DOI:10.16356/j.1005-2615.2025.02.018

基于困难感知元学习的跨域人脸伪造检测

金世辰,谭晓阳
(南京航空航天大学计算机科学与技术学院,南京 211106)

摘要: 随着面部伪造技术的快速迭代,能够应对未见过的伪造方法的鲁棒检测机制需求变得日益重要。然而,当前的方法主要针对特定的伪造技术设计,这在应对更广泛的检测挑战时存在局限性。为了解决这些问题,本文提出了一种用于跨域人脸伪造检测的难度感知元学习(Difficulty-aware meta-learning, DAML)方法。在元训练阶段,本文方法利用与伪造图像无关的元学习(Model-agnostic meta-learning, MAML)方法来训练模型。通过利用目标域中的少量数据,可以调整参数以适应新任务。为了解决与模型无关的元学习方法中的不稳定训练问题,本文引入了一种难度感知机制,在训练阶段动态调整不同任务的学习权重。在多个公开的基准数据集上进行了广泛的实验,实验结果表明,本文方法优于RECCE、Xception、RFM等方法,在适应未见过的目标域方面表现更好。

关键词: 人脸伪造检测;元学习;跨领域;动态调整学习权重;泛化性
中图分类号: TP391 **文献标志码:** A **文章编号:** 1005-2615(2025)02-0371-07

Difficulty-Aware Meta-Learning for Cross-Domain Face Forgery Detection

JIN Shichen, *TAN Xiaoyang*
(College of Computer Science and Technology, Nanjing University of Aeronautics & Astronautics, Nanjing 211106, China)

Abstract: With the rapid iteration of facial forgery techniques, robust detection mechanisms that can handle unseen forgery methods are increasingly crucial. However, current approaches are primarily tailored to specific forgery techniques, posing limitations in addressing this broader detection challenge. To address these issues, this paper proposes a difficulty-aware meta-learning (DAML) method for cross-domain face forgery detection. During the meta-training phase, our method utilizes a model-agnostic meta-learning (MAML) approach, using historical forgery images to optimize the meta-leaner. By leveraging a small amount of data in the target domain, we can adjust the parameters to fit new tasks. To tackle the issue of unstable training in model-agnostic meta-learning methods, we introduce a difficulty-aware mechanism to dynamically adjust the learning weights for different tasks during the training phase. We conduct extensive experiments on several publicly available benchmark datasets. The experimental results demonstrate that our method outperforms several methods, such as: RECCE, Xception, RFM, etc., achieving better adaptability to unseen target domains.

Key words: face forgery detection; meta-learning; cross domain; dynamic adjustments of learning weights; generalization

近年来,随着深度学习的广泛应用,人脸伪造技术[1-4]迅速发展。这一进步使得生成的假人脸图像和视频越来越逼真,难以通过肉眼辨别。恶意行为者利用这些技术进行身份冒充、散布虚假信息、伪造法

基金项目: 国家自然科学基金(6247072715)。
收稿日期: 2024-08-21;**修订日期:** 2025-01-01
通信作者: 谭晓阳,男,教授,博士生导师,E-mail:x.tan@nuaa.edu.cn。

引用格式: 金世辰,谭晓阳. 基于困难感知元学习的跨域人脸伪造检测[J]. 南京航空航天大学学报(自然科学版),2025,57(2):371-377. JIN Shichen, TAN Xiaoyang. Difficulty-aware meta-learning for cross-domain face forgery detection[J]. Journal of Nanjing University of Aeronautics & Astronautics(Natural Science Edition),2025, 57(2):371-377.

律证据,并对政治、经济和司法等社会各个领域[5]构成了重大威胁。因此,迫切需要开发有效的检测方法,以减少人脸伪造技术被恶意利用的危害。

早期的人脸伪造检测方法通常将人脸图像直接输入模型,并使用卷积神经网络(Convolutional neural networks, CNNs)[1]作为主干结构,这在图像分类领域是常见的做法。这些方法将人脸伪造检测视为二分类任务,主要关注区分真实与伪造图像类别,要求模型只需将图像分类为真实或伪造。然而,这些检测器往往缺乏对图像像素中具体伪造特征的深入理解。由于相机成像原理的原因,真实图像的像素分布具有明显的特征[2],而伪造方法导致伪造区域与真实区域之间的分布差异。相比之下,近期的人脸伪造检测方法则专注于识别伪造图像中的篡改痕迹,如深度视觉特征[6]、生成对抗网络(Generative adversarial network,GAN)指纹[7]以及手工特征[8-9]。

这些方法在基准数据集上训练后能够实现较高的检测准确率。然而,这些基于伪造特征设计的检测方法[10]依赖于训练过程中使用的伪造技术。因此,当检测未知方法生成的伪造图像时,这些方法的检测性能显著下降。解决这一问题的基本方法是通过微调[11]。在检测前,可以手动标注部分检测目标中的数据,并利用这些数据微调已训练的模型参数。然而,数据的手动标注既困难又昂贵,微调过程中数据不足可能导致过拟合问题。

为了解决在人脸检测中的跨域问题[12],即将来自不同数据分布的图像视为不同域,提出了一种对抗域适应方法。该方法引入一个判别器网络,用于判断特征数据是来自源域还是目标域。然后,通过对抗方式训练特征提取器和判别器,以实现特征空间的对齐,从而达到特征空间对齐的目的。然而,这种方法面临着源域和目标域之间不同类别分布错位的问题。此外,迁移学习[13]和多任务学习[14-15]也被用于增强模型在跨域任务中的表现,但对于未知域样本及其标签的获取通常伴随着高昂的成本。

另一种解决该问题的方法是引入元学习[12],它可以提高模型快速适应新任务的能力。模型无关元学习(Model-agnostic meta-learning,MAML)[16]是一种非常受欢迎的方法,因为它能够在不依赖于特定模型结构的情况下增强模型的泛化性和适应性。该方法通过在多个任务之间共享学习经验和模式,使模型更好地理解新任务的特征和模式,从而能够快速适应不同的数据分布或新的伪造技术。然而,在元训练的梯度更新阶段,为所有任务的损失使用相同的权重可能会导致梯度不稳定问题[17]。

为了解决这些挑战,本文提出了一种新颖的跨域人脸伪造检测的难度感知元学习方法(Difficulty-aware meta-learning, DAML)。DAML采用了类似于MAML的模型架构,并利用元学习的优势,在调整目标域的模型时避免过拟合的风险。整体框架如图1所示。

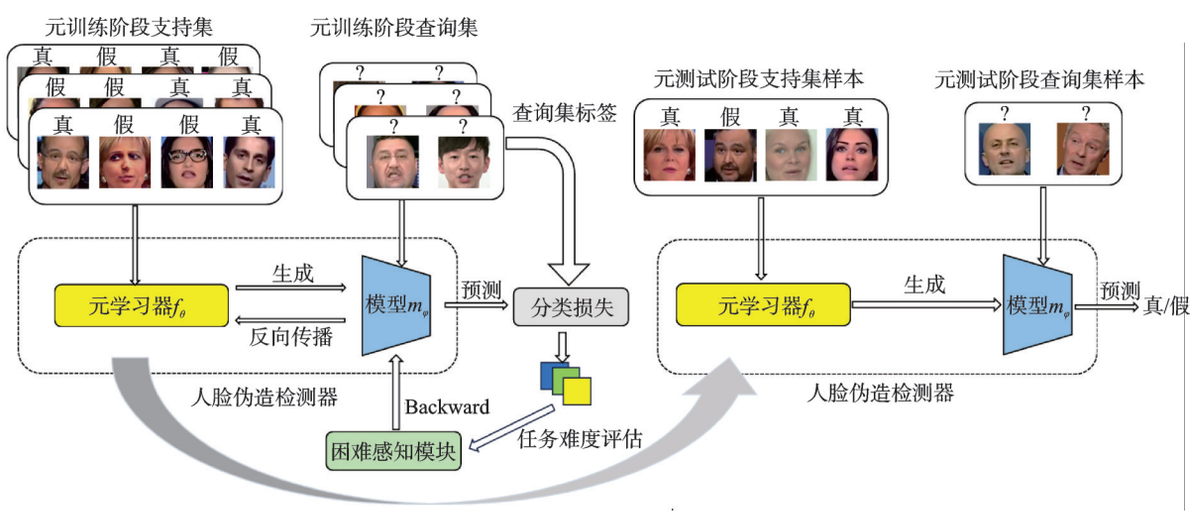

图1 人脸伪造检测的难度感知元学习框架

Fig.1 Framework of DAML for face forgery detection

此外,DAML在元训练过程中引入了一种难度感知机制,在梯度更新阶段动态调整不同任务的损失权重,这样可以稳定模型训练的梯度,使模型能够有效地学习和调整具有不同难度级别的伪造图像。因此,DAML提高了模型在新领域中检测伪造图像的泛化能力和性能。

为了评估本文提出的方法,在多个基准数据集上进行了广泛的跨域人脸伪造检测实验,包括

FaceForensics++[3]、DFDC[18]、Celeb-DF[19]和WildDeepfake[20]。实验结果表明，本文方法能够有效地适应新领域，并优于KECCE、Xception、RFM等方法。

1　相关工作

1.1　人脸伪造检测

近年来，人脸伪造检测已成为一个重要的研究领域。早期研究主要集中在创建特征提取器，如加速鲁棒特征(Speeded-up robust features，SURFs)[21]、方向梯度直方图(Histogram of oriented gradients，HOG)[22]和局部二值模式(Local binary pattern，LBP)[23]，通过从图像中提取实时特征来检测伪造人脸。随着深度学习的进展，研究的重点转向了使用图像分类的主干网络，如VGGNet[24]和XceptionNet[25]。这些方法利用数据驱动的方法来识别真实与篡改面部特征之间的差异，并进行二分类[2]，突出类别级别的差异。

随着伪造技术的进步，伪造面部变得越来越难以视觉区分。因此，研究人员开始探索伪造方法在图像中留下的特定痕迹，如局部噪声、频率细节、纹理特征[26-27]以及高层语义特征[28]。然而，这些方法往往严重依赖于针对特定伪造技术标注的数据，导致在面对新的伪造方法时表现显著下降。

为了解决这些问题，研究人员转向了领域自适应技术[29]。领域自适应是一种常用的方法，包括一个鉴别器网络，它区分源域和目标域的特征，而特征提取器通过对抗学习来对齐这些特征空间。文献[10]提出了一种无监督任务来增强跨域伪造人脸检测的鲁棒性。

除了领域适应，基于迁移学习的微调方法也得到了探索。例如，文献[30]描述了一种将源数据集的知识转移到目标数据集的技术。

1.2　元学习

近年来，元学习因其在快速适应新数据挑战方面的潜力，受到了学术界的广泛关注。元学习主要关注于设计有效的策略，以实现快速适应。在元学习领域，研究人员确定了两大类主要方法：(1)学习策略和政策；(2)学习最佳的初始模型参数。在人脸伪造检测任务的背景下，许多研究已经利用了元学习方法。例如，文献[12,31-32]的研究工作已经将元学习技术应用于提升人脸伪造检测系统的效能。

2　本文方法

2.1　跨域人脸伪造检测

在人脸伪造检测任务中，假设对于每个输入图像 $x_i \in \mathbf{R}^{h\times w\times 3}$，分类器需要输出相应的真实性标签 y_i，其中 $y_i \in \{0,1\}$。对于每一个独立的源域 D_s，其中均包含标记数据 $\{x_i, y_i\} \in D_s$。人脸伪造检测模型判别结果 $Y=\{y_1, y_2, \cdots, y_n\}$ 的条件概率为

$$P(Y|X,d;\theta)=\prod_{i=1}^{N} P(y_i|x_i,d;\theta) \tag{1}$$

式中：θ 表示一组模型参数；d 表示源域中提取的任务。本文方法不受特定模型架构和领域划分方法的影响，因此可以推广到不同伪造方法的检测任务上。

2.2　跨域人脸伪造检测的困难感知元学习

本节提出了将模型无关的元学习方法应用于跨领域人脸伪造检测。利用MAML来获取初始模型参数 θ^0，从而使模型能够在有限的训练数据下快速适应新的领域。在元训练阶段，引入了一个专门的组件，用于动态评估学习任务的重要性。该组件在元训练过程中主动监测学习任务的重要性，从而优化模型的泛化能力，增强其在不同领域中高效适应的能力。

定义1组源任务集合 $T=\{T_{d_1}, T_{d_2}, \cdots, T_{d_k}\}$，其中每个任务 T_{d_i} 表示1个特定领域 d_i 上的检测任务；k 表示训练源领域的数量。在每一轮元学习迭代中，从集合 T 中随机选择一个任务 T_{d_i}。随后，从任务 T_{d_i} 中抽取2组相互独立的数据批次：一组为支持批次 $D_{d_i}^{s}$；另一组为查询批次 $D_{d_i}^{q}$。首先，使用 $D_{d_i}^{s}$ 来更新模型参数 θ，更新过程如下

$$\theta'_{d_i}=\theta-\alpha\nabla_{\theta}L_{D_{d_i}^{s}}(\theta) \tag{2}$$

式中：α 表示学习率，L 表示交叉熵损失(Cross-entropy loss，CE loss)。

$$L_{D_{d_i}^{s}}(\theta)=-\sum_{D_{d_i}^{s}}\log P(Y|X,\theta) \tag{3}$$

在 $D_{d_i}^{q}$ 数据集上评估更新后的参数，并利用从多个源任务的多个训练循环中聚合的评估结果计算的梯度来更新原始模型参数 θ。在MAML算法中，原始模型参数 θ 的评估过程为

$$\theta=\theta-\beta\nabla_{\theta}\sum_{d_i}L_{D_{d_i}^{q}}(\theta'_{d_i}) \tag{4}$$

可以直观地观察到，在这种方法中，参数 θ 的梯度更新通过对所有查询任务 $D_{d_i}^{q}$ 的损失进行求和来计算。这种配置的目的是确保在参数更新过程中每个任务都得到平等对待，使得最终的模型能够在各种潜在任务中进行泛化。

在训练过程中，本文模型会针对每个任务进行不同的学习，从而导致损失的差异。模型在简单任务上表现良好，但在困难任务上表现不佳，困难任务的损失较高。通过更多关注困难任务，可以促使模型更有效地学习这些难以捕捉的特征和模式，从

而提高模型在各种潜在任务中的泛化能力。这种关注困难任务的策略有助于模型对数据分布的复杂性有更全面的理解,从而提升模型的鲁棒性和泛化性能。

因此,引入了一个动态难度感知系数$1\Big/\left(1+\mathrm{e}^{-L_{D_{d_i}^{\mathrm{q}}}(\theta'_{d_i})}\right)$来调整每个任务的损失。该函数$1/(1+\mathrm{e}^{-x})$是单调递增的。当$x>0$时,它会从0.5逐渐逼近1。当任务的学习表现较差(即经验损失较大时),损失调整趋向于1;相反,当表现较好(即经验损失较小时),调整趋向于0.5。这种修改方法增加了对困难任务的学习权重,同时降低了对简单任务的学习权重,使得模型能够更快地在不同任务中达到稳定的平均性能。困难感知模型参数θ的评估过程如下

$$\theta=\theta-\beta\nabla_{\theta}\sum_{d_i}\left(\frac{L_{D_{d_i}^{\mathrm{q}}}(\theta'_{d_i})}{1+\mathrm{e}^{-L_{D_{d_i}^{\mathrm{q}}}(\theta'_{d_i})}}\right) \tag{5}$$

式中β表示元学习率。此外,该方法将损失调整到[0.5,1]的范围,这可能在最初阶段稍慢,但在随后的迭代中加快了梯度下降速度。这使得模型能够更灵活地适应任务,从而增强了模型的泛化能力和适应性。这种动态调整机制在元训练过程中有效地优化了模型参数,提升了整体性能。具体细节见算法1。

算法1 Difficulty-aware meta-learning algorithm

Require: T: set of source tasks

Require: α, β: learning rate

(1) Randomly initialize θ

(2) while not done do

(3) Sample batch of tasks $T_{d_i}\sim T$

(4) for all T_{d_i} do

(5) $(D_{d_i}^{\mathrm{s}},D_{d_i}^{\mathrm{q}})\sim D_{d_i}$

(6) Evaluate $\nabla_{\theta}L_{D_{d_i}^{\mathrm{s}}}(\theta)$ using $D_{d_i}^{\mathrm{s}}$

(7) Compute adapted parameters with gradient descent:

(8) $\theta'_{d_i}=\theta-\alpha\nabla_{\theta}L_{D_{d_i}^{\mathrm{s}}}(\theta)$

(9) end for

(10) Update meta θ: $\theta=\theta-\beta\nabla_{\theta}\sum_{d_i}\left(\frac{L_{D_{d_i}^{\mathrm{q}}}(\theta'_{d_i})}{1+\mathrm{e}^{-L_{D_{d_i}^{\mathrm{q}}}(\theta'_{d_i})}}\right)$

(11) end while

在元参数更新过程中涉及到二阶偏导数的求解。然而,由于更新频率较高,使用二阶偏导数会消耗大量的计算资源和内存。因此,本文选择了先前研究中使用的一阶近似方法,这有效地节省了计算资源。在完成元训练阶段后,对新目标任务T_{d}的少量样本进行任务特定学习,以获得任务特定模型θ_{d}。

3 实验过程与结果

3.1 实验设置

3.1.1 数据集

实验中使用了4个基准数据集来评估所提方法与现有方法在面部伪造检测中的表现。这些数据集包括:FaceForensics++(FF++)、DFDC、Celeb-DF和WildDeepfake[33]。其中,FaceForensics++是最广泛使用的数据集之一,涵盖了4种主要的伪造技术:Deepfakes(DF)、Face2Face(F2F)、FaceSwap(FS)和NeuralTextures(NT)。DFDC数据集包含了来自3 426名演员的超过100 000个视频片段。Celeb-DF数据集包括590个真实视频和5 639个使用改进的DeepFake方法生成的伪造视频。另一方面,WildDeepfake数据集包含了从707个伪造视频中提取的7 314个面部序列。尽管规模相对较小,但该数据集提供了更丰富的视频场景和个体活动,以及各种背景、光照条件等现实条件因素。

3.1.2 评估指标

为了评估本文的方法,使用了包括准确率(Accuracy, ACC)、ROC曲线下面积(Area under the curve, AUC)和交叉熵损失(CE loss)在内的常见指标。这些指标是通过不同方法[11-12,30,34-35]在数据集上获得。为了确保公平比较,所有指标都在单张图像层面进行评估,并且所有被比较的方法都遵循了先前研究中设定的设置。

3.1.3 实验细节

本文方法使用文献[16]中使用的四卷积块结构作为骨干网络。每个模块包含一个3×3的卷积核和64个滤波器,之后是批量归一化、ReLU激活函数以及一个2×2的最大池化层。在元训练过程中,使用了Adam优化器,元学习率设置为1e-3,每150个周期将其值减小为原来的1/10。训练过程总共进行了3 000次迭代,每个批次的大小为4,每个批次包括从源领域采样的4个任务。在元测试阶段,本文对目标任务训练集进行了微调,使用的学习率为1e-4。

3.2 实验结果

3.2.1 域内评估

为了评估面部伪造检测方法的基线性能,本文在FF++数据集上进行了一系列实验。这些实验涵盖了几种最先进的面部伪造检测方法,并使用相

同的性能指标进行评估。

如表 1 所示，在 FF＋＋数据集中使用两种不同的压缩率下，本文方法在不同实验设置中均表现优于或与几种其他面部伪造检测方法持平。值得注意的是，在这个具有挑战性的任务中，本文方法在低图像质量（压缩率 c40）时表现尤为出色，优于基于伪造特征的各种检测方法。这是因为高压缩率在一定程度上会掩盖图像中的伪造痕迹，而本文方法基于图像分类，更侧重于类别间的差异，而不是单独图像的细节。

表 1　在 FF++数据集上的比较

Table 1　Comparing within FF++ dataset　%

Method	FF＋＋(c23)		FF＋＋(c40)	
	ACC	AUC	ACC	AUC
Multi-task[36]	85.65	85.43	81.30	75.59
Xception[2]	95.73	96.30	86.86	89.30
RECCE[10]	97.06	**99.32**	91.03	**95.02**
RFM[2]	95.69	98.79	87.06	89.83
Add-Net[11]	96.78	97.74	87.50	91.01
F^3-Net[4]	97.52	98.10	90.43	93.30
MultiAtt[15]	**97.60**	99.29	88.69	90.40
DAML(Ours)	97.24	99.25	**91.23**	92.42

3.2.2　跨域评估

为了验证本文方法在未知伪造方法下的泛化能力，本文在不同数据集上进行了交叉验证实验。具体而言，在 FF＋＋数据集上训练模型后，分别在 Celeb-DF、DFDC 和 WildDeepfake 数据集上进行测试。结果如表 2 所示。根据呈现的结果，DAML 在相同设置下优于所列方法，表现出较强的泛化能力。

表 2　在 FF++数据集上训练后的模型进行跨领域检测

Table 2　Cross-domain detection leveraging FF++ trained models　%

Method	Celeb-DF		DFDC		WildDeepfake	
	ACC	AUC	ACC	AUC	ACC	AUC
Xception[2]	74.23	63.87	71.17	62.34	69.41	62.55
RECCE[10]	76.85	66.82	73.59	**69.13**	72.28	64.79
RFM[2]	69.63	65.63	68.37	66.01	61.75	57.75
Add-Net[11]	72.74	67.25	68.55	66.81	71.79	64.57
MultiAtt[15]	67.09	62.85	62.49	63.17	65.62	57.49
DAML (Ours)	**77.14**	**68.13**	**75.07**	67.89	**73.98**	**69.03**

此外，本文还进行了跨伪造技术检测实验。具体来说，本文在 FF＋＋(c40)上使用特定伪造方法训练模型，然后在该数据集中的其他伪造方法上进行测试，如表 3 所示。本文办法显著优于针对特定伪造技术设计的方法，展示了本文提出的方法在处理未知伪造任务中的有效性。

表 3　不同伪造技术下的 AUC 测试

Table 3　Testing across different forgery techniques in terms of AUC　%

Method	Train	DF	F2F	FS	NT	Cross AVG
Xception[2]		99.37	57.89	49.67	65.79	57.78
RECCE[10]	DF	99.51	69.92	73.29	**77.23**	73.48
DAML(Ours)		**99.63**	**71.63**	**74.51**	75.64	**73.93**
Xception[2]		67.27	97.14	51.38	55.71	58.12
RECCE[10]	F2F	74.91	97.64	65.39	**71.92**	70.74
DAML(Ours)		**77.52**	**98.14**	**69.25**	70.82	**72.53**
Xception[2]		53.37	49.92	97.75	52.95	52.08
RECCE[10]	FS	79.35	62.16	98.02	59.93	67.15
DAML(Ours)		**81.21**	**69.72**	97.67	**62.51**	**71.14**
Xception[2]		58.94	50.21	61.70	**99.80**	56.95
RECCE[10]	NT	69.07	**78.89**	64.38	95.14	70.78
DAML(Ours)		**82.79**	77.67	**69.85**	98.27	**76.77**

3.3　消融实验

为了验证本文提出的难度感知机制的有效性，分析了在训练过程中遇到的挑战性任务，这些任务从 FF＋＋数据集中提取，表现出在同一训练任务批次中较高的损失。图 2 展示了在元训练阶段，使用不具有难度感知机制的 MAML 模型和使用难度感知机制的 DAML 模型的损失变化曲线。两个模型均使用相同量的数据进行训练。如图 2 所示，与没有引入任何机制的 MAML 相比，本文的方法通过融入自适应调整显著提升了训练过程的整体稳定性。使用难度感知机制可以动态调整任务的学习权重，从而有效应对训练过程中遇到的挑战性任务。这种自适应调整策略不仅提高了模型在面对不同难度任务时的适应能力，还加快了模型的收敛速度，最终增强了模型的泛化能力和稳定性。本文为利用 $\frac{L_{D_{d_i}^{\mathrm{q}}}(\theta'_{d_i})}{1+\mathrm{e}^{-L_{D_{d_i}^{\mathrm{q}}}(\theta'_{d_i})}}$ 作为元训练梯度更新能够使得元训练梯度更好地集中于模型尚未充分学习的任务，从而减少对已学习良好的任务的关注。此外，使用相同的数据对 MAML 和 DAML 架构进行了训练，采用相

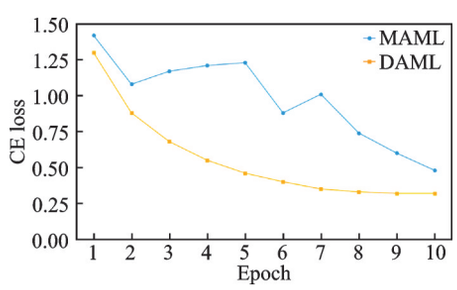

图 2　MAML 与 DAML 在 FF＋＋上损失曲线对比的消融实验

Fig.2　Ablation studies in terms of loss curves comparing with MAML and DAML on FF＋＋ dataset

同的训练程序,并在相同的目标数据集上进行评估。结果如图3所示,包含困难感知机制显著提高了模型在验证集上的表现。这表明,根据样本的困难程度加权可以使模型优先关注挑战性较大的样本,从而提升其从困难实例中学习的能力,并增强模型对未见过的伪造技术的泛化能力。

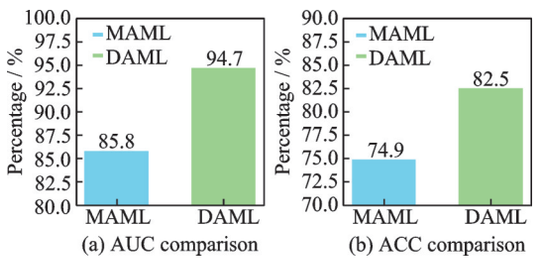

图3 MAML与DAML在FF++数据集上的ACC和AUC消融对比

Fig.3 Ablation in terms of ACC and AUC of MAML and DAML on FF++ dataset

4 结 论

本文提出了一种新颖的困难感知元学习方法,以解决面部伪造检测任务中的跨领域泛化问题。创新的困难感知调整机制在训练阶段实现了不同任务间的动态平衡,从而实现了对未知方法的快速泛化。在广泛使用的基准数据集上进行的大量实验验证了本文方法在应对跨领域面部检测任务挑战中的有效性。

参考文献:

[1] WANG J, LI Z, ZHANG C, et al. Fighting malicious media data: A survey on tampering detection and deepfake detection [EB/OL]. (2022-12-12) [2024-08-01]. https://arxiv.org/abs/2212.05667.

[2] TOLOSANA R, VERA-RODRIGUEZ R, FIERREZ J, et al. DeepFakes and Beyond: A survey of face manipulation and fake detection[J]. Information Fusion, 2020,63: 131-148.

[3] ROSSLER A, COZZOLINO D, VERDOLIVA L, et al. FaceForensics++: Learning to detect manipulated facial images[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul,Korea(South): IEEE, 2019.

[4] CAO W, WANG T, DONG A, et al. TransFS: Face swapping using Transformer[C]//Proceedings of 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG). Waikoloa, HI,USA: IEEE, 2023.

[5] LYU S. DeepFake detection: Current challenges and next steps[C]//Proceedings of 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). London: IEEE, 2020.

[6] GANDHI A, JAIN S. Adversarial perturbations fool deepfake detectors[C]//Proceedings of 2020 International Joint Conference on Neural Networks (IJCNN). Glasgow, United Kingdom: IEEE, 2020.

[7] ANKOLEKAR A, MADAPPA R, SAVAKIS A E. Can simpler be better? Review of methods for the detection of GAN-generated imagery[C]//Proceedings of Pattern Recognition and Tracking XXXIII. Bellingham, WA, USA: SPIE, 2022.

[8] LI H, LI B, TAN S, et al. Identification of deep network generated images using disparities in color components[J]. Signal Processing, 2020,177: 107616.

[9] HE P, LI H, WANG H. Detection of fake images via the ensemble of deep representations from multi color spaces[C]//Proceedings of 2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE, 2019.

[10] CAO J, MA C, YAO T, et al. End-to-end reconstruction-classification learning for face forgery detection[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 3577-3590.

[11] VERISSIMO S, GADELHA G, BATISTA L, et al. Transfer learning for face anti-spoofing detection [J]. IEEE Latin America Transactions, 2023,21(3): 530-536.

[12] SUN K, LIU H, YE Q, et al. Domain general face forgery detection by learning to weight[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022,36(2): 2638-2646.

[13] COZZOLINO D, THIES J, RÖSSLER A, et al. ForensicTransfer: Weakly-supervised domain adaptation for forgery detection[EB/OL]. (2018-12-06)[2024-08-01]. https://arxiv.org/abs/1812.02510.

[14] DANG H, LIU F, STEHOUWER J, et al. On the detection of digital face manipulation[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020.

[15] NGUYEN H H, FANG F, YAMAGISHI J, et al. Multi-task learning for detecting and segmenting manipulated facial images and videos[C]//Proceedings of 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS). Tampa, FL,USA: IEEE, 2019.

[16] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [EB/OL]. (2017-03-09) [2024-08-01]. https://arxiv.org/abs/1703.03400.

[17] ANTONIOU A, EDWARDS H, STORKEY A. How to train your MAML[EB/OL]. (2019-10-19) [2024-08-01]. https://arxiv.org/abs/1910.08854.

[18] DOLHANSKY B, HOWES R, PFLAUM B, et al. The deepfake detection challenge (DFDC) preview dataset[EB/OL]. (2018-12-06) [2024-08-01]. https://arxiv.org/abs/1812.02510.

[19] LI Y, YANG X, SUN P, et al. Celeb-DF: A large-scale challenging dataset for deepfake forensics [C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA,USA: IEEE, 2020.

[20] ZI B, CHANG M, CHEN J, et al. WildDeepfake [C]//Proceedings of the 28th ACM International Conference on Multimedia. [S.l.]: ACM, 2020.

[21] IPARRAGUIRRE J, BALMACEDA L, MARIANI C. Speeded-up robust features (SURF) as a benchmark for heterogeneous computers[C]//Proceedings of 2014 IEEE Biennial Congress of Argentina (ARGENCON). Buenos Aires, Argentina: IEEE, 2014.

[22] RAHMAD C, ASMARA R A, PUTRA D R H, et al. Comparison of Viola-Jones Haar cascade classifier and histogram of oriented gradients (HOG) for face detection[J]. IOP Conference Series: Materials Science and Engineering, 2020,879(1): 012038.

[23] AHONEN T, HADID A, PIETIKAINEN M. Face description with local binary patterns: Application to face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 2037-2041.

[24] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]//Proceedings of International Conference on Learning Representations, International Conference on Learning Representations. San Diego, CA, USA: ICLR, 2015.

[25] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017.

[26] QIAN Y, YIN G, SHENG L, et al. Thinking in frequency: Face forgery detection by mining frequency-aware clues[C]//Proceedings of Computer Vision-ECCV 2020, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2020: 86-103.

[27] LI J, XIE H, LI J, et al. Frequency-aware discriminative feature learning supervised by single-center loss for face forgery detection[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 5039-5049.

[28] ZHAO H, WEI T, ZHOU W, et al. Multi-attentional deepfake detection[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 5039-5049.

[29] SICILIA A, ZHAO X, HWANG S J. Domain adversarial neural networks for domain generalization: When it works and how to improve[J]. Machine Learning, 2023, 112: 3605-3624.

[30] PAN S J, YANG Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010,22(10): 1345-1359.

[31] ZHAO H, LIU B, HU Y, et al. Hybrid domain meta-learning network for face forgery detection and localization in deepfakes[C]//Proceedings of 2023 International Joint Conference on Neural Networks (IJCNN). Gold Coast, Australia: IEEE, 2023.

[32] XU N, FENG W. MetaFake: Few-shot face forgery detection with meta learning[C]//Proceedings of Proceedings of the 2023 ACM Workshop on Information Hiding and Multimedia Security. New York, USA: ACM, 2023.

[33] ZI B, CHANG M, CHEN J, et al. WildDeepfake: A challenging real-world dataset for deepfake detection [EB/OL]. (2021-05-07)[2024-08-01]. https://arxiv.org/abs/2105.03256.

[34] WANG B, LI Y, WU X, et al. Face forgery detection based on the improved siamese network[J]. Security and Communication Networks, 2022(1): 1-13.

[35] LI H, LIU Y, LIN Y, et al. UniHDA: Towards universal hybrid domain adaptation of image generators [EB/OL]. (2024-01-23)[2024-08-01]. https://arxiv.org/abs//2401.12596.

[36] HALIASSOS A, VOUGIOUKAS K, PETRIDIS S, et al. Lips don't lie: A generalisable and robust approach to face forgery detection[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021.

（编辑：刘彦东）