

# 通过引入 ML,谷歌 ARCore 实时自拍 AR 的逼真效果更进 一步

人工智能3月12日



ARCore 是谷歌于去年 2 月份正式推出的 增强现实 (AR) 软件开发工具包, 给开发者提供了一个开发 AR APP 的平台。不到一个月前, 谷歌正式推出 ARCore 1.7 版本, 为其添加了前置摄像头 AR 自拍能力以及动画效果支持; 而日前, 谷歌又为其引入了机器学习技术, 使其实时自拍 AR 的逼真效果更进一步, 该技术相关的文章发布在谷歌 AI 的官方博客上, 雷锋网 AI 科技评论编译如下。

增强现实 (AR) 通过将数字内容与信息叠加到物质世界的真实场景中, 来让人类实现超越现实的感官体验。例如, 谷歌地图的 AR 功能, 能够在现实场景中叠加方向

信息来为你指路。借助于 Pixel 相机中的 Playground 模式，你可以使用 AR 从不同的视角看世界。并且通过借助于最新发布的「YouTube Stories」以及 ARCore 全新的面部增强（Augmented Faces）API，你可以在自拍上添加动画面具、眼镜、帽子、皮肤等特效。

实现这种 AR 功能的关键挑战之一，就是以合适的方式将虚拟内容锚定到现实世界：这一过程需要一套能够追踪到每一次微笑、皱眉或假笑的高动态表面几何结构的独特的感知技术。



ARCore 的 3D 网格以及它能实现的一些特效

为此，我们引入了机器学习（ML）技术来推断近似的 3D 表面几何结构，并且仅需要用到一个单摄像头输入而无需使用专用的深度传感器。这种方法利用面向移动 CPU 界面推断或其可用的全新移动 GPU 功能的 TensorFlow Lite，实现了逼真的实时 AR 效果。与「YouTube Stories」全新的创作者特效所利用的技术一样，这项技术已通过最新的 ARCore SDK 和 ML Kit Face Contour Detection API 向广泛的开发者社区开放。

## 用于自拍 AR 的机器学习 workflow

我们的机器学习 workflow 由两个协同工作的实时深度神经网络模型组成：一个是探测器，它在整张图像上运行并计算出面部位置；另一个是通用的 3D 网格模型，它在探测器计算出来的面部位置上运行，并通过回归预测近似的表面几何结构。精确地裁剪面部能够极大地减少对一般的数据增强的要求，例如由旋转、平移和比例变换组成的仿射变换（affine transformations）等，同时让网络将大部分性能投入于预测坐标来提高其准确性——这对于锚定合适的虚拟内容至关重要。

裁剪完所关注的位置后，该神经网络一次仅应用于一个单帧，并利用加窗平滑（windowed smoothing）来减少面部处于静态时的噪音，同时避免在大幅移动期间出现的延迟。

针对 3D 网格，我们采用了迁移学习，并训练了一个具有多个目标的网络：该网络能够同时地基于合成、渲染的数据预测 3D 网格坐标以及基于类似于 MLKit 所提供的带有注释的、现实世界的的数据预测 2D 语义轮廓。最终的网络为我们提供了基于合成数据乃至现实世界数据的合理的 3D 网格预测。所有模型都在源自于地理学多样化的数据集的数据上进行训练，并接着在平衡、多样化的测试集上进行定性和定量性能的测试。

3D 网格网络接收经裁剪的视频帧作为输入。由于它不依赖于额外的深度输入，因而也可以应用于预先录制的视频。该模型不仅输出 3D 点的位置，还输出在输入中存

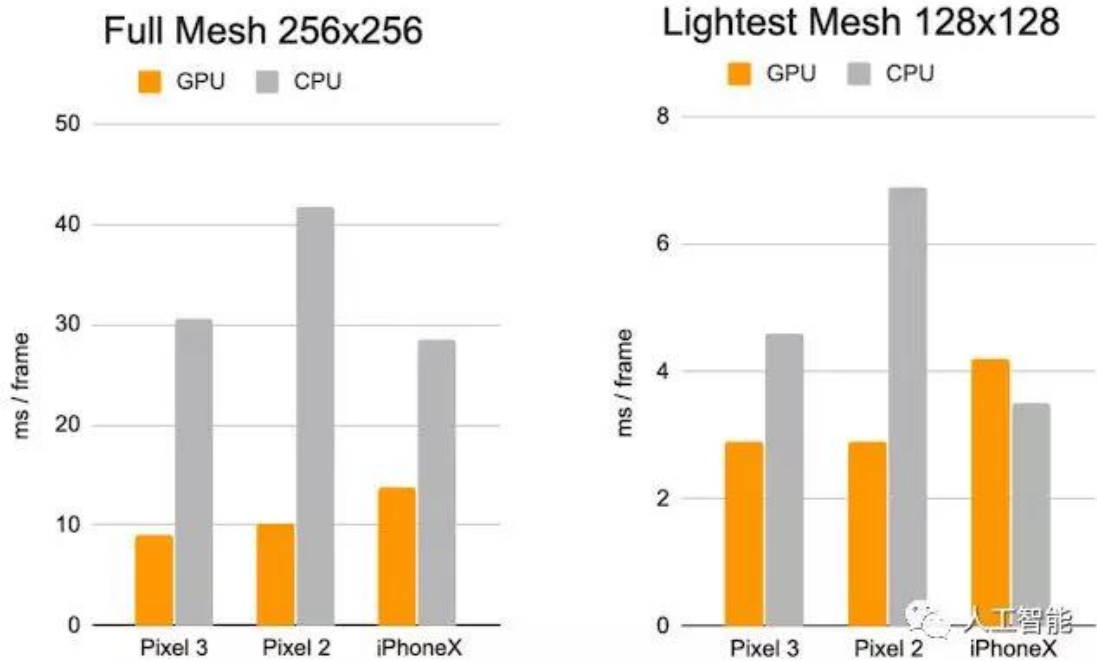
在并合理对齐的面部概率。一种常见的替代方法就是为每个地标预测出 2D 热图，但这种方法并不适用于深度预测，并且对于如此多的数据点来说，使用这种方法耗费的计算成本非常高。

通过迭代自展和精炼预测，我们进一步提高了模型的准确性和鲁棒性，同时也将数据集发展成为愈加具有挑战性的实例，如鬼脸，拍摄角度倾斜以及面部遮挡等。此外，数据集增强技术也扩展了可用的 ground truth 数据，并开发出了对摄像头瑕疵等人工产品问题或极端光照条件具有弹力的模型。

数据集扩展和改善路径

### **专为硬件定制的界面**

我们使用 TensorFlow Lite 进行设备内置的神经网络推理。新推出的 GPU 后端加速能够有效提升性能，并显著降低功耗。此外，我们还设计了具有不同性能和效率特性的一系列模型架构，来覆盖到更加广泛的消费级硬件。较简单网络之间的最重要的区别就在于残差块 ( Residual Block ) 设计和可接受的输入分辨率 ( 最简单模型中的输入分辨率为  $128 \times 128$  像素，而最复杂模型中的输入分辨率为  $256 \times 256$  )。同时，我们还改变了层数以及子采样率 ( 输入分辨率随网络深度的减少而减少的速度 )。



每一帧的推断时间：CPU vs. GPU

这一系列的优化的结果就是：使用更简单的模型带来了实质性的加速，同时也将 AR 特效质量的降低幅度最小化。

这些成果最终则能够通过以下方式，让用户在 YouTube、ARCore 以及其他客户端体验到更加逼真的自拍 AR 效果：

- 
- 通过环境映射模拟光反射，来实现眼镜的逼真渲染；
- 
- 
- 通过将虚拟的目标阴影投射到面部网格上，来实现了自然光照效果；
- 
-

对面部遮挡进行建模，来隐藏面部后面的虚拟目标部分，如下所示的虚拟眼镜案例。

- 
- 

另外，我们还通过以下方式实现了逼真的妆容效果：

- 对应用到嘴唇上的镜面反射进行建模；
- 
- 利用亮度感知材料来实现面部着色。
- 



案例对比：5 个目标在不同光线下的真实化妆效果以及 AR 化妆效果

我们很高兴将这项新技术分享给创作者、用户以及开发者。感兴趣的读者可以通过下

载 最 新 的 ARCore SDK ( 下 载 地 址 :

<https://developers.googleblog.com/2019/02/new-ui-tools-and-richer-creative-canvas.html> ) 来将这项技术用起来。未来，我们还计划将这项技术广泛应用到更多的谷歌产品线中。

via [Google AI](#) :

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

雷锋网(公众号：雷锋网)