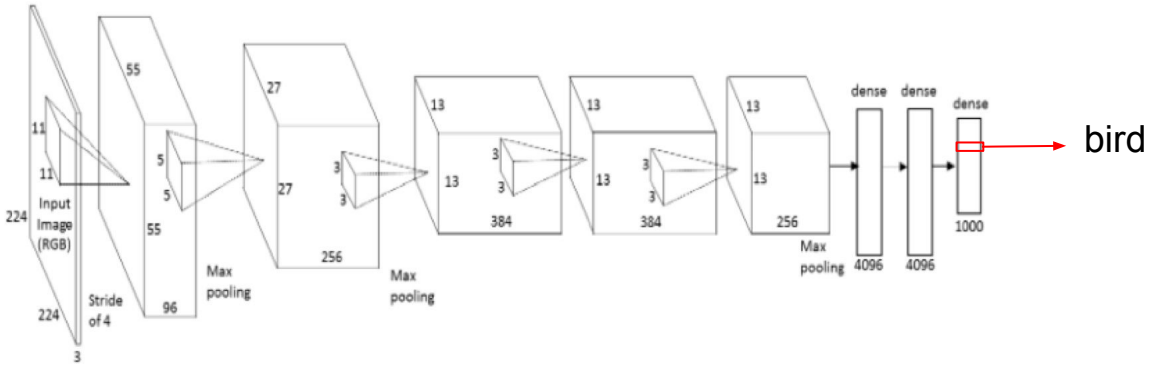
Attention Models: Motivation

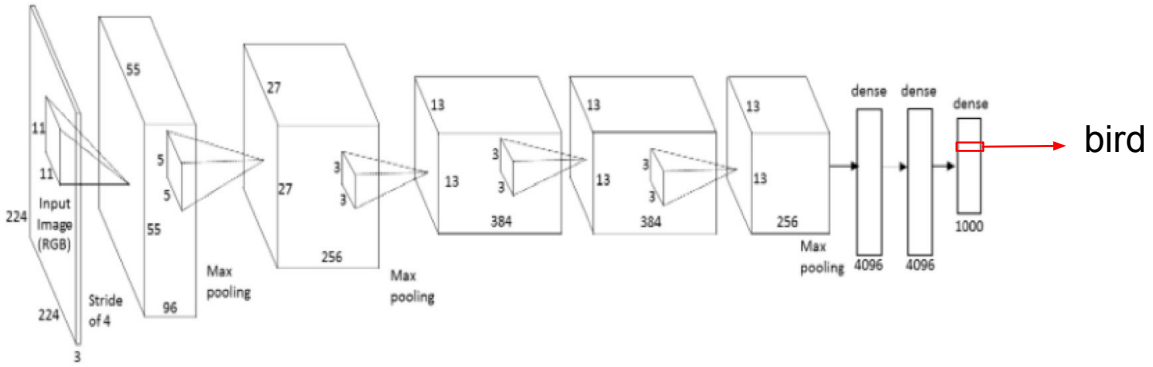
Image:
H x W x 3



The whole input volume is used to predict the output...

Attention Models: Motivation

Image:
H x W x 3



The whole input volume is used to predict the output...

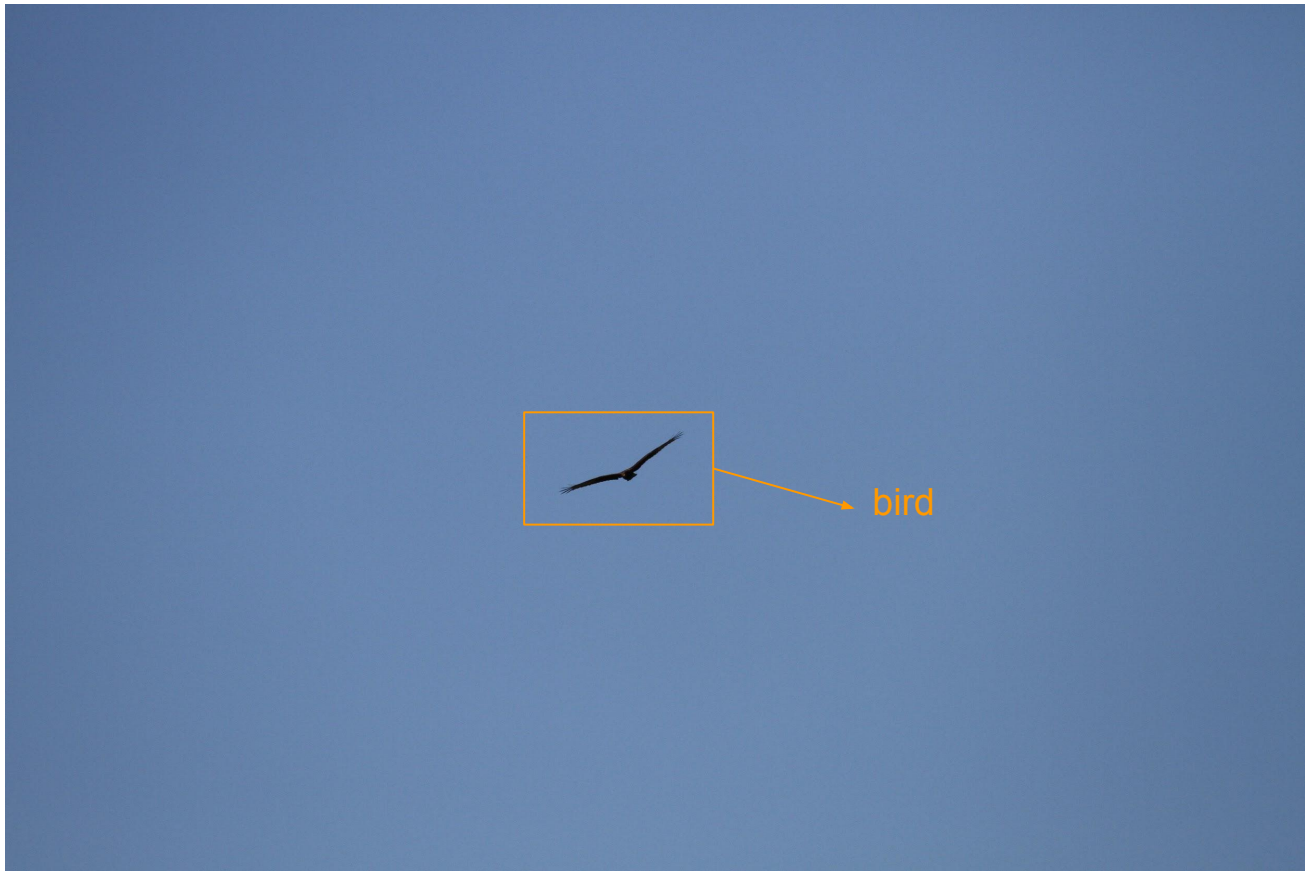
...despite the fact that not all pixels are equally important

Attention Models: Motivation

Attention models can relieve computational burden

Helpful when processing big images !

Attention Models: Motivation



Attention models can relieve computational burden

Helpful when processing big images !

Attention Models

Attend to different parts of the input to optimize a certain output

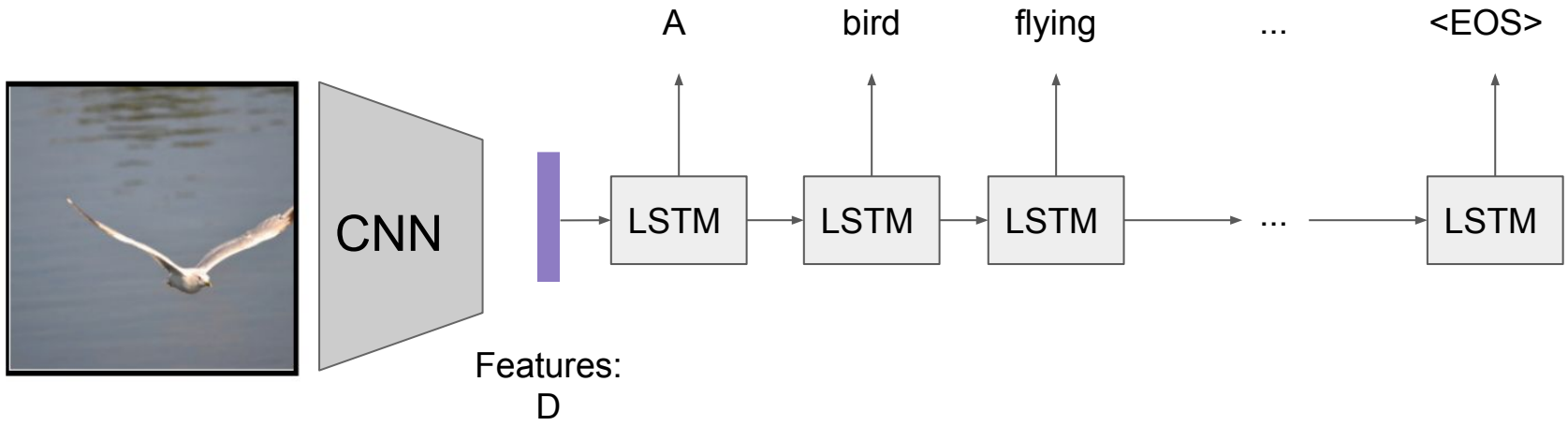
Attention Models

Attend to different parts of the input to optimize a certain output

Input: Image; Output: Text

A bird flying over a body of water

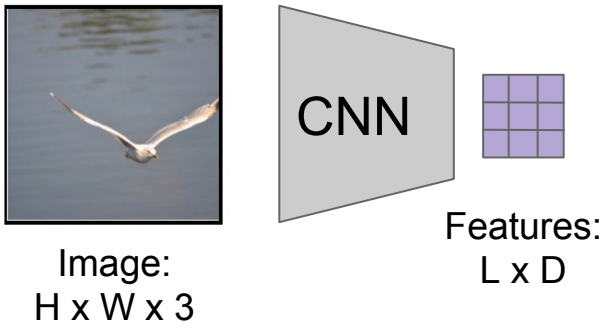
LSTM Decoder for Image Captioning



The LSTM decoder “sees” the input only at the beginning !

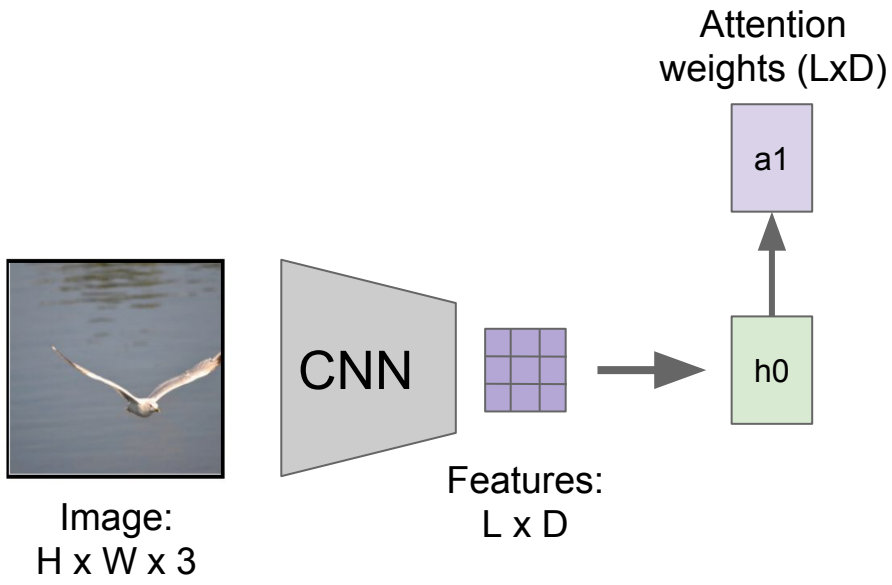
Vinyals et al. Show and tell: A neural image caption generator. CVPR 2015

Attention for Image Captioning



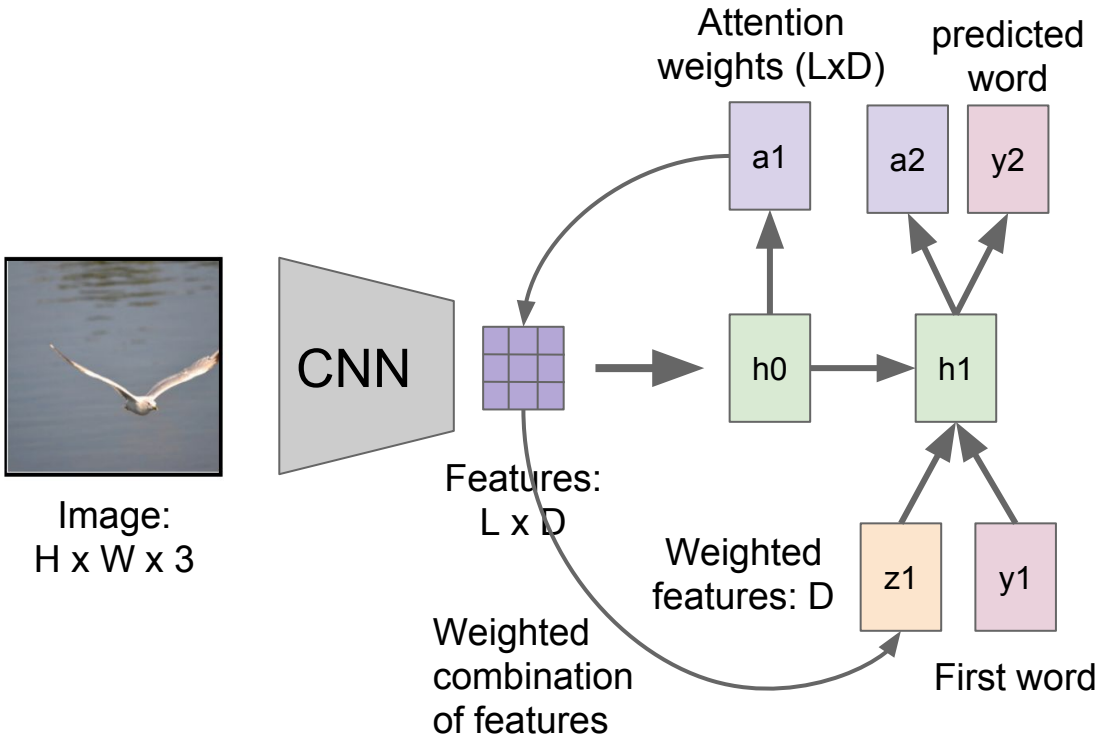
Slide Credit: CS231n

Attention for Image Captioning



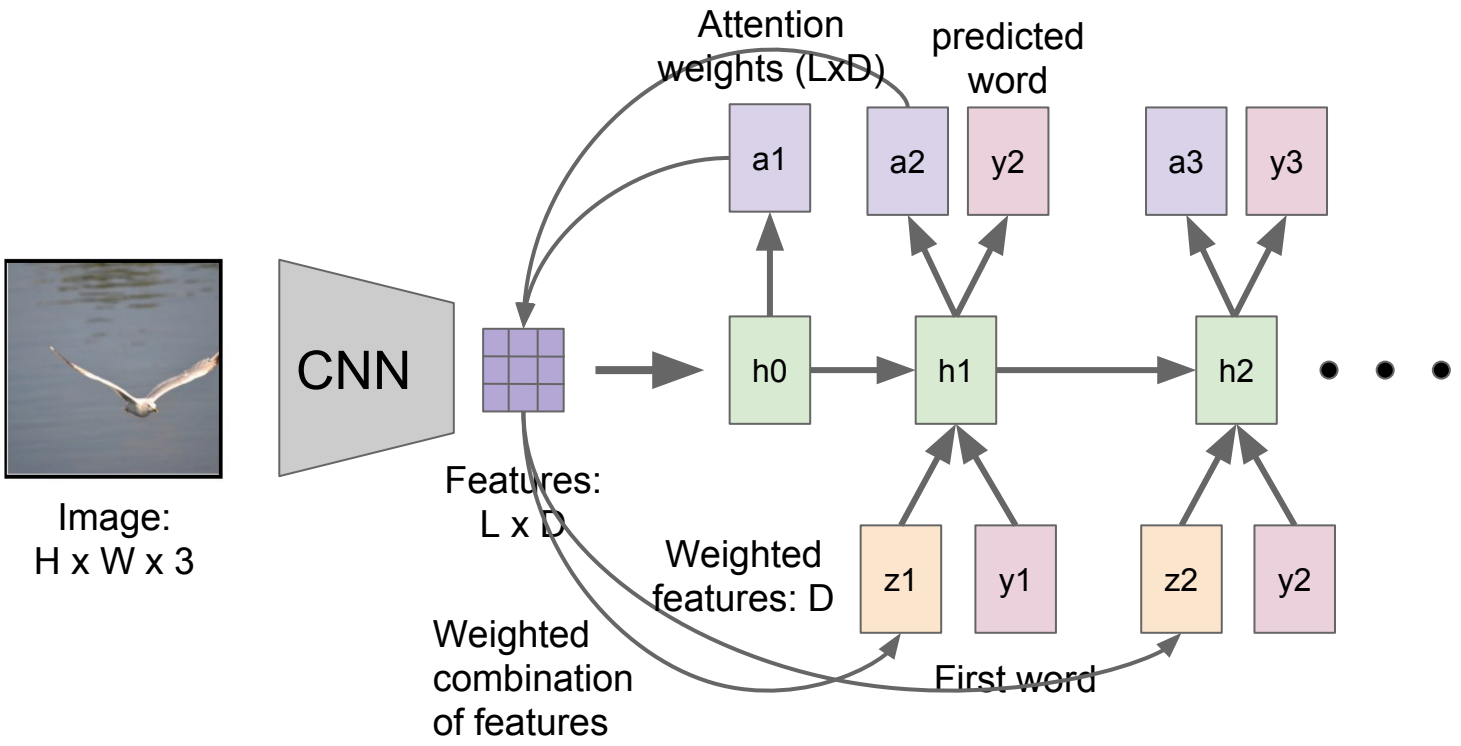
Slide Credit: CS231n

Attention for Image Captioning



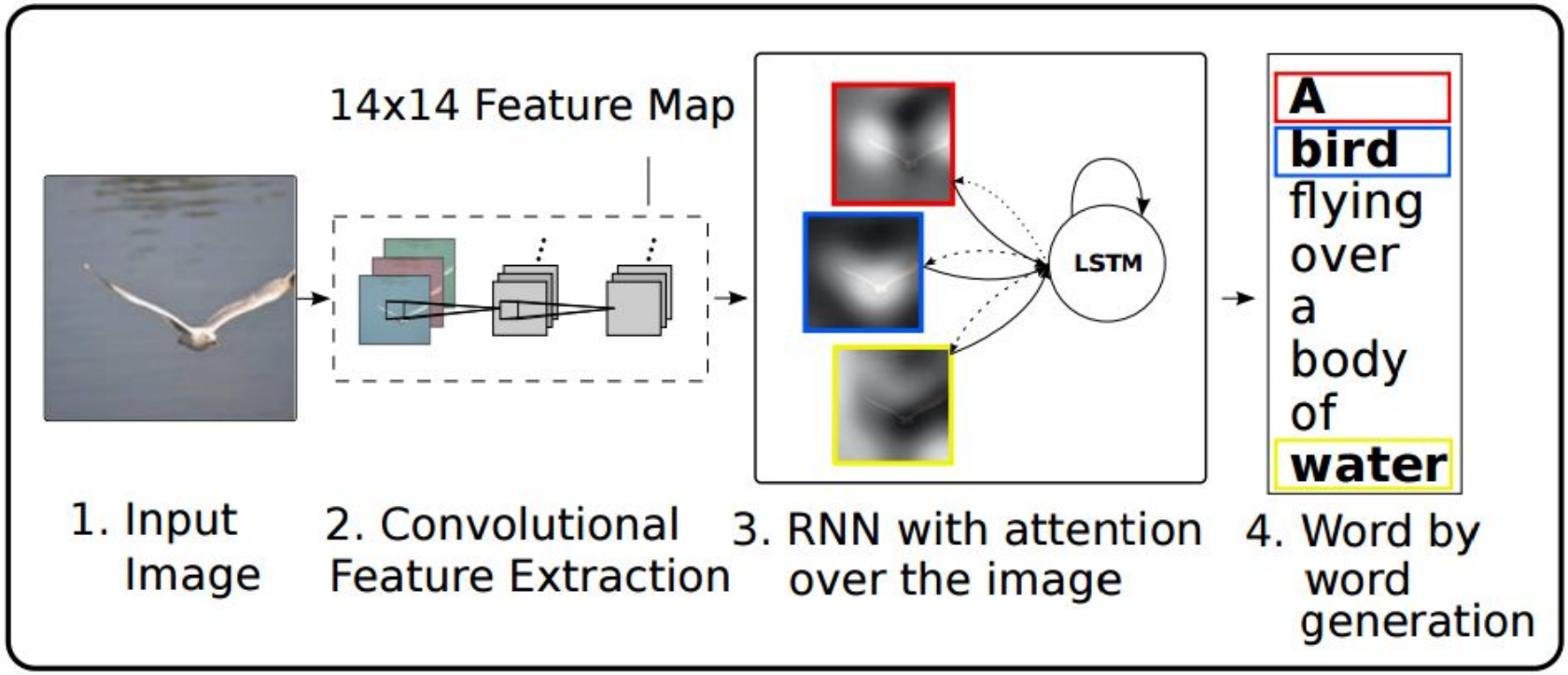
Slide Credit: CS231n

Attention for Image Captioning



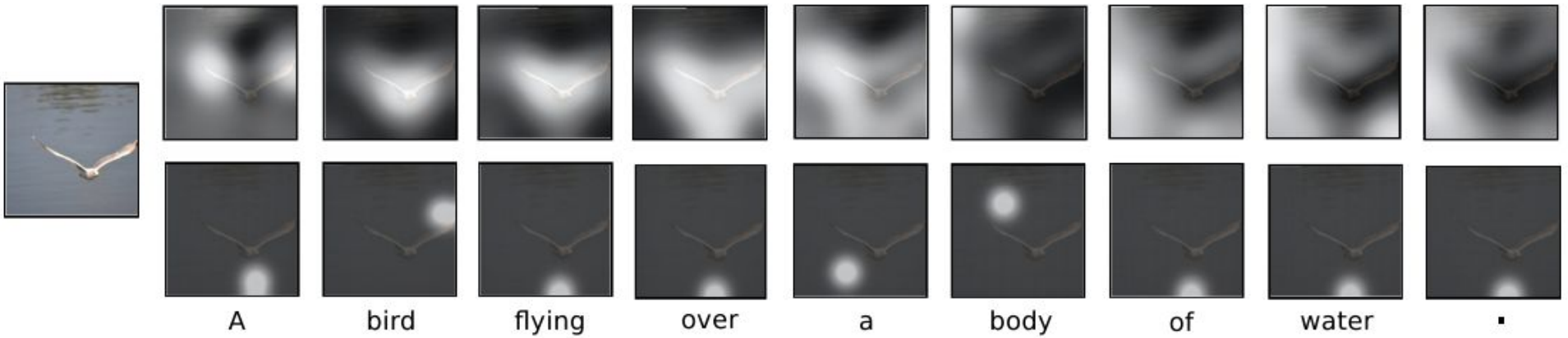
Slide Credit: CS231n

Attention for Image Captioning



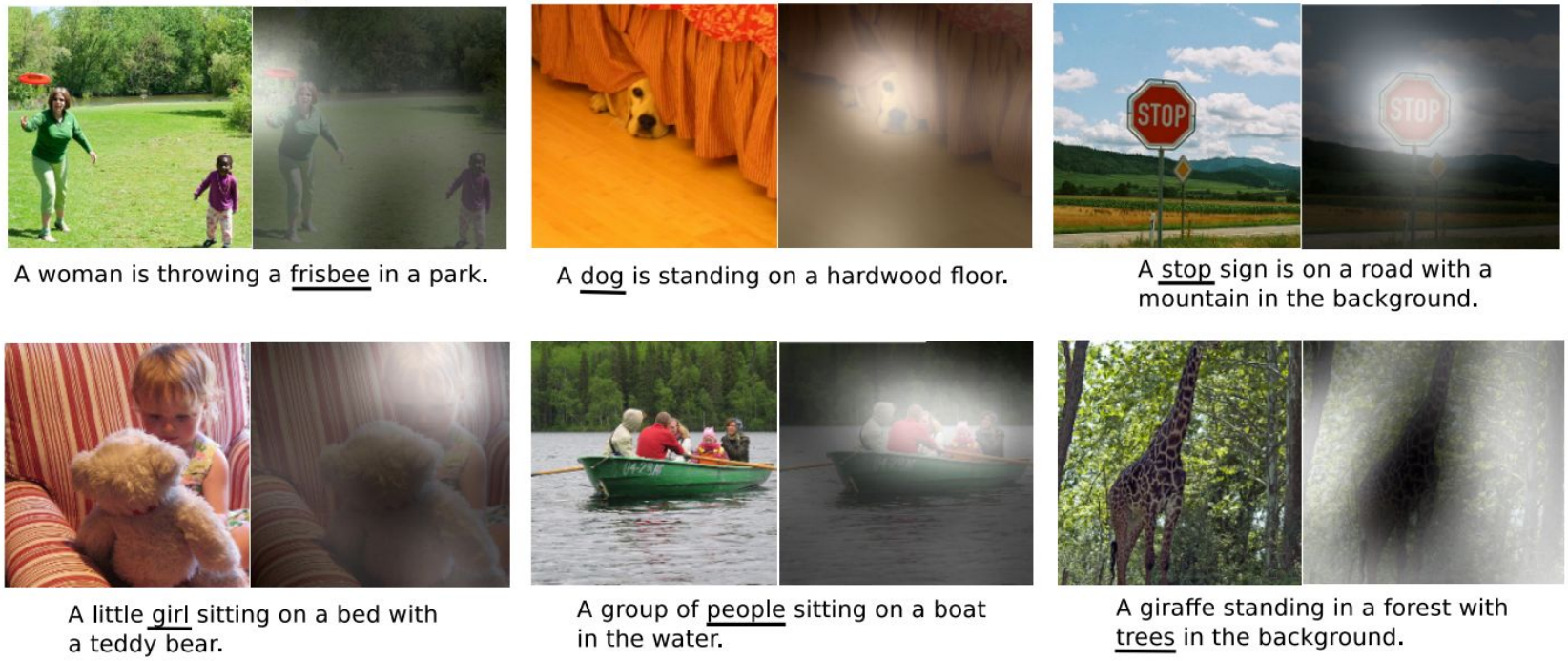
Xu et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. ICML 2015

Attention for Image Captioning



Xu et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. ICML 2015

Attention for Image Captioning



Xu et al. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. ICML 2015