

138 亿年在手中！谷歌用手机体验宇宙大爆炸，还教你用机器学习实现 AR 自拍特效

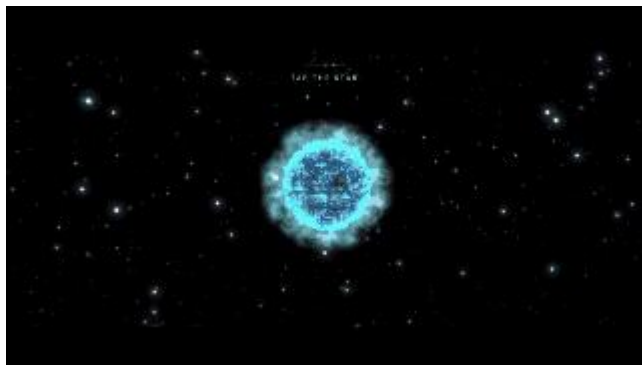
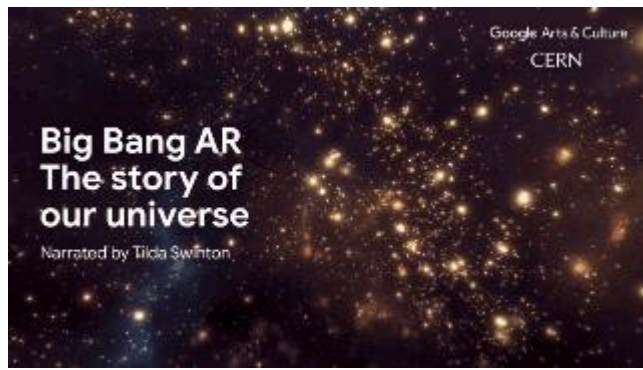
人工智能 3 月 10 日



谷歌艺术与文化推出新作品，手机上体验宇宙大爆炸，并可以作为背景进行自拍。而谷歌博客解释了如何使用 ARCore 和机器学习，只用一个手机摄像头就可以实现给自拍加 AR 特效的工作原理。

谷歌与欧洲研究机构 CERN 合作创建了一个增强现实应用程序，可以让用户通过手机体验宇宙大爆炸 Big Bang，感受上帝创世纪时候的宏伟壮阔。

应用程序将带给你一场 360 度的奇幻旅程，从宇宙诞生开始，到第一颗恒星诞生，直到我们的太阳系，以及地球的出现。



更有意思的是，你还可以以大爆炸作为背景，进行自拍！可以说是超炫酷了。各大应用商店搜索 Google Arts & Culture 下载。



如果你使用的是 Android 手机，你需要确认一下是否装了 ARCore。

ARCore 是谷歌推出的搭建增强现实（Augmented Reality，简称 AR）应用程序的软件平台，类似苹果的 ARKit，它可以利用云软件和设备硬件的进步，将数字对象放到现实世界中。

用 ARCore 实现 AR 效果

增强现实，是一种实时地计算摄影机影像的位置及角度并加上相应图像的技术，这种技术的目标是在屏幕上把虚拟世界套在现实世界并进行互动。

这种技术最早于 1990 年提出。随着随身电子产品运算能力的提升，增强现实的用途越来越广。比如下面这个动图，通过 AR 设备你可以看到实际并不存在的蓝色光圈。

AR 已经被广泛应用于自拍、短视频、直播 app 中，可以实现一些好玩的效果。而其中最关键的挑战在于将虚拟内容适当地锚定到现实世界中。这个过程需要一套独特的感知技术，能够跟踪每个微笑，皱眉或傻笑背后的高动态表面几何。

所以今天我们就为大家介绍一下，开发者如何借助最新版本 ARCore 的全新 Augmented Faces API，使机器学习（ML）来推断近似的 3D 表面几何形状，从而只需要一个摄像机输入，无需专用的深度传感器就可以实现给自拍添加动画特效，比如添加眼镜、3D 帽子等物品。



3D 网格及其实现的一些效果

这种方法提供了实时速度的 AR 效果，使用 TensorFlow Lite 进行移动 CPU 推理或其可用的新移动 GPU 功能。

此技术与 YouTube Stories 的新 creator effect 效果相同，并且通过最新的 ARCore SDK 版本和 ML Kit Face Contour Detection API，可供更广泛的开发人员社区使用。

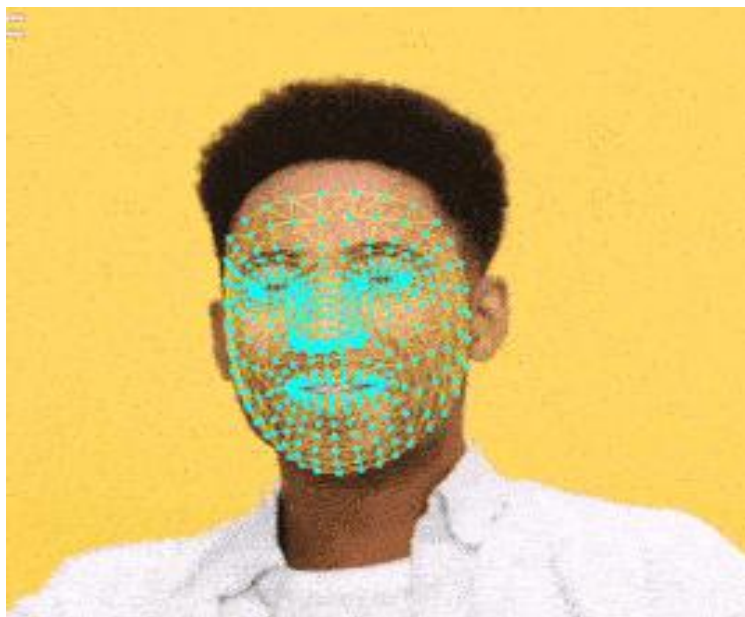
自拍 AR 的 ML pipeline

我们的 ML pipeline 由两个一起工作的实时深度神经网络模型组成。一个探测器，在整个图像上运行，并计算面部位置；一个通用 3D 网格模型，在这些位置上运行并通过回归预测近似表面几何。

精确地裁剪面部可以大大减少对共同数据增强的需求，例如由旋转，平移和比例变化组成的仿射变换。

它允许网络将其大部分容量用于坐标预测准确度，这对于实现虚拟内容的正确锚定至关重要。

一旦需要的位置被裁剪，神经网络每次仅应用于单个帧，使用窗口平滑以便在面部静止时减少噪声，同时避免在快速移动时滞后。



3D 网格

对于我们的 3D 网格，我们采用了传递学习并训练了一个具有多个目标的网络。

网络同时预测合成、渲染数据上的 3D 网格坐标，以及类似于 MLKit 提供的带注释的真实世界数据的 2D 语义轮廓。

由此产生的网络不仅在合成上，而且在现实世界数据上为我们提供了合理的 3D 网格预测。

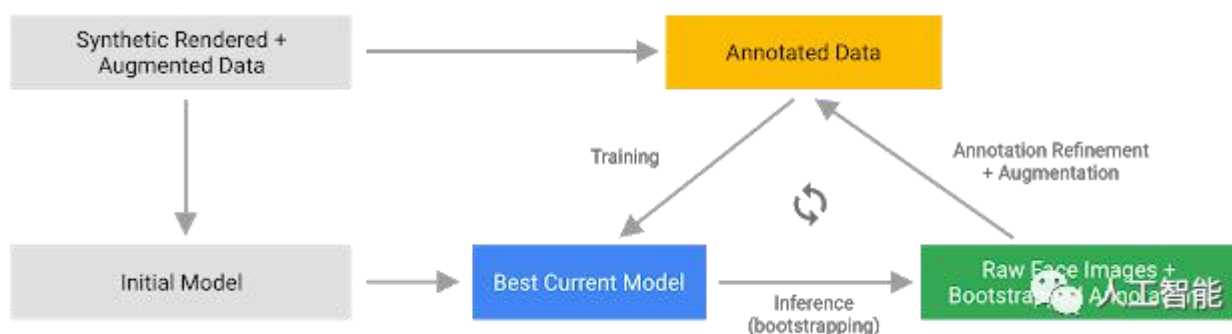
所有模型都接受来自地理上不同数据集的数据培训，随后在平衡，多样化的测试集上进行测试，以获得定性和定量性能。

3D 神经网络接收裁剪的视频帧作为输入。它不依赖于额外的深度输入，因此它也可以应用于预先录制的视频。

该模型输出 3D 点的位置，以及在输入中存在并合理对齐的面部概率。一种常见的替代方法是预测每个地标的 2D 热图，但它不适合深度预测，并且对于这么多点具有高计算成本。

通过迭代引导和细化预测来进一步提高模型的准确性和鲁棒性。这样我们就可以将我们的数据集增长到越来越具有挑战性的案例，例如鬼脸，斜角和遮挡。

数据集增强技术还扩展了可用的地面实况数据，开发了模型对相机缺陷或极端光照条件等工件的弹性。



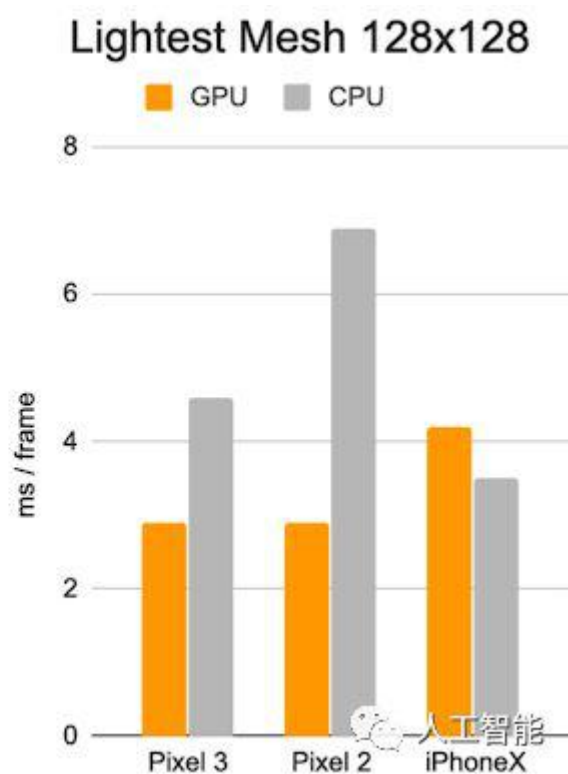
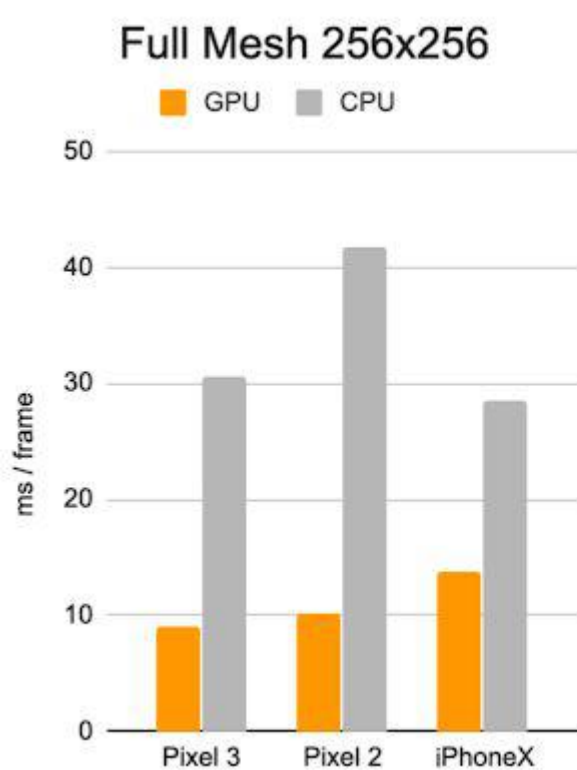
数据集扩展和改进 pipeline

我们使用 TensorFlow Lite 进行设备上的神经网络推理。新推出的 GPU 后端加速可在可用的情况下提升性能，并显著降低功耗。

此外，为了涵盖广泛的消费类硬件，我们设计了各种具有不同性能和效率特性的模型架构。

较轻的网络最重要的区别是残余块布局 and 可接受的输入分辨率（最轻的模型中为 128x128 像素，而最复杂的模型中为 256x256）。

我们还改变了层数和子采样率（输入分辨率随网络深度减小的速度）。



每帧的推理时间：CPU 与 GPU

这些优化的结果是使用较轻型号的显著加速，AR 效果质量的降低最小。



比较最复杂（左）和最轻的模型（右）。

在轻型模型上，时间一致性以及唇部和眼睛跟踪略微降低

这些努力的最终结果是通过以下方式为用户体验在 YouTube，ARCore 和其他客户中提供令人信服的，逼真的自拍 AR 效果：

通过环境映射模拟光反射，实现眼镜的逼真渲染

通过将虚拟对象阴影投射到面网格上来实现自然光照

对面部遮挡建模以隐藏面部后面的虚拟对象部分，例如虚拟眼镜，如下图所示



YouTube Stories 基于 3D 网格的逼真虚拟眼镜

此外，我们通过以下方式实现高度逼真的妆效：

建模在嘴唇和嘴唇上应用的镜面反射

通过使用亮度感知材料进行面部绘画

案例研究将不同光照条件下 5 个主题的真实化妆与 AR 妆容进行比较。

未来，谷歌计划将此技术扩展到更多谷歌产品中。

参考链接：

<https://www.engadget.com/2019/03/06/google-tilda-swinton-big-bang-ar-app/>

<https://ai.googleblog.com/2019/03/real-time-ar-self-expression-with.html>