第48卷 第2期
Vol.48 No.2
计 算 机 工 程
Computer Engineering
2022年2月
February 2022
·图形图像处理·
文章编号：1000-3428(2022)02-0194-07 文献标志码：A 中图分类号：TP391.41

基于自注意力胶囊网络的伪造人脸检测方法

李 柯，李邵梅，吉立新，刘 硕
（解放军战略支援部队信息工程大学 信息技术研究所，郑州 450003）

摘 要：当前以换脸为代表的伪造视频泛滥，给国家、社会和个人带来潜在威胁，有效检测该类视频对保护个人隐私和维护国家安全具有重要意义。为提高视频伪造人脸检测效果，基于可解释性好的胶囊网络，以Capsule-Forensics检测算法为基础，提出一种结合自注意力胶囊网络的伪造人脸检测方法。使用部分Xception网络作为特征提取部分，降低模型的参数量，在主体部分引入带注意力机制的胶囊结构，使模型聚焦人脸区域，将综合多维度的Focal Loss作为损失函数，提高模型对难分样例的检测效果。实验结果表明，与Capsule-Forensics算法相比，该方法能够减少模型参数量和计算量，在多种伪造类型数据集上均具有较高的准确率。

关键词：伪造人脸检测；胶囊网络；模型可视化；注意力机制；损失函数

开放科学（资源服务）标志码（OSID）：

中文引用格式：李柯，李邵梅，吉立新，等.基于自注意力胶囊网络的伪造人脸检测方法[J].计算机工程，2022，48(2)：194-200，206.

英文引用格式：LI K，LI S M，JI L X，et al.Method of face forgery detection based on self-attention capsule network[J].Computer Engineering，2022，48(2)：194-200，206.

Method of Face Forgery Detection Based on Self-Attention Capsule Network

LI Ke，LI Shaomei，JI Lixin，LIU Shuo
（Institute of Information Technology，People's Liberation Army Strategic Support Force Information Engineering University，Zhengzhou 450003，China）

【Abstract】In recent years，face forgery is abused in fake videos，imposing a potential threat on the national，social and individual level，so face forgery detection is of great significance to individual privacy protection and national security. To improve the performance of face forgery detection for fake videos，a face forgery detection method that combines a self-attention capsule network with the Capsule-Forensics algorithm is proposed.This method uses part of the Xception network for feature extraction，which reduces the number of parameters. Then a capsule structure with attention mechanism is introduced into the main part to make the model focus on the facial area.Finally，the comprehensive multi-dimensional Focal Loss is used as the loss function to improve the detection effect of the model for indistinguishable samples. Experimental results show that compared with the Capsule-Forensics algorithm，this method can reduce the number of model parameters and the amount of computation，while displaying a higher accuracy on multiple forgery data sets.

【Key words】face forgery detection；capsule network；model visualization；attention mechanism；loss function
DOI：10. 19678/j. issn. 1000-3428. 0060267

0 概述

随着深度学习技术的发展，以换脸为代表的人脸伪造视频开始在社交媒体上广泛传播，包括基于计算机图形学的方法FaceSwap、Face2Face[1]以及基于深度学习的方法DeepFakes、NeuralTextures[2]等主流的人脸伪造生成方法。该技术的滥用将对个人隐私和国家政治安全带来巨大的威胁。

针对当前以换脸为代表的伪造视频的危害，国内外学者对视频中伪造人脸的检测方法进行了大量研究。目前，视频伪造人脸检测的方法主要包含基于帧间特征的检测方法和基于帧内特征的检测方法两类。基于帧间特征的检测方法主要关注视频前后帧中的时序信息，文献[3]将视频每一帧的特征向量

基金项目：国家自然科学基金青年科学基金项目(62002384)。
作者简介：李 柯(1995—)，男，硕士研究生，主研方向为图像处理、深度学习、计算机视觉；李邵梅，副教授；吉立新，研究员、博士生导师；刘 硕，硕士研究生。
收稿日期：2020-12-14 **修回日期：**2021-01-20 **E-mail：**lishaomei_may@126.com

经过处理后，输入到LSTM网络中得到帧间特征，作为分类的依据。文献[4]利用视频相邻帧中人脸图像存在差异的特性提出一种检测方法。基于帧间特征的检测方法无法判断单帧的真伪，同时对待检测视频的长度存在需求，对模型训练条件的要求也较高，因此在实际应用中存在限制。

基于帧内特征的检测方法是采用传统卷积神经网络（Convolutional Neural Network，CNN）进行特征学习，训练分类器区分伪造与真实人脸图片。文献[5]利用伪造视频在合成时需要进行仿射变换才能匹配，从而导致人脸区域与周围环境分辨率不一致的特点训练分类器。文献[6]利用生成视频时存在的人为视觉上的缺陷进行判别，具体可包括合成图片存在的左右眼虹膜颜色不一致、在处理光照时对入射光的错误处理导致产生特定的伪影、对头发建模错误导致的空洞等问题，通过提取这些特征，训练较小的分类器完成分类任务。文献[7]发现在使用2D面部图像估计3D头部姿态（比如头的方向和位置）合成伪造图片时会存在误差，进而利用该特征进行判别。文献[8]使用集成多种CNN的网络进行视频中的伪造人脸检测。

针对CNN在图像处理中没有考虑部件间的相对位置、角度等信息以及可解释性较差等不足，文献[9]提出了基于动态路由的胶囊网络。随后胶囊网络被应用于计算机视觉的多个领域，文献[10]使用胶囊网络进行肺部CT图像分割检测，文献[11]将基于二元分类的胶囊网络应用于医学图像的乳腺癌检测，文献[12]尝试将胶囊网络运用于伪造视频检测领域，提出算法模型Capsule-Forensics，在多种类型的视频伪造人脸检测上均具有良好的性能。

为进一步提高胶囊网络对视频伪造人脸的检测效果，本文在Capsule-Forensic的基础上，提出一种基于自注意胶囊网络的视频伪造人脸检测算法Xception-Attention-Capsules。该算法使用部分Xception网络作为特征提取部分，在Primary Capsule部分构建Attention-Capsules结构，并在此基础上提出新的损失函数Dimension-Focal Loss。最后运用模型可视化技术对本文算法的检测结果进行分析。

1　Capsule-Forensics视频伪造人脸检测算法

胶囊网络中的胶囊结构是一组向量或矩阵的集合，不同输出向量代表图像中出现的特定实体的各种属性，例如输出向量的模长表示实体存在可能性的大小。胶囊网络使用向量输出代替值输出，可以在识别物体类别的同时保留其位置和姿态信息；使用动态路由算法[9]更新胶囊层间的权重，使得向量值能在不同的胶囊层间进行传递。因此，与CNN等传统神经网络相比，胶囊网络具有更好的泛化能力，同时对图形变换具有更强的鲁棒性。

NGUYEN等[12]提出基于胶囊网络的视频伪造人脸检测算法Capsule-Forensics，其网络结构如图1所示。Capsule-Forensics特征提取部分使用部分VGG-19网络；Primary Capsules部分使用了N个相同的胶囊结构，中间部分引入了Stats Pooling（Stattistic Pooling）；在Primary Capsules和Classifier Capsules间运用动态路由算法进行权重更新；Classifier Capsules中真假胶囊将分别输出一组N维向量，将两组N维向量每个维度的两两组合进行Softmax运算，最后取N个Softmax结果的均值作为最终的预测类别。训练过程中使用交叉熵（Cross Entropy，CE）损失函数。

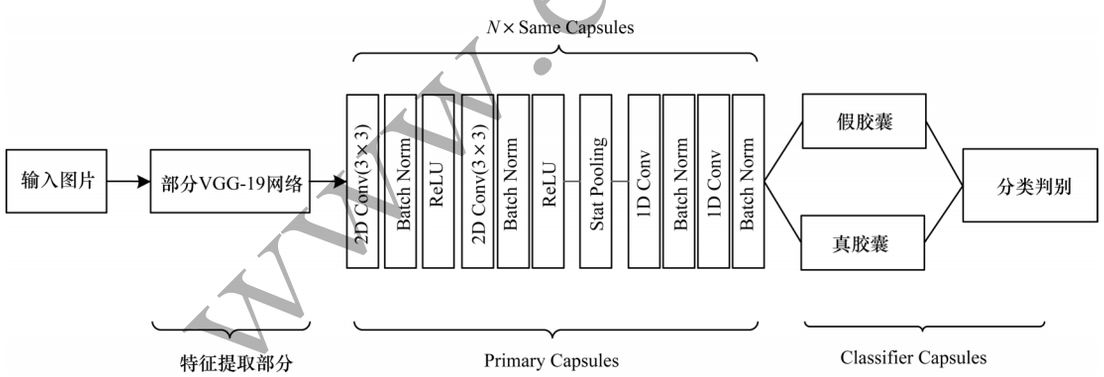

图1　Capsule-Forensics网络结构

Fig.1　Capsule-Forensics network structure

通过分析，本文认为该网络有以下不足：一是特征提取部分参数量较大，模型训练速度较慢；二是Primary Capsules中相同的胶囊结构不利于模型关注到人脸不同位置的特征；三是交叉熵损失函数未考虑数据集中正负样本不均衡的问题。

2　基于自注意力胶囊网络的伪造人脸检测技术

2.1　Xception-Attention-Capsules结构组成

本文设计的检测模型Xception-Attention-Capsules的网络结构如图2所示，分为特征提取部分、Primary Capsules、Classifier Capsules 3个部分。

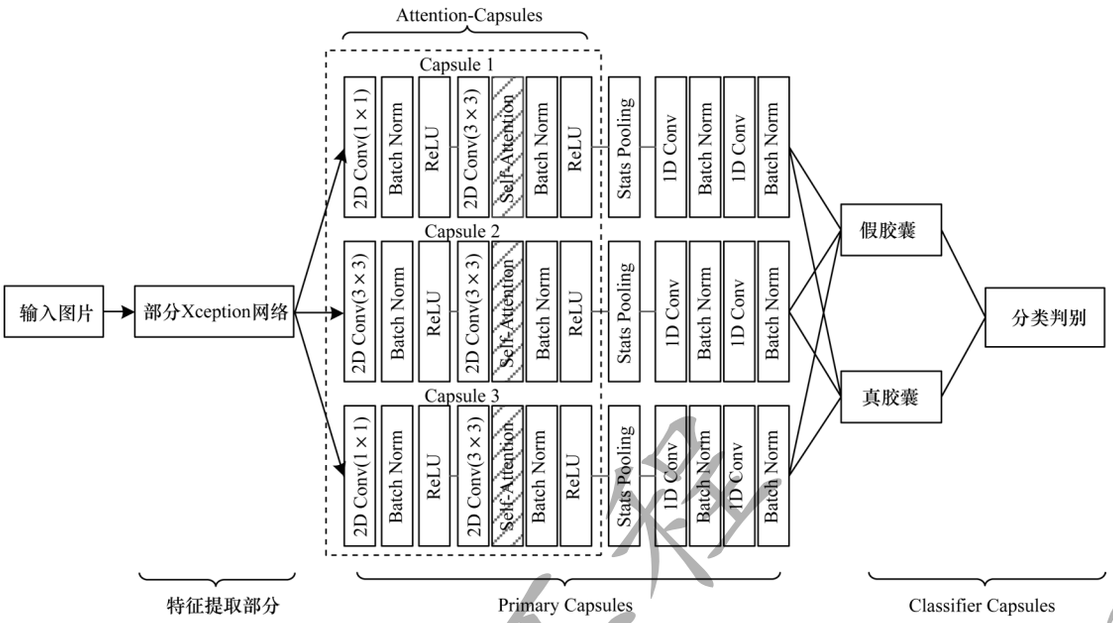

图 2 Xception-Attention-Capsules网络结构

Fig.2 Xception-Attention-Capsules network structure

相较Capsule-Forensics,本文进行了以下3个部分的改进:

1)特征提取部分使用部分Xception网络,该改进使模型在仅损失微小精度的情况下,相较原模型减少了85.7%的参数量,训练速度提高了57.3%。

2)Primary Capsules部分提出Attention-Capsules,该部分使用3个不同结构的胶囊模块,其中Capsule 1使用步长为1的1×1卷积核;Capsule 2使用步长为1的3×3卷积核;Capsule 3使用步长为2的1×1卷积核,以保留更多的人脸信息。每个胶囊模块中均加入了自注意力层Self-Attention。Primary Capsules和Classifier Capsules间使用动态路由算法[9]更新参数。

3)Classifier Capsules部分沿用了文献[12]类别预测的方法。通过对比分析输出向量维度为2维、4维、8维、16维的检测效果,本文将输出维度设置为4,使模型在准确率和收敛速度间的平衡达到最优。另外,提出新的损失函数Dimension-Focal Loss。

2.2 基于部分Xception网络的特征提取

Xception[13]主要由残差网络与深度可分离卷积组成。Xception包括36个卷积层,共包含3个flow,分别是Entry flow、Middle flow、Exit flow;包括14个Block,其中Entry flow中4个、Middle flow中8个、Exit flow中2个;中间12个Block都包含线性残差连接。同时,模型对输入数据的每一个通道分别进行空间逐层卷积,再对其结果进行逐点卷积。如图3所示,Xception模型第1步通过1×1卷积进行通道分离,第2步独立绘制每个输出通道的空间相关性,用3×3卷积单独处理,最后合并。

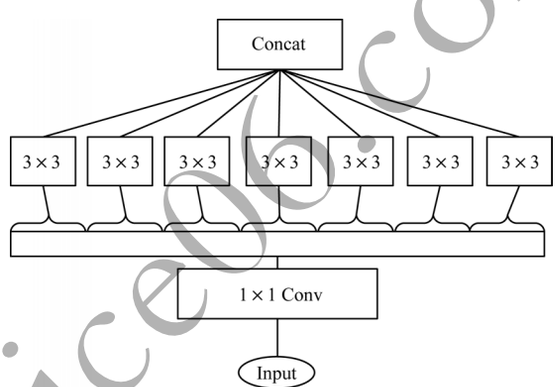

图 3 Xception基本模块

Fig.3 Xception basic module

作为伪造视频检测数据集FaceForensics++[14]的基线模型,Xception在该数据集上具有良好的效果。为此,本文选取Xception中Entry flow的前3个Block模块作为Xception-Attention-Capsules的特征提取部分。

2.3 基于自注意力机制的Attention-Capsules

伪造人脸大部分的操作作用于面部五官区域,面部平坦区域的信息对分类的帮助较少。文献[15]指出在神经网络中引入注意力机制,有助于使模型关注具有丰富细节的面部五官部分。

在传统神经网络中,由于每个卷积核的尺寸有限,因此每次卷积操作只能获得像素点周围很小一块邻域,不易捕捉到距离较远的特征[16]。为此,本文采用多个胶囊模块,每个模块中使用大小不同的卷积核,利用不同结构的胶囊模块关注到不同尺寸的特征。另外,本文提出在胶囊中加入Self-Attention层[17],通过计算图像中任意两个像素点之间的关系,

从而学习到某一像素点和其他所有位置(包括较远位置)的像素点之间的关系,这样可以捕获到更大范围内像素间的关系,获取图像的全局几何特征。

Self-Attention自注意力模块结构如图4[17]所示,其原理是将输入分别通过1×1的卷积层;然后在输出特征图的矩阵上进行运算得到注意力图;最后的注意力模块部分将逐渐学习到如何将注意力特征图加在原始的特征图上,从而得到增加了注意力部分的特征图。

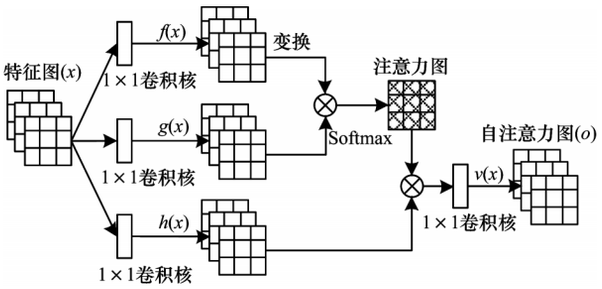

图4　自注意力模块结构

Fig.4　Self-attention module structure

2.4　Dimension-Focal Loss损失函数

伪造人脸数据集普遍存在正负样本不均衡问题,且不同的伪造生成方法生成的视频检测难易也存在差别。因此,本文在Focal Loss[17]损失函数的基础上提出了Dimension-Focal Loss损失函数。该方法使用真假胶囊每个批次输出维度损失的均值之和作为该批次最终的损失。相较于文献[9]仅将最终预测结果与标签做交叉熵的损失计算方法,该方法可以考虑到每个维度的输出值,并将其体现到损失函数上,更精准地指导Classifier Capsule进行学习。

Dimension-Focal Loss的计算步骤如下:分别计算单个样本的单维度Focal Loss损失,如(1)所示:

$$L_{fl}^{ij}=\begin{cases}-\alpha(1-y')^{\gamma}\log_a y', y=1\\-(1-\alpha)y'^{\gamma}\log_a(1-y'), y=0\end{cases}\tag{1}$$

其中:α表示平衡因子;γ表示难易样本权重;L_{fl}^{ij}表示计算的单个样本的单维度损失;i表示每批次对应的单个样本编号;j表示当前运算的维数;y表示预测标签;y'表示经过激活函数Softmax的输出。

在得到L_{fl}^{ij}后,最终损失的计算如式(2)和式(3)所示:

$$L_{\text{mean}}^{j}=\frac{1}{n}\sum_{i=1}^{n}L_{fl}^{ij}\tag{2}$$

$$\text{loss}=\sum_{j=1}^{m}L_{\text{mean}}^{j}\tag{3}$$

其中:n为批次大小;m为当前输出胶囊的总维数;j表示当前运算的维数;L_{mean}^{j}代表批次维度平均损失;loss代表最终损失。

3　实验结果与分析

3.1　数据集

为评估本文方法的性能,本文在公开数据集FaceForensics++[14]上进行实验。FaceForensics++包含从国外视频网站采集的1 000个真实视频。每条原始视频时长约为15 s,然后使用计算机图形学和深度学习方法生成伪造视频。使用的伪造生成方法包含Face Swap、Fac2Face、Fac2Face、NeuralTextures 4种类型,每种方法各生成1 000条假视频。FaceForensics++数据集将所有视频依据H.264编码进行压缩,共生成了3个版本,其中Raw版本表示未进行压缩,HQ(High Quality)版本的压缩参数为23,LQ(Low Quality)版本的压缩参数为40。

3.2　实验设置

3.2.1　数据处理

由于该数据集中视频的伪造区域集中在人脸面部,因此提前使用人脸检测方法检测并提取面部区域,将其单独保存作为训练集和测试集的输入有助于使分类算法集中于被篡改的面部区域,减弱背景的影响,提高检测算法的收敛速度和分类精度。本文使用RetinaFace[18]作为人脸检测算法,首先将每条视频流按帧进行截取,然后利用算法识别每帧中得分最高的人脸,以人脸为中心裁剪后并缩放至320×320大小(该尺寸易于执行裁剪和缩放)分别保存在对应的视频文件夹中。

在训练模型的过程中,对其增加一定概率的水平翻转和垂直翻转以及随机的图片切割。目的是增加数据集的随机扰动,同时避免模型对图像部分区域的过度关注,以提高模型的泛化能力。最后将图片缩放至指定的大小,以符合模型的输入要求,本文统一使用224×224的输入。

本文在对比测试中分别对不同压缩版本的伪造视频进行了单独实验和混合实验。在优化方法对比测试中使用了网络上最为普遍的HQ版本进行实验。视频级则使用了LQ和HQ两个版本的视频进行测试,每个视频每隔10帧抽样进行检测,综合10帧的检测结果对视频的真伪进行判别。参考文献[12]中的实验数据划分,分别设置训练集720条,验证集和测试集各为140条,本文实验中FaceForensics++中每个类别的训练集和测试集数量如表1所示,其中新增Mix类别为4种伪造类型的混合。

表1　FaceForensics++中每个类别训练集和测试集图片数

Table 1　Uumber of images in the training set and test set of each category in FaceForensics++

类别	训练集	测试集
Mix	151 200	5 235
Deep Fakes	32 027	5 235
Face Swap	32 027	5 235
Fac2Face	32 027	5 235
NeuralTextures	32 027	5 235

3.2.2　评价指标

为方便与其他算法对比,本文实验使用加权平均准确率(Weighted Average Accuracy,WAA)作为评价指标。在正负样本不平衡的数据集中,该指标可以更好地评估模型的分类能力。

计算公式如式(4)和式(5)所示：

$$P_i=\frac{T_{\mathrm{TP}}}{T_{\mathrm{TP}}+F_{\mathrm{FP}}} \tag{4}$$

$$W_{\mathrm{WAA}}=\sum_{i=1}^{n}\frac{c_i}{C}P_i \tag{5}$$

其中：TP(True Positive)表示将假样例正确识别为假样例的数量；FP(False Positive)表示将真样例错误识别为假样例的数量；P_i表示类别i的准确率；C表示样本的总数；c_i表示第i类类别的样本总数；n表示类别总数。

3.2.3 实验配置

本文实验平台为Ubuntu 16.04操作系统，使用了4块Nvidia TitanX显卡，所有代码都在PyTorch框架下实现。使用在ImageNet上预训练的权重对Capsule-Forensics和Xception-Attention-Capsules的特征提取部分进行初始化。

3.3 实验结果

3.3.1 Xception-Attention-Capsules消融实验

本文首先评估对Capsule-Forensics改进的合理性。训练和测试使用HQ版本，实验结果如表2所示。表2中的第2、3、4行分别为Xception-Capsules、Xception-Capsules+A、Xception-Capsules+F的结果，第5行为Xception-Attention-Capsules的结果。

1)Xception-Capsules：在Capsule-Forensics模型的基础上仅替换特征提取部分为Xception部分(如2.2节所述)。

2)Xception-Capsules+A：在Xception-Capsules的基础上增加Attention-Capsules(如2.3节所述)。

3)Xception-Capsules+F：在Xception-Capsules的基础上增加Dimension-Focal Loss(如2.4节所述)。

表2 Xception-Attention-Capsules消融实验的加权平均准确率

Table 2 Weighted average accuracy rate of Xception-Attention-Capsules ablation experiment %

模型	Mix	Deep Fakes	Face Swap	Fac2Face	NeuralTextures
Xception-Capsules	94.38	98.32	95.09	96.76	91.92
Xception-Capsule+A	95.63	98.91	96.42	97.63	92.02
Xception-Capsules+F	95.58	98.47	95.12	96.87	92.22
Xception-Attention-Capsules	96.32	98.94	96.96	97.92	93.29

如表2所示，增加不同优化机制的模型准确率Xception-Attention-Capsules>Xception-Capsules+A>Xception-Capsules+F>Xception-Capsules。可以看出，在增加Attention-Capsules和Dimension-Focal Loss优化方法后模型的准确率均有提升，其中Attention-Capsules对模型准确率的提升更加有效。在综合两种方法后，模型的准确率在5种类型数据集上分别有1.94%、0.62%、1.87%、1.16%、1.37%的提升，证明本文提出的改进方法是合理有效的。

3.3.2 Xception-Attention-Capsules与其他模型对比

为验证模型的改进效果，分别在FaceForensics++不同压缩率下进行了对比实验。不同伪造类别图片级和视频级实验结果如表3~表5所示，分别表示视频RAW、HQ、LQ版本的结果。

表3 FaceForensics++ RAW数据集上的加权平均准确率

Table 3 Weighted average accuracy rate on FaceForensics++ RAW datasets

模型	加权平均准确率/%					参数量
	Mix	Deep Fakes	Face Swap	Face2Face	NeuralTextures	
文献[12]模型	97.28	99.83	98.42	99.15	93.81	2 796 889
文献[14]模型	96.37	99.12	98.51	99.12	94.56	—
文献[15]模型	97.31	98.28	98.57	98.39	92.34	—
文献[19]模型	95.18	97.23	95.85	96.77	89.23	—
文献[20]模型	96.50	97.35	98.12	97.48	93.08	—
本文模型	98.35	99.49	98.89	98.94	94.76	400 821

表4 FaceForensics++HQ数据集上的加权平均准确率

Table 4 Weighted average accuracy rate on FaceForensics++HQ datasets %

模型	Mix	Deep Fakes	Face Swap	Face2Face	NeuralTextures	Video
文献[12]模型	95.14	98.84	96.38	98.21	91.32	—
文献[14]模型	94.45	98.77	96.57	97.76	92.25	—
文献[15]模型	94.47	96.63	95.69	97.82	89.38	—
文献[19]模型	92.57	94.57	92.87	92.46	86.25	—
文献[20]模型	93.26	96.50	96.51	95.63	91.89	—
本文模型	96.32	98.94	96.96	97.92	93.29	96.60

表5　FaceForensics++LQ数据集上的加权平均准确率

Table 5　Weighted average accuracy rate on FaceForensics++LQ datasets　%

模型	Mix	Deep Fakes	Face Swap	Face2Face	NeuralTextures	Video
文献[12]模型	89.65	97.58	91.26	95.78	80.63	—
文献[14]模型	89.13	97.24	93.15	96.83	81.51	—
文献[15]模型	87.40	94.27	92.98	95.80	80.62	—
文献[19]模型	82.45	93.38	84.58	88.64	75.38	—
文献[20]模型	86.59	93.58	92.19	92.64	79.92	—
本文模型	91.25	97.69	93.24	96.03	81.63	93.20

从以上实验结果可以看出:

1)本文改进的胶囊网络Xception-Attention-Capsules相较原模型参数量减少85.7%,在LQ数据集上分别提高1.6%、0.11%、1.98%、0.25%、1.0%。HQ数据集除Face2Face类型下降0.29%,其余类型分别提高1.18%、0.1%、0.58%、1.9%,表明本文改进方法在多数情况下能得到更高的准确率。

2)相较其他模型,本文模型在Mix、Face Swap、NeuralTextures数据集的不同压缩率下均具有最高的准确率。

3)在压缩率上升时,各模型在NeuralTextures和Face2Face伪造方法的准确率均有较大下降,说明这两种伪造类型的低质量图片较难判别。

4)对比本文方法在视频级和图片级测试的结果,视频级结果的检测准确率较图片级有所上升,说明本文模型在视频级判别时通过综合多帧的判别结果,可以提高判别的准确率。

3.4　有效性可视化分析

3.4.1　决策区域可视化分析

为分析本文Attention-Capsules中不同胶囊模块在鉴别伪造人脸时是否关注人脸不同位置,本文使用Grad-CAM[19]对FaceForensics++数据集中几种伪造方法生成的样本进行可视化分析。

对Face Forensics++的每种伪造方法,分别使用训练好的模型通过Grad-CAM生成热力图[21],结果如图5所示,其中,第1行是输入图片,其余行分别是3个胶囊的激活图,第1列是原始图片样例,其余列分别对应原始样例的4种伪造类型。

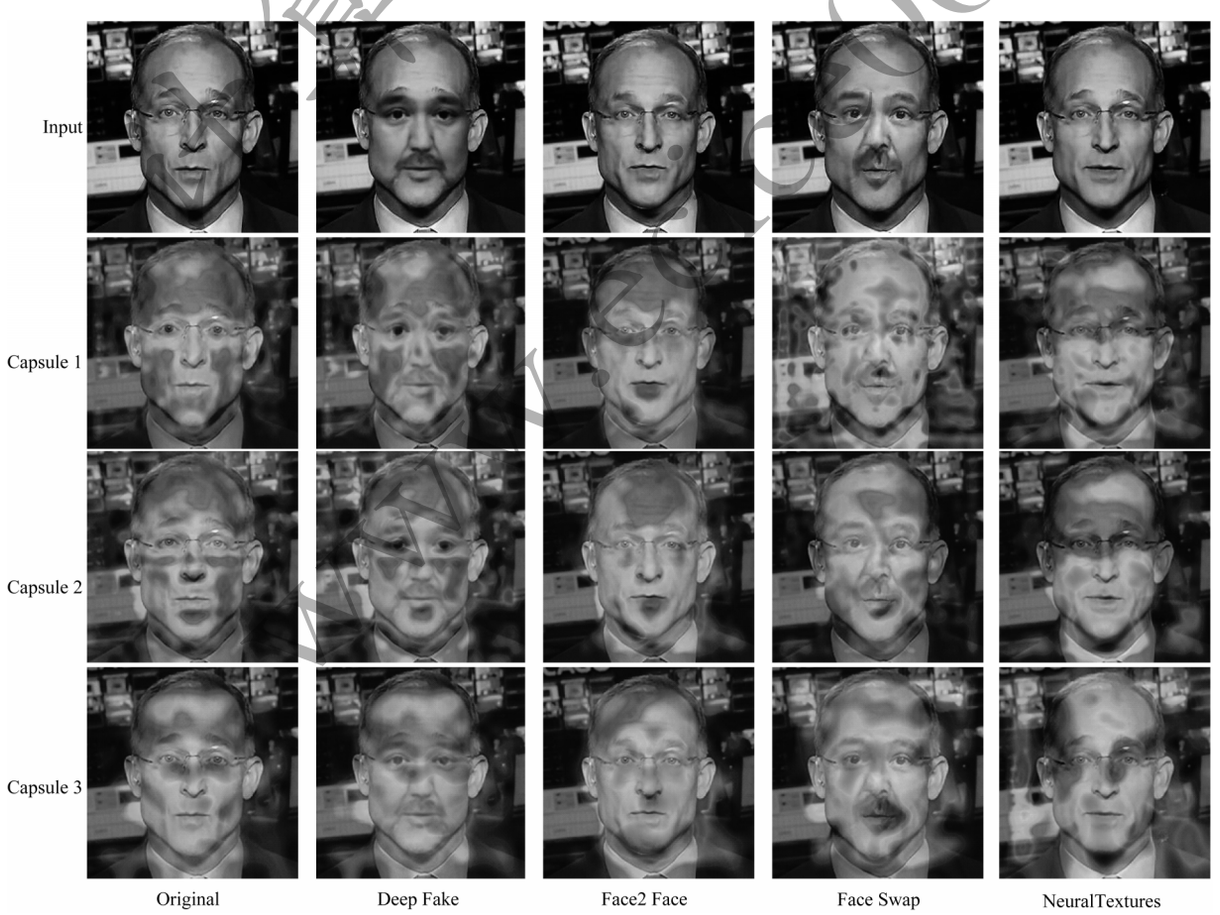

图5　不同胶囊在FaceForensics++中伪造方法的Grad-CAM热力图

Fig.5　Grad-CAM heat map of different capsules in FaceForensics++manipulated methods

根据图5可以观察到Origion样例和Deepfake样例主要的激活区域在眼睛和鼻子下方，这与图像的伪造区域一致。Face2Face样例的激活区域主要为两侧下巴和鼻尖区域，分析原因为Face2Face使用基于计算机图形学的方法，对五官区域的形变支持较差，容易在鼻部区域留下较大的伪造痕迹，同时其下巴区域的激活符合原始伪造区域的位置。Face Swap样例的激活区域集中在嘴部和眼睛区域，同时面部的其他区域具有不同程度的激活，分析原因为该技术进行的是全脸替换，因此其面部区域均得到了激活，同时其嘴部和眼睛区域符合原始伪造区域的位置。NeuralTextures样例的激活区域集中在脸的下半部分，因为该技术基于纹理渲染合成，主要用于伪造人物的说话口型，因此其嘴部区域得到了激活。

总体而言，3个胶囊均激活了脸部的关键区域，该区域由该类伪造类型的主要修改区域决定。但是对于脸部非关键部分，其激活区域的位置和范围存在一定的差别，说明3个胶囊会关注到人脸的不同区域，这与本文使用不同大小的卷积核和步长所预期的效果一致。由于3个胶囊在识别出关键区域的同时，不同的胶囊层之间学习到的特征可以相互补充进行判别。相较与传统CNN通常只能激活一整块热区，本文模型的激活区域更加准确，这有助于最后的分类结果，因此，可以一定程度上解释本文的改进方法在不同的伪造类型上均表现出较好的分类效果。

3.4.2 胶囊输出可视化分析

为说明本文模型是否将真伪人脸样本进行有效的区分，利用T-SNE[22]（T-distributed Stochastic Neighborhood Embedding）将同一个样本真假胶囊的共8个输出维度的数据降维至2个维度，进行可视化分析。

使用HQ版本的Mix测试集的测试数据进行T-SNE可视化，其结果如图6所示。其中方框表示真样本，圆形表示假样本，可以看出，大部分的假样本集中分布在左边区域，真样本集中分布在右边区域，可见本文模型输出的8维向量值对真伪样本具有较好的可区分性。

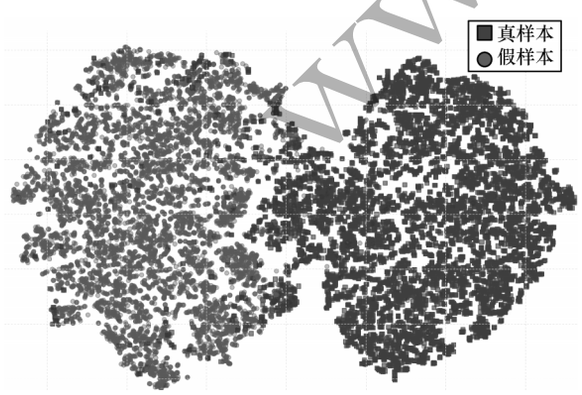

图6 模型输出的T-SNE可视化结果

Fig.6 T-SNE visualization results of model output

4 结束语

为提高胶囊网络对视频伪造人脸的检测能力，本文以Capsule-Forensics为基础，提出一种结合自注意力胶囊网络的伪造人脸检测方法。使用部分Xception网络替代VGG-19作为特征提取部分，并给出基于自注意力机制的Attention-Capsules，使模型更加关注面部的五官区域，降低背景的影响，在此基础上提出考虑不同维度输出和度量学习的损失函数Dimension-Focal Loss。在FaceForensics++上的实验结果表明，本文方法对多数伪造方法生成的伪造人脸均具有较高的检测准确率。

参考文献

[1] THIES J, ZOLLHOFER M, STAMMINGER M, et al. Face2Face: real-time face capture and reenactment of RGB videos[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 2387-2395.

[2] THIES J, ZOLLHOFER M, NIEßNER M. Deferred neural rendering: image synthesis using neural textures[J]. ACM Transactions on Graphics, 2019, 38(4): 1-12.

[3] GUERA D, DELP E J. DeepFake video detection using recurrent neural networks[C]//Proceedings of the 15th IEEE International Conference on Advanced Video and Signal Based Surveillance. Washington D. C., USA: IEEE Press, 2018: 1-6.

[4] 张怡暄，李根，曹纭，等. 基于帧间差异的人脸篡改视频检测方法[J]. 信息安全学报，2020，5(2)：49-72.
ZHANG Y X, LI G, CAO Y, et al. A method for detecting human-face-tampered videos based on Interframe difference[J]. Journal of Cyber Security, 2020, 5(2): 49-72. (in Chinese)

[5] LI Y, LYU S. Exposing deepfake videos by detecting face warping artifacts[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2019: 46-52.

[6] MATERN F, RIESS C, STAMMINGER M. Exploiting visual artifacts to expose deep fakes and face manipulations[C]//Proceedings of 2019 IEEE Winter Applications of Computer Vision Workshops. Washington D. C., USA: IEEE Press, 2019: 83-92.

[7] YANG X, LI Y Z, LYU S, Exposing deep fakes using inconsistent head poses[EB/OL]. [2020-11-10]. https://arxiv. org/pdf/1811. 00656v3. pdf.

[8] TARIQ S, LEE S, KIM H, et al. Detecting both machine and human created fake face images in the wild[C]//Proceedings of the 2nd International Workshop on Multimedia Privacy and Security. Vancouver, Canada: CCS Press, 2018: 81-87.

[9] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[EB/OL]. [2020-11-10]. https://arxiv. org/abs/1710. 09829v1.

[10] LALONDE R, BAGCI U. Capsules for object segmentation[EB/OL]. [2020-11-10]. https://arxiv. org/pdf/1804. 04241. pdf.

（下转第206页）

（上接第200页）

[11] IESMANTAS T, ALZBUTAS R. Convolutional capsule network for classification of breast cancer histology images [C]//Proceedings of International Conference on Image Analysis and Recognition. Berlin, German: Springer, 2018: 853-860.

[12] NGUYEN H H, YAMAGISHI J, ECHIZEN I. Capsule-forensics: using capsule networks to detect forged images and videos[C]//Proceedings of IEEE International Conference on Acoustics. London, UK: IEEE Press, 2019: 2307-2311.

[13] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1800-1807.

[14] ROSSLER A, COZZOLINO D, VERDOLIVA L, et al. FaceForensics++: learning to detect manipulated facial images [EB/OL]. [2020-11-10]. https://arxiv. org/pdf/1901. 08971. pdf.

[15] BONETTINI N, DANIELE E, MANDELLI S, et al. Video face manipulation detection through ensemble of CNNs[EB/OL]. [2020-11-10]. https://arxiv. org/pdf/2004. 07676. pdf.

[16] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[EB/OL]. [2020-11-10]. https://arxiv. org/pdf/1711. 07971v3. pdf.

[17] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 2999-3007.

[18] DENG J K, GUO J, ZHOU Y X, et al, RetinaFace: single-stage dense face localisation in the wild[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 5202-5211.

[19] AFCHAR D, NOZICK V, YAMAGISHI J, et al. MesoNet: a compact facial video forgery detection network [C]// Proceedings of 2018 IEEE International Workshop on Information Forensics and Security. Washington D. C., USA: IEEE Press, 2018: 1-7.

[20] TAN M, QUOC V. EfficientNet: rethinking model scaling for convolutional neural networks [C]//Proceedings of International Conference on Machine Learning. New York, USA: ACM Press, 2019: 6105-6114.

[21] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

[22] MAATEN L V D. Accelerating T-SNE using tree-based algorithms[J]. Journal of Machine Learning Research, 2014, 15(1): 3221-3245.

编辑　索书志