什么是自编码器?

自编码器是一种特别的神经网络，两头宽，中间窄，像个哑铃的样子。

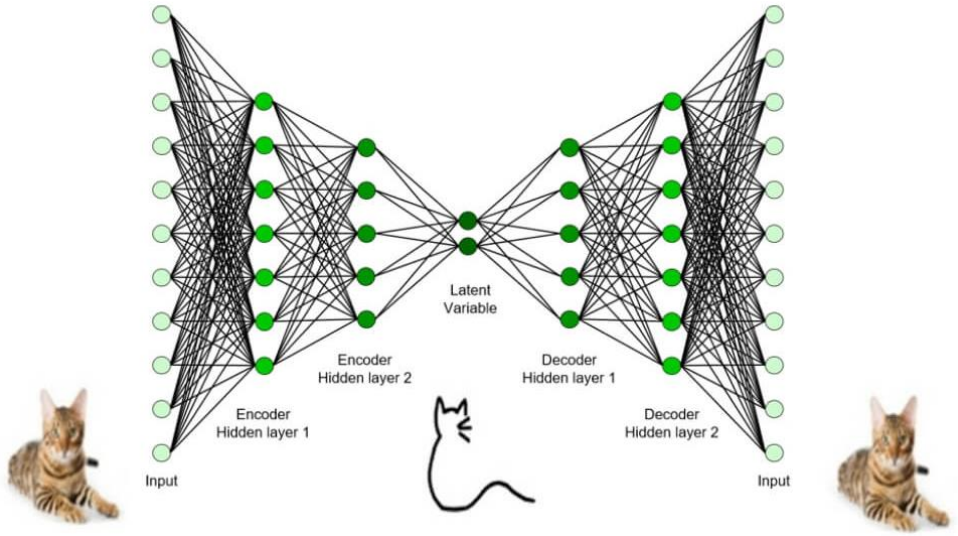

图 1：AutoEncoder 结构

这个神经网络的学习目标是努力在输出层还原输入数据。如图所示，输入一幅猫的图片，这个自编码器会努力在输出层还原这只猫。关键是，因为网络的中间层比较窄，必然会过滤掉一些信息，因此也称为信息瓶颈层。显然，为了尽可能恢复输入图片，网络在瓶颈层的编码中会尽可能地保留图片中的重要特征，例如猫的轮廓。 因此，自编码器是一个强大的特征提取器，可以从数据中自动发现显著特征。自动从数据中发现特征这件事有重要意义，它不仅使人们摆脱了手工设计特征的压力，而且有望发现数据中隐藏的规律。 如图所示，用自编码器把数字图片映射到一个平面上，可以清楚地看到不同数字之间的相邻关系和转换过程。

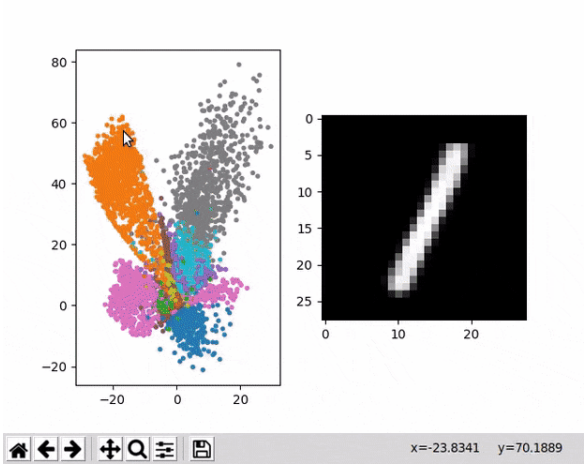

图 2：AutoEncoder 发现隐空间[4]

除此之外，自编码器还有很多其它用处。例如，因为编码只保留重要特征，干扰特征会被滤除，可以用这一特性去除噪音，提高数据质量。

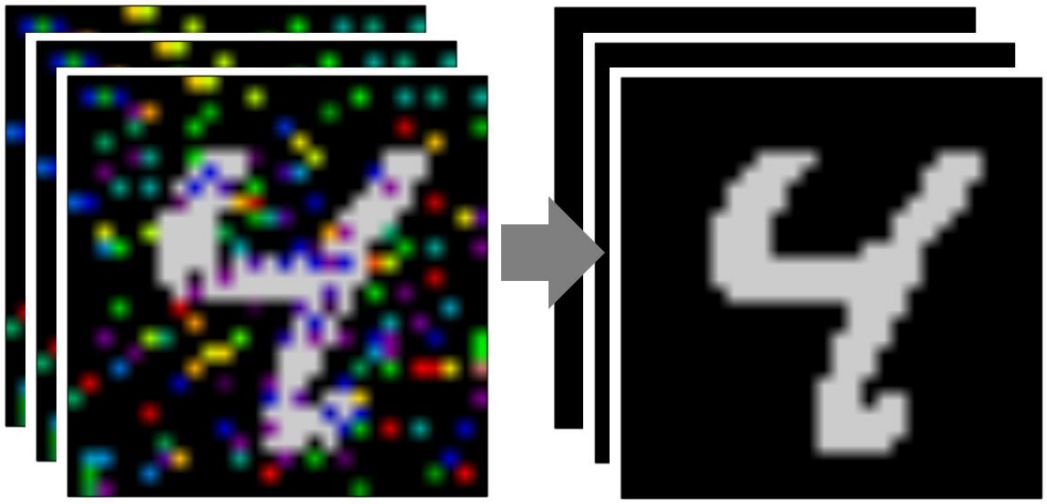

图 3： 利用 AutoEndocder 去噪 [3]

另外，因为自编码器发现了显著特征，我们可以通过修改这些特征来改变数据的表现形式。如图所示，自编码器发现了汽车的十几个特征，通过修改这些特征，可以轻松改变车的样子。

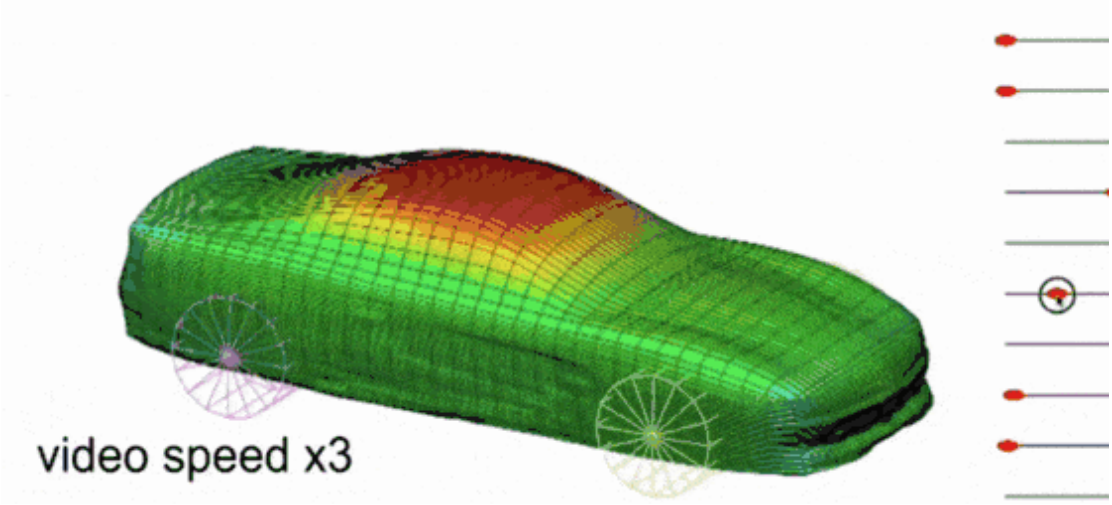

图 4：通过修改 AutoEncoder 隐空间特征改变图片属性[2]

最后，因为自编码器对异常数据会产生较大的恢复误差，利用这一现象可以进行异常检测。如下图所示，用健康人脑部扫描图片训练出的自编码器，当输入一幅病人的脑部扫描图片时，在输出端会产生较大的残差，从而实现疾病的自动检测。

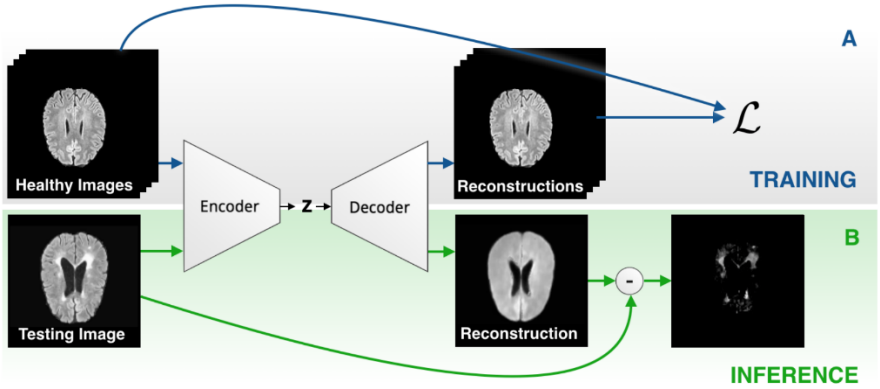

图 5：基于 AutoEncoder 的异常检测[1]

[1]https://www.researchgate.net/publication/340499853_Autoencoders_for_Unsupervised_Anomaly_Segmentation_in_Brain_MR_Images_A_Comparative_Study
[2] https://towardsdatascience.com/unsupervised-learning-part-2-b1c130b8815d
[3] https://towardsdatascience.com/autoencoder-zoo-669d6490895f
[4] https://gfycat.com/inconsequentialesteemeddartfrog