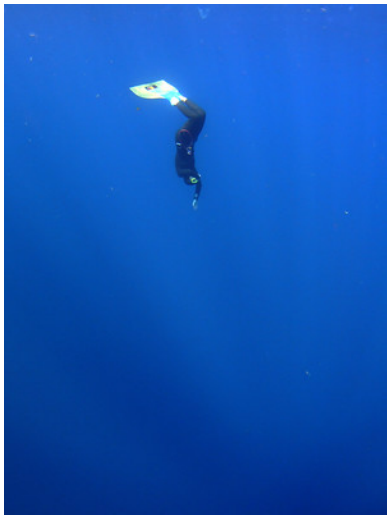
Images

Following images on Flickr are freely available (All creative commons).



(a) man in black wetsuit is jumping into the ocean

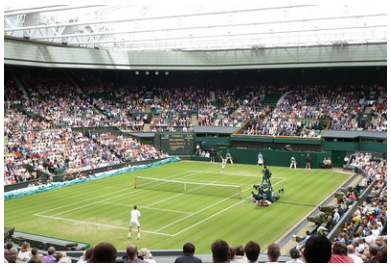
(b) two children are playing on the floor

(c) brown and white dog is running on the grass

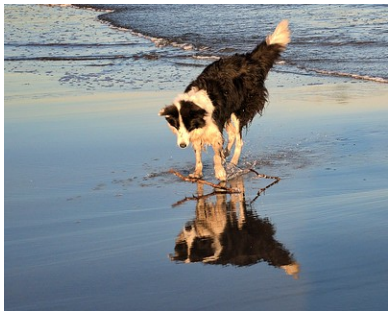
(d) two children are playing in the pool

(e) man in blue shirt is playing guitar

(f) girl in pink shirt is standing on the grass with his arms in his mouth



(a) group of people are playing with their arms in the air



(b) dog is running through the water

(c) two boys playing soccer in field of the grass

(d) two people are riding on the street

(e) two people are skiing down snowy hill

(f) two girls are standing on the sidewalk

(a) group of people are standing in front of the crowd of the crowd

(b) man in white shirt is playing tennis

(c) group of people are riding bicycles on the street

(d) black and white dog is running on the grass

(e) man in blue shirt is walking on beach

(f) woman in red shirt is standing in front of the camera

The decoder is based on a small 16-layer VGG model for feature extraction. Larger models, as well as more sophisticated configurations of the entire network, are likely to achieve better performance.

Word vectors are learned as part of fitting the model, thus it is suggested to use word vectors pre-trained on much larger corpus of text. The vocabulary extracted from the *Flickr8k_Dataset* is very limited and, indeed, the model seems to work "quite decently" only on people and dogs.