Newton-coupled Dual-Teacher Semi-supervised Learning Framework

Hongyang He Xinyuan Song Yan Zhong Daizong Liu Xuanyu Liu Victor Sanchez

University of Warwick Wuhan University Peking University Manifolda.AI

Introduction

SSL usually transfers supervision from teacher to student through pseudo-labels.This signal is mostly zero-order. It tells the student what class to predict, but it does not describe the local geometry of the loss surface. As a result, the student can suffer from unstable optimization, strong loss oscillation, and limited generalization when labeled data are scarce.as shown in Figure 1.

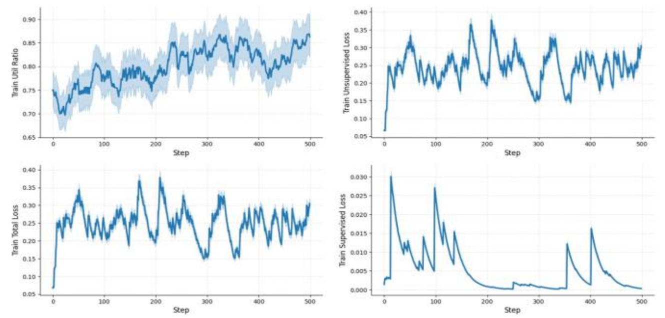

*Figure 1.*Training dynamics of FixMatch on CIFAR-100

Most existing SSL methods also rely on a single-teacher structure. This makes the student highly dependent on one representational bias. When the teacher provides noisy or sharp supervision, the student may follow unstable gradients and converge to less robust solutions.

We propose TTN, a Newton-coupled dual-teacher SSL framework.TTN uses two complementary self-supervised teachers, MAE and DINOv3, to provide structural and semantic supervision. Their pseudo-labels and local curvature cues are fused to guide the student with a second-order learning signal.

Our Method Overview

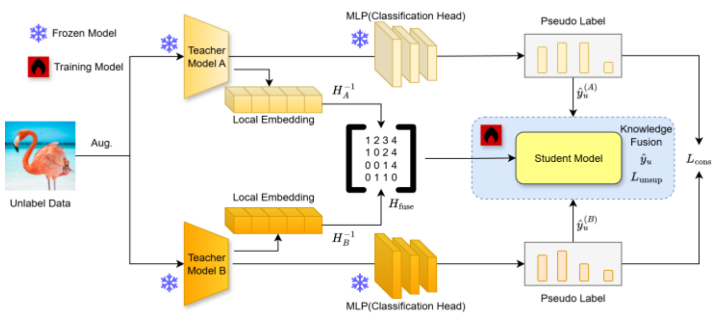

Figure 2. Overview of the proposed TTN framework. TTN contains three main components: two frozen self-supervised teachers, a confidence-aware fusion module, and a Newton-guided student optimizer.

Dual-teacher supervision.

MAE provides fine-grained patch-level structural cues, while DINOv3 provides global semantic regularity. Both teachers generate soft pseudo-labels for the same weakly augmented unlabeled sample.

Confidence-weighted pseudo-label fusion.

For each teacher, TTN computes a confidence score from the maximum softmax probability.The two predictions are then fused in the log-probability space. This forms a KL-barycenter-style soft pseudo-label that gives larger weight to the more confident teacher. ***Curvature fusion.*** TTN also fuses local Hessian information from the two teachers. The fused curvature is used as a second-order preconditioning metric. This helps suppress unstable update directions and improves optimization smoothness. ***Newton-guided student update.*** Instead of using a standard first-order update, TTN updates the student with a Newton-step style rule. The student gradient is preconditioned by the fused curvature matrix, which allows the student to follow a more stable and geometry-aware learning trajectory.

Analysis

Model	Type	Method	1%	10%	25%	Param. (M)
FixMatch (Sohn et al., 2020)	Consistency-based	Pseudo-labeling	52.6±0.32	68.7±0.28	74.9±0.25	25.0
UDA (Xie et al., 2020)	Consistency-based	Distribution alignment	51.2±0.35	67.5±0.30	73.8±0.27	25.0
FlexMatch (Zhang et al., 2021)	Confidence-aware	Adaptive threshold	53.5±0.30	70.2±0.26	75.3±0.24	25.0
Meta Pseudo Label (Pham et al., 2021)	Meta-learning	Meta pseudo-labeling	55.0±0.29	71.8±0.24	76.4±0.22	28.0
SimCLRv2+KD (Chen et al., 2020)	Self-supervised + KD	Distillation	54.5±0.31	69.7±0.27	75.5±0.24	30.0
Semi-ViT (ViT-B) (Cai et al., 2022)	Self-training	Self-labeled	74.1±0.21	81.6±0.18	84.2±0.16	86.0
Semi-ViT (ViT-L) (Cai et al., 2022)	Self-training	Self-labeled	77.3±0.20	83.3±0.17	85.1±0.15	307.0
Semi-ViT (ViT-H) (Cai et al., 2022)	Self-training	Self-labeled	78.9±0.18	84.6±0.16	86.2±0.14	632.0
REACT (ViT-L) (Liu et al., 2023)	Robust SSL	Distribution calibration	**81.6**±0.15	85.1±0.13	86.8±0.12	307.0
SemiFormer (Weng et al., 2022)	Semi-supervised ViT	Confidence teacher	75.8±0.22	82.1±0.19	84.5±0.17	86.0
DINO (ViT-L) (Caron et al., 2021)	Self-supervised	Linear head	78.1±0.19	82.9±0.17	84.9±0.15	307.0
Semi-SST (ViT-H) (Zhao et al., 2025)	Semi-supervised ViT	Self-adaptive thresholding	80.7±0.25	84.9±0.13	–	632.0
Co-Training (Rothenberger & Diochnos, 2023)	Co-training	Two-view mutual labeling	80.1±0.17	85.1±0.15	–	45.0
MCT (Rothenberger & Diochnos, 2023)	Meta Co-training	Meta feedback	80.7±0.16	85.8±0.14	–	47.0
TRiCo (He et al., 2025a)	Game-theoretic SSL	TRiCo-Training	81.2±0.14	**85.9**±0.12	88.3±0.11	95.0
TTN (ours)	Dual-teacher SSL	Newton-guided update	**81.4**±0.13	85.7±0.12	**88.5**±0.10	90.0

Table 1. Top-1 accuracy (%, mean ± std) on ImageNet

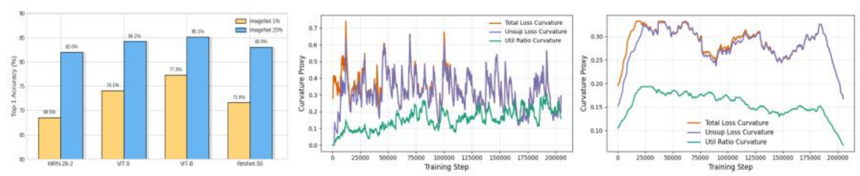

*Figure 3.*Visualization of curvature dynamics and backbone performance.

1

$$\hat{y}_u^{(A)} = T_A(x_u^w), \qquad \hat{y}_u^{(B)} = T_B(x_u^w)\,\omega_A = \max_c \hat{y}_u^{(A)}[c],$$

$$\omega_B = \max_c \hat{y}_u^{(B)}[c]\,\tilde{\omega}_A = \frac{\omega_A}{\omega_A + \omega_B}, \qquad \tilde{\omega}_B = 1 - \tilde{\omega}_A$$

2

$$log\hat{y}_u = \tilde{\omega}_A log\hat{y}_u^{(A)} + \tilde{\omega}_B log\hat{y}_u^{(B)}$$

$$\hat{y}_u[c] = \frac{\left(\hat{y}_u^{(A)}[c]\right)^{\tilde{\omega}_A}\left(\hat{y}_u^{(B)}[c]\right)^{\tilde{\omega}_B}}{\sum_k (\hat{y}_u^{(A)}[k])^{\tilde{\omega}_A}\left(\hat{y}_u^{(B)}[k]\right)^{\tilde{\omega}_B}}$$

$$H_{fuse} = \left(\tilde{\omega}_A H_A^{-1} + \tilde{\omega}_B H_B^{-1}\right)^{-1}$$

$$\theta_s \leftarrow \theta_s - \eta H_{fuse}^{-1} \nabla_{\theta_s} L_{total}$$

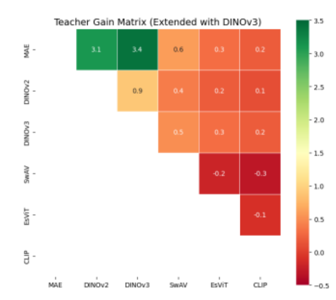

Figure 4. Teacher gain matrix on ImageNet (25% labeled data setting).