**Middle Text Generation: COCO Image Captions**

我们使用的另一个真实数据集是COCO Image Captions Dataset（Chen et al.2015），这是一个包含图像描述对组的数据集

我们将图像的文字描述作为要生成的文本

请注意，COCO数据集不是长文本数据集，其中大多数句子大约10个单词

COCO图像标题训练数据集由20,734个单词和417,126个句子组成

我们删除频率低于10的单词以及包含它们的句子

我们为训练集随机抽取80,000个句子，为测试集随机抽取5,000个句子

COCO数据集上的BLEU得分结果表明LeakGAN在中长文本生成任务中的表现明显优于基线模型

两个数据集上的性能表明，LeakGAN生成的句子具有比SeqGAN更高的全局一致性和更好的可读性