Chpt2 Pretrained networks

Augest 4, 2021

Cover:

1. Runing pretrained image-recognition models
2. An intro to GANs and CycleGAN
3. Captioning models that can produce text descriptions of images

In this chapter, we will explore three popular pretrained models:

1. A model that can label an image according to its content
2. A model that can fabricate a new image from a real image
3. A model that can describe the content of an image using proper English sentences.

2.2.1 The GAN game

GAN stands for **generative adversarial network**, where **generative** means something is being created, **adversarial** means the two networks are competing to outsmart the other. These networks are one of the most original outcomes of recent deep learning research.

The *generator* network takes the role of the painter in scenario, tasked with producing realistic-looking images, starting from an arbitrary input. The *discriminator* network is the amoral art inspector, needing to tell whether a given image was fabricated by the generator or belongs in a set of real images. This two-network design is atypical for most deep learning architecture but, when used to implement a GAN game, can lead to incredible results.

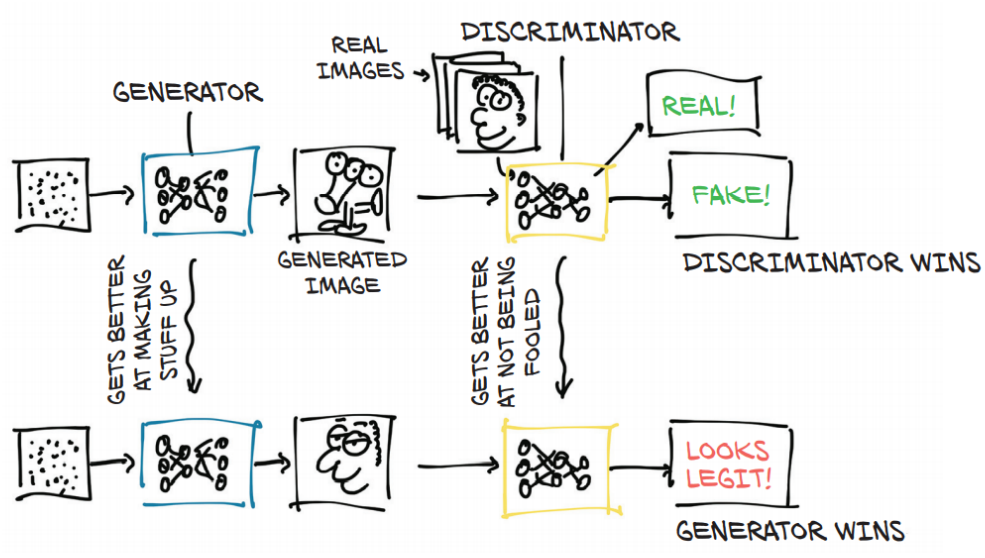

Figure 2.5 Concept of a GAN game

2.2.2 CycleGAN

A CycleGAN can turn images of one domain into images of another domain, without the need for us to explicitly provide matching pairs in the training set.

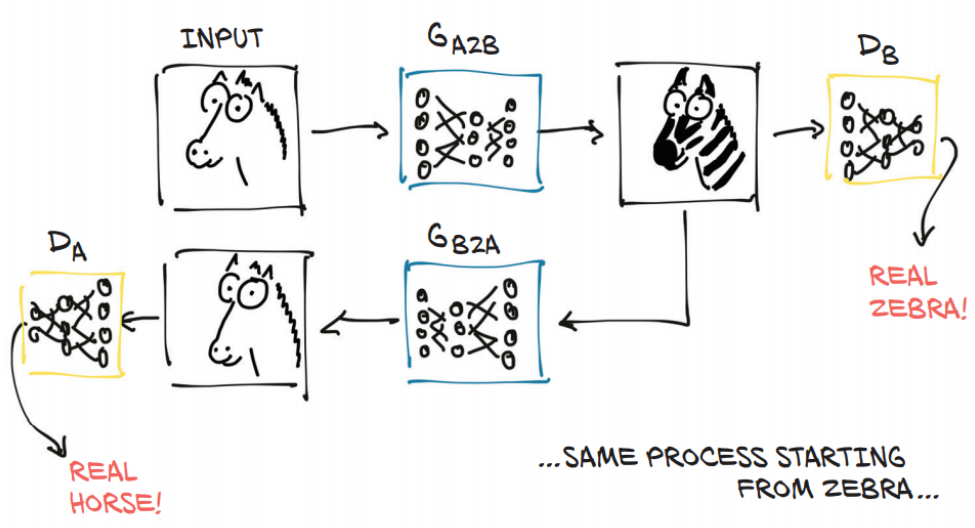

Figure 2.6 A CycleGAN trained to the point that it can fool both discriminator networks

The first generator learn to produce an image conforming to a target distribution (zebra) starting from an image belonging to a different distributions (horse).

2.3 A pretrained network that describes scenes

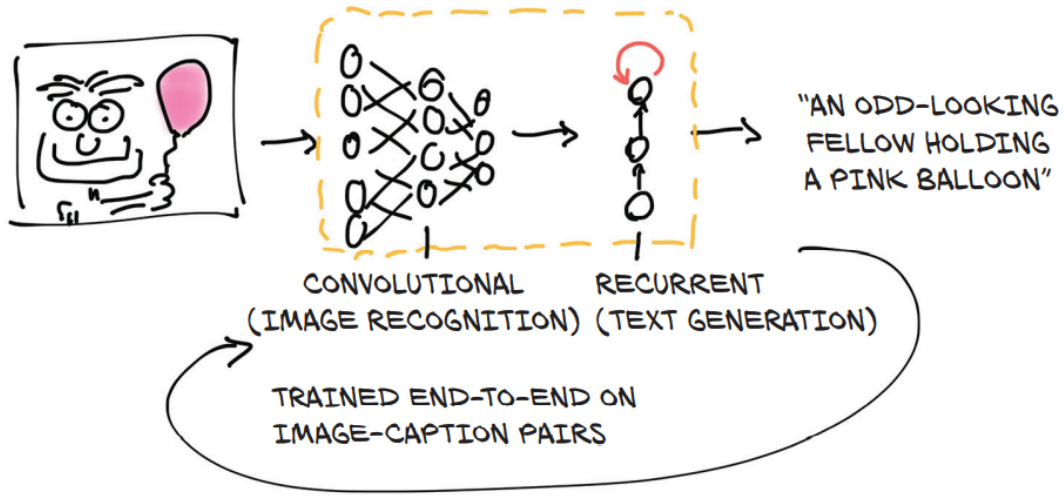

Figure 2.9 Concept of a captioning model

The captioning model has two connected halves. The first half of the model is a network that learns to generate "descriptive" numerical representations of the scene, which are then taken as input to the second half. That second half is a *recurrent neural network* that generates a coherent sentence by putting those numerical descriptions together. The two halves of the model are trained together on image-caption