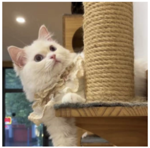
Towards Understanding Why FixMatch Generalizes Better Than Supervised Learning

Jingyang Li

National University of Singapore

Joint work with: Jiachun Pan, Vincent Y. F. Tan
Kim-Chuan Toh, Pan Zhou

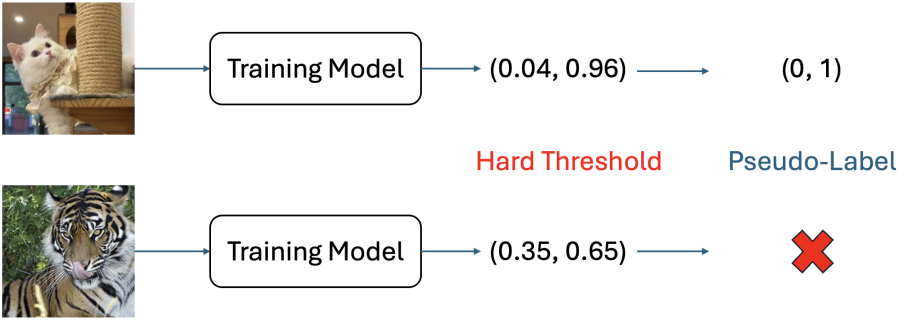
Semi-Supervised Learning

Image			
Label	Dog	Cat	None

Labeled Data

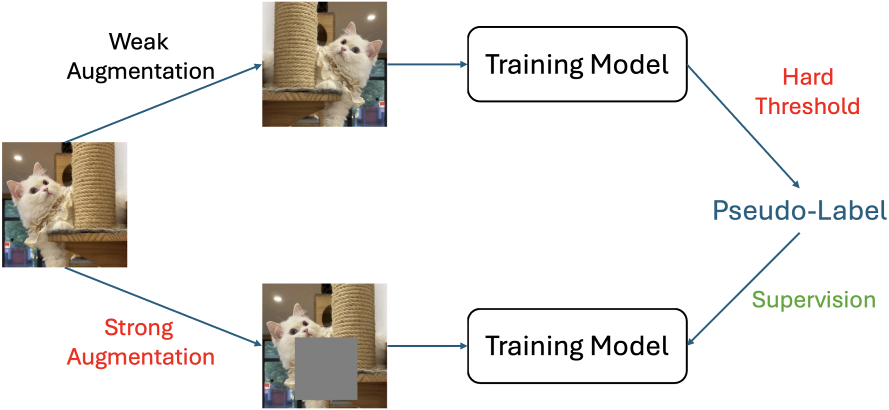
Unlabeled Data

FixMatch[1]: Pseudo-Label



[1] Kihyuk Sohn et al. "FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 596–608.

FixMatch: Consistency Regularization



Theoretical Question

Why FixMatch Generalizes Better Than Supervised Learning?

A Feature Learning Perspective

Problem Setup

Consider a k-class classification problem, we assume each class $i \in [k]$ has two semantic features, $v_{i,1}$ and $v_{i,2}$, capable of independently ensuring correct classification.

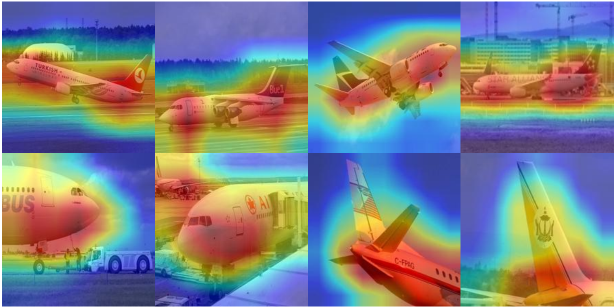


Figure: Visualization of pretrained ResNet-50 using Grad-CAM on airplane images from ImageNet.

Data Distribution

We consider the multi-view data assumption:

- A **multi-view image** $(X, y) \sim \mathcal{D}_m$ has patches with two semantic features $v_{y,1}$ and $v_{y,2}$ plus some noises.
- A **single-view image** $(X, y) \sim \mathcal{D}_s$ contains only one semantic feature $v_{y,1}$ or $v_{y,2}$ plus noises.

Supervised learning (SL), due to the "winning lottery" phenomenon, only learns one feature per class, resulting in only near 50% test accuracy on single-view images[2].

[2] Zeyuan Allen-Zhu and Yuanzhi Li. "Towards Understanding Ensemble, Knowledge Distillation and Self-Distillation in Deep Learning". In: *The Eleventh International Conference on Learning Representations*. 2023.

Feature Learning of FixMatch

FixMatch's hard threshold divides its entire training process into two phases:

- **Phase I**: The network relies primarily on labeled data, as it cannot yet generate confident pseudo-labels.
- **Phase II**: Having learned one feature per class, the network can generate confident pseudo-labels for multi-view and single-view images containing the learned feature. Unlabeled data is involved and dominates the training loss due to its volume.

Can Phase II helps feature learning?

Feature Learning in Phase II

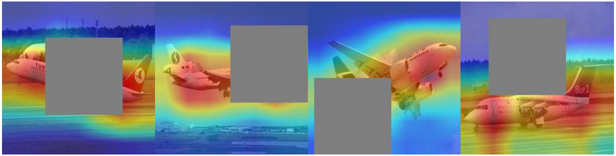
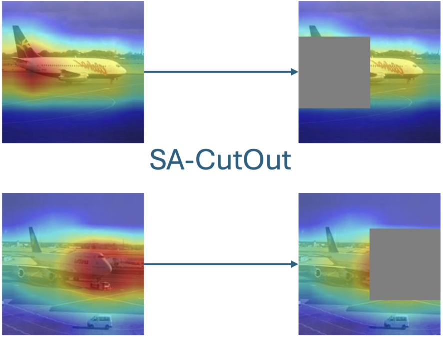


Figure: The effect of strong augmentation on images.

Strong augmentation has probability π to **remove partial semantic feature**, which generates $\frac{\pi}{2} \cdot N_{u,m}$ samples containing only the missed features. This portion of samples dominate the training loss and force the network to learn missed features in Phase II.

SA-FixMatch

Semantic-Aware CutOut (SA-CutOut) for better data efficiency.



Feature Learning

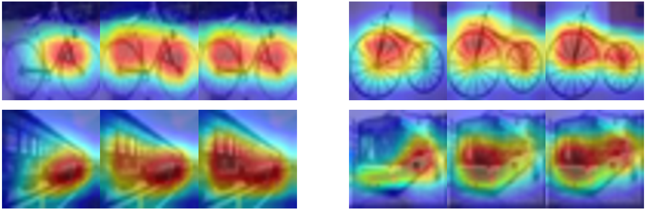


Figure: Visualization of WRN-28-8 via Grad-CAM on CIFAR-100. Each group of three images corresponds to models trained with SL (left), FixMatch (middle), and SA-FixMatch (right).

Test Accuracy

Table: Comparison of Test Accuracy (%) using entire training dataset as unlabeled data.

Dataset	STL-10			ImageNet
Label Amount	40	250	1000	100K
SL	23.61 $\pm$ 1.62	38.83 $\pm$ 1.12	64.08 $\pm$ 0.47	44.62 $\pm$ 1.16
FixMatch	70.00 $\pm$ 4.02	88.73 $\pm$ 0.92	93.45 $\pm$ 0.19	50.80 $\pm$ 0.73
SA-FixMatch	71.81 $\pm$ 4.23	89.45 $\pm$ 1.19	94.04 $\pm$ 0.19	52.18 $\pm$ 0.32

Table: Comparison of Test Accuracy (%) with the same training dataset for SSL and SL.

Dataset	STL-10			CIFAR-100			ImageNet
Data Amount	40	250	1000	400	2500	10000	100K
SL	19.93	44.06	67.29	9.87	40.98	63.48	41.82
FixMatch	38.88	64.70	79.15	18.58	47.20	67.94	43.34
SA-FixMatch	40.25	65.85	79.74	19.72	47.71	68.30	44.88

FixMatch-like SSLs

Table: Comparison of Test Accuracy (%) of FixMatch-like SSLs with CutOut and SA-CutOut on STL-10 with 40 labeled data.

Dataset	STL-10			
Algorithm	FlexMatch	FreeMatch	Dash	SoftMatch
CutOut	72.13 $\pm$ 5.66	75.29 $\pm$ 1.29	67.51 $\pm$ 1.47	78.55 $\pm$ 2.90
SA-CutOut	75.91 $\pm$ 5.59	77.91 $\pm$ 2.01	78.41 $\pm$ 1.91	84.04 $\pm$ 4.67

Thank you for your attention!

li_jingyang@u.nus.edu