

一位中国博士把整个 CNN 都给可视化了，可交互有细节，每次卷积 ReLU 池化都清清楚楚

AI有道 今天

AI 有道

资源、干货、教程、前沿

郭一璞 发自 云凹非寺

量子位 报道 | 公众号 QbitAI

CNN是什么？美国有线电视新闻网吗？

每一个对AI抱有憧憬的小白，在开始的时候都会遇到CNN（卷积神经网络）这个词。

但每次，当小白们想了解CNN到底是怎么回事，为什么就能聪明的识别人脸、听辨声音的时候，就懵了，只好理解为玄学：

结构 [编辑]

卷积层 [编辑]

卷积层是一组平行的特征图（feature map），它通过在输入图像上滑动不同的卷积核并运行一定的运算而组成。此外，在每一个滑动的位置上，卷积核与输入图像之间会运行一个元素对应乘积并求和的运算以将感受野内的信息投影到特征图中的一个元素。这一滑动的过程可称为步幅 Z_s ，步幅 Z_s 是控制输出特征图尺寸的一个因素。卷积核的尺寸要比输入图像小得多，且重叠或平行地作用于输入图像中，一张特征图中的所有元素都是通过一个卷积核计算得出的，也即一张特征图共享了相同的权重和偏置项。

线性整流层 [编辑]

线性整流层（Rectified Linear Units layer, ReLU layer）使用线性整流（Rectified Linear Units, ReLU） $f(x) = \max(0, x)$ 作为这一层神经的激励函数（Activation function）。它可以增强判定函数和整个神经网络的非线性特性，而本身并不会改变卷积层。

事实上，其他的一些函数也可以用于增强网络的非线性特性，如双曲正切函数 $f(x) = \tanh(x)$ ， $f(x) = |\tanh(x)|$ ，或者Sigmoid函数 $f(x) = (1 + e^{-x})^{-1}$ 。相比其它函数来说，ReLU函数更受青睐，这是因为它可以将神经网络的训练速度提升数倍^[3]，而并不会对模型的泛化准确度造成显著影响。

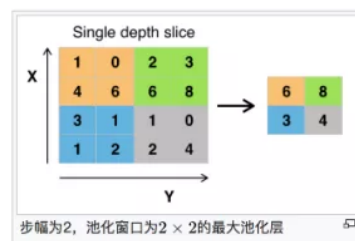
池化层 [编辑]

池化（Pooling）是卷积神经网络中另一个重要的概念，它实际上是一种非线性形式的降采样。有多种不同形式的非线性池化函数，而其中“最大池化（Max pooling）”是最为常见的。它是将输入的图像划分为若干个矩形区域，对每个子区域输出最大值。

直觉上，这种机制能够有效地原因在于，一个特征的精确位置远不及它相对于其他特征的粗略位置重要。池化层会不断地减小数据的空间大小，因此参数的数量和计算量也会下降，这在一定程度上也控制了过拟合。通常来说，CNN的网络结构中的卷积层之间都会周期性地插入池化层。池化操作提供了另一种形式的平移不变性。因为卷积核是一种特征发现器，我们通过卷积层可以很容易地发现图像中的各种边缘。但是卷积层发现的特征往往过于精确，我们即使高速连拍拍摄一个物体，照片中的物体的边缘像素位置也不大可能完全一致，通过池化层我们可以降低卷积层对边缘的敏感性。

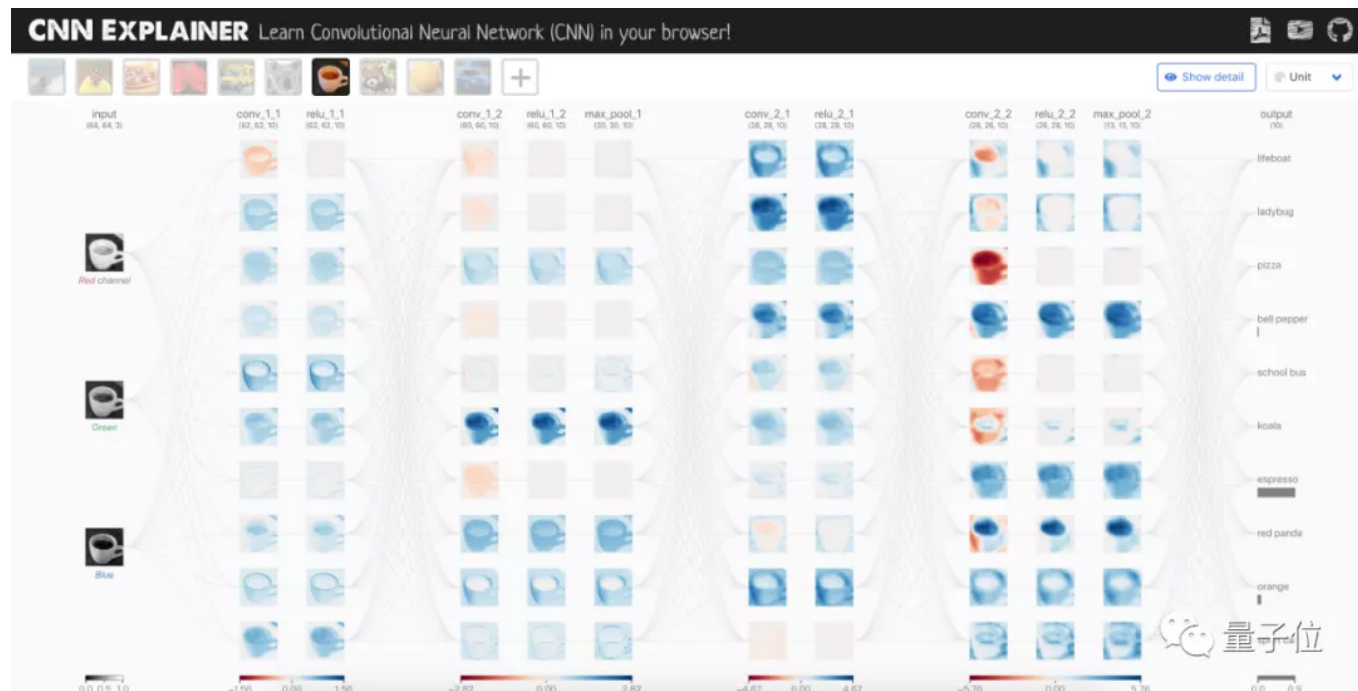
池化层每次在一个池化窗口（depth slice）上计算输出，然后根据步幅移动池化窗口。下图是当前最常用的池化层，步幅为2，池化窗口为 2×2 的二维最大池化层。每隔2个元素从图像划分出 2×2 的区块，然后对每个区块中的4个数取最大值。这将会减少75%的数据量。

$$f_{X,Y}(S) = \max_{a,b=0}^1 S_{2X+a,2Y+b}$$



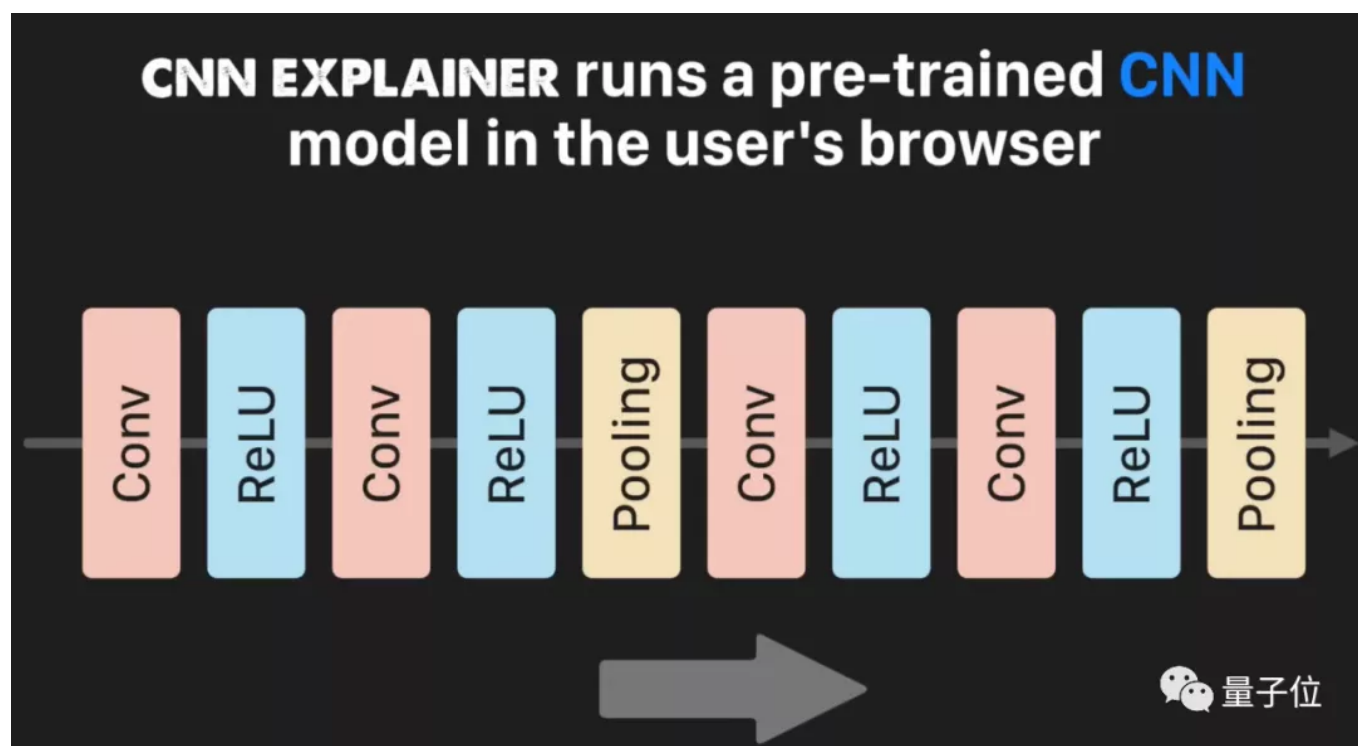
量子位

好吧，维基百科解决不了的问题，有人给解决了。



这个名叫**CNN解释器**在线交互可视化工具，把CNN拆开了揉碎了，告诉小白们CNN究竟是怎么回事，为什么可以辨识物品。

它用TensorFlow.js加载了一个10层的预训练模型，相当于在你的浏览器上跑一个CNN模型，只需要打开电脑，就能了解CNN究竟是怎么回事。



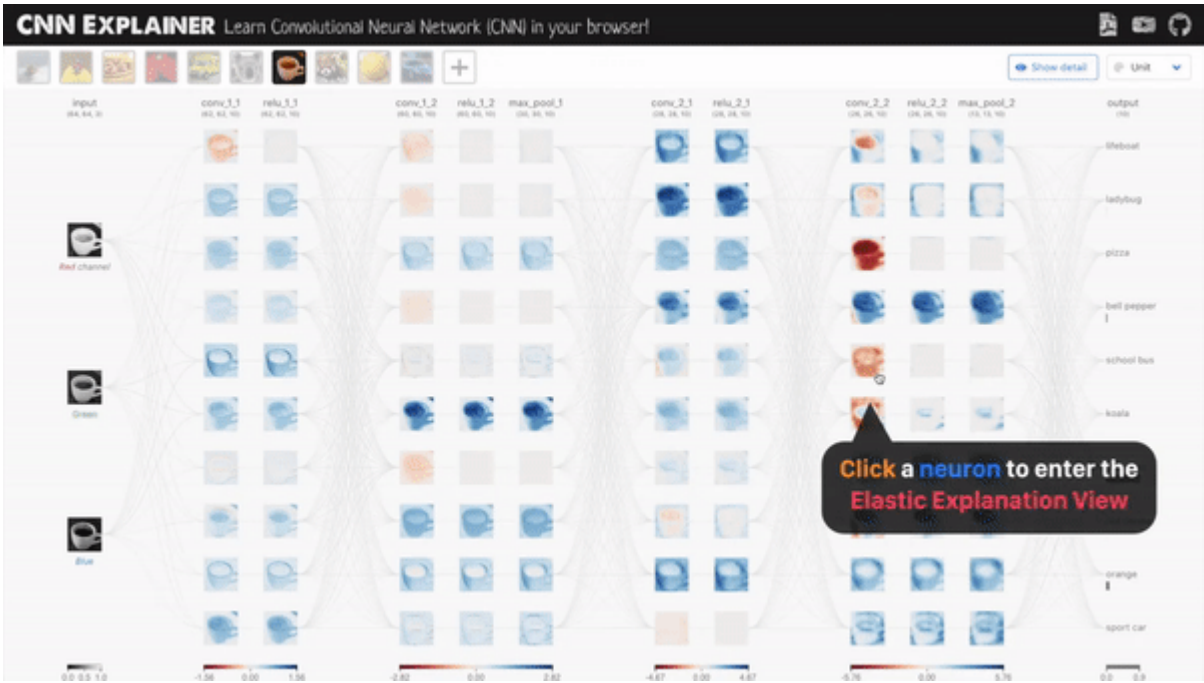
而且，这个网页工具还可以实现交互，只要点击其中任何一个格子——就是CNN中的“神经元”，就能显示它的输入是哪些、经过了怎样细微的变化。

甚至，连每一次卷积运算都能看得清。

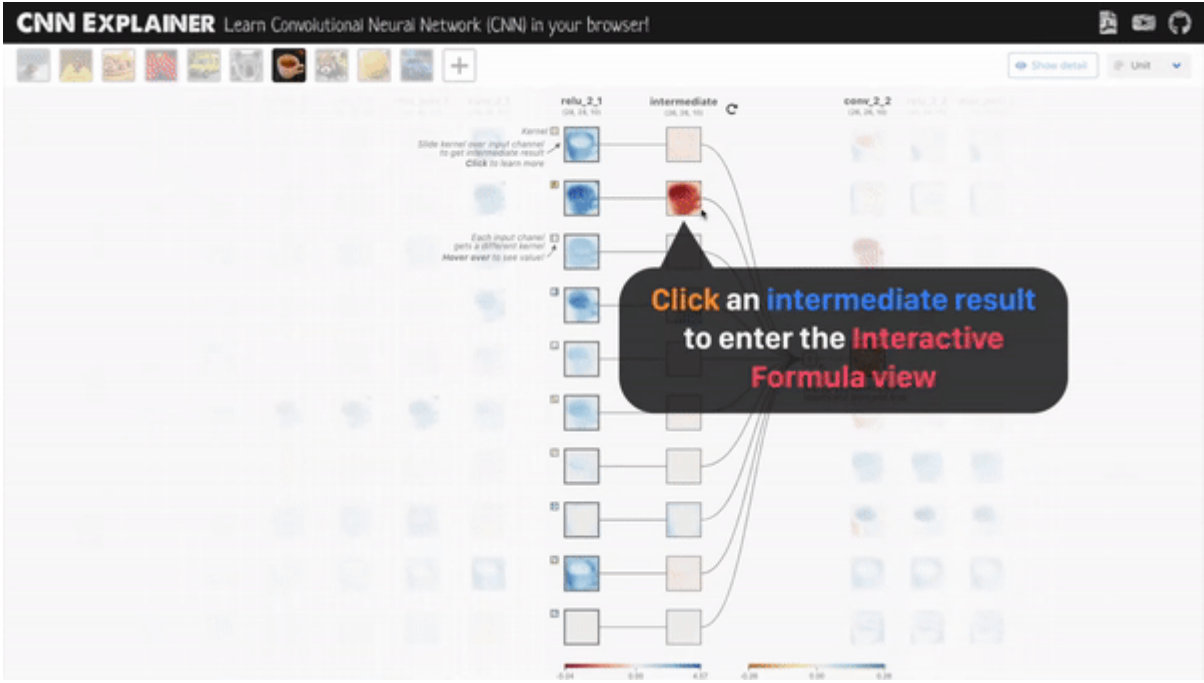
看清卷积

这个CNN解释器的使用方式也非常简单：鼠标戳戳戳就好了。

单击神经元，进入弹性解释视图，就可以看到卷积核滑动的过程的动画模拟：



点击一个正在卷积的过程图，就可以看到更具体的过程：



可以看到底层的卷积运算过程， 3×3 的卷积核是如何经过运算被变成1个数字的。

看清ReLU和最大池化层

点击一个ReLU层的神经元，可以看具体过程，ReLU函数是这样工作的：



点击一个池化神经元，也可以看具体最大池化层是怎样工作的：

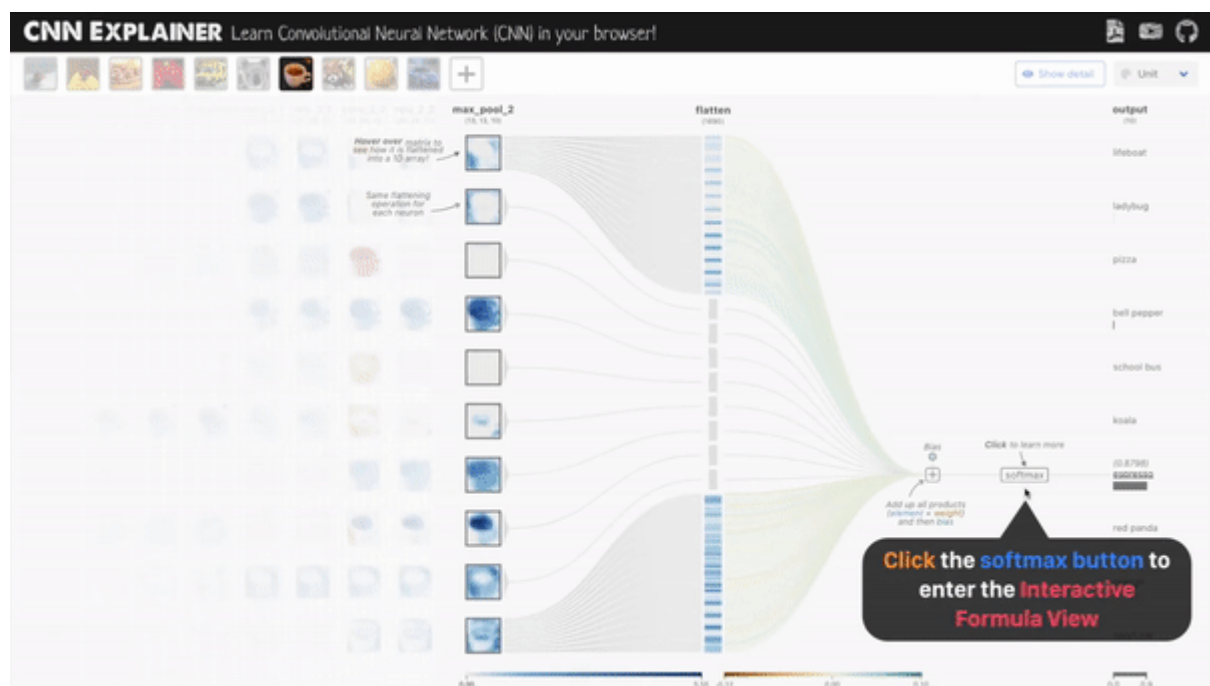


看清CNN是怎么输出预测的

点击最右侧的输出神经元，进入弹性解释视图：



可以查看Softmax函数的详情：



识别「狸克」试试？

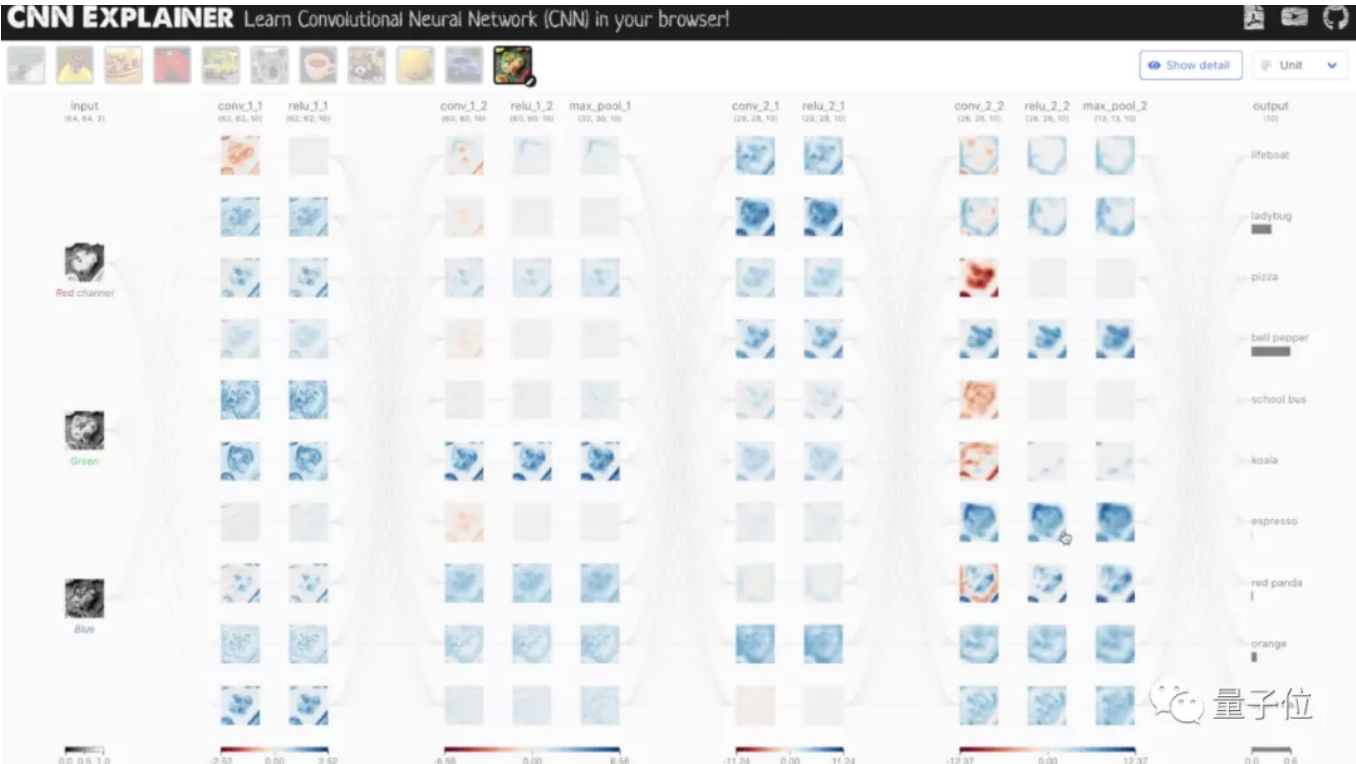
CNN解释器默认有10张图片，你也可以增加自己的自定义图片。

比如这个：



芝士焗灯笼椒？灯笼椒披萨？还是什么鬼？

复制图片链接或者上传图片之后，它经过了10层处理，得出结论：

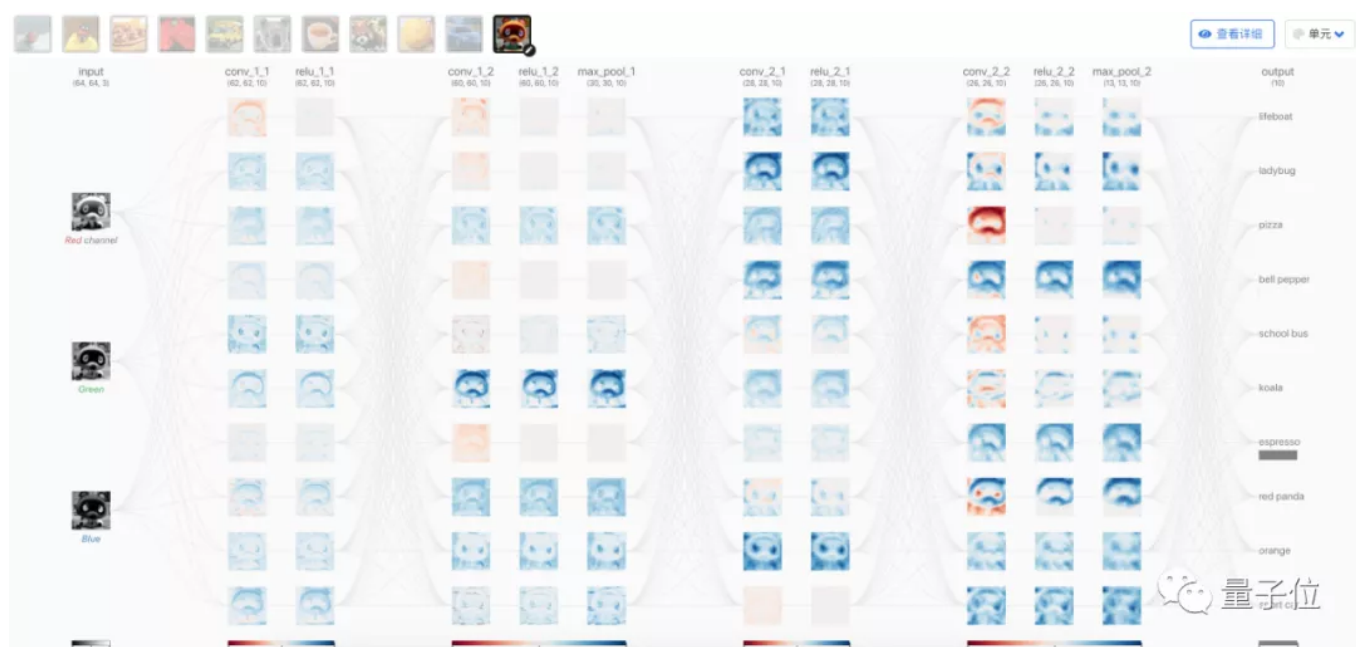


是灯笼椒，但也可能是臭虫。

但是，只能归类于右侧原有的10个品类里，比如放一只狸克：



就会被识别成espresso。

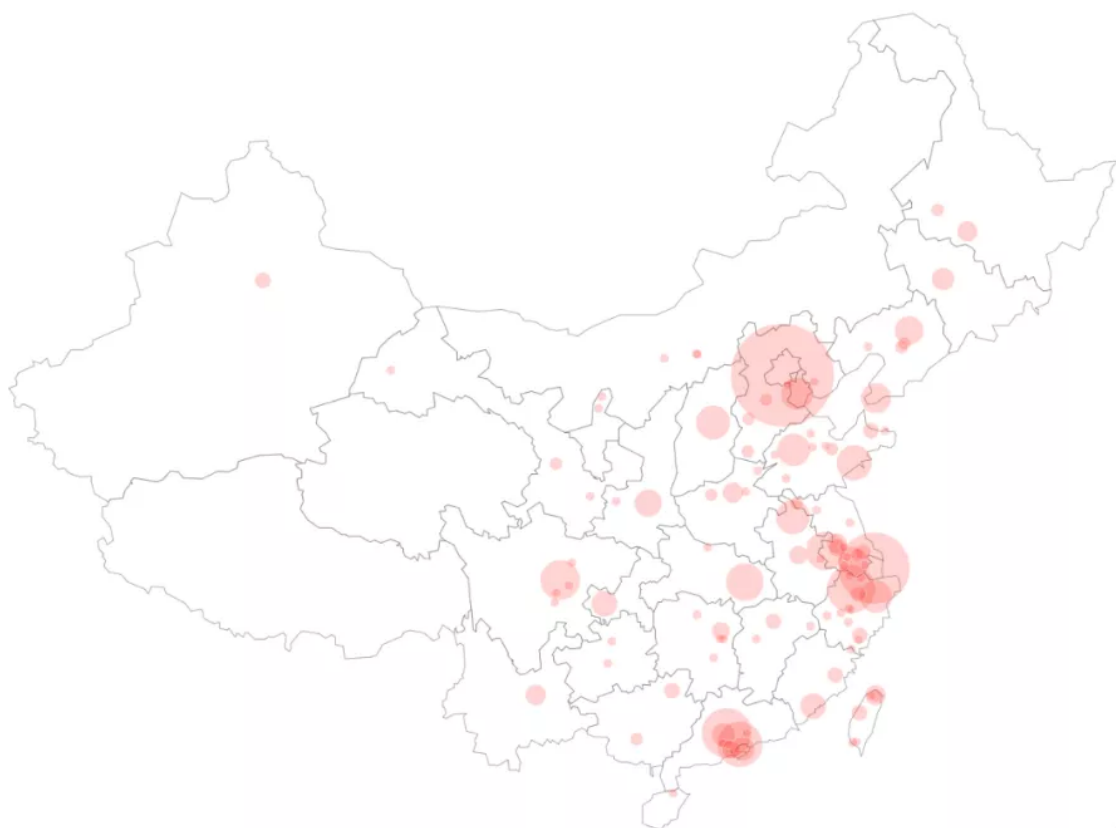


佐治亚理工中国博士出品

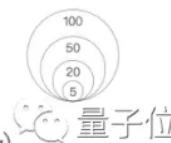
最后，这个CNN解释器的作者是一位中国小哥，佐治亚理工的Zijie Wang，去年刚开始读机器学习博士，本科毕业于威斯康星大学麦迪逊分校，是一位GPA 3.95/4.00的大学霸。



他也曾经做过一些其他有趣的数据可视化项目，比如威斯康星大学麦迪逊分校的中国本科生都来自哪里：



Chinese Students on UW-Madison Dean's List (2018 Spring)



传送门

CNN解释器

<https://poloclub.github.io/cnn-explainer/>

GitHub

<https://github.com/poloclub/cnn-explainer>

论文

<https://arxiv.org/abs/2004.15004>

作者系网易新闻·网易号“各有态度”签约作者

推荐阅读 （点击标题可跳转阅读）

[干货 | 公众号历史文章精选](#)

[我的深度学习入门路线](#)

[我的机器学习入门路线图](#)

算法工程师必备！

AI有道年度技术文章电子版PDF来啦！

扫描下方二维码，添加 **AI有道小助手微信**，可申请入群，并获得2020完整技术文章合集PDF（一定要备注：**入群 + 地点 + 学校/公司**。例如：**入群+上海+复旦**。



长按扫码，申请入群

（添加人数较多，请耐心等待）

最新 AI 干货，我在看 

[阅读原文](#)