第 59 卷 第 16 期/2022 年 8 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于 CutMix 算法和改进 Xception 网络的深度伪造检测研究

耿鹏志[1]，唐云祁[1*]，樊红兴[2]，朱新同[1]

[1]中国人民公安大学侦查学院，北京 100038；

[2]中国科学院自动化研究所智能感知与计算研究中心，北京 100190

摘要 随着深度伪造技术的发展，生成的图片视频质量越来越逼真，给社会带来了巨大的安全风险。针对现有的检测方法参数量大、网络较深、模型结构复杂等情况，首先对取证领域的经典检测模型 XceptionNet 进行优化，提出一种轻量化的取证模型 Xcep_Block8，在减少模型参数量的同时，仍保持较高的检测精度。其次，针对类别不均衡问题，通过提高较少类别样本的采样概率，较好地解决了正负样本不均的情况。最后使用混合式数据增强方法 CutMix 增强样本之间的线性表达。实验结果表明，所提模型的测试结果较基线结果提升约 1.01 个百分点，同时在参数量方面较其他方法也有一定优势。

关键词 机器视觉；深度伪造；伪造检测；Xception 网络；混合式数据增强

中图分类号 **文献标志码** **DOI**：10.3788/LOP202259.1615007

Deep Forgery Detection Using CutMix Algorithm and Improved Xception Network

Geng Pengzhi[1], Tang Yunqi[1*], Fan Hongxing[2], Zhu Xintong[1]

[1]*School of Criminal Investigation, People's Public Security University of China, Beijing* 100038, *China*;

[2]*Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing* 100190, *China*

Abstract The rapid development of deep forgery technology has improved the quality of generated pictures and videos to mirror reality. However, it has brought huge security risks to society. In view of the large parameters used in existing detection methods, deep network, complex model structure, etc., this paper first optimizes the classic detection model XceptionNet in the forensics field and proposes a lightweight forensic model Xcep_Block8 that reduces the model parameters while maintaining high detection accuracy. Second, we improve the solution of the unevenness of positive and negative samples by increasing the sampling probability of samples with fewer categories to solve the problem of unbalanced categories. Finally, we employ the hybrid data enhancement method, CutMix, to improve the linear expression between samples. The experimental results show that the test results of the proposed model are about 1.01 percentage points higher than the baseline results. Additionally, it has certain advantages compared with other methods in terms of parameter quantity.

Key words machine vision; DeepFake; DeepFake detection; Xception network; CutMix

1 引 言

数字图像已经成为人类日常生活不可分割的一部分。但是近几年来，随着计算机视觉技术的发展，深度伪造技术随之诞生，它的出现导致图像编辑技术向智能化、自动化方向发展。近期臭名昭著的 DeepFake 就是该技术的产物，用它制作的虚假视频和图片不仅给当事人财产和名誉带来了巨大的损失，同时让人们对数字视频图像产生了信任危机，降低了电子证据的司法公信力。鉴于此，国内外科研团队已经对深度伪造

收稿日期：2021-07-08；**修回日期**：2021-07-21；**录用日期**：2021-07-28

基金项目：中央高校基本科研业务费项目（2021JKF203）、上海市现场物证重点实验室开放课题基金（2021XCWZK04）

通信作者：*tangyunqi@ppsuc.edu.cn

检测进行了深入研究。同时为了更好地促进深度伪造检测技术的发展，FaceBook豪掷百万美金举办了Deepfake Detection Challenge（DFDC）比赛[1]，南洋理工大学举办了DeeperForensics Challenge比赛[2]，并且国内也举办了相应的比赛，如GeekPwn举办了CAAD虚假人脸AI识别大赛[3]，中国科学院自动化研究所举办了Deepfake Game Competition（DFGC）[4]。在推进检测技术发展的同时，各国也在立法层面做出了相应的努力，如中国的《网络音视频信息服务管理规定》[5]和美国的《2019年深度伪造报告法案》[6]等。

目前，针对DeepFake的检测技术可以分为基于时序信息的检测技术和基于单张图像的检测技术。在时序信息检测方面，Agarwal等[7]发现伪造视频存在面部表情不一致以及位置差异，所以对脸部进行建模，之后通过使用支持向量机（SVM）对面部提取运动特征进行分类。由于伪造视频和自然拍摄的视频在连续帧间的运动存在一定差异，两者的光流信息不同，所以Amerini等[8]使用VGG网络对帧间的光流差异信息进行检测。由于生成算法在不同光源条件下生成的伪造视频存在一定的缺陷性，Güera等[9]使用循环神经网络处理伪造视频的序列信息，主要通过不同帧间的光源差异进行检测。在图像检测方面，针对DeepFake的检测技术又可分为传统手工建模技术和数据驱动方法。传统手工建模技术中，Li等[10]对眨眼信息进行建模，Matern等[11]对人脸的全局、光照、几何位置等不一致进行特征提取，Yang等[12]使用面部和头部不一致性进行伪造检测，但是以上方法只能在一些特定的数据集中起作用。由于深度伪造视频是通过生成对抗网络（GAN）技术生成的，有研究人员从GAN生成的缺陷技术出发，如Nataraj等[13]、Li等[14]通过对图像和色度空间特征进行学习来检测伪造图像，Xuan等[15]通过噪声、滤波等后处理操作破坏低级特征，使模型学习更鲁棒的高级伪造特征。数据驱动方法主要使用卷积神经网络去拟合伪造数据集。目前已有许多经典的网络结构，如Afchar等[16]设计的MesoNet、Nguyen等[17]设计的胶囊网络、Tan等[18]设计的EfficientNet、Dang等[19]通过注意力机制来增强对伪造视频的学习、Rössler等[20]使用的XceptionNet[21]。其中XceptionNet由于出色的检测性能，已成为该领域的经典检测模型。目前已有研究人员通过模型集成[22-23]、细粒度分类[24]等方法对XceptionNet进行改进，进而提升检测效果。但这些方法均以增加网络结构、提升模型复杂度为前提来提升检测效果，导致检测模型参数量较大。

基于此，本文对XceptionNet结构进行优化。总体来说，主要有3方面贡献：XceptionNet较为复杂且参数量较大，高达20.81×10^6，本文对该网络进行优化，选取Block8作为特征提取层，使得参数量减少了11.63×10^6；针对取证领域数据集中类别不平衡的情况，通过随机采样真实图片增加类别较少的样本来解决正负样本分布不均的问题，在网络训练中正负样本的比例大致为1∶1，并通过实验证明了该方法的有效性；使用混合式数据增强方法CutMix[25]对FaceForensics++数据集[20]进行处理，增强样本之间的线性表达，增加了模型鲁棒性，提升了模型的检测精度。

2 基于CutMix算法和改进Xception网络的深度伪造检测模型

2.1 CutMix数据增强策略

CutMix数据增强策略是由Yun等[25]于2019年提出的，属于混合式数据增强方法，可以有效地解决过拟合问题。其主要思路是对图像A的一部分区域进行随机裁剪，然后将裁剪区域填充到图像B的对应区域中，进而达到扩充数据样本的效果，效果如图1所示。

图1 CutMix增广图像示例。(a)(b)原始样本；(c)增广后的样本

Fig. 1 CutMix augmented image example. (a)(b) Original samples; (c) augmented sample

CutMix对特征和标签同时进行线性插值，进而增强训练样本的线性表达，具体方法如下：

$$\begin{cases} x = M \odot x_{\mathrm{A}} + (1-M) \odot x_{\mathrm{B}} \\ y = \lambda y_{\mathrm{A}} + (1-\lambda) y_{\mathrm{B}} \end{cases}, \tag{1}$$

式中：$M \in \{0,1\}^{W\times H}$为二进制掩码，用于表示图像中删除和填充的位置；$W$和$H$分别代表图像的宽和高；$\odot$表示像素相乘；$\lambda \in [0,1]$，其取值遵循Beta分布$\beta(\alpha,\alpha)$，$\alpha \in (0,\infty)$，本文$\alpha=1.0$。

在对掩码M进行采样的过程中，首先需要选取裁剪区域$C=(r_x,r_y,r_w,r_h)$，其中r_x，r_y是裁剪框C左上

角的随机坐标点，$r_x \sim (0, W)$，$r_y \sim (0, H)$。r_w 和 r_h 的计算公式为

$$\begin{cases} r_w = W\sqrt{1-\lambda} \\ r_h = H\sqrt{1-\lambda} \end{cases}, \tag{2}$$

即裁剪框的比例为 $\frac{r_w r_h}{WH} = 1-\lambda$。在确定好裁剪框 C 后，设置 M 中的区域 C 的掩码为 0，其余掩码为 1，之后通过式(1)运算得到图像 x 和标签 y。

2.2　类别平衡化处理

在深度伪造取证领域，由于生成方法的多样性，真实视频数量小于伪造视频，这会使检测模型在训练过程中会遇到正负样本不均衡的问题。如果使用随机采样会加剧这种不平衡现象，这不仅会导致模型在训练中提取的真实数据特征较少，让模型容易过拟合，并使预测的结果更倾向于伪造分类，进而造成错检情况的发生。针对该问题，使用类别平衡化策略，通过增大真实图片的采样频率来补充训练数据，最后使在网络训练中正负样本的采样概率都大致为 0.5，以解决正负样本分布不均的问题，具体原理如图 2 所示。

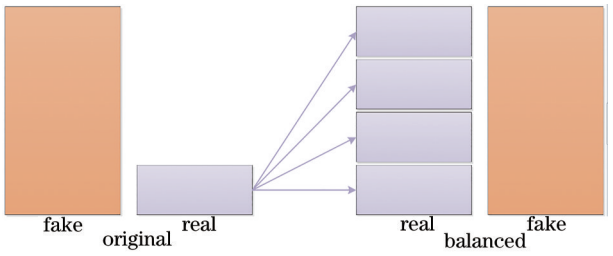

图 2　不平衡数据集的采样器

Fig. 2　Sampler for unbalanced data sets

2.3　网络结构

2.3.1　XceptionNet 模型

XceptionNet 作为对 Inception[26] 的改进，主要使用深度可分离卷积替换 Inception 中的卷积。深度可分离卷积是由深度卷积(Depthwise Convolution)和逐点卷积(Pointwise Convolution)两部分组成的，与能同时学习空间和通道相关性的一般卷积滤波器不同的是，它可以将通道和空间相关性分离开来，即先完成空间映射，然后再进行通道相关性的学习。相比于 Inception V3，XceptionNet 在没有增加网络参数量的前提下，提升了模型的性能。

2.3.2　Xception 网络模型的改进

XceptionNet 是通过深度可分离卷积的线性叠加构成的，可分为 Entry flow、Middle flow、Exit flow 三部分，共计 14 个模块，除了第一个和最后一个模块，其余模块均使用残差连接，具体结构如表 1 所示。

表 1　XceptionNet 结构

Table 1　XceptionNet structure

Input size	Operator		Number of channels
299×299×3	Entry flow	Conv1, 2×2	32
149×149×32		Conv2, 3×3	64
147×147×64		Block1	128
74×74×128		Block2	256
37×37×256		Block3	728
19×19×728	Middle flow	Block4, 3×3	728
19×19×728		Block5, 3×3	728
19×19×728		Block6, 3×3	728
19×19×728		Block7, 3×3	728
19×19×728		Block8, 3×3	728
19×19×728		Block9, 3×3	728
19×19×728		Block10, 3×3	728
19×19×728		Block11, 3×3	728
19×19×728	Exit flow	Block12	1024
10×10×1024		SeparableConv2d, 3×3	1536
10×10×1536		SeparableConv2d, 3×3	2048
10×10×2048		Pool, 1×1	

XceptionNet 较为复杂，在伪造检测任务中，可能存在一定的参数冗余情况。所以对 XceptionNet 进行改进，通过删减 XceptionNet 层数，降低模型的参数量，实验结果表明，当网络层数删减到 Block8 时，即将 Block8 作为特征提取层，检测精度较好，同时模型参数量大大减少，因此选取 Block8 作为特征输出层，并将网络命名为 Xcep_Block8。相比于原始网络结构的 2048 输出通道数，改进后的网络最后输出的通道数为 728。所提网络结构如图 3 所示。

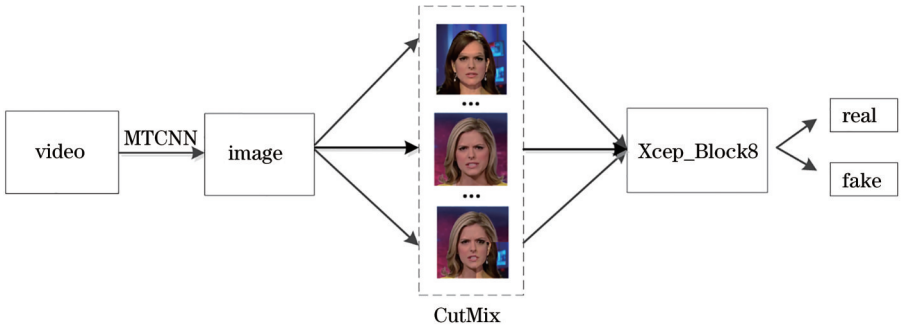

图 3　所提网络结构

Fig. 3　Proposed network structure

3　实验结果及分析

3.1　实验准备

3.1.1　实验环境配置

实验平台为 linux 操作系统，GPU 为 2 块 TITAN X（Pascal）显卡，搭配 Intel（R）Xeon（R）CPUE5-2650v4@2.20 GHz 处理器。代码均在 Pytorch1.2 框架下实现。为提高检测性能，所用算法均使用迁移学习策略，即使用 ImageNet 预训练模型。图片大小设置为 299×299。学习率调整策略为 Adam。超参数设置为：学习率为 0.0001，WeightDecay（权重衰减）为 0.001，Batchsize 为 40，epoch 为 15。并且为了保证实验结果的稳定性，每次实验均设置相同的随机种子。

3.1.2　实验数据

为了更好地评价所提方法的有效性，在 FaceForensics++数据集上进行实验。该数据集于 2019 年提出，是人脸的面部伪造视频数据集。该数据集包括 Face2Face、FaceSwap、DeepFakes 和 NeuralTextures 四种篡改方法，前 2 种伪造数据集是基于计算机图形方法的，后两种数据集是基于深度学习方法的。每一种篡改方法有 1000 个视频，每一个篡改视频均有对应的篡改关系，加上原始视频，共计有 5000 个视频。

首先根据官方给定的 Json 文件对数据集进行划分，将数据集划分为 Real 和 Fake 两组，每组又分别划分成对应的训练集、测试集和验证集。这样做有两方面好处，一方面可以避免学习高级人脸身份信息，另一方面可以使实验结果更具有一定的说服力。之后先对每个视频等间隔截取 10 帧，然后使用多任务卷积神经网络（MTCNN）[27]获取人脸框，并向外扩张 0.3 保存。最终共制作 50000 张图片作为实验样本，实验数据如表 2 所示。

3.1.3　评价指标

同当前主流方法一致，将深度伪造检测视为一个二分类问题。为了更好地评价所提方法的性能，使用的评价指标为 Accuracy 和 Logloss（BCELoss），公式为

$$A=\frac{N_{\text{right}}}{N_{\text{all}}},\tag{3}$$

表 2　本文的数据集

Table 2　Description of dataset

Dataset	Number of fake images	Number of real images
Train dataset	28800	7200
Test dataset	5600	1400
Validation dataset	5600	1400

$$L_{\text{Log}}=-\frac{1}{n}\sum_{i=1}^{n}\left[y_i\ln\left(y_i'\right)+\left(1-y_i\right)\ln\left(1-y_i'\right)\right],\tag{4}$$

$$L_{\text{total}}=\lambda L_{\text{ori}}+(1-\lambda)L_{\text{CutMix}},\tag{5}$$

式中：N_{right} 为正样本数量，即概率大于 0.5 的样本；N_{all} 和 n 均为样本总量；y_i' 为检测伪造图片的置信度，范围为（0，1）；y_i 为测试图像的真实标签，如果为真实人脸，则为 1，反之，为 0；L_{ori} 是对原始标签的损失优化；L_{CutMix} 是 BCELoss 对增强后的数据的损失优化；λ 是式（1）中的参数，作为比例系数，其数值服从 Beta 分布。

同时为了更好地验证所提方法的优越性，还使用 area under curve（AUC）和参数量（Parameters）作为评价指标，AUC 值可以更直观地区分不同方法的优劣性，模型参数量可评价模型文件的大小。

3.2　实验结果分析

3.2.1　模型的优化实验

XceptionNet 模型较为复杂且网络结构较深，参数量较大，所以针对该问题，对 XceptionNet 进行了优化。从表 1 可知，原 XceptionNet 最后几层的输出维度分别为 2048、1536、1024，远大于 728 维度。所以为降低参数量，优化 XceptionNet 模型，通过删除所选取特征层后的网络层，保留作为特征提取层之前的网络层（以该层名称进行命名），分别为 Block6、Block7、Block8、Block9、Block10，并在对应的特征提取层后使用全局平均池化（global average pooling）和全连接层进行特征映射。搭建好网络后分别在 FaceForensics++数据集上进行训练，最终实验结果如表 3 所示。

从实验结果可以看出：优化后的模型的检测精度并没有出现大幅度下降，这说明过多网络层数可能对提升深度伪造检测精度作用不是很大；Block8 作为特征提取层，检测效果最好，精度为 0.8721，相比于

表 3　模型的优化实验

Table 3　Model optimization experiment

Description	Logloss	Accuracy	Parameters /10^6
Block6	0.5539	0.8554	5.95
Block7	0.5386	0.8631	7.56
Block8	0.5258	0.8721	9.18
Block9	0.5185	0.8684	10.79
Block10	0.5398	0.8687	12.41
XceptionNet	0.5497	0.8757	20.81

XceptionNet，虽然检测精度降低 0.36，但其全连接层特征维度由[2048，1]变为[728，1]，特征维度大大减少，并且参数量不到 XceptionNet 的 1/2。所以选取 Block8 作为特征提取层，在不影响检测精度的情况下，使得模型参数量大大减少。

3.2.2　与其他算法的比较

为进一步证明所提优化模型的有效性，选取了该领域内基于深度学习的经典算法作为比较对象。为了保证实验结果有效性，所有方法均在本文的数据集上进行重新训练，并保持实验参数和实验条件一致。最终实验结果如表 4 所示。

表 4　所提模型与其他经典算法的对比

Table 4　Comparison between the proposed model and other classical algorithms

Method	Model	Logloss	Accuracy	Parameters /10^6
Method in Ref. [18]	EfficientNet_b3	0.5840	0.8803	12.23
Method in Ref. [28]	ResNet50	0.5413	0.8684	25.56
Method in Ref. [10]	SPPNet	0.8092	0.8660	25.64
Method in Ref. [20]	XceptionNet	0.5497	0.8757	20.81
Proposed method	Xcep_Block8	0.5258	0.8721	9.18

从表 4 可以看出：EfficientNet 作为神经架构搜索设计的模型，检测精度高达 0.8803，比 FaceForensics＋＋中的基线模型高出 0.46 个百分点，同时在参数量方面也有一定的优势；所提方法的检测精度为 0.8721，与其他方法的检测精度相差不大，可以较好地对低分辨率的篡改图像进行检测，但是参数量仅有 9.18×10^6，低于其他方法，同时在 AUC 值方面，如图 4 所示，也证明了所提优化模型的有效性。

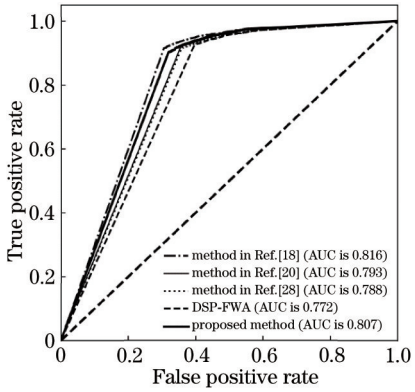

图 4　不同模型在验证集上的 ROC 曲线和 AUC 值

Fig. 4　ROC curves and AUC values of different models on the validation set

3.2.3　样本类别不平衡的消融实验

随着计算机视觉技术的发展，出现了多种伪造生成技术，使得伪造视频与真实视频的数量存在不均衡的情况，这在 FaceForensics＋＋数据集中表现明显，在该数据集中真实视频为 1000 个，而伪造视频数量多达 4000 个，存在明显的样本不均衡情况。所以针对该问题，使用过采样策略进行优化，该方法主要将真实样本的采样数量扩大 3 倍，使得真假类别被采样的几率相同，结果如表 5 所示。从表 5 可以看出，使用类别均衡方法训练的检测模型，即 Xcep_Block8 和 XceptionNet，检测精度均有一定程度的提升，检测精度分别提升 0.3 个百分点和 0.22 个百分点，同时 Logloss 评价指标也有明显的下降，分别下降 0.0929 和 0.0772，证明了类别均衡方法的有效性。

表 5　针对样本类别不平衡的改进

Table 5　Improvements for the imbalance of sample categories

Description	Logloss	Accuracy
Xcep_Block8	0.5258	0.8721
Xcep_Block8＋Over sampling	0.4329	0.8751
XceptionNet	0.5497	0.8757
XceptionNet＋Over sampling	0.4725	0.8779

3.2.4　混合式数据增强策略的检测效果

为了弥补训练数据集不足，提升检测模型的效果，选取混合式数据增强策略 CutMix。CutMix 随机地将样本的区域填充为其他数据中的区域像素值，而非像 Cutout 一样简单地将像素设置为一个固定值（该方法会使得训练数据中有许多无效信息），之后将分类标签均按比例分配，最终对伪造图片和真实图片进行混合处理，使得数据的标签不仅仅为真假两类，最终增强了对训练样本的线性表达。为进一步突出 CutMix 算法的有效性，选择了另一种混合式数据增强策略 Mixup[29] 和 Cutout 进行对比。Mixup 将随机的真假图片按比例混合，Cutout 将像素设置为一个固定值。由于混合系数由 Beta 分布计算所得，所以对混合系数的超参数进行了探究，超参数主要有 2 个，分别为 α 和 p。其中 α 是一个服从 Beta 分布的超参数，该参数控制了特征和标签之间的差值强度，将其设置为 0.5 和 1.0。

p为增强的概率，设置为0.7，0.8，0.9，1。实验结果如表6和图5所示，实验结果表明，当CutMix超参数条件为α=1和p=0.9，或者Mixup超参数为α=0.5和p=0.7时，检测效果最好。

表6　不同参数设置对混合式数据增强结果的影响

Table 6　Influence of different parameter settings on hybrid data enhancement results

Description	CutMix		Mixup	
	Logloss	Accuracy	Logloss	Accuracy
α=0.5, p=0.7	0.3271	0.8760	0.3196	0.8780
α=0.5, p=0.8	0.3270	0.8759	0.3080	0.8724
α=0.5, p=0.9	0.3334	0.8734	0.3117	0.8737
α=0.5, p=1	0.3097	0.8819	0.3170	0.8731
α=1, p=0.7	0.3210	0.8743	0.3151	0.8683
α=1, p=0.8	0.3376	0.8767	0.3162	0.8704
α=1, p=0.9	0.3122	0.8822	0.3147	0.8686
α=1, p=1	0.3185	0.8773	0.3211	0.8667

表7和图6分别为数据增强后的实验结果和效果图。从表7可以看出：两类方法均可以降低Logloss，提升了模型的检测效果；其中CutMix提升较大，相比于基线模型Xcep_Block8，检测精度提升0.71个百分点，LogLoss下降0.1207，与表4中效果最好的EfficientNet_b3相比，Logloss下降0.2718，检测精度提升0.0019；相比于基线模型Xcep_Block8，Cutout和Mixup的检测精度提升约0.09和0.29个百分点，低于CutMix，这可能是由于CutMix既有Mixup混合式的特点，同时添加的区域也保留了遮挡式数据增强的特点，提高了训练的效率，实验结果证明了CutMix算法的有效性。

表7　数据增强的实验结果

Table 7　Experimental results of data augmentation

Description	Logloss	Accuracy
Baseline(Xcep_Block8)	0.4329	0.8751
＋Cutout(size is 50)	0.5143	0.8760
＋Cutout(size is 80)	0.5228	0.8750
＋Cutout(size is 110)	0.4674	0.8744
＋Mixup	0.3196	0.8780
＋CutMix	0.3122	0.8822

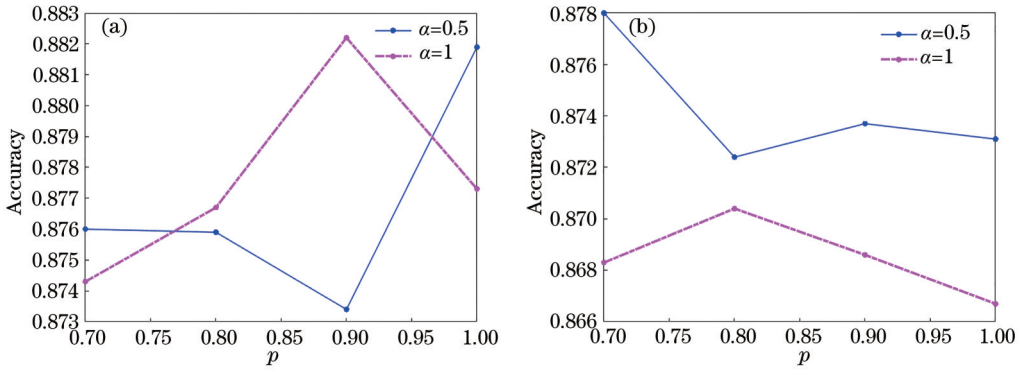

图5　超参数α和概率p对检测模型的影响。(a) CutMix；(b) Mixup

Fig. 5　Influence of hyper-parameter α and probability p on the detection model. (a) CutMix; (b) Mixup

图6　数据增强的效果

Fig. 6　Results of data enhancement

4　结　　论

对深度伪造检测经典算法XceptionNet进行了优化，并将所提模型与其他主流方法相比，在准确率相差不大的情况下，参数量大大减少，证明所提模型的有效性。之后通过随机采样真实图片增加类别较少的样本，来解决正负样本分布不均衡的问题。与此同时，为进一步增强检测效果和模型的鲁棒性，使用混合式数据增强方法CutMix训练检测模型，并对其超参数进行了探究，提升了模型的检测效果。未来的工作是设计更加轻量的伪造检测模型以及更适合于深度伪造检测的数据增强算法。

参 考 文 献

[1] FaceBook. Deepfake detection challenge(DFDC) [EB/OL]. [2021-03-02]. https://www.kaggle.com/c/deepfake-detection-challenge.

[2] ChallengeDeeperForensics 2020. @ECCV SenseHuman Workshop[EB/OL]. [2021-03-02]. https://competitions.codalab.org/competitions/25228.

[3] Geekpwn. 虚假人脸AI识别大赛[EB/OL]. [2021-03-02]. http://www.geekpwn.org/zh/index.html.
Geekpwn. Fake face AI recognition contest[EB/OL]. [2021-03-02]. http://www.geekpwn.org/zh/index.html.

[4] Peng B. DeepFake Game Competition (DFGC)@IJCB 2021[EB/OL]. [2021-03-02]. http://dfgc2021. iapr-tc4.org/.

[5] 国家互联网信息办公室、文化和旅游部、国家广播电视总局联合印发《网络音视频信息服务管理规定》[J]. 有线电视技术, 2019, 26(12): 8-9.
The State Internet Information Office, the Ministry of Culture and Tourism, and the State Administration of Radio and Television jointly issued the "regulations on the administration of network audio and video information services"[J]. CATV Technology, 2019, 26(12): 8-9.

[6] House-Energy and Commerce. Deep falsified report act of 2019[EB/OL]. [2021-01-02]. https://www.congress.gov/bill/116th-congress/house-bill/3600/.

[7] Agarwal S, Farid H, Gu Y, et al. Protecting world leaders against deep fakes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, June 14-19, 2020. New York: IEEE, 2020: 38-45.

[8] Amerini I, Galteri L, Caldelli R, et al. Deepfake video detection through optical flow based CNN[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), October 27-28, 2019, Seoul, Korea(South). New York: IEEE Press, 2019: 1205-1207.

[9] Güera D, Delp E J. Deepfake video detection using recurrent neural networks[C]//2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), November 27-30, 2018, Auckland, New Zealand. New York: IEEE Press, 2018: 18455926.

[10] Li Y Z, Chang M C, Lyu S W. In Ictu oculi: exposing AI created fake videos by detecting eye blinking[C]//2018 IEEE International Workshop on Information Forensics and Security (WIFS), December 11-13, 2018, Hong Kong, China. New York: IEEE Press, 2018: 18431664.

[11] Matern F, Riess C, Stamminger M. Exploiting visual artifacts to expose deepfakes and face manipulations[C]//2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), January 7-11, 2019, Waikoloa, HI, USA. New York: IEEE Press, 2019: 83-92.

[12] Yang X, Li Y Z, Lyu S W. Exposing deep fakes using inconsistent head poses[C]//2019 IEEE International Conference on Acoustics, Speech and Signal Processing, May 12-17, 2019, Brighton, UK. New York: IEEE Press, 2019: 8261-8265.

[13] Nataraj L, Mohammed T M, Manjunath B S, et al. Detecting GAN generated fake images using co-occurrence matrices[J]. Electronic Imaging, 2019, 31(5): 532.

[14] Li H D, Li B, Tan S Q, et al. Identification of deep network generated images using disparities in color components[J]. Signal Processing, 2020, 174: 107616.

[15] Xuan X S, Peng B, Wang W, et al. On the generalization of GAN image forensics[EB/OL]. (2019-02-27)[2021-02-03]. https://arxiv.org/abs/1902.11153.

[16] Afchar D, Nozick V, Yamagishi J, et al. MesoNet: a compact facial video forgery detection network[C]//2018 IEEE International Workshop on Information Forensics and Security (WIFS), December 11-13, 2018, Hong Kong, China. New York: IEEE Press, 2018: 18439519.

[17] Nguyen H H, Yamagishi J, Echizen I. Capsule-forensics: using capsule networks to detect forged images and videos[C]//2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 12-17, 2019, Brighton, UK. New York: IEEE Press, 2019: 2307-2311.

[18] Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[EB/OL]. (2019-05-28)[2021-03-06]. https://arxiv.org/abs/1905.11946.

[19] Dang H, Liu F, Stehouwer J, et al. On the detection of digital face manipulation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 5780-5789.

[20] Rössler A, Cozzolino D, Verdoliva L, et al. FaceForensics++: learning to detect manipulated facial images[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 19398261.

[21] 陈德刚, 艾孜尔古丽, 尹鹏博, 等. 基于改进Xception迁移学习的野生菌种类识别研究[J]. 激光与光电子学进展, 2021, 58(8): 0810023.
Chen D G, Azragul, Yin P B, et al. Research on identification of wild mushroom species based on improved Xception transfer learning[J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810023.

[22] 李旭嵘, 于鲲. 一种基于双流网络的Deepfakes检测技术[J]. 信息安全学报, 2020, 5(2): 84-91.
Li X R, Yu K. A Deepfakes detection technique based on two-stream network[J]. Journal of Cyber Security, 2020, 5(2): 84-91.

[23] 耿鹏志, 樊红兴, 张翌阳, 等. 基于篡改伪影的深度伪造检测方法[J]. 计算机工程, 2021, 47(12): 156-162.
Geng P Z, Fan H X, Zhang Y Y, et al. Deep forgery detection method based on tampering artifacts[J]. Computer Engineering, 2021, 47(12): 156-162.

[24] 周涛, 吕晓琪, 任国印, 等. 基于集成卷积神经网络的面部表情分类[J]. 激光与光电子学进展, 2020, 57(14): 141501.
Zhou T, Lü X Q, Ren G Y, et al. Facial expression classification based on ensemble convolutional neural network[J]. Laser & Optoelectronics Progress, 2020, 57

(14): 141501.

[25] Yun S, Han D, Chun S, et al. CutMix: regularization strategy to train strong classifiers with localizable features [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6022-6031.

[26] 王凯旋, 李卓容, 王晓宾, 等. 刑事案件现场图自动分类算法[J]. 激光与光电子学进展, 2020, 57(4): 041009.
Wang K X, Li Z R, Wang X B, et al. Automated classification method for crime scene sketches[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041009.

[27] Zhang K P, Zhang Z P, Li Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.

[28] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[29] Zhang H Y, Cisse M, Dauphin Y N, et al. Mixup: beyond empirical risk minimization[EB/OL]. (2017-10-25)[2020-06-15]. https://arxiv.org/abs/1710.09412.