

# 机器如何做梦？

先看两幅动图。



【gif 动图】 <https://i.gifer.com/JnGh.gif>



【fig 动图】 <https://www.primogif.com/p/cTvanfGrVCTK0>

是不是很魔幻？

这些图是 Google 开发的 DeepDream 工具[1]生成的。DeepDream 翻译过来是“深度做梦”，让计算机进入“梦境”。让我们看看这个梦是怎么做出来的。

我们知道，一个训练好的深度神经网络可以逐层进行特征提取，在低层提取较细节的特征，

在高层提取较全局的特征。如图 1 所示的一个用于动物识别的卷积神经网络（CNN），在第一层提取一些线条特征，第二层提取一些局部轮廓特征，到了第三层将提取整体轮廓特征。这些特征被保存在神经网络的卷积核中，当图片出现某一特征时，在相应位置上对应的卷积核将得到激发。例如，当图片中出现一条马腿，则对应马腿的卷积核（一般位于中间层）在图片中马腿出现的位置将得到一个强激发。当然，一匹马有马头、马尾、马身等很多部件，这些部件对应的特征都会在合适的位置激发相应的卷积核。正是依靠这些层次性的特征激发，神经网络才能最终得到输入图片是一匹马或一条狗的判断。

值得说明的是，在不同数据上训练得到的 CNN 所存储的特征（卷积核）是不一样的，用于动物识别的 CNN 存储的可能是头、尾、腿等特征，用于花草识别的 CNN 存储的可能是花瓣，花蕊，叶片等特征。从这个角度看，CNN 其实是记忆了训练数据的主要特征，这些特征分层次地组织起来，由局部到全局，由具体到抽象。这有点儿像我们的大脑，把见过的形象分层次地组织并记忆下来。DeepDream 正是基于这一记忆结构实现做梦过程的。

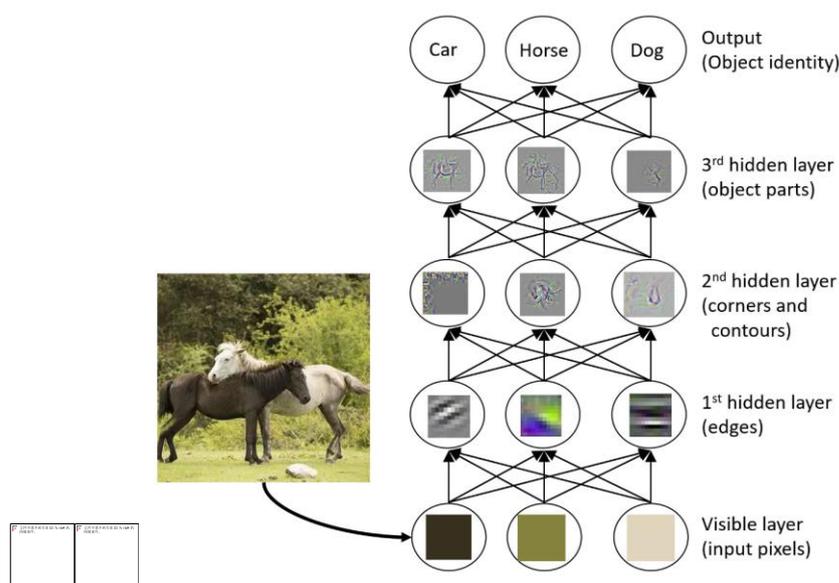


图 1：用于动物识的深度卷积神经网络[2]

具体是如何做的呢？输入任意一张图片，比如一张蓝天白云的照片，在 CNN 的某一层寻找激发最强的卷积核。这个激发的存在说明输入图片在相应位置存在某一个显著特征，如一张马脸或一个花瓣。虽然马和花在图中并不存在，但马脸、花瓣这些局部特征可能隐藏在云朵图像里，否则不会被 CNN 发现。找到这个激发值后，对图片进行调整，使得这个激发值更大，就可以将对应的特征更明显地展示出来，如图 2 所示。这有点儿像我们盯着云朵时给自己的心理暗示，发现一朵云象一匹马，然后盯着这匹马不停地看，越看就会感觉越象。这就是我们说的特征强化过程。

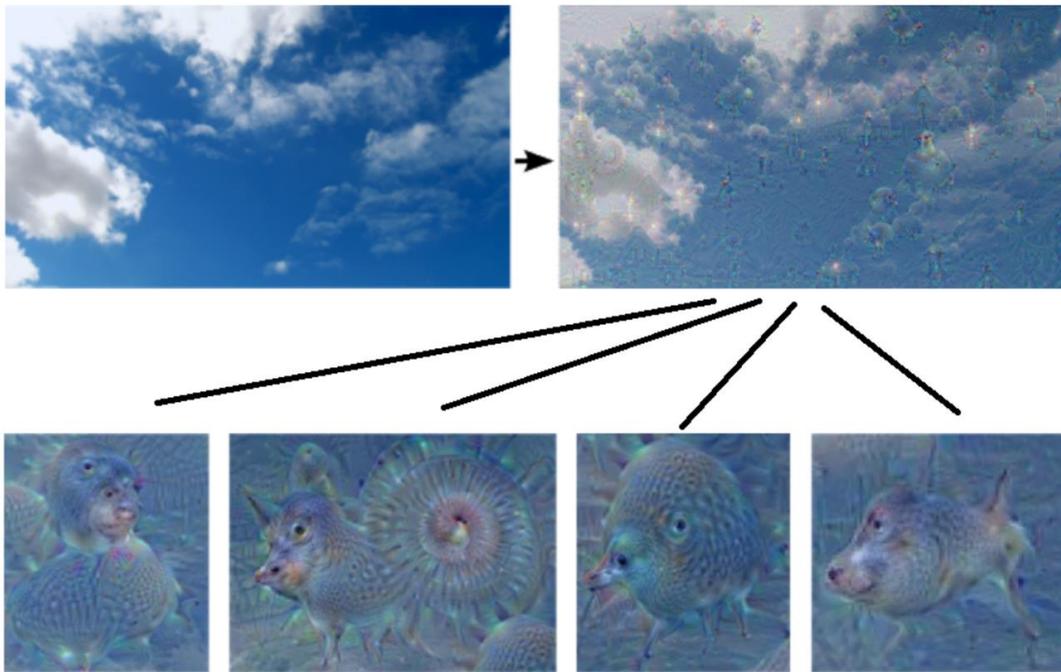


图 2：对局部特征进行强化后，在图中更明显地展示出来[1,3].

特征强化过程只针对一张图。如果我们在强化过程中“拉近镜头”，即把图片做放大和裁剪，就可以得到一个动态的“做梦”过程。如图 3 所示，初始输入图片完全是噪声，经过若干次强化和放大后，渐渐形成梦境。

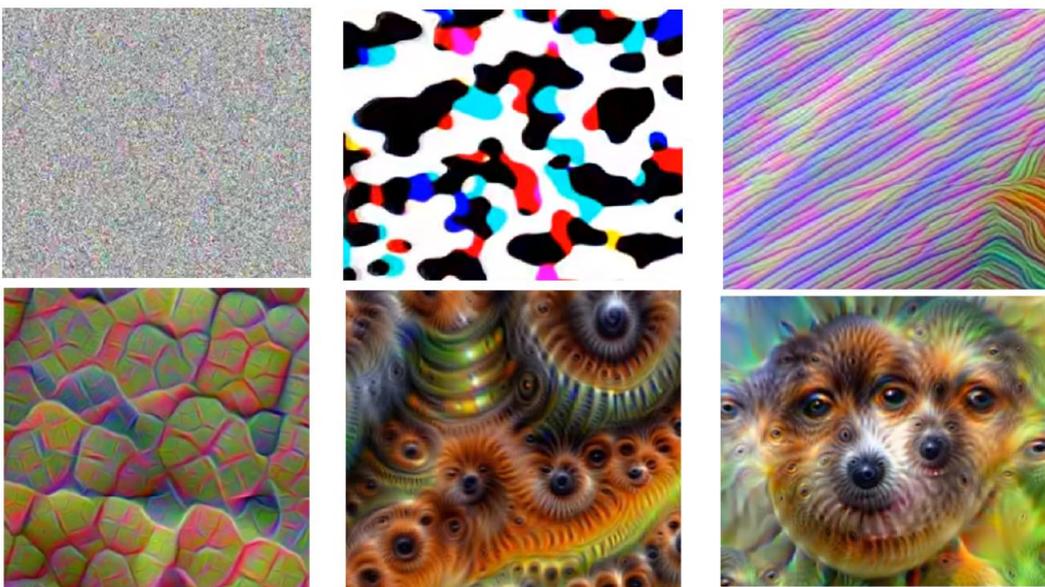


图 3：经过反复局部放大和特征增强，从一幅白噪声生成动态的梦境[1,3]。

总结起来，深度神经网络之所以会做梦，主要原因是 CNN 通过卷积核记住了不同层次的特

征模式，将这些模式通过一种随机的方式进行组合和强化，就形成了梦境般的效果。DeepDream 事实上展示了神经网络对特征的层次性记忆能力。

[1]<https://www.wired.co.uk/article/google-deepdream-dreamscope>

[2]Zhihua Zhou, Ji Feng, Deep Forest, 2020.6.

[3]王东，利节，许莎，《人工智能》，2019,10.