Application topics

Chengzong Cai

February 2019

1 Image Caption

Image caption is a task that combines CNN and RNN together to address the problem of image understanding.The basic idea of image caption is to use pretrained CNN to produce a vector including high level features of input image,then feed it to RNN as a hidden state vector and RNN output a distribution of words that describe the image.

Caption.png

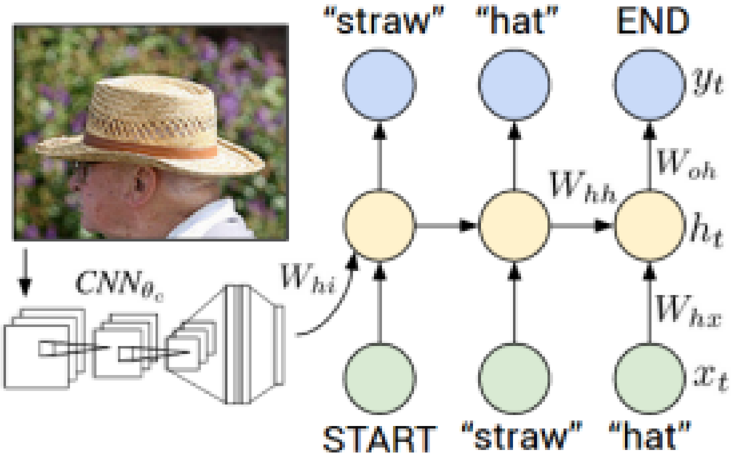

Figure 1: Image Caption

1.1 Attention Mechanism

Thinking about the way we describe a image,before we speak,we would focus on a certain part of the image.And this inspires those scientists.They design the network in Figure 2 style. Instead of using feature in FC layers,they use features f that come from conv layers.And feed the features and the hidden state given by

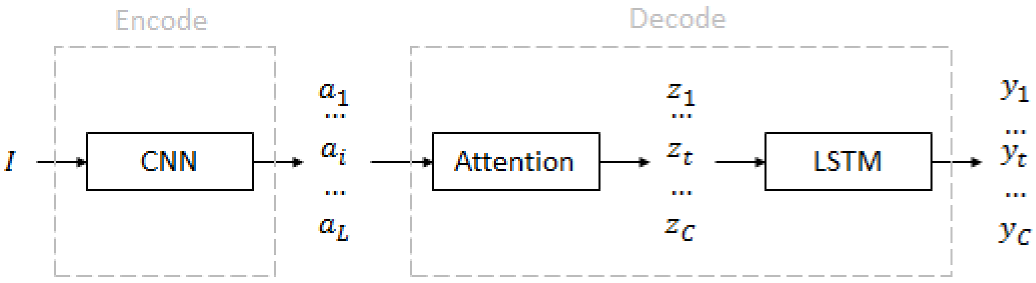

Figure 2: Image Caption with Attention

the following LSTM network to the attention layer to generate attention weight matrix α.And multiply α with f to generate the inputs of LSTM network.And then LSTM network outputs words that describe the image.

2 Detection and Segmentation

2.1 Segmentation

The basic idea of segmentation is to label each pixel of the input image.We now often use fully convolutional network using transposed convolution. A fully convolutional network contains ordinary convolutional layers.Without using FC layers,FCN uses transposed convolutional layers to generate the segmentation maps.A transpoed layer generates high-res image using a set of filters and low-res feature maps which is regarded as the weight of those filters.

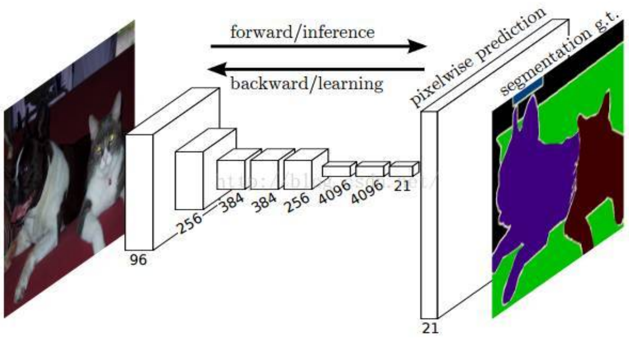

Figure 3: FCN

2.2 Detection

Detection is a task that combines both classification and location.In detection tasks,the algorithm first find out the location in where objects may exist.And

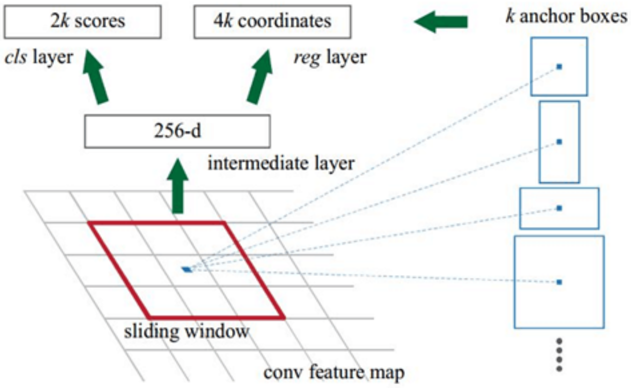

Figure 4: Faster R-CNN

then the algorithm do classification to determine the what the object is.So far we freqeuntly use R-CNN series, YOLO and SSD series.I will introduce Faster R-CNN of R-CNN series in the following section.

2.2.1 Faster R-CNN

Instead of searching proposal area in raw images,Faster R-CNN searches RoI in feature maps extracted by convolutional layers by using RPN(Region proposal network).And then it applies RoI pooling layer to make sure the size of all RoIs is the same with each other.And then it does classification and regression.The multitask loss contains RPN object classification loss,RPN boxes regression loss,final object classification loss and final boxes regression loss

3 Generative Models

See this section in GAN intro.pdf written few months ago