

作者：许铁-巡洋舰科技

链接：<https://www.zhihu.com/question/36500536/answer/281943900>

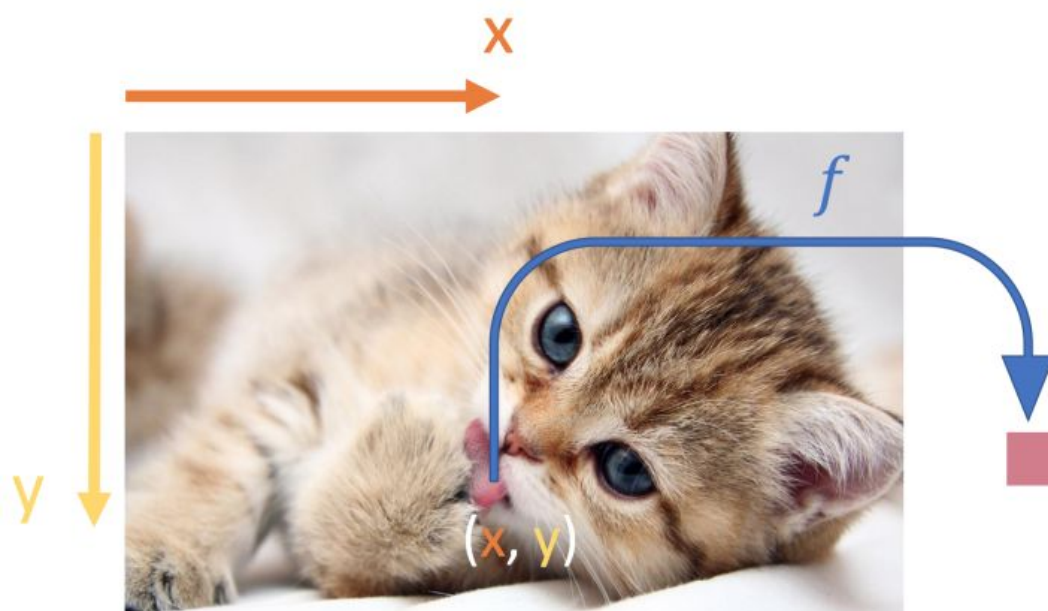
来源：知乎

著作权归作者所有。商业转载请联系作者获得授权，非商业转载请注明出处。

用深度学习玩图像的七重关卡

第一个重境界： 图像识别

如果你开始了解深度学习的图像处理，你接触的的第一个任务一定是图像识别：
比如把你的爱猫输入到一个普通的CNN网络里，看看它是猫咪还是狗狗。



v2-a58afcaf16004d32827dbce090b25233_hd.jpg

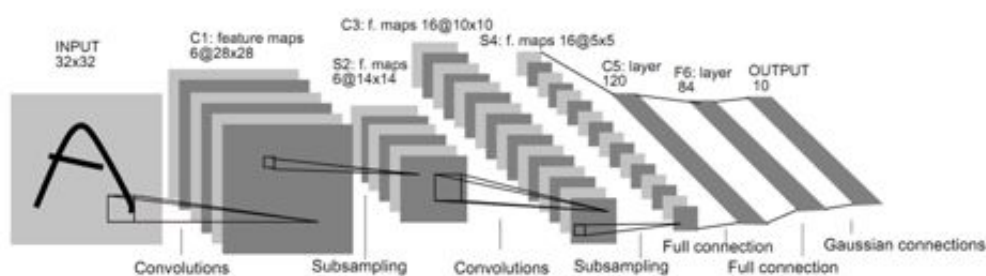
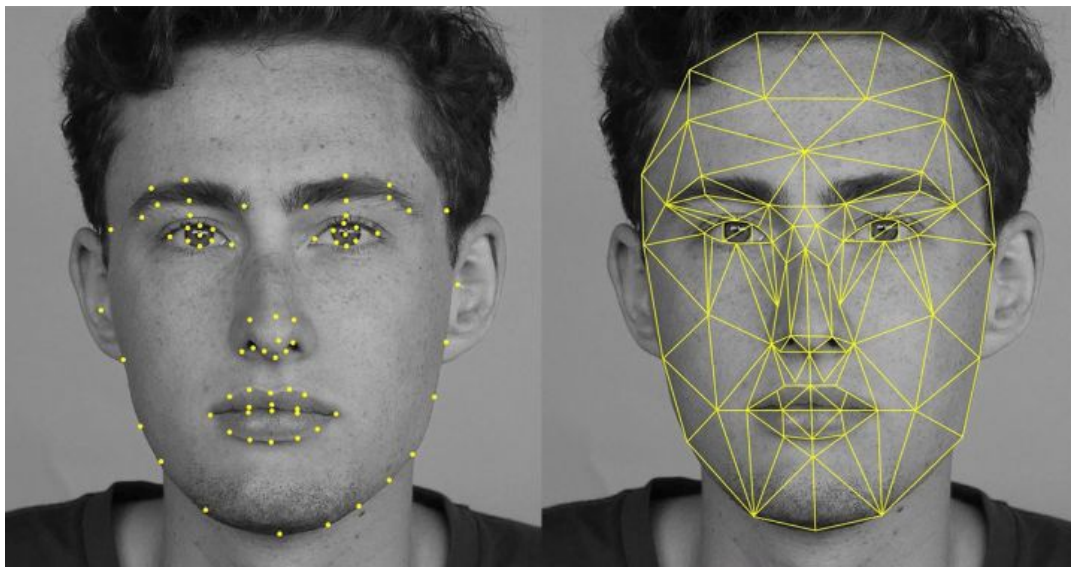


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

v2-849cbfd7dc6fc55efe1d8bbd18691a91_hd.jpg

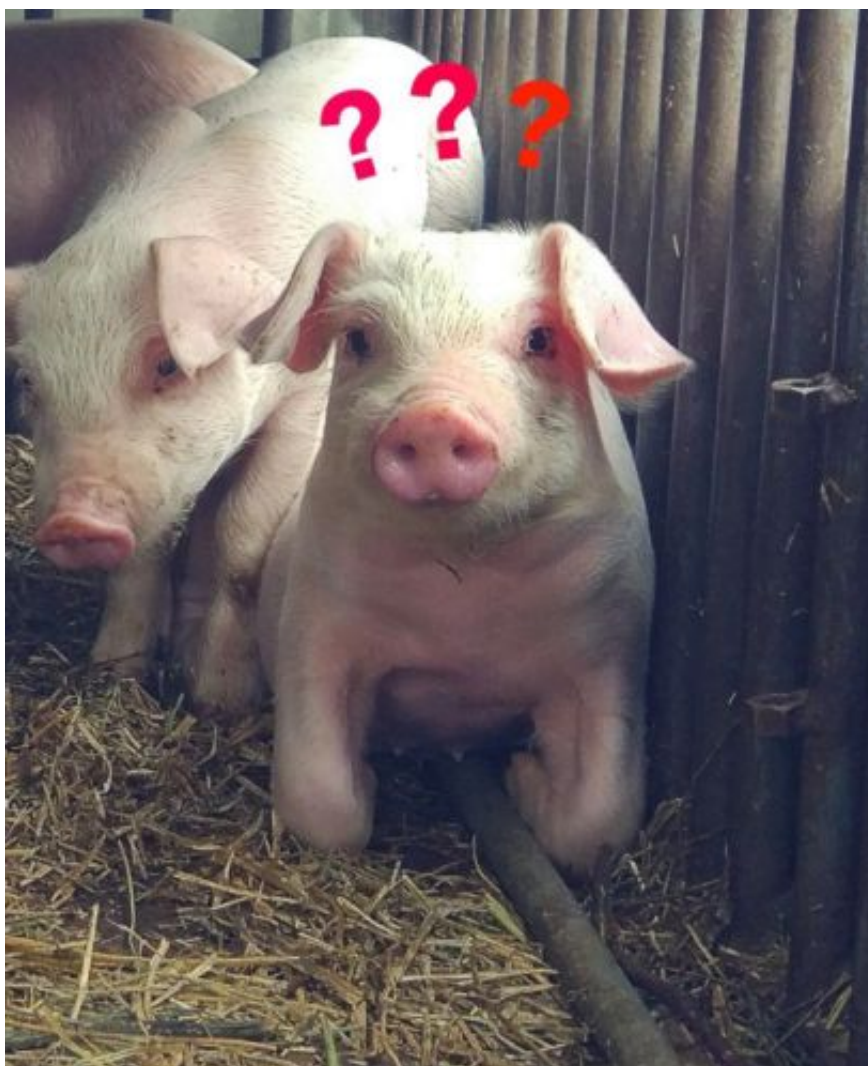
一个最普通的CNN，比如像这样几层的CNN鼻祖Lenet，如果你有不错的数据集（比如kaggle猫狗大战）都可以给出一个还差强人意的分类结果(80%多准确率)，虽然不是太高。

当然，如果你再加上对特定问题的一些知识，也可以顺便识别个人脸啥的，开个startup叫face 减减什么：



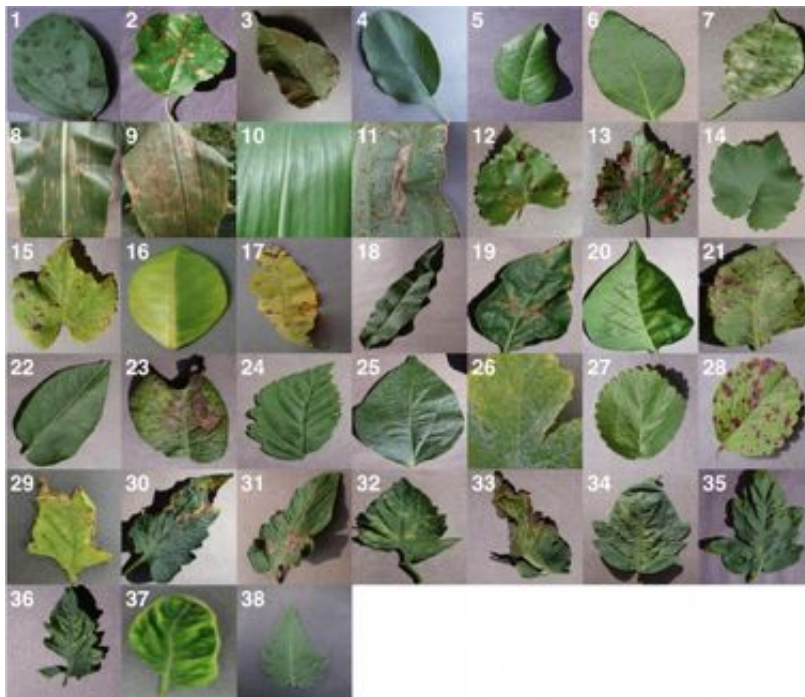
v2-6ba9173a9650e1a6c893411f22428ea7_hd.jpg

会玩的，也可以顺便识别个猪脸什么哒（我觉得长得都一样哦），这样搞出来每个猪的身份，对于高质量猪肉的销售，真是大有裨益的。



v2-1848e686a2d5ef1bec243ec6a7a58725_hd.jpg

或者看看植物都有个什么病害什么的，像这样不同的病斑，人都懒得看的，它可以给你看出来。植物保护的人可以拿着手机下田了。

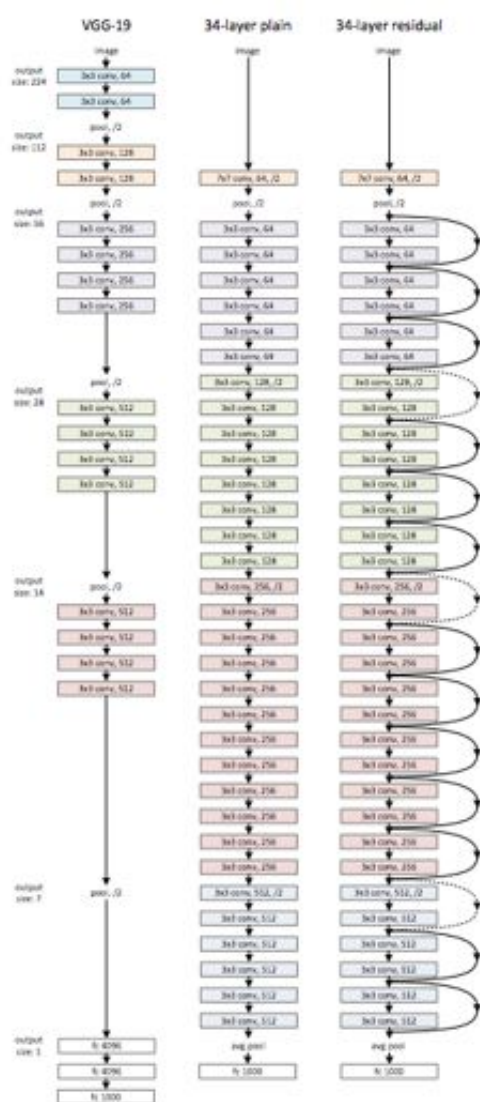


v2-bd7203140b1965b0294959d19e2b153c_hd.jpg

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015.

虽然植物保护真的很好用，分类问题做就了还真是挺无聊的。

我们进化的方向，也就是用更高级的网络结构取得更好的准确率，比如像下图这样的残差网络（已经可以在猫狗数据集上达到99.5%以上准确率）。分类做好了你会有一种成为深度学习大师，拿着一把斧子眼镜里都是钉子的幻觉。分类问题之所以简单，一要归功于大量标记的图像，二是分类是一个边界非常分明的问题，即使机器不知道什么是猫什么是狗，看出点区别还是挺容易的，如果你给机器几千几万类区分，机器的能力通过就下降了（再复杂的网络，在imagenet那样分1000个类的问题里，都很难搞到超过80%的准确率）。



v2-6a9a9615c9d32153415b1c4864deb08b_hd.jpg

He, Kaiming, et al. "Identity mappings in deep residual networks." European Conference on Computer Vision. Springer International Publishing, 2016.

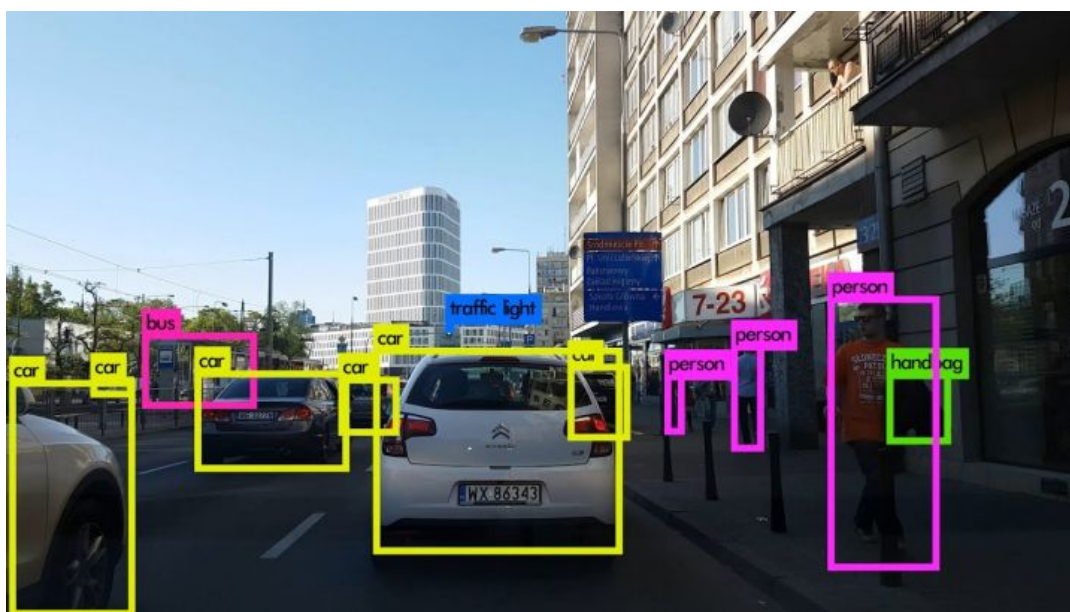
第二重境界： 物体检测

很快你发现，分类的技能在大部分的现实生活里并没有鸟用。因为现实中的任务啊， 往往是这样的：



v2-6b27f8050e2f99c97aad1e0834431bd1_hd.jpg

或者这样的：

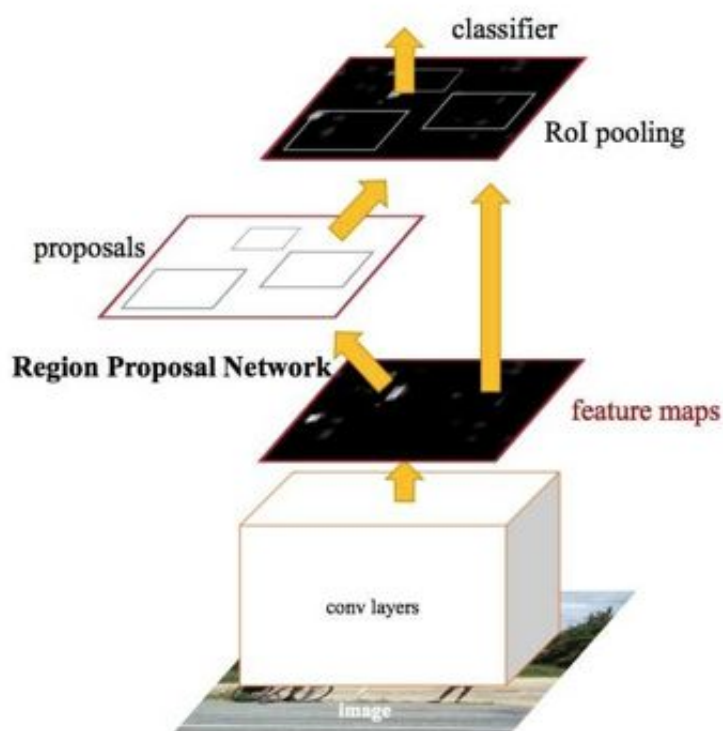


v2-468921caa00e188e31ea8af23d9a999a_hd.jpg

那么多东西在一起，你拿猫狗大头照训练的分类网络一下子就乱了阵脚。即使是你一个图片里有一个猫还有一个狗，甚至给猫加点噪声，都可以使你的分类网络分寸大乱。

现实中，哪有那么多图片，一个图里就是一个猫或者美女的大图，更多的时候，一张图片里的东西，那是多多的，乱乱的，没有什么章法可言的，你需要

自己做一个框，把你所需要看的目标给框出来，然后，看看这些东西是什么。于是你来到机器视觉的下一层挑战 - 目标检测（从大图中框出目标物体并识别），随之而来的是一个新的网络架构，又被称为R - CNN，图片检测网络，这个网络不仅可以告诉你分类，还可以告诉你目标物体的坐标，即使图片里有很多目标物体，也一一给你找出来。

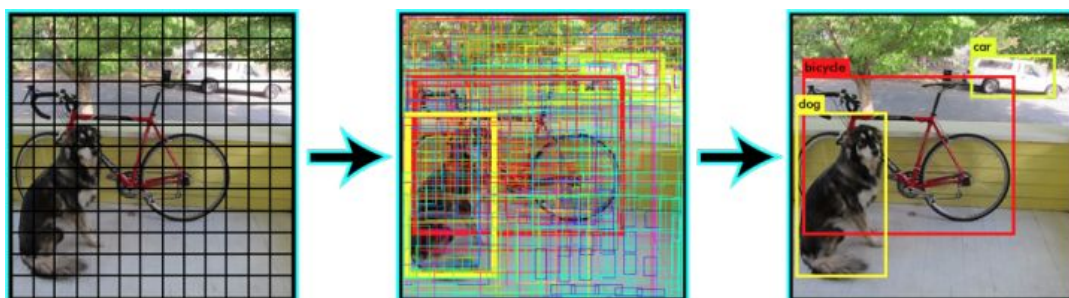


v2-0fdb2e4aaa97b5e942d244f8b42ca75_hd.jpg

Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.

万军斩你首级那是杠杠的，在众多路人甲中识别嫌疑犯，也是轻而易举，安防的人听着要按捺不住了。

今年出现的YOLO算法更是实现了快速实时的物体检测，你一路走过就告诉你视线里都有什么在哪里，要知道这在无人驾驶里是何等的利器。



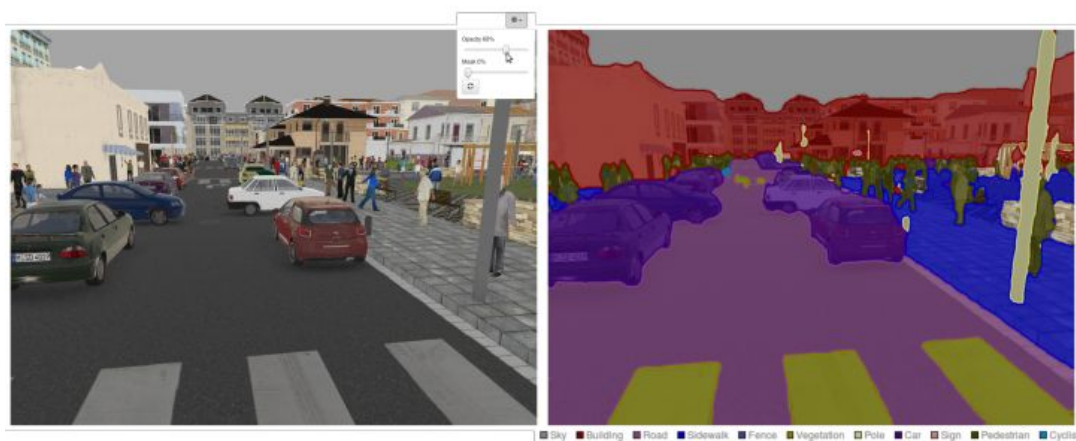
v2-7b16197312853cc22bae869434e7eb29_hd.jpg

YOLO快速检测法Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

当然，到这里你依然最终会觉得无聊，即使网络可以已经很复杂，不过是一个CNN网络（推荐区域），在加上一层CNN网络做分类和回归。能不能干点别的？

第三重境界：图像切割

啊哈，这就来到了第三个关卡，你不仅需要把图片中边边角角的物体给检测出来，你还要做这么一个猛料的工作，就是把它从图片中扣出来。要知道，刚出生的婴儿分不清物体的边界，比如桌上有苹果这种事，什么是桌子，什么是苹果，为什么苹果不是占在桌子上的？所以，网络能不能把物体从一个图里抠出来，事关它是否真的像人一样把握了视觉的本质。这也算是对它的某种“图灵测试”。而把这个问题简化，我们无非是在原先图片上生成出一个原图的“mask”，面具，有点像photoshop里的蒙版的東西。



所谓抠图



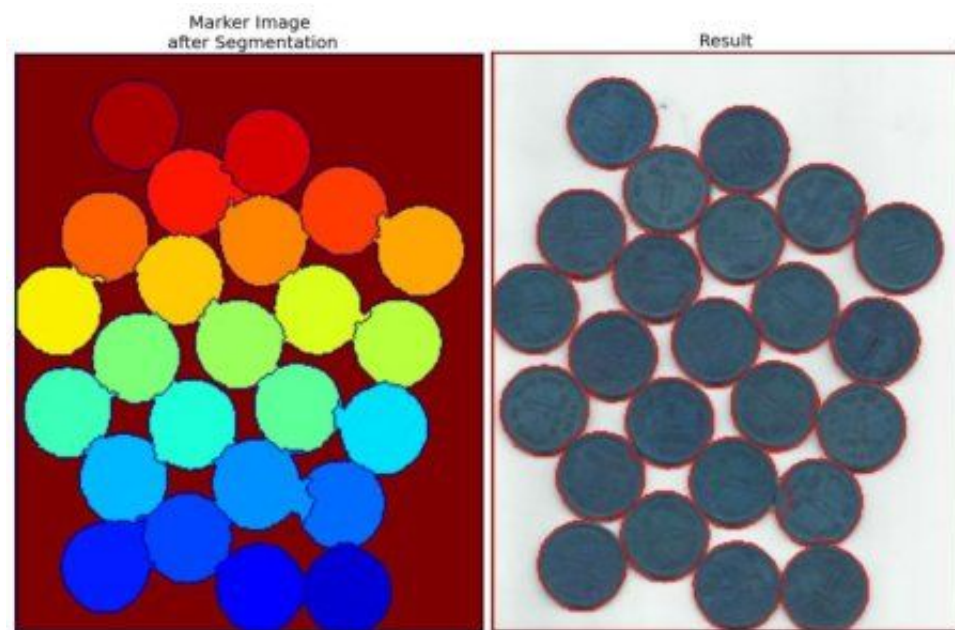
注意，这个任务里，我们是要从一个图片里得到另一个图片哦！生成的面具是另一个图片，这时候，所谓的U型网络粉墨登场，注意这是我们的第一个生成式的模型。它的组成单元依然是卷积，但是却加入了maxpooling的反过程升维采样。

这个Segmentation任务，作用不可小瞧哦，尤其对于科研口的你，比如现在私人卫星和无人机普及了，要不要去看看自己小区周围的地貌，看是不是隐藏了个金库？清清输入，卫星图片一栏无余。哪里有树，哪里有水，哪里有军事基地，不需要人，全都给你抠出来。



v2-c0cc0b25312d40370c65a0e6d52586f5_hd.jpg

如果你要数个细胞啥的，都是挺容易的，给它变成这样的轮廓不就你得了。



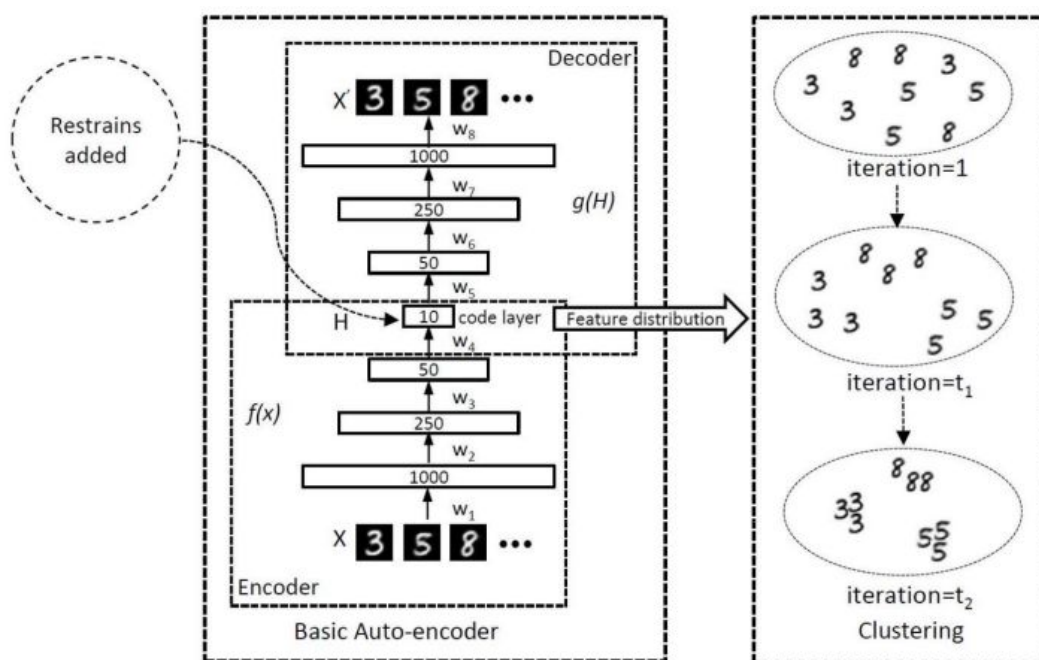
v2-90f49f5010c4da6cfedea74cae741878_hd.jpg

第四重境界：图像匹配

我们开始fashion起来，如果你是淘宝服装小店的老板，想让客户输入一张服装的图片，然后得到一组推荐的服装，来个以图搜图的功能怎么搞呢？注意啊，我可以从网络上爬一大堆图出来，但是这些数据是没有标注的。怎么办？铁哥告诉你还是有的搞，这个搞法，就是聚类。

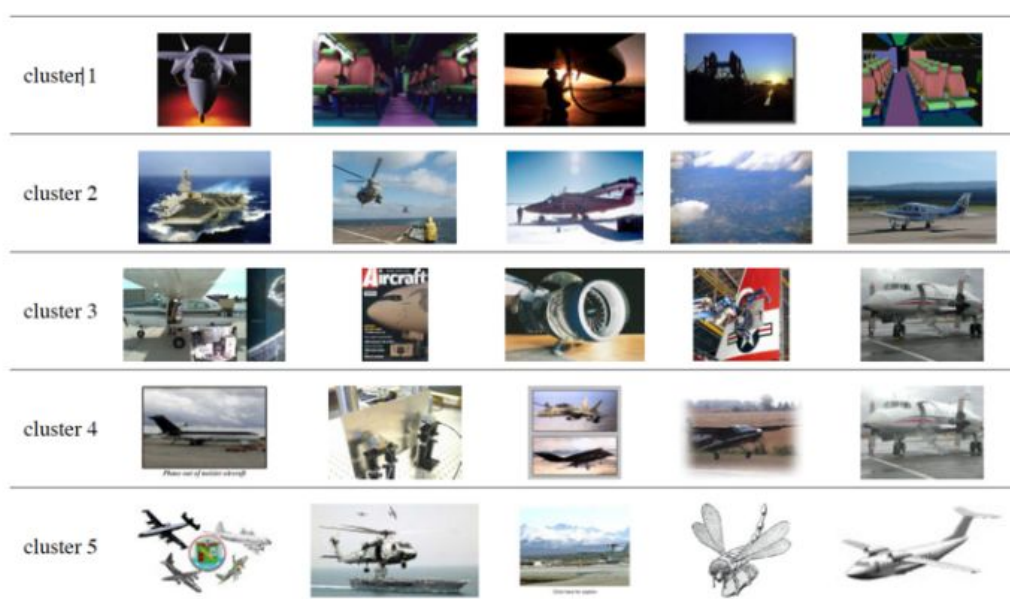
铁哥教你最简单的一招聚类哦，那就是，把图片统统放进卷积网络，但是我们

不提取分类，而只是提取一些网络中间层的特征，这些特征有点像每个图片的视觉二维码，然后我们对这些二维码做一个k-means聚类，也会得到意想不到的效果。为什么要深度？因为深度提取的特征，那是与众不同的。然后以图搜图呢？不过是找到同一聚类里的其它图片啊。



v2-61d190a1f40022be1d9db12bc93b622f_hd.jpg

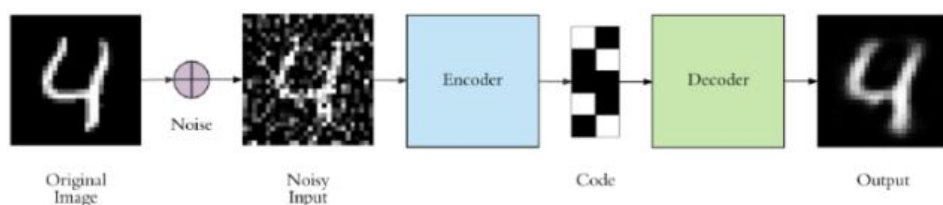
在聚类的基础上，就可以做个搜索！



v2-48f3a8a1f9a2d0da87ceeca5c709f443_hd.jpg

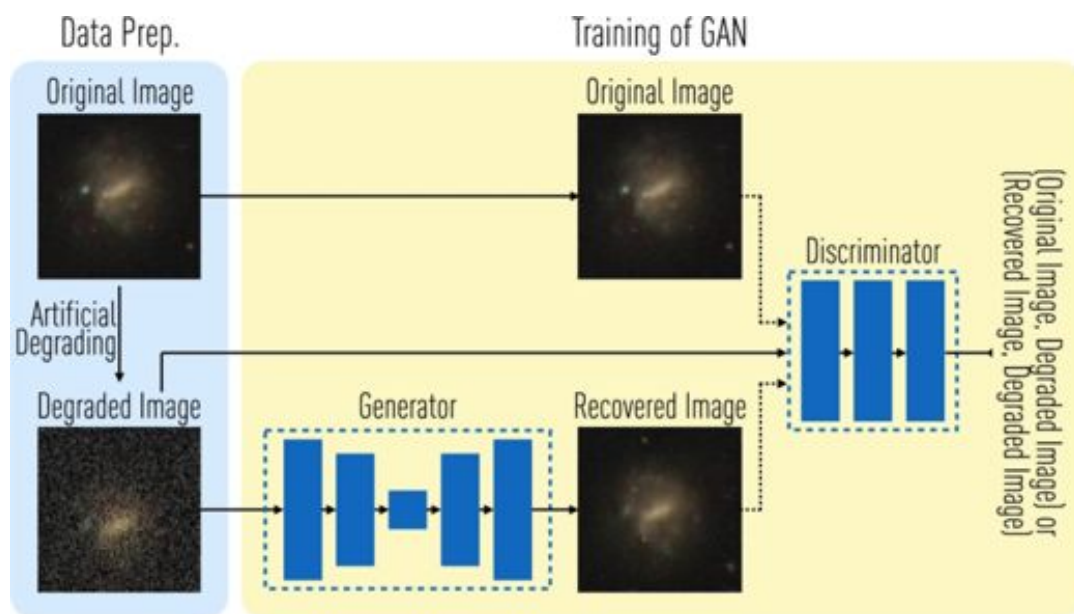
第五层境界：图像加强还原

我们开始晋升为仰望星空的人，之前那些分类赚钱的应用太无聊了。机器视觉搞科学怎么港？作为一群仰望星空后观察细胞的人，我们最常发现的是我们得到的天文或者细胞图片的噪声实在太大了，这简直没法忍啊，然后，深度学习给了你一套降噪和恢复图像的方法。一个叫auto-encoder的工具，起到了很大的作用，刷的一下，图像就清楚了。

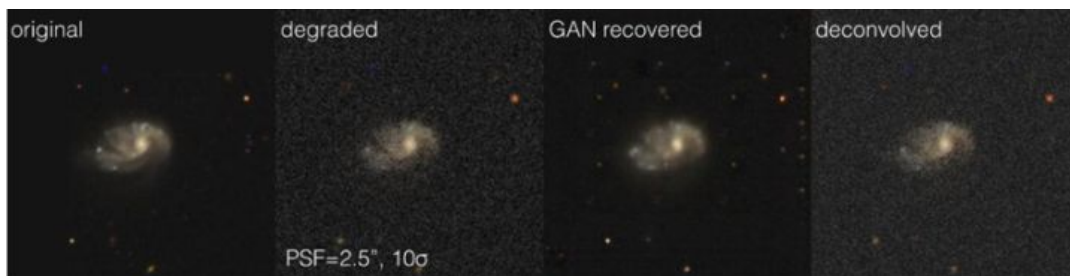


v2-f1fd4c19b92e80d4b9530663fe6f988c_hd.jpg

这还不是最酷炫的，那个应用了博弈理论的对抗学习，也可以帮你谋杀噪点！如果你会对抗所谓GAN，也是一种图像生成的工具，让网络去掉噪声的图片，与没有噪声的自然图片，连卷积网络都判别不出来，对，就是这样！



v2-0b7ad6cc47d3938baca681a7b4287f49_hd.jpg

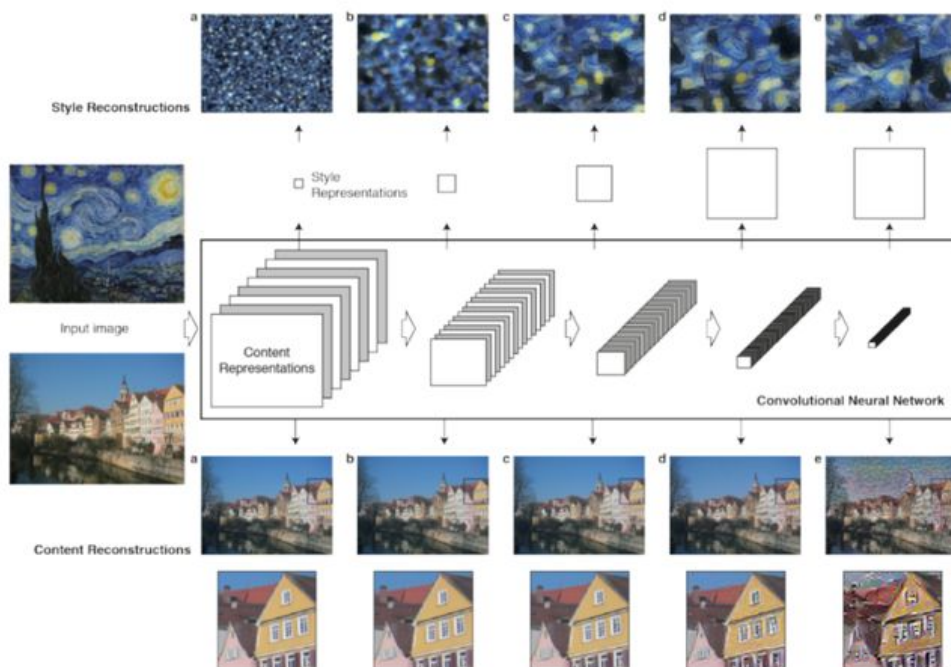


v2-0250e3df0b1eebc26b21400c175f801e_hd.jpg

Schawinski, Kevin, et al. "Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit." Monthly Notices of the Royal Astronomical Society: Letters 467.1 (2017): L110–L114.

第六重境界：图像风格迁移

在工业界赚够了钱，科学也太nerd了，我们来玩艺术思考哲学，第一招，图像风格迁移，请见[铁哥之前的文章](#)：



v2-2c3b818a09ae02b2caf046a0bb217005_hd.jpg

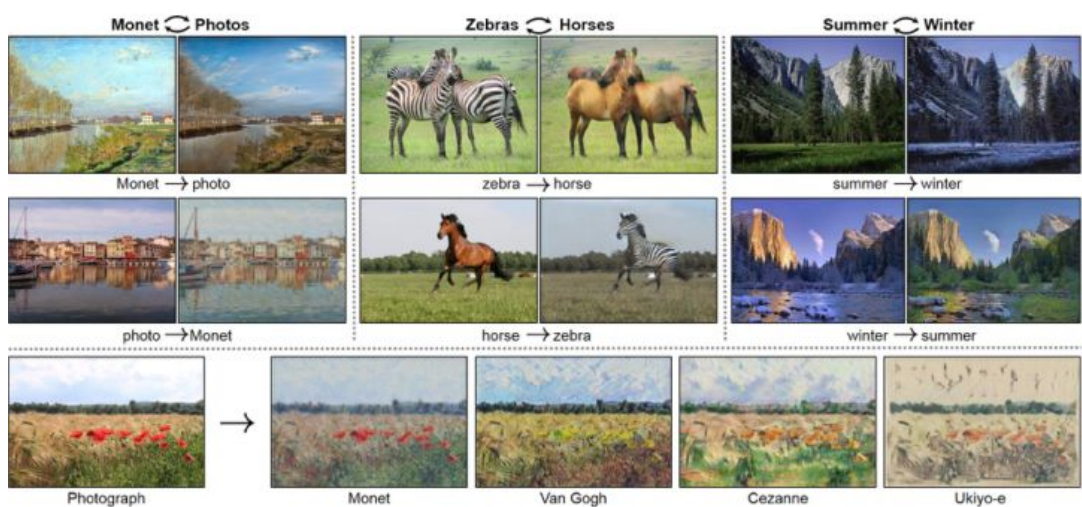


v2-770eaa7de6e5c36ebc6d384ae2857566_hd.jpg



v2-7a70630dc486f449880e34e2cba4cf60_hd.jpg

然而真正能玩好这一事项的，还是那个刚刚提过的对抗学习GAN，比如大名鼎鼎的CycleGAN，几乎可以实现一种你自定义的“图像翻译”功能，而且你不用做标注哦，拿出冬天和夏天的两组图片，它会自动的在两组图片中找出对应来。



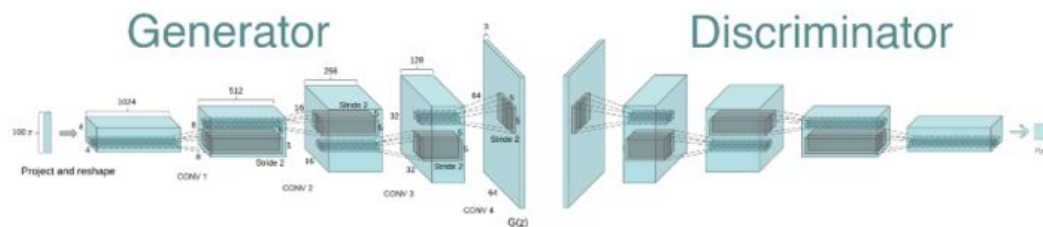
v2-2c96ea0f07d33a609534fb82914f6b8c_hd.jpg

Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." arXiv preprint arXiv:1703.10593 (2017).

第七重境界：

图像翻译也懒的玩了，你神经网络不是号称能够理解图像，看你来个无中生有，在噪声里生成图片来？

对，依然是GAN，而且是最基础的卷积GAN (DCGAN)就可以给你干出来。看看GAN所幻想的宾馆情景，你能想到是计算机做的图吗？哈哈！



v2-0428ac3bc49293a0ee1cd9920540cfb7_hd.jpg

Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems. 2014.



v2-83501eb2e0ab45cdcd81cb670e77ef3f_hd.jpg

写到这里，我自己都觉得GAN是非常有前途的，有前途的，有前途的，以前我还以为只是好玩呢。

这里展示的七级浮屠，也不过深度学习被人类discover的冰山一角，醉卧沙场君莫笑，古来征战几人回。

给你一个稍微清晰一些的大纲：



v2-83b96d41790d4dd26c68a697a25a458b_hd.jpg

如果对基础理论部分有不熟悉，请返回文章[你不能不知道的CNN](#)，当然它只是冰山一角，了解更多并挨个实战请关注：巡洋舰的[深度学习实战课程](#)，手把手带你进行深度学习实战，课程涵盖机器学习，深度学习，深度视觉，深度自然语言处理，以及极具特色的深度强化学习，看你能不能学完在你的领域跨学科的应用深度学习惊艳你的小伙伴，成为身边人眼中的大牛。刚刚讲的方法都将在课程里详细展开。

目前课程线下版本已经基本报名完毕（特殊申请可加一到两个名额），为了缓解众多异地学员的需求，我们提出一个线上加线下的课程简版，课程包括全部课程视频，notebook作业，和一个课程模块的来京线下实践机会，名额限5名，预报从速，详情请联系陈欣（cx13951038115）。