# 检测计算机在社交媒体中生成伪造的面孔的感知判断

Suzan Anwar（1,2,email），Mariofanna Milanova （1），Mardin Anwer（2,3），和Anderson Banihirwe（1）

1.阿肯色大学，小石城，美国小石城

[sxanwar@ualr.edu](mailto:sxanwar@ualr.edu)

2.萨拉赫丁大学，伊拉克埃尔比勒

3.黎巴嫩法国大学，伊拉克埃尔比勒

## 摘要

人们对开发用于图像表示的方法的学习越来越感兴趣，特别专注于训练深度神经网络来合成图像。生成对抗网络（GAN）被用于进行面部老化，生成新视点或更改面部属性比如肤色。对于专门针对面部取证，已经有一些方法提出将计算机生成的面孔与自然面孔区别开来以及检测脸部修饰。我们建议研究基于感知判断的技术来检测通过深度学习建筑产生的图像/视频。这项研究的主要目标是：（1）开发技术以，基于面部表情分析，区分计算机生成的脸部和拍摄的脸部；（2）开发基于熵的技术以进行，计算机生成（CG）人脸中的伪造检测。结果显示了原始视频和更改后的视频中的情绪差异。这些计算的结果很大且具有统计意义。结果表明，与原始视频的值相比，更改后的视频的熵值降低了。原始帧的直方图具有重尾分布，而在帧改变的情况下：由于图像垂直和水平边缘值很小，直方图更清晰。

## 关键词：

视频处理， ASM ，面部表情，基于熵的直方图

## 1引言

随着计算机视觉和图形学的发展，生成具有逼真的合成人脸的图像/视频变得有可能了。像Google，百度，Nvidia，Adobe和诸如Voicecey之类的初创公司最近为制作音频或视频做出资助贡献。这些公司已经发布了自己动手做的软件和在GitHub上可用开源工具，例如DeepFake。当前正在使用的方法可以生成实时处理的Face2Face视频，可以基于音频输入合成视频，或可以人工制作静态图像动画。新技术允许用户编辑面部表情。在虚假新闻讨论的背景下，这获得了令人难以置信的关注。结果引起了人们对脸部互换技术可用于传播误导性信息的关注。

根据最近的出版物[1]，“目前，没有任何工具可以一直工作”，Mitre DCorp的一个研究员Mikel Rodriguez说。使用深度学习技术面部图像合成方法的概况在[2]中有所表述。大部分的用于交换面部的技术会生成输出作为面部图像或3D面罩。例如，实际上不可能区分真正的保罗·沃克和电影《速度与激情7》中由计算机生成的他。拍摄期间的演员的死亡导致导演使用以前记录的数字3D扫描数据，为未完成的场景重建沃克先生的脸。另一个例子是《实况足球2》，由Konami开发和发行的视频游戏。自2012版本以来，足球运动员的图像渲染地是如此逼真，以至于看起来几乎像真实的人一样[3]。

NVIDIA在2017年6月创建了一个使用CelebA-HQ的名人照片数据库的GAN来生成实际上不存在的人的图像。2018年，NVIDIA提出了一种新的GAN，可以增加生成的图像的变化[4]。GAN用于应用人脸老化，产生新观点或改变人脸属性，例如肤色[5]。

虚拟世界已被建设性地使用，以造福社会。但是，也存在安全方面的顾虑，例如网络恐怖主义活动，儿童色情侦查和洗钱等经济犯罪。

虚幻的图像和视频可用于伤害他人或获取政治和/或经济优势。例如，有关外星人，灾难，政治家或商人的虚假图像或视频，可能会造成混乱或改变人们的观点。Facebook，Twitter，Flickr或YouTube等社会媒体平台是广泛传播这些伪造的图像和视频的理想环境。

为了应对这种威胁，在将来，CG操作检测软件将需要变得更加复杂和有用。这项技术以及强大的训练，明确可接受的准则，使媒体组织能够把握住反对故意进行图像处理的路线，从而保持其作为真理的传播者的声誉。创造新技术的挑战是：

1.用于制作图像/视频的算法基于卷积神经网络，广泛用于物体识别。在深度学习方法中，特性是从训练样本中自动学习，而不是手动设计。但是，基于深度学习的方法大多使用监督学习。尽管基于深度学习的方法很有希望，但它们在数字图像取证领域还不成熟；在这方面仍有大量工作要做。

2.缺乏共享数据集，维护和可用性。来到一个一切都已连接（IoT）的世界，需要从流媒体设备收集数据，例如Roku或AppleTV和无人驾驶飞机（UAV），以及使用新的深度学习架构的计算机生成图像。

我们建议研究基于感知判断的技术来检测由深度学习架构产生的图像/视频的操纵。这项研究的主要目标是：

• 开发技术以区分计算机生成的和基于面部表情分析的摄影面孔。假设是人类表达的面部表情和假冒面孔产生的面部表情是不同的。人类可以产生各种各样高强度范围的面部表情。

• 开发基于熵的CG人脸伪造检测技术。假设是自然图像具有一些特殊性质不同于来自其他类型的图像。

## 2相关工作

鉴于需要对数字图像/视频内容进行自动实时验证，研究人员提出了几种技术。数字图像处理检测方法有两个主要类别：主动方法和被动方法。主动方法涉及各种图像内容的水印或指纹，并将其嵌入数字图像中[6]。随着社交网络中使用的图像数目的上升，不可能要求都在分发前先给所有互联网上的数字图像加水印。因此，被动取证方法已成为更流行的选择。

被动方法通过分析在数字图像修改阶段出现的特定的固有线索或图案​​来检测数字图像的变化。被动方法不依赖任何先前或预设的信息，并且在图像取证中具有更广泛的应用。这些技术已成功应用于跟踪真假新闻。在[7]迹线分为三组：图像采集中剩余的迹线，图像存储中剩余的迹线，和图像编辑中剩余的迹线。最近由计算机图形软件生成的新类别成为流行图像。

对于专门针对面部取证的方法，已提出了一些方法来区分计算机生成的面孔与自然面孔 [8]并检测脸部修饰[9]。在生物识别技术中，提出了两个预先训练的深层CNN，VGG19和AlexNet来检测变形的脸[10]。在[11]中，作者提出使用两流网络交换操作检测两个不同的面部：一个流检测图像补丁之间的低级别不一致，而另一个流明确检测篡改的脸。

慕尼黑工业大学的研究人员已经开发出了一种学习算法，可潜在地识别互联网上的伪造的人脸互换视频。他们使用大量的他们自己制造的面部互换来训练算法，创建了此类可用图像的最大数据库。然后他们训练了称为XceptionNet的算法来检测人脸交换[12]。

在[13]中展示了，用于检测和分类原始视频和经处理视频的不同算法。实际上，对于人类和计算机，这都是一项艰巨的任务，尤其是当视频被压缩且分辨率较低时（通常在社交媒体上发生）。作者还介绍了一个名为“面部取证”（ “Face Forensics”）的大型视频数据集。

一些伪造检测方法还使用统计特性来检测伪造。这项技术基于使用自然图像统计的方法。自然图像有一些不同于其他类型图像[14]的特殊属性。在[15] 中，通过分析特征点集的视频中，人脸表情的变化，来区分CG面孔和真实的面孔。

## 3面部表情

在实验中，[16]中提出的FaceXpress软件用于识别FaceForansic数据集中每个帧的情绪。该软件以使用Viola-Jones检测器检测脸部开始，然后使用多分辨率跟踪器Active Shape Model（ASM）跟踪器[17]指示116个脸部标志。FaceXpress 使用Active Shape Model追踪器检测面部三角测量点（见图1）。通过测量检测到的面部三角测量点的长度来获得属性。对于某些属性，例如嘴巴宽度，嘴巴高度，和眼间隙中点到眉毛中点之间的距离，通过使用马哈拉诺比斯（Mahalanobis）距离获得。其他属性，例如前额的垂直边缘区域和中前额水平边缘的区域通过高斯核滤波获得。跟踪点的位置用于计算面部区域的变化，例如，眼眉皱纹，额头皱纹，脸颊皱纹，眼睛与眉毛之间的距离，以及口的垂直和水平测量[18]（见图2）。最后，一个支持向量机（SVM）用于识别检测七个普遍情绪的面部表情：惊喜，愤怒，幸福，悲伤，恐惧，厌恶和中立。

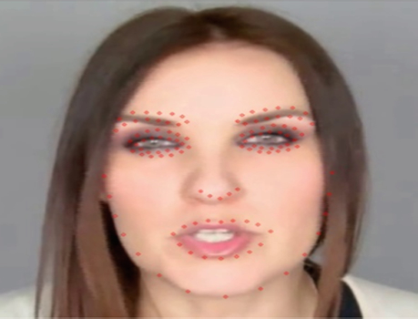


图1.使用ASM跟踪器进行面部标志检测



图2.用于检测感兴趣区域的面部属性

## 4基于熵的直方图

直方图处理包括通过修改其直方图来更改图像。为了绘制图像平面的直方图，对原始图像和从FaceForensics数据集中更改了的图像（请参阅第5.1节）均进行归一化处理是必要的。此过程称为对比增强，在这儿，例如压缩，描述和分段等信息的强度转换功能被提取。要计算图像的直方图，以下离散函数适用于强度级别在[0，L-1]范围内。

h(rk) = nk (1)

r k是强度值，

n k是强度为r k的图像中的像素数，

h(rk)是灰色等级r k的数字图像的直方图。

通过假设一个M ×N图像，像素总数用于归一化图像直方图。归一化计算与在图像中的出现概率r k有关。下面给出了归一化直方图的公式：

p(rk)= nk/MN, k = 0,1,2,…,L-1 (2)

p(nk)计算强度为r k的图像表情发生的概率。

所有归一化直方图分量的总和应等于1 [19]。原始视频和更改后的视频中相同帧的直方图不同（请参见

图3）。



图3.在原始帧和变更帧上应用图像直方图的结果

原始帧的直方图具有重尾分布。如果帧发生变化，由于图像的垂直和水平边缘的值很小，因此直方图更加清晰。另外，从图像直方图可以确定图像均值信息或熵。计算图像熵的目的是找到其自动聚焦。对于任何具有概率密度函数f（x）的随机变量X的熵定义为：

H(X) = -E[logf(X)] = -∫f(x)logf(x)dx (3)

变量的范围分为n个间隔（l k，u k），k = 1，2，…，n。的

上面的熵定义和以直方图表示的密度之间的关系如以下方程式所示：

H(X) = - （4）

直方图的第k个bin与上述宽度求和的第k个项之间的关系如下式表示：

wk = uk - lk (5)

bin概率p k，k = 1，2，…。n定义为：

pk = (6)

可以近似为w k f（x k），其中：

f（x k）是矩形的面积，

x k是在区间（l k，u k）值，

到等式的第k个积分。（4）可以近似为w k f（x k）log（x k），就bin概率使用的该表达式将熵重写为：

H(X) = - (7)

上面的表达式是由Harris [20]给出的离散分布和一个Rich和Tracy [21]给出的直方图，如果w k =1。当w k为常数且不等于1时，我们使用：

H(X) = - (8)

## 5结果

### 5.1 FaceForensic视频数据集

在本文中，我们使用了faceForansics视频数据集[13]，由从YouTube收集了大约1004个视频中的大约500,000个面部帧。数据集已使用最新的人脸编辑方法（包括分类和分段）进行了处理。在根据目标表情从口腔数据库中选择口腔内部，使用原始的face2face重演方法。

### 5.2面部表情

FaceXpress软件生成一个csv文件，其中包含对每个原始和更改视频的帧识别的情感。为了评估原始视频和更改后的视频之间的情感差异，均方误差（MSE）是在原始视频和更改后的视频[22]（来自FaceForensic数据集）之间的均方误差。原始和修改后视频的csv文件中存储的产生的情绪之间的MSE指标是计算出来的。对于原始视频和更改后的视频中的每一帧，情绪不同是平方的和平均的。表1显示了MSE对FaceForensic数据集中的一些视频帧的计算结果，结果令人恐惧且值得关注。通过在某些FaceForensic的视频上应用FaceXpress软件，在原始视频和变更视频中情感表达中，结果显示出明显的差异（请参阅图4）。



图4.在某些FaceForensic视频上应用FaceXpress的结果

表1. MSE计算结果

|  |  |
| --- | --- |
| 视频 | MSE均方误差 |
| v1 | 43.12965 |
| v2 | 494.748 |
| v3 | 420.4217 |
| v4 | 224.6175 |
| v5 | 327.7263 |
| v6 | 96.96217 |
| v7 | 322.4336 |
| v8 | 172.4224 |
| v9 | 367.1219 |
| v10 | 366.6556 |
| v11 | 306.3773 |
| v12 | 123.6393 |
| v13 | 22.05341 |
| v14 | 162.6497 |
| v15 | 38.28153 |
| v101 | 168.4492 |
| v102 | 226.1895 |
| v148 | 113.8129 |
| v149 | 29.48472 |
| v150 | 364.4547 |

### 5.3基于熵的直方图

评估图像质量的另一种方法是对原始视频和更改后的视频计算熵值。对于原始视频和更改后的视频的相同帧，我们在等式 8应用了熵公式。表2显示了：与原始视频他们的值相比，更改后的帧的熵值降低了。

表2.计算三个选定帧的熵值的结果

|  |  |  |
| --- | --- | --- |
| 帧 | 熵值 | |
| 原始的 | 修改的 |
| 1 | 3.9122 | 3.8919 |
| 2 | 3.9114 | 3.8922 |
| 3 | 3.9106 | 3.8924 |

## 6结论

在本文中，我们应用了两种不同的方法来测试FaceForensic原始和更改后的视频数据集中的情感的质量和差异。在第一种方法中，我们使用FaceXpress软件识别视频中的情绪以进行比较。原始视频和更改后的视频之间的情感差异是使用MSE测量方法计算的。MSE值的结果得出结论：原始视频和更改后的视频之间的情感差异清晰明显。在第二种方法中，我们计算由帧的直方图生成的熵值，来测试视频的质量。第二种方法的结果显示：与原始视频的值相比，更改后的视频的熵值降低了。原始帧的直方图具有重尾分布，而在帧改变的情况下：由于图像垂直和水平边缘的值极小，直方图更加清晰。

## 参考文献

1. Nordrum, A.: Forging voices and faces. Spectrum IEEE 14–15, May 2018. https://doi.org/10.1109/mspec.2018.8352562

2. Lu, Z., Li, Z., Cao, J., He, R., Sun, Z.: Recent progress of face image synthesis (2017). https://arxiv.org/abs/1706.047173

3. Face-Swapping Porn: How a Creepy Internet Trend Could Threaten Democracy, Rolling Stone, 4.18.18. https://www.rollingstone.com/culture/features/face-swapping-porn-how-creepy-trend-could-threaten-democracy-w518929

4. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive Growing of GANs for Improved Quality, Stability, and Variation, ICLR 2018. http://research.nvidia.com/publication/2017-10\_Progressive-Growing-of

5. Antipov, G., Baccouche, M., Dugelay, J.-L.: Face Aging With Conditional Generative Adversarial Networks, May 2017. https://arxiv.org/abs/1702.01983

6. Milanova, M., Kountchev, R., Ford, C., Kountcheva, R.: Watermarking with inverse difference pyramid decomposition. In: International Signal Processing Conference, Dallas, USA, pp. 346–362 (2003)

7. Lin, X., et al.: Recent advances in passive digital image security forensics: a brief review. Engineering 4, 29–39 (2018)

8. Rahmouni, N., Nozick, V., Yamagishi, J., Echizeny, I.: Distinguishing computer graphics from natural images using convolution neural networks. In: IEEE Workshop on Information Forensics and Security, pp. 1–6 (2017)

9. Bharati, A., Singh, R., Vatsa, M., Bowyer, K.: Detecting facial retouching using supervised deep learning. IEEE Trans. Inf. Forensics Secur. 11(9), 1903–1913 (2016)

10. Raghavendra, R., Raja, K., Venkatesh, S., Busch, C.: Transferable Deep-CNN features for detecting digital and print-scanned morphed face images. In: IEEE Computer Vision and Pattern Recognition Workshops, pp. 10–18 (2017)

11. Zhou, P., Han, X., Morariu, V., Davis, L.: Two-stream neural networks for tampered face detection. In: IEEE Computer Vision and Pattern Recognition Workshops, pp. 1831–1839 (2017)

12. https://www.engadget.com/2018/04/11/machine-learning-face-swaps-xceptionnet/Perceptual Judgments to Detect Computer Generated Forged Faces 47

13. Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., Nießner, M.: FaceForensics: a large scale video dataset for forgery detection in human faces. CV Cornell University Library, March 2018

14. Souza, D., Yampolskiy, R.: Natural vs artificial face classification using uniform local directional patterns and wavelet uniform local directional patterns. In: IEEE CVPRW, pp. 27–33 (2014)

15. Dang-Nguyen, D.-T.: Discrimination of Computer Generated versus Natural Human Faces, February 2014. http://eprints-phd.biblio.unitn.it/1168/

16. Anwar, S., Milanova, M.: Real time face expression recognition of children with autism. IAEMR 1(1) (2016)

17. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., et al.: Active shape models-their training and application. Comput. Vis. Image Underst. 61, 38–59 (1995)

18. Ekman, P., Friesen, W.: Facial Action Coding System. Consulting Psychologists Press, Palo Alto (1978)

19. Vij, K., Singh, Y.: Enhancement of images using histogram processing techniques. Int. J. Comput. Tech. Appl. 2(2), 309–313. ISSN: 2229-6093

20. Harris, B.: Entropy. In: Balakrishnan, N., Read, C.B., Vidakovic, B. (eds.) Encyclopedia of Statistical Sciences, vol. 3, 2nd edn., pp. 1992–1996. Wiley, New York (2006)

21. Rich, R., Tracy, J.: The relationship between expected inflation, disagreement, and uncertainty: evidence from matched point and density forecasts. Staff Report No. 253, Federal Reserve Bank of New York. (Revised version published in Review of Economics and Statistics 92(2010), 200–207 (2006)

22. http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL\_COPIES/VELDHUIZEN/node18.html