Towards Semi-Supervised Deep Facial Expression Recognition with An Adaptive Confidence Margin

摘要

大多数半监督学习方法只选择部分未标记的数据来训练模型，其置信度得分通常高于预定义的阈值 (即置信边际)。我们认为，通过充分利用所有未标记的数据，可以进一步提高识别性能。在本文中，我们学习了一种自适应置信区间 (Ada-CM)，以充分利用所有未标记的数据进行半监督深度面部表情识别。将所有未标记样本在每个训练 epoch 的置信度得分与自适应学习的置信度边际进行比较，将其划分为两个子集:(1) 子集 I 包含置信度得分不低于边际的样本;(2) 子集 II，包括置信度分数低于边际值的样本。对于子集 I 中的样本，我们约束它们的预测以匹配伪标签。同时，子集 II 的样本参与特征级对比目标，学习有效的面部表情特征。

关键词: 半监督学习；深度学习；面部表情识别；置信度得分；自适应置信区间

1 引言

人脸表情识别是计算机视觉领域的一个重要研究方向，其意义在于通过分析人脸表情来理解情感状态和情感变化。这一技术在各个领域都具有广泛的应用，从情感分析到人机交互，以及安全领域的应用。在过去的几十年里，计算机视觉领域取得了显著的进展，其中之一是人脸表情识别。该技术涉及从图像或视频中检测和识别人脸上的表情，如快乐、愤怒、悲伤等。这不仅有助于了解人的情感状态，还在各个领域具有重要应用。人脸表情识别对于情感分析、人机交互、医疗和安全领域具有重要作用。总之，人脸表情识别在多个领域都具有重要的应用潜力。它有助于理解和解释人的情感和行为，提高了用户体验，增强了安全性，同时也为科学研究提供了有趣的工具。

研究者注意到，半监督学习领域已经涌现出多种方法来改进未标记数据的利用。这些方法包括自生成对抗网络（GANs）、伪标签、自监督学习等。虽然这些方法已经取得了一些成功，但在某些情况下，它们可能仍然受到高置信度未标记数据的局限，忽略了低置信度数据。这引发了研究者的兴趣，提出了一种新的自适应置信度学习方法（Ada-CM），以更全面地利用未标记数据，不受数据置信度的限制。因此，研究的出发点是要解决半监督学习中的数据利用问题，特别是那些由于低置信度而被忽视的未标记数据。通过将未标记数据分为高置信度和低置信度的两个子集，并为这两个子集采取不同的策略，研究者希望提高深度面部表情

识别的性能，并使模型更具鲁棒性。这一出发点反映了对提高模型性能和减少对已标记数据依赖的渴望，同时也考虑了半监督学习领域的最新发展趋势。

2 相关工作

在先前的研究中，通常采用置信度阈值生成伪标签以识别高置信度数据，并将其与有标签数据一同用于模型训练 [1]。然而，这些方法中的置信度阈值往往是人为设置并保持不变的。通常，这样的阈值被设置为较高的数值，例如 0.95，以确保选择高置信度数据。本研究认为，采用固定阈值的方法在模型学习特定表情分类方面存在缺陷。通过问卷调查，我们确认了不同分类之间样本的置信度差异较大。对于每种表情分类采用相同而且固定的置信度阈值，可能导致一些类内样本由于其较低的置信度分数而无法被选择用于训练模型的问题 [2]。

2.1 面部表情识别

已经提出了许多面部表情识别方法 [3–5]。在 FER 领域，主要有两大研究方向，即手工特征和基于深度学习的方法。在传统方法中，早期的尝试侧重于实验室内 FER 数据集的纹理信息 [6, 7]，例如 CK+ 和 Oulu-CASIA。受到大规模非受限 FER 数据集的启发，DFER 算法设计了有效的卷积神经网络或损失函数，以实现更卓越的性能。从一开始，Li 等人 [8] 提出了一种保持局部性的损失，以学习更具判别性的面部表情特征。在受到注意机制的启发时，Wang 等人 [9] 提出了基于区域的注意网络，以捕捉重要的面部区域。Li 等人 [10] 还探讨了部分遮挡的面部表情识别。此外，一些工作考虑了 DFER 中的不一致标注问题。同时，Xue 等人 [11] 首次探讨了基于关系感知的 Transformer DFER 的表示。上述方法均以全监督方式进行 FER。与此不同的是，Florea 等人 [12] 提出了 MixMatch [13] 的扩展，即 Margin-Mix，并利用未标记的样本解决了密集区域问题。确实，Margin-Mix 通过嵌入类中心确定未标记样本的人工标签，而不是通过置信度边际。此外，中心更新是昂贵且耗时的。据我们所知，尚未提出基于阈值的伪标签方法用于 SS-DFER 任务。在我们的工作中，我们设计了自适应置信度边际，用于生成高质量的未标记样本的伪标签，这些样本具有较高的置信度分数。

2.2 半监督学习

近年来，半监督学习方法已成功应用于一些具有挑战性的问题 [14–16]。现有的 SSL 方法采用一致性正则化 [17, 18]、熵最小化 [19, 20] 和传统正则化 [13] 等手段来利用未标记的数据。在其中，伪标签是一种 SSL 方法，可以从模型预测中获得硬标签。特别是，基于阈值的方法选择具有高置信度预测的未标记样本。FixMatch 和 UDA 基于固定阈值获取伪标签，并利用弱和强的数据增强来实现一致性正则化。此外，一些工作已经研究了动态阈值 [35, 40]。例如，Xu 等人提出了一种通用方法，动态选择具有高置信度预测的样本。在我们的工作中，首次尝试学习适应性置信度边际用于 SS-DFER。此外，我们还首次尝试学习所有未标记的样本，这也是 SSL 领域的首次尝试。

3 本文方法

3.1 本文方法概述

所提出的 Ada-CM 首先运行所有给定的标记数据，并根据不同面部表情的学习难度自适应地更新置信区。置信率在训练周期中逐渐提高。然后，它预测弱增强未标记数据的置信度分数，将其与学习的置信度进行比较，将所有未标记的样本分为两个子集：子集 I，包括置信度分数高的样本（即置信度分数不低于边际）和子集 II，包括置信度分数低（即置信度分数低于边际）的样本。对于子集 I 中的样本，Ada-CM 利用强增强的未标记样本和弱增强版本的伪标签来计算交叉熵损失。此外，对于子集 II，进行特征级对比目标，通过应用 InfoNCE 损失来学习有效特征。如图 2所示：

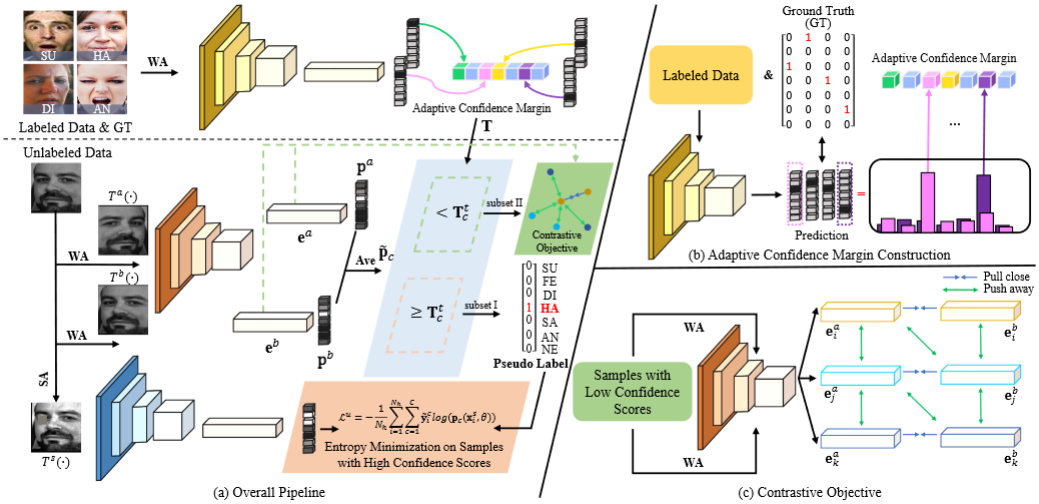


图 1. 方法示意图

3.2 有标签数据

从图中简要概括其中关键步骤：

（1）通过有标签数据计算自适应阈值，论文中对传统方法（获取有标签数据的预测概率求平均值）进行了改良，认为获取全部有标签数据含有噪声标签，使得样本的某些置信度分数不可信，所以论文只使用预测结果正确的数据样本来计算阈值。（相当于剔除了预测不正确的那部分数据样本对阈值的影响）。

$$T_c = \frac{1}{N_{st}^c} \sum_{i=1}^{N_{st}} \mathbb{1}(\hat{y}_i = c) \cdot s_i \tag{1}$$

T_c 在这一步中代表类型为 c 的表情类别的置信度阈值（暂时，下一步将添加变换）。

N_{st}^c 表示预测正确样本中类型为 c 的数量。

$\hat{y}_i$ 表示第 i 个样本的预测分类结果。

s_i 表示第 i 个样本的置信度分数。

$\mathbb{1}(\hat{y}_i = c)$ 表示当第 i 个样本预测分类结果为 c 时，该式取 1，否则取 0。

$\sum_{i=1}^{N_{st}} \mathbb{1}(\hat{y}_i = c) \cdot s_i$ 表示预测结果为 c 的样本的置信度分数之和。

该公式即为预测结果为 c 的样本的置信度分数的平均值，也为论文提出的基于正确预测的置信度阈值计算方法。

论文还提出对该自适应置信度阈值在逐渐增大的训练迭代轮次中的变换方法：

$$T_c^t = \frac{BT_c}{1+\gamma^{-t}} \tag{2}$$

该公式可以描述为：随着训练轮次 t 的逐渐增大，γ^{-t} 逐渐变小，分数的分母变小，整体的值变大，也就达成了置信度阈值随着训练过程不断增大的目标。

论文中设计该阈值变换的原因为，随着不断训练（训练轮次 t 增大），模型的效果越来越好，该置信度阈值也应该不断增大。

3.3 无标签数据

一共有两个分支，分别为：A. 两个弱增强无标签数据输入模型中训练，共享参数，对获得的两个预测概率求平均值，得到的预测概率与有标签数据得到的置信度阈值进行比较，大于阈值的归于 subset I，生成伪标签；小于阈值的归于 subset II，这个 subset 中使用 infoNCE 损失来学习低置信度数据的特征。B. 一个强增强无标签数据输入模型中训练，根据 A 中 subset I 得到伪标签计算交叉熵损失。

3.4 损失函数定义

对于有标签数据，训练模型使用了普通的交叉熵损失。

$$\mathcal{L}_{CE}^{s} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_i^c log(p_c(x_i,\theta)), \tag{3}$$

subset I 得到伪标签计算交叉熵损失。

$$\mathcal{L}^{u} = -\frac{1}{N_h}\sum_{i=1}^{N_h}\sum_{c=1}^{C} \tilde{y}_i^c log(p_c(x_i^s,\theta)), \tag{4}$$

总的损失函数为：

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_{CE}^{s} + \lambda_2\mathcal{L}^{u} + \lambda_3\mathcal{L}^{c}, \tag{5}$$

式中：

L_{CE}^{s}—有标签数据交叉熵损失

L^{u}—强增强无标签数据与伪标签的交叉熵损失

L^c—置信度低于阈值的数据 InfoNCE 损失

$_1 = 0.5, _2 = 1, _3 = 0.1$

4 复现细节

4.1 与已有开源代码对比

使用了 https://github.com/hangyu94/Ada-CM 的开源代码作为基础，在其上实验了：

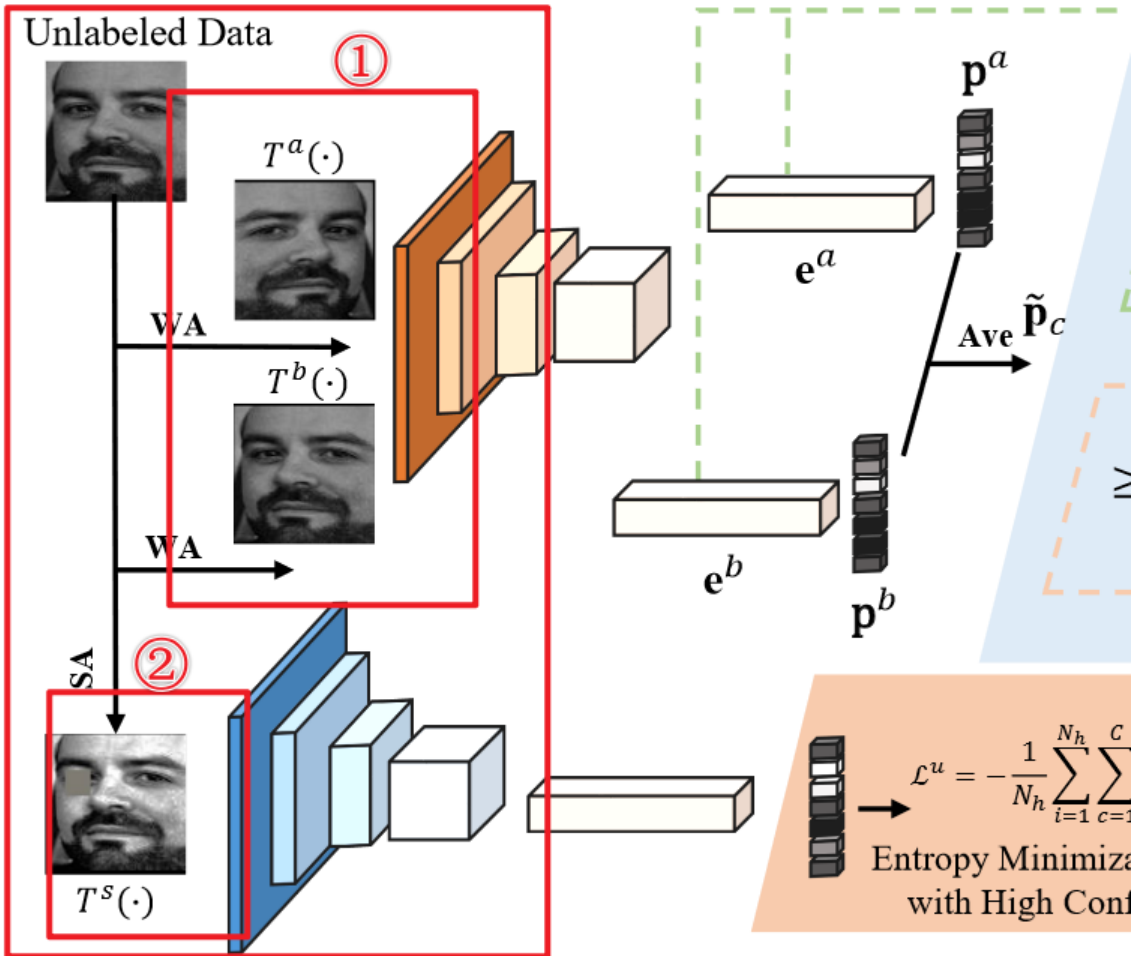


图 2. 修改示意图

4.2 对 1 的部分进行改进

论文中使用了两个弱增强数据预测概率的平均值 P_{avg2}，尝试使用三个弱增强数据预测概率的平均值 P_{avg3}。结果：模型前期的训练效果较好，但最高预测准确率仍然不能超过 P_{avg2} 的准确率（83.47%）。

4.3 对 2 的部分进行改进

和 1 相似，使用两个强增强的数据输入模型中，共享权重。结果：模型整体准确率下降，与原准确率相比下降了 0.5% 左右（82.88%）。

4.4 对自适应阈值的变换进行改进

$$T_c^t = \frac{BT_c}{(1 + \frac{t+1}{t})} \tag{6}$$

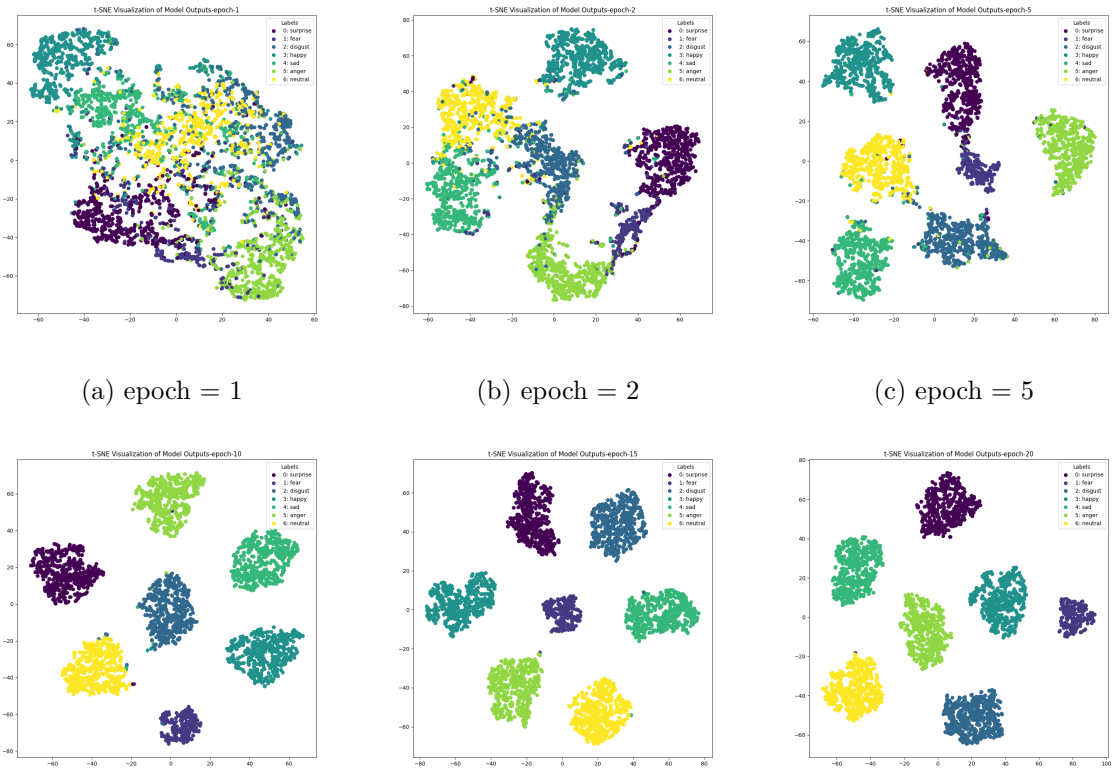
（1）得到的结果为 82.85%

$$T_c^t = \frac{BT_c}{(1 + \frac{t+1}{2t})} \tag{7}$$

（2）得到的结果为 82.82%

$$T_c^t = \frac{BT_c}{(1 + e^{log_2(t+1)}} \tag{8}$$

（3）得到的结果为 83.11%



(a) epoch = 1　(b) epoch = 2　(c) epoch = 5

(d) epoch = 10　(e) epoch = 15　(f) epoch = 20

图 3. 模型训练过程

5 实验结果分析

5.1 实验结果

5.2 结果分析

如表 1，六种方法中，不管是在 RAF-DB 数据集中，还是在 AffectNet 数据集里，在不同标签数的数据下训练的模型准确率均排在前列。虽然该算法在同一数据集的各种条件下的

表 1. 同数据集准确率 (%)

Method	RAF-DB	AffectNet
Baseline	80.28	62.52
PseudoLabeling	80.35	61.58
MixMatch	81.06	63.22
ReMixMatch	84.00	65.45
FixMatch	83.21	65.37
Ada-CM	83.70	65.96

表 2. 跨数据集准确率 (%)

Method	R -> A	R -> C	A -> R	A -> C
Baseline	29.88	70.87	17.73	47.57
PseudoLabeling	29.45	64.40	19.10	47.57
MixMatch	27.89	60.52	21.25	59.55
ReMixMatch	33.53	69.58	21.48	61.49
FixMatch	32.89	69.57	22.16	63.98
Ada-CM	33.18	72.49	22.23	64.08

准确率并非最高，但是它的表现一直稳居前列，且与最高准确率之间的差距很小。这表明本算法能够高效地利用数据集，并展现出强大的适应能力。总体来说，本算法能够更有效地利用可用数据。

如表 2，六种方法中，RAF-DB->AffectNet（表示 RAF-DB 数据集中训练的算法模型在 AffectNet 的测试数据集中进行识别，下文均以此方式表示）的情况下，仅比最高准确率的 ReMixMatch 差 0.35%，排第二。除此之外，RAF-DB->CK+、AffectNet->RAF-DB、AffectNet->CK+ 三种情况下，本算法的准确率均为最高。通过对六种算法的跨数据集准确率表现进行对比，本算法能够在跨数据集测试下能够获得较为优良的实际表现，在多种比较算法中准确率排名前列，表明了本算法的适应性较好，改善了原始方法的低适应性问题。

在 RAF-DB 数据集上训练模型过程如图 3所示，在不断地迭代中，模型的分类效果逐步加强。

6 总结与展望

在该工作中，提出了一种新的自适应置信区间 (Adaptive Confidence Margin，Ada-CM) 用于半监督深度人脸表情识别，它自适应地利用所有未标记样本 (即置信度高的子集 I 中的样本和置信度低的子集 II 中的样本) 来训练模型。该工作提出的 Ada - CM 从两个方面显著提高了性能。一方面，对置信度超过学习到的置信边缘的未标注样本直接进行伪标注，以匹配强增强版本的预测。另一方面，对比目标被用于学习子集 II 中样本之间的面部表情特征。

此外，还能探索无监督的数据增强方法，减少对标注数据的依赖。通过自动生成合成的

未标记样本，提高模型在缺乏标签数据时的性能，降低在实际应用中的数据收集成本。结合生成对抗网络（GANs）等对抗性学习技术，提高数据增强的质量和多样性。通过引入对抗性元素，使生成的样本更具真实性，有助于提高模型对真实世界复杂场景的适应性。

参考文献

[1] Dong Hyun Lee. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. 2013.

[2] Hangyu Li, Nannan Wang, Xi Yang, Xiaoyu Wang, and Xinbo Gao. Towards semi-supervised deep facial expression recognition with an adaptive confidence margin. 2022.

[3] Hangyu Li, Nannan Wang, Xinpeng Ding, Xi Yang, and Xinbo Gao. Adaptively learning facial expression representation via c-f labels and distillation. *IEEE Transactions on Image Processing*, PP(99):1–1, 2021.

[4] Li Shan and Deng Weihong. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. *IEEE Transactions on Image Processing*, PP:1–1, 2018.

[5] Jiahui She, Yibo Hu, Hailin Shi, Jun Wang, Qiu Shen, and Tao Mei. Dive into ambiguity: Latent distribution mining and pairwise uncertainty estimation for facial expression recognition. 2021.

[6] Yuxiao Hu, Zhihong Zeng, Lijun Yin, Xiaozhou Wei, Xi Zhou, and Thomas S Huang. Multi-view facial expression recognition. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–6. IEEE, 2008.

[7] Yuan Luo, Cai-ming Wu, and Yi Zhang. Facial expression recognition based on fusion feature of pca and lbp with svm. *Optik-International Journal for Light and Electron Optics*, 124(17):2767–2770, 2013.

[8] Abhinav Dhall, Roland Goecke, Simon Lucey, and Tom Gedeon. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, pages 2106–2112. IEEE, 2011.

[9] Kai Wang, Xiaojiang Peng, Jianfei Yang, Debin Meng, and Yu Qiao. Region attention networks for pose and occlusion robust facial expression recognition. *IEEE Transactions on Image Processing*, 29:4057–4069, 2020.

[10] Yong Li, Jiabei Zeng, Shiguang Shan, and Xilin Chen. Occlusion aware facial expression recognition using cnn with attention mechanism. *IEEE Transactions on Image Processing*, 28(5):2439–2450, 2018.

[11] Fanglei Xue, Qiangchang Wang, and Guodong Guo. Transfer: Learning relation-aware facial expression representations with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3601–3610, 2021.

[12] Corneliu Florea, Mihai Badea, Laura Florea, Andrei Racoviteanu, and Constantin Vertan. Margin-mix: Semi-supervised learning for face expression recognition. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, pages 1–17. Springer, 2020.

[13] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32, 2019.

[14] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020.

[15] Zhenyu Wang, Yali Li, Ye Guo, Lu Fang, and Shengjin Wang. Data-uncertainty guided multi-phase learning for semi-supervised object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4568–4577, 2021.

[16] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34:18408–18419, 2021.

[17] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with stochastic transformations and perturbations for deep semi-supervised learning. *Advances in neural information processing systems*, 29, 2016.

[18] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. *Advances in neural information processing systems*, 33:6256–6268, 2020.

[19] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems*, 17, 2004.

[20] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896. Atlanta, 2013.