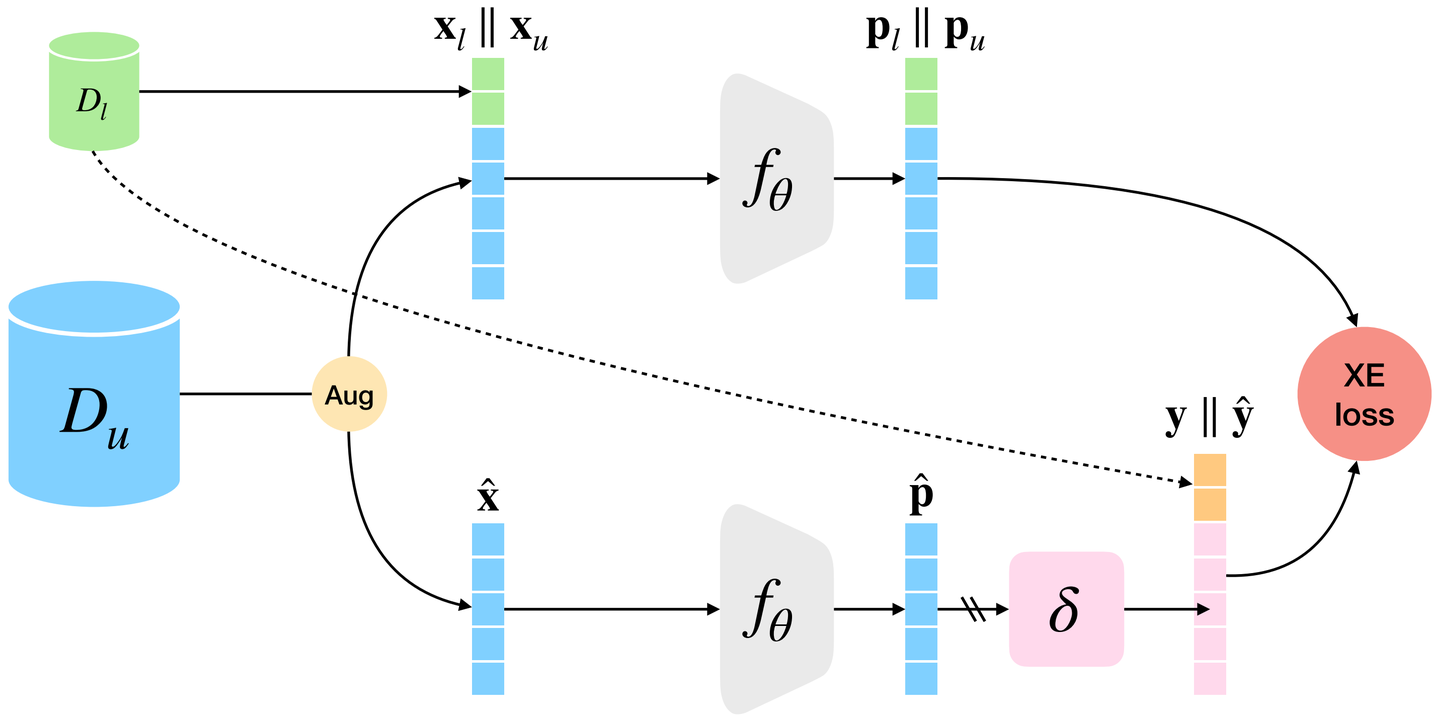


XE loss

$$\ell(\mathbf{p}, \bar{\mathbf{y}}), \quad \bar{\mathbf{y}} = \begin{cases} \mathbf{y}, & \mathbf{x} \in D_l \\ \hat{\mathbf{y}}, & \mathbf{x} \in D_u \end{cases}$$

Clustering-based

Semi-supervised



XE loss

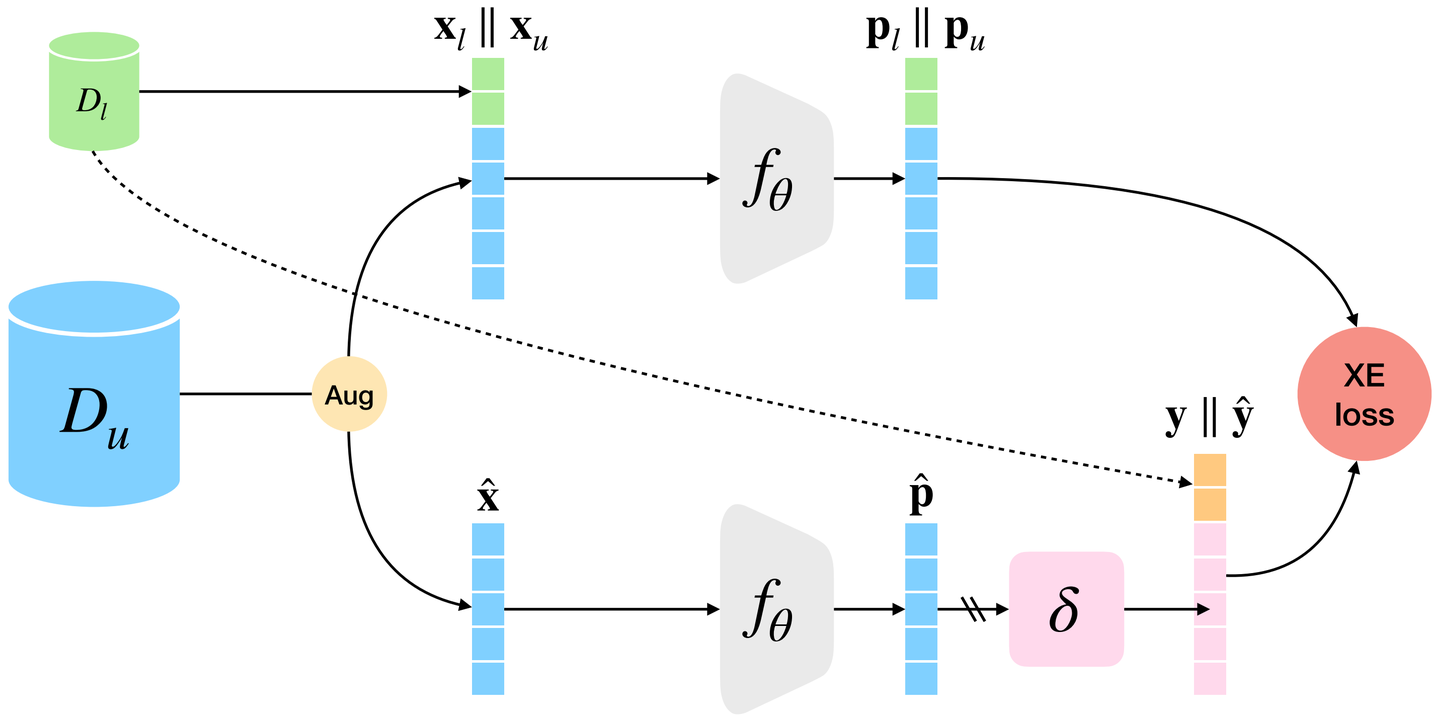
$$\ell(\mathbf{p}, \bar{\mathbf{y}}), \quad \bar{\mathbf{y}} = \begin{cases} \mathbf{y}, & \mathbf{x} \in D_l \\ \hat{\mathbf{y}}, & \mathbf{x} \in D_u \end{cases}$$

Instantiations

Suave (SwAV) and Daino (DINO)

Suave

Semi-supervised SwAV



Clustering

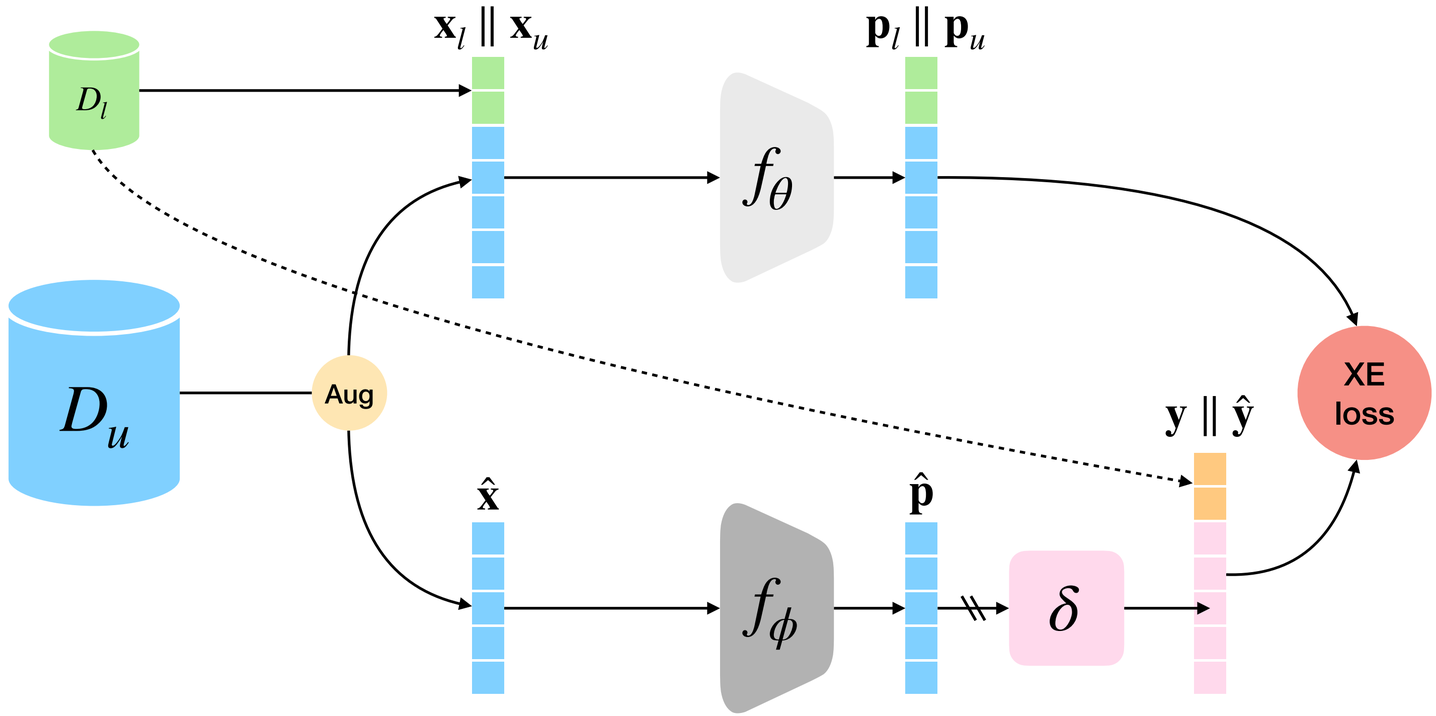
$$\hat{\mathbf{y}}_1 = \delta(\hat{\mathbf{p}}_1, \mathbf{c}, \epsilon)$$

$$[\hat{\mathbf{p}}_2, \ldots, \hat{\mathbf{p}}_B]$$

$$\hat{\mathbf{Y}} = \max_{\mathbf{Y} \in \Gamma} \mathrm{Tr}(\mathbf{Y}\hat{\mathbf{P}}) + \epsilon\, \mathrm{H}(\mathbf{Y})$$

Daino

Semi-supervised DINO



Clustering

$$\hat{\mathbf{y}} = \delta(\hat{\mathbf{p}}, \mathbf{c}, \epsilon)$$

$$\gamma \leftarrow \mu\gamma + (1-\mu)\frac{1}{B}\sum_{b=0}^{B}\hat{\mathbf{p}}_b$$

$$\hat{\mathbf{y}} = \sigma\left(\frac{\hat{\mathbf{p}} - \gamma}{\epsilon}\right)$$

CIFAR100

Experiments

Method	Acc@1		
	400	2500	10000
Π-Model [42]	-	42.8	62.1
Mean Teacher [57]	-	46.1	64.2
MixMatch [12]	32.4	60.2	72.2
UDA [64]	53.6	72.3	77.5
ReMixMatch [11]	55.7	72.6	77.0
FixMatch [55]	50.1	71.4	76.8
Dash [66]	55.2	72.8	78.0
CoMatch [44]	60.0	73.0	78.2
Meta Pseudo Labels [50]	55.8	72.3	77.5
FlexMatch [73]	60.1	73.5	78.1
FixMatch+DM [46]	59.8	74.1	79.6
NP-Match [60]	61.1	74.0	78.8
ConMatch [40]	61.1	74.6	-
SimMatch [76]	62.2	74.9	79.4
CCSSL [67]	61.2	75.7	80.1
Daino	61.1	75.2	79.2
Suave	**64.6**	**77.0**	**81.6**

400 (0.8%), 2500 (5%), 1000 (20%)

CIFAR100

Experiments

Method	Acc@1		
	400	2500	10000
Π-Model [42]	-	42.8	62.1
Mean Teacher [57]	-	46.1	64.2
MixMatch [12]	32.4	60.2	72.2
UDA [64]	53.6	72.3	77.5
ReMixMatch [11]	55.7	72.6	77.0
FixMatch [55]	50.1	71.4	76.8
Dash [66]	55.2	72.8	78.0
CoMatch [44]	60.0	73.0	78.2
Meta Pseudo Labels [50]	55.8	72.3	77.5
FlexMatch [73]	60.1	73.5	78.1
FixMatch+DM [46]	59.8	74.1	79.6
NP-Match [60]	61.1	74.0	78.8
ConMatch [40]	61.1	74.6	-
SimMatch [76]	62.2	74.9	79.4
CCSSL [67]	61.2	75.7	80.1
Daino	61.1	75.2	79.2
Suave	**64.6**	**77.0**	**81.6**

400 (0.8%), 2500 (5%), 1000 (20%)

SwAV and DINO w/ 100% labels 64.9% and 66.8%

ImageNet

Experiments

Suave

Method	Epochs	Batch size		Acc@1	
		Unlab.	Lab.	10%	1%
with similar batch size and number of epochs					
S^4L-Rotation [72]	200†	256	256	61.4	-
FM-DA (MoCo v2) [44]	(800) 400	640	160	72.2	59.9
PAWS [6]	100	256	1680	70.2	-
CoMatch (MoCo v2) [44]	(800) 400	640	160	73.7	67.1
FM-EMAN (MoCo-EMAN) [13]	(800) 300	320	64	74.0	63.0
SimMatch [76]	400	320*	64*	74.4	**67.2**
DebiasPL (MoCo-EMAN) [62]	(800) 50	640	128	-	65.3
	(800) 200	640	128	-	66.5
Suave	(100) 100	256	128	73.6	63.8
	(200) 100	256	128	74.3	65.0
	(800) 100	256	128	**75.0**	66.2
with larger batch size or number of epochs					
UDA [55]	∼480	15360	512	68.8	-
Meta Pseudo Labels [50]	∼800	2048	2048	73.9	-
PAWS [6]	100	4096	6720	73.9	63.8
	200	4096	6720	75.0	66.1
	300	4096	6720	75.5	66.5
DebiasPL (MoCo-EMAN) [62]	(800) 300	1280	256	-	67.1

Daino

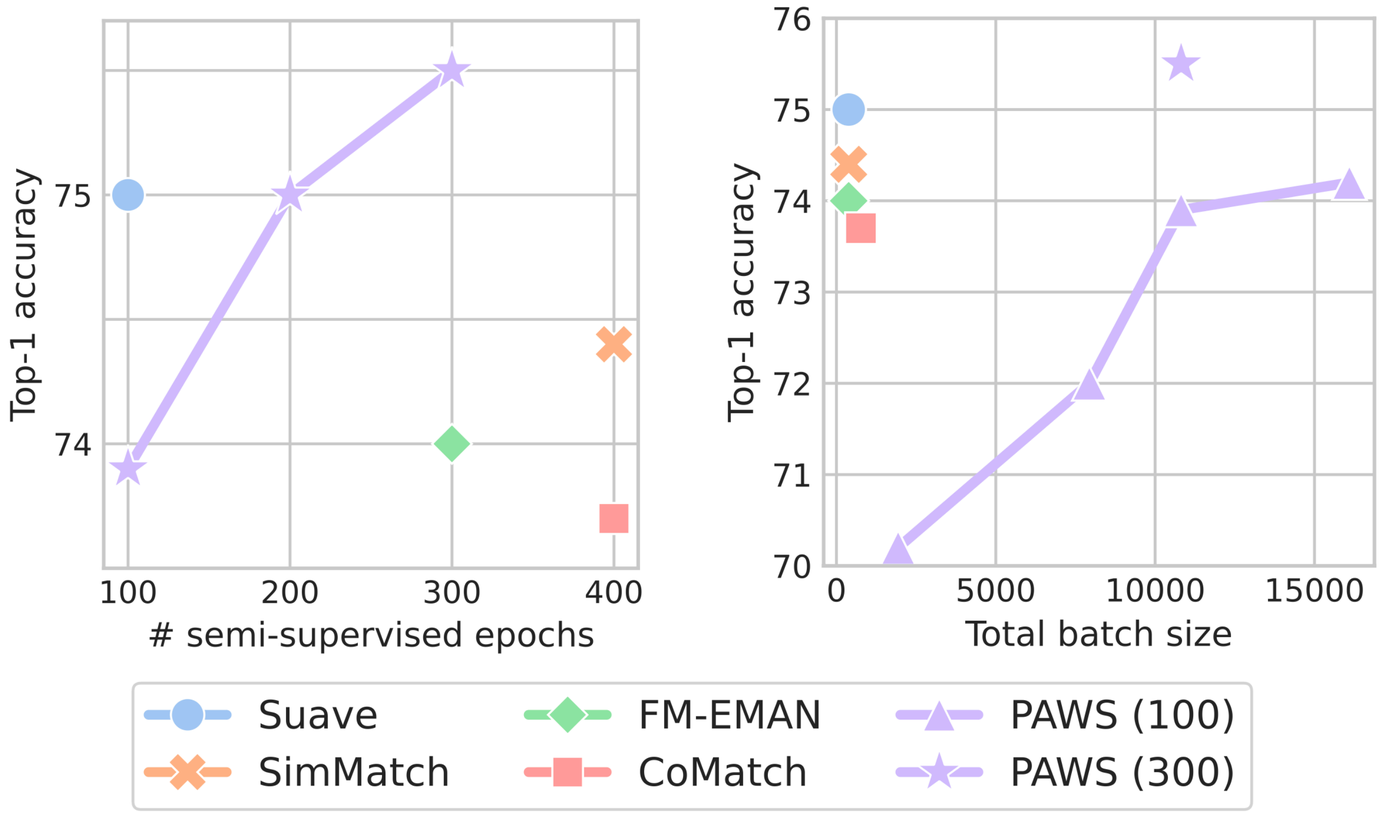
Method	Epochs	Batch size		Acc@1	
		Unlab.	Lab.	10%	1%
DINO [16]	(800)	1024	-	72.2	64.5
MSN [5]	(800)	1024	-	-	**67.2**
Daino	(800) 60	1024	512	**76.6**	67.1

Efficiency

Batch size and # of epochs

Efficiency

Batch size and # of epochs



Semi-supervised learning made simple with self-supervised clustering

CVPR 2023 - TUE-AM-303







Fini Enrico*, Astolfi Pietro*, Alahari Karteek, Alameda-Pineda Xavier, Mairal Julien, Nabi Moon, Ricci Elisa

University of Trento, INRIA, SAP AI research