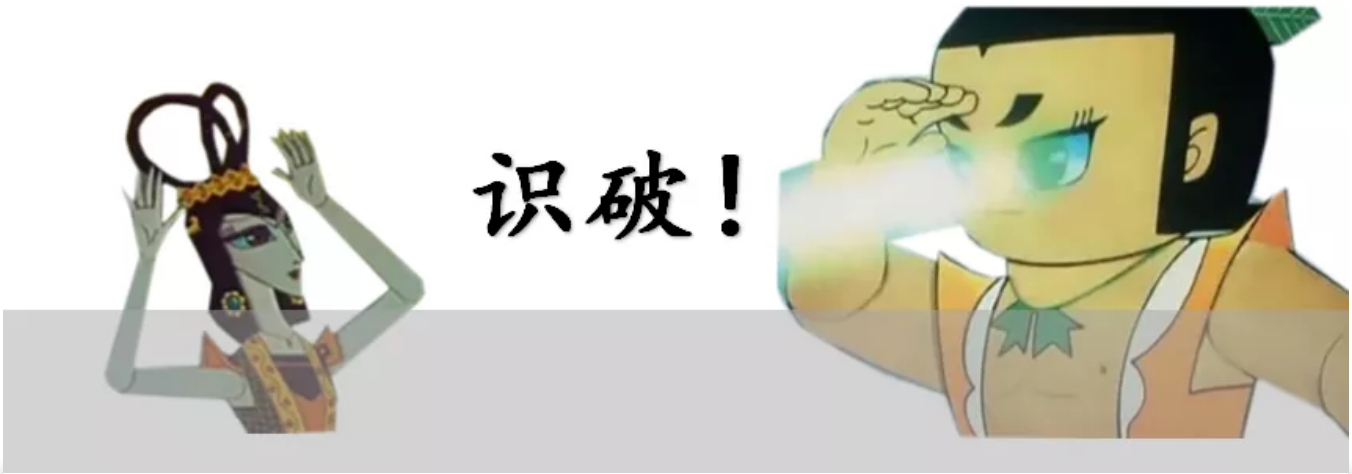
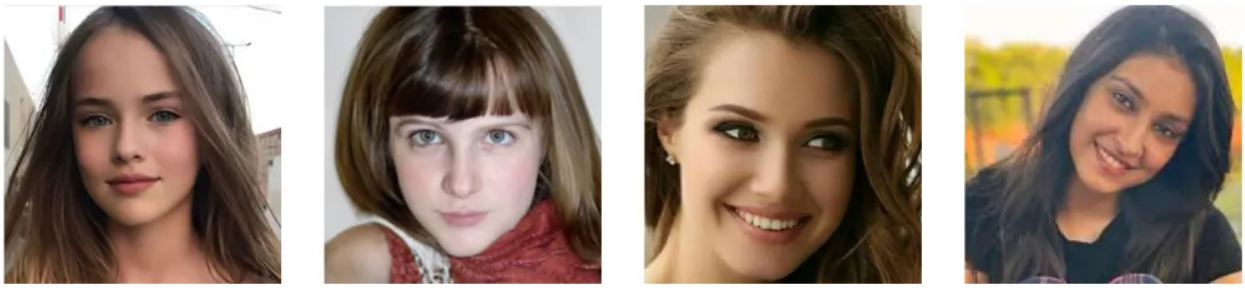
GAN 的内在漏洞！只看眼睛就能找出虚拟人脸?

原创 小戏 夕小瑶的卖萌屋 2021-09-15 12:05



文 | 小戏

让我们先来看一组漂亮的小姐姐：



一　　二　　三　　四

如果说，**这四张照片里有一张并不是真人，而是由 GAN 生成的虚拟人像**，大家可以看出是哪一张吗?

很抱歉，请在手机微信登录投票

上图中哪个小姐姐不是真人（单选）

一

二

三

四

答案揭晓！在这一组照片中，由 GAN 生成的虚拟人像是二号小姐姐。可以看到，无论如何，**目前由 GAN 及其衍生技术所生成的虚拟人像已经完全可以达到以假乱真的程度，其生成的人脸很难被人类从视觉上进行分辨**。这自然是人像生成领域的一大进步，然而，这种可以被以极低成本大量生产的虚拟人像很容易被滥用于诸如虚假信息欺诈、社交媒体头像等等地方。

在这些场景下，**如何从大量图片信息中分类出真实人脸与虚拟人像便成为了一个新的问题**。其实初想或许会觉得这是一件很容易的二分类的问题，可以如果仔细一想**一个二分类的判别器很难在 GAN 的训练机制下对分类真实人脸与虚拟人像取得良好的分类效果与鲁棒性**。

在这样的背景下，来自 **UAlbany** 的学者们另辟蹊径，提出了一种基于物理的方法，**通过暴露出 GAN 模型本身与真实物理世界交互的缺陷来巧妙识别出真实人脸与虚拟人像的方法，即通过识别瞳孔的形状来判断人脸的真实与否**。让我们来看看这篇论文吧！

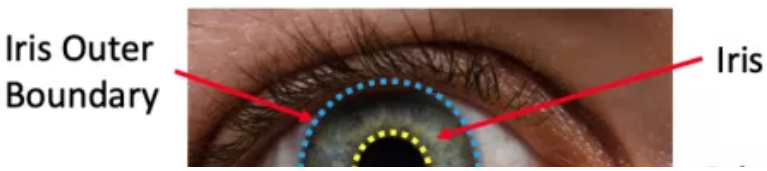
论文题目：

EYES TELL ALL: IRREGULAR PUPIL SHAPES REVEAL GAN-GENERATED FACES

论文链接：

https://arxiv.org/abs/2109.00162

Arxiv访问慢的小伙伴也可以在 **【夕小瑶的卖萌屋】**订阅号后台回复关键词 **【0915】** 下载论文PDF~



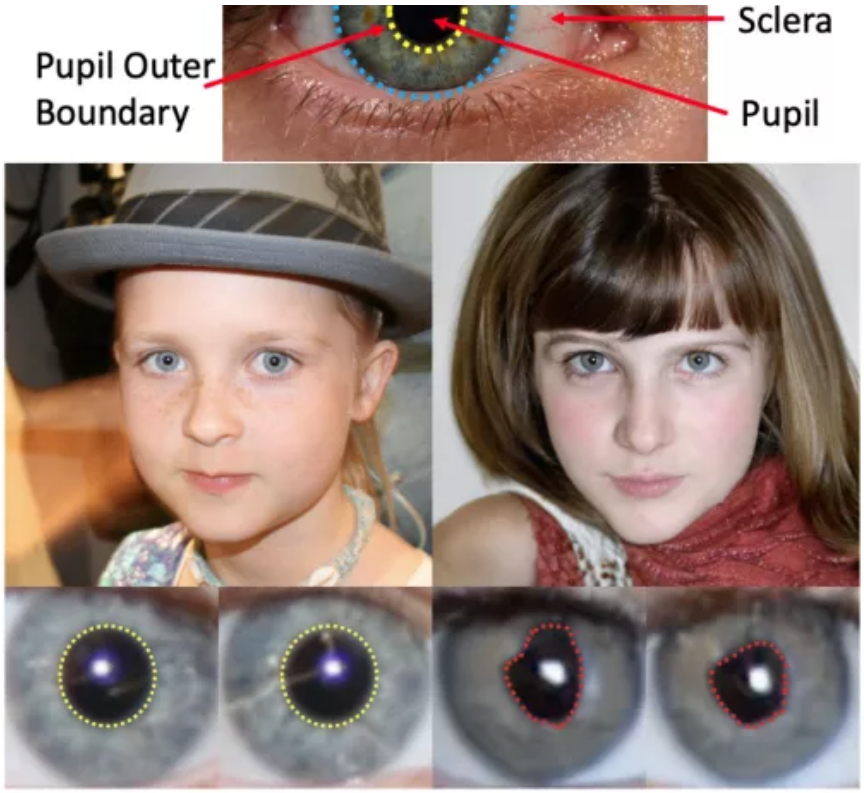


Fig. 1: ***Top:*** *Anatomy structures of a human eye.* ***Bottom:*** *Examples of pupils of real human (left) and GAN-generated (right). Note that the pupils for the real eyes have a strong circular or elliptical shapes (**yellow**) while those for the GAN-generated pupils are with irregular shapes (**red**). And also the shapes of both pupils are very different from each other in the GAN-generated face image.*

从上图可以看到，我们的眼睛中心是虹膜与瞳孔，白色的区域是巩膜。**对于一个健康的成年人而言，瞳孔的形状一般是圆形的**。如上图下方左侧的图像，从正面看瞳孔趋于正圆。而论文作者发现，**使用 GAN 等技术生成的人脸，其瞳孔形状是不规则的**，放大由 GAN 生成的虚假人像可以清楚的看到，其瞳孔的形状呈现了明显的不规则。

论文作者推断，出现这种现象的根本原因在于，**类似 GAN 等模型实质上缺乏对人眼结构的真正理解，换而言之，GAN 等模型在生成人像时，仍然缺乏从人类生理结构出发的约束**。而这种机制上的缺陷为判别真实人像与虚拟人像提供了可能。

总的来说，论文提出的虚拟人像检测方法分为三步，如下图所示，对于一张输入的人像（a），首先需要定位到人像的瞳孔部分，得到如下图（b）的结果，接下来论文使用 **EyeCool** 算法从（b）中提取得到瞳孔掩膜(**Pupil Mask**)，勾勒出瞳孔边界，如下图（c）所示，同时，论文提出了一种基于最小二乘的椭圆拟合方法，得到理想情况下真实人像的椭圆形瞳孔掩膜（d）（这里使用椭圆而非正圆的原因在于由于人像拍摄角度的印象，导致一般而言图片人像瞳孔趋于椭圆）。最后通过改进的考虑边界的 **IoU** 算法（**BIoU**）计算得到图像与理论上真实瞳孔形状之间的差异，从而判断输入人像是否是真实人像。

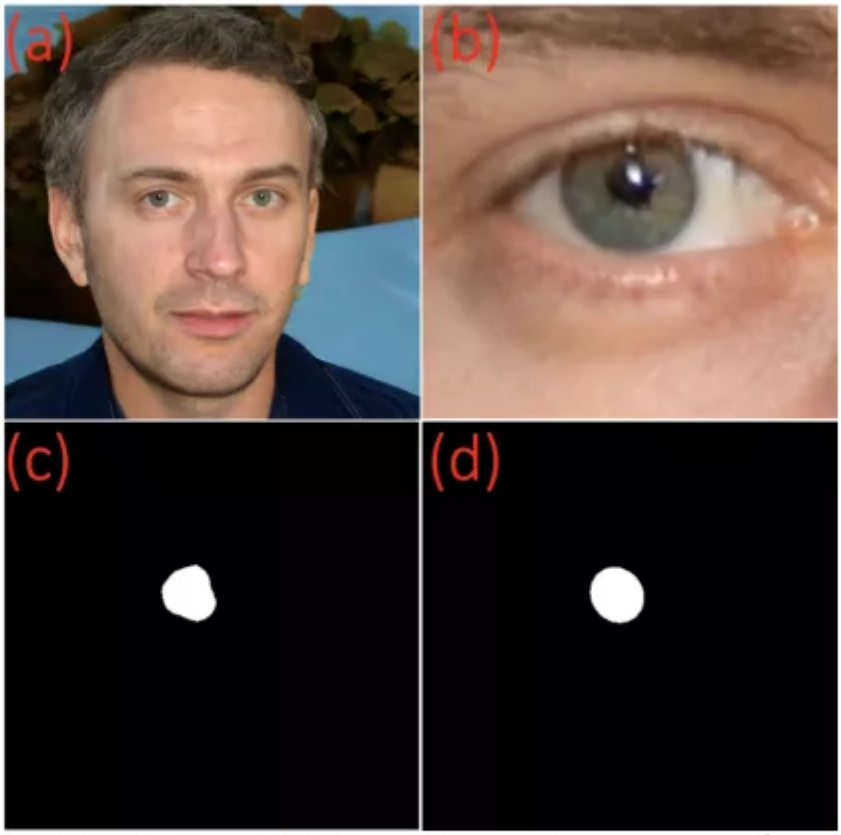


Fig. 2: *(a) The input high-resolution face image, (b) The the cropped eye image using landmarks, (c) Predicted pupil mask of image (b), (d) Ellipse fitted pupil mask of (c). Note that this example is a GAN-generated face.*

根据这种方法，作者选用 **Flickr-Faces-HQ (FFHQ)** 数据集的一千张人脸作为真实人脸，使用 **StyleGAN2** 创建了一千张虚拟人脸进行实验。

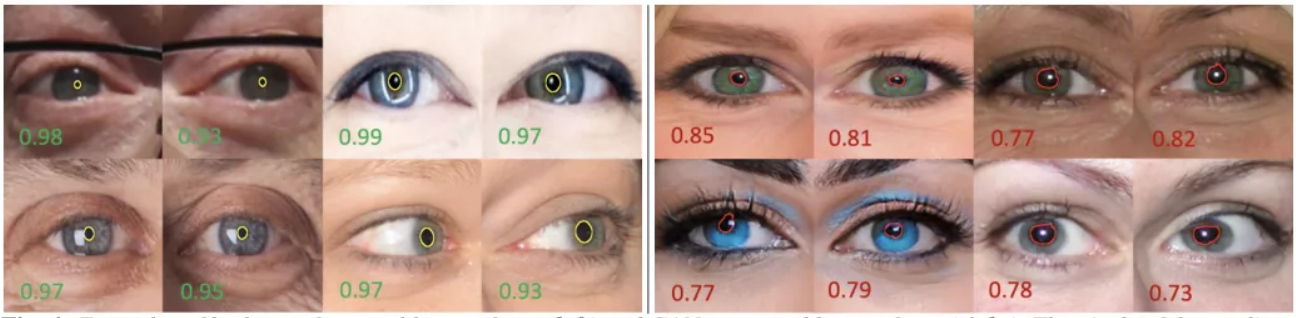
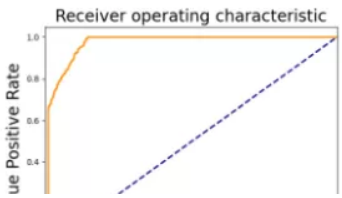
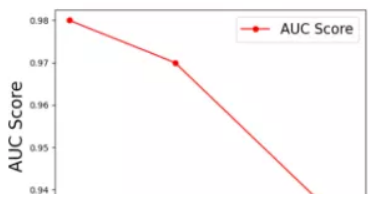


Fig. 4: *Examples of both eyes from real human faces (**left**) and GAN generated human faces (**right**). The pixels of the predicted pupil mask within a distance $d = 4$ from the prediction boundary contours are highlighted. The Boundary IoU score ($d = 4$) between the predicted pupil mask and the ellipse-fitted pupil mask for each pupil is shown on the images.*

实验发现使用瞳孔形状可以有效的区分真实人脸与虚拟人脸，**算法的 AUC 分数达到了0.94，其评估指标——即 BIoU 值在真实人脸与虚拟人脸之间的分布也呈现了较大的差异。**





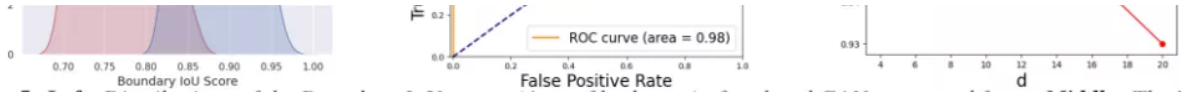


Fig. 5: ***Left:*** *Distributions of the Boundary IoU scores (Ave. of both eyes) of real and GAN-generated faces.* ***Middle:*** *The ROC curve is based on the Boundary IoU scores. The $d = 4$ for both figures.* ***Right:*** *BIoU hyper-parameter analysis, where x axis indicates the variation of hyper-parameter d and y axis is the AUC score.*

我们可以看到，这篇论文提出了一个简单有效的方法区分真实人像与虚拟人像，这种方法在保证了准确率的同时，又提供了很好的可解释性，甚至抛开算法，**这个思路对我们使用肉眼判断人像真实与否都有很好的实践价值**。

目前，无论是图像生成还是文字生成，其实质上都是一种自下而上的重复模仿，或多或少都缺乏一些如这篇论文所描述的一样真实人脸的生理约束或是自然语言领域的语法句法。**如何在进行生成时能更多的考虑这样自上而下的先验信息，使得这种先验信息不仅可以作为评价真实与否的方式方法，更能作为图像或文字生成时的内在约束，或许是更为有意义的问题吧！**

喜欢此内容的人还喜欢

在嘴巴里放入124 个传感器，谷歌眼镜创始人新项目：用舌头发信息

大数据文摘

架构升级，ARM v9与v8版本有何不同?

架构师技术联盟

改名Meta俩月，脸书放弃虚拟现实操作系统：负责人跳槽谷歌

机器之心