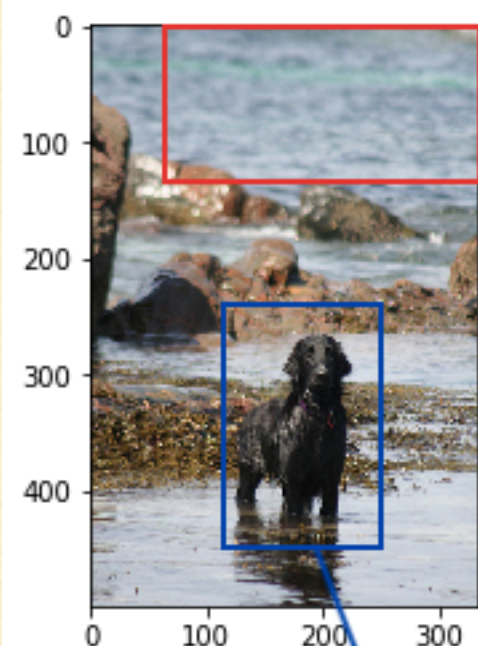

IMAGE CAPTIONING

TEACHING COMPUTERS TO DESCRIBE IMAGES

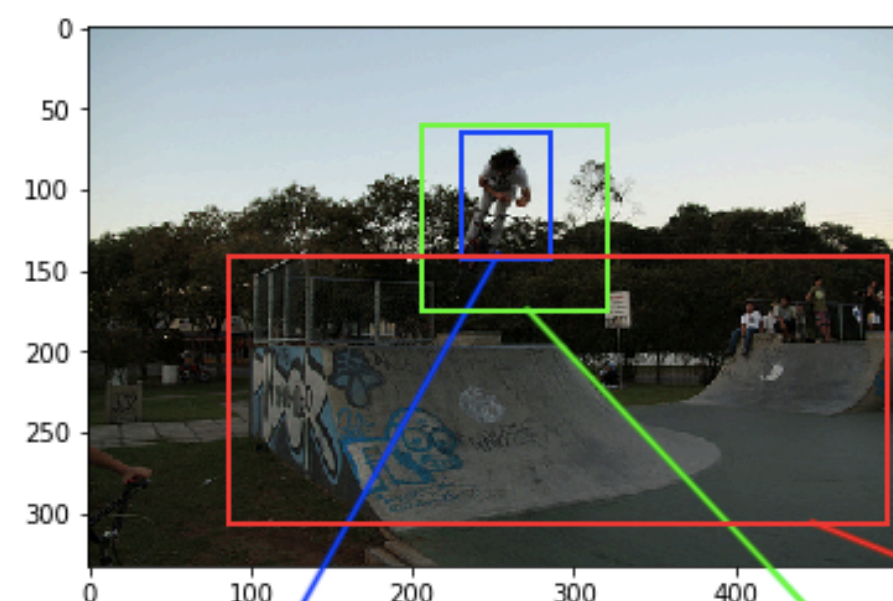
3602838407_bf13e49243



prediction: black dog is running through the water
Ground Truth: black dog in water
Ground Truth: black dog standing in shallow area of
Ground Truth: dog stands in tide pool
Ground Truth: small black dog in the ocean with some
Ground Truth: small black dog plays in the water

model8

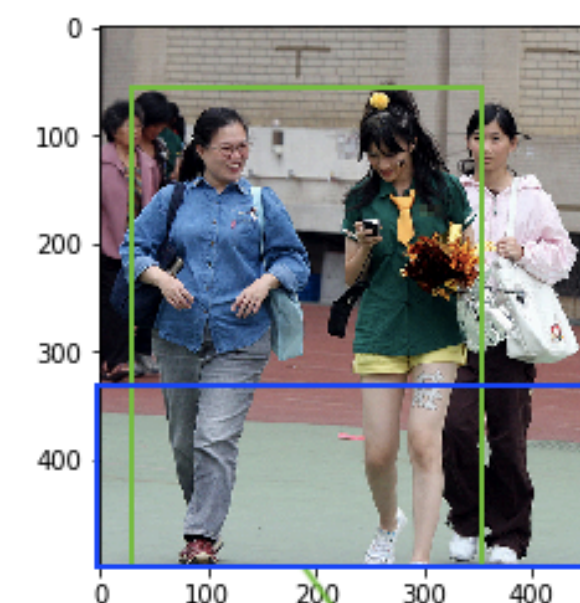
3662963630_8f097e38d4



prediction: man in black shirt is skateboarding on ramp
Ground Truth: biker leaps off halfpipe littered with gr
Ground Truth: bmx rider launches off quarterpipe
Ground Truth: cyclist airborne having jumped his bike o
Ground Truth: kid does trick on bike at skate park
Ground Truth: teenage boy on bicycle is jumping off ram

model17

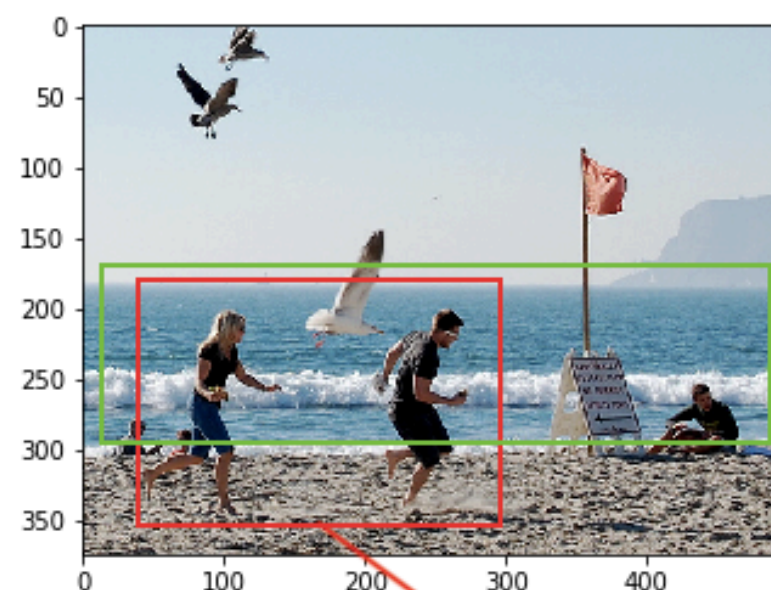
3115901702_f07aa0ef74



prediction: two girls in skirts walking on the sidewalk
Ground Truth: woman walks as she looks at her cellphone and others look on
Ground Truth: several ladies talking together
Ground Truth: three girls walking one is checking her cellphone
Ground Truth: three girls walk together and talk
Ground Truth: three high school girls walk and talk on astroturf

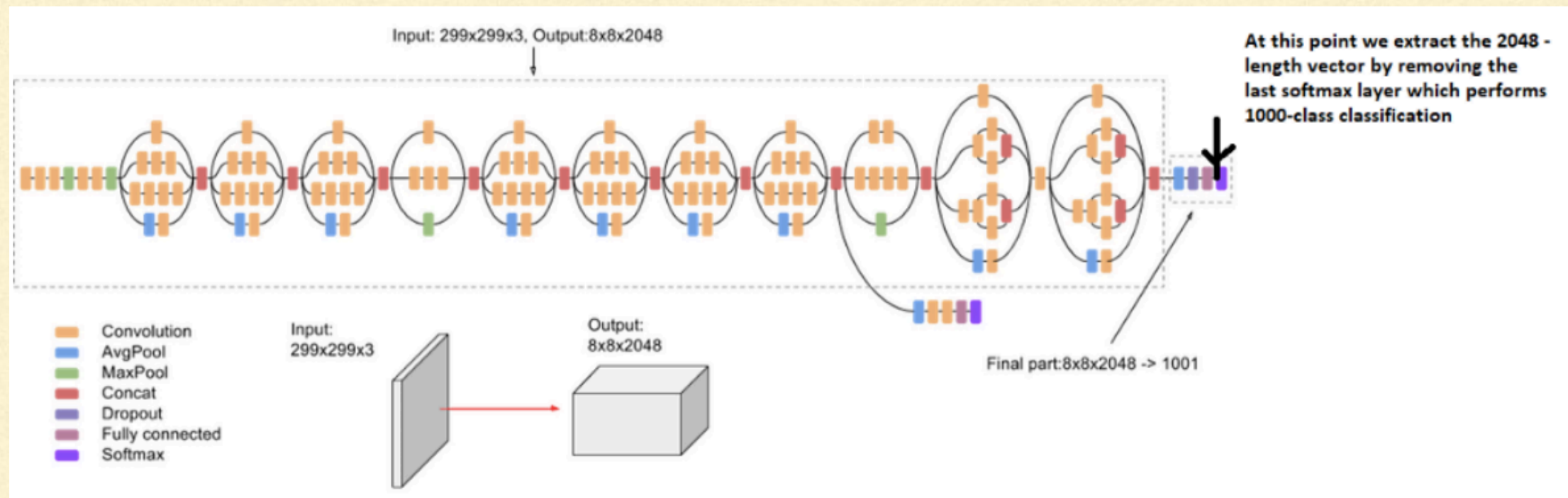
model11

3158327361_6f1a518228



prediction: two people are standing on the edge of body of water and looking
distance
Ground Truth: couple of people running on the shore near the beach
Ground Truth: man and woman hopping acrouss hot sand
Ground Truth: people running away from seagulls on the beach
Ground Truth: two people run in the sand at the beach
Ground Truth: two people run along beach with seagulls

IMAGE CLASSIFICATION MODEL MINUS THE LAST SOFTMAX LAYER



InceptionV3

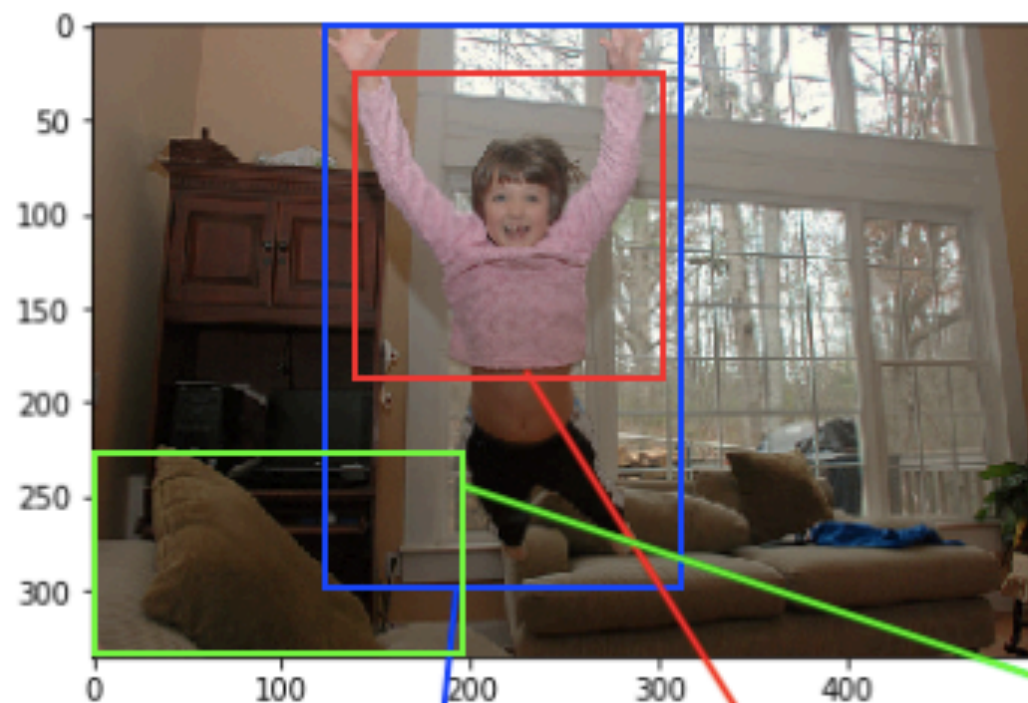
InceptionV3 - dimension of feature vector is (2048x1)

ResNet152V2 - dimension of feature vector is (2048x1)

InceptionResNetV2 - dimension of feature vector is (1536x1)

+ ImageNet weights

model
2210368267_0615754b48



prediction: little girl in pink dress is jumping on bed

Ground Truth: girl in pink shirt jumping from one couch to another in well apporportioned living room

Ground Truth: little girl in pink cardigan is jumping from the couch

Ground Truth: little girl in pink shirt has jumped into the air in her house

Ground Truth: little girl jumps through the air to land on the sofa

Ground Truth: girl jumps from couch to couch

P(girl | startseq)

P(in | startseq girl)

P(pink | startseq girl in)

P(shirt | startseq girl in pink)

.

.

P(endseq | startseq girl in pink shirt jumping...living room)

P(little | startseq)

P(girl | startseq little)

P(in | startseq little girl)

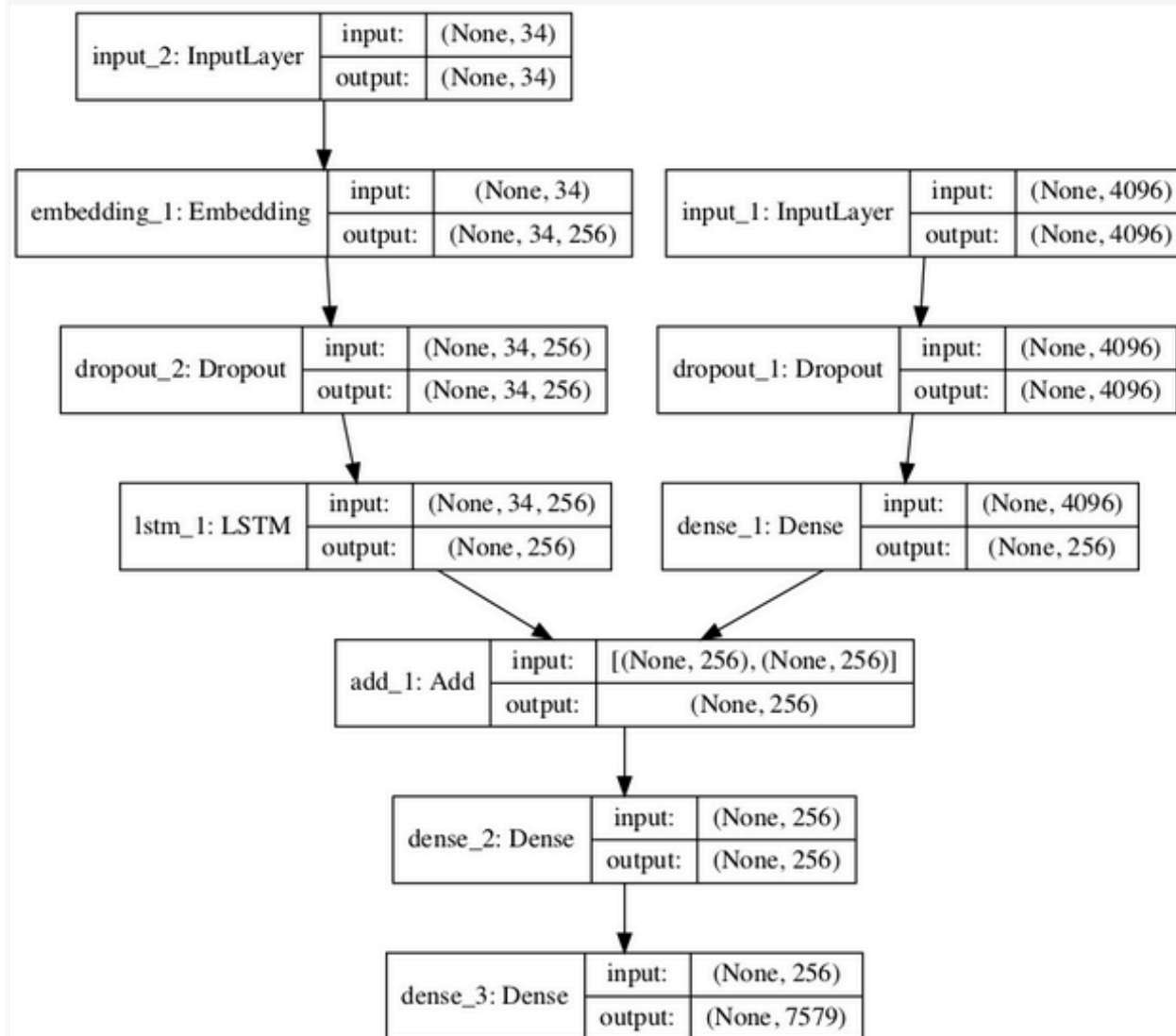
P(pink | startseq little girl in)

.

.

P(endseq | startseq little girl in pink cardigan is...the couch) ...

IMAGE MODEL + LANGUAGE MODEL



Plot of the Caption Generation Deep Learning Model

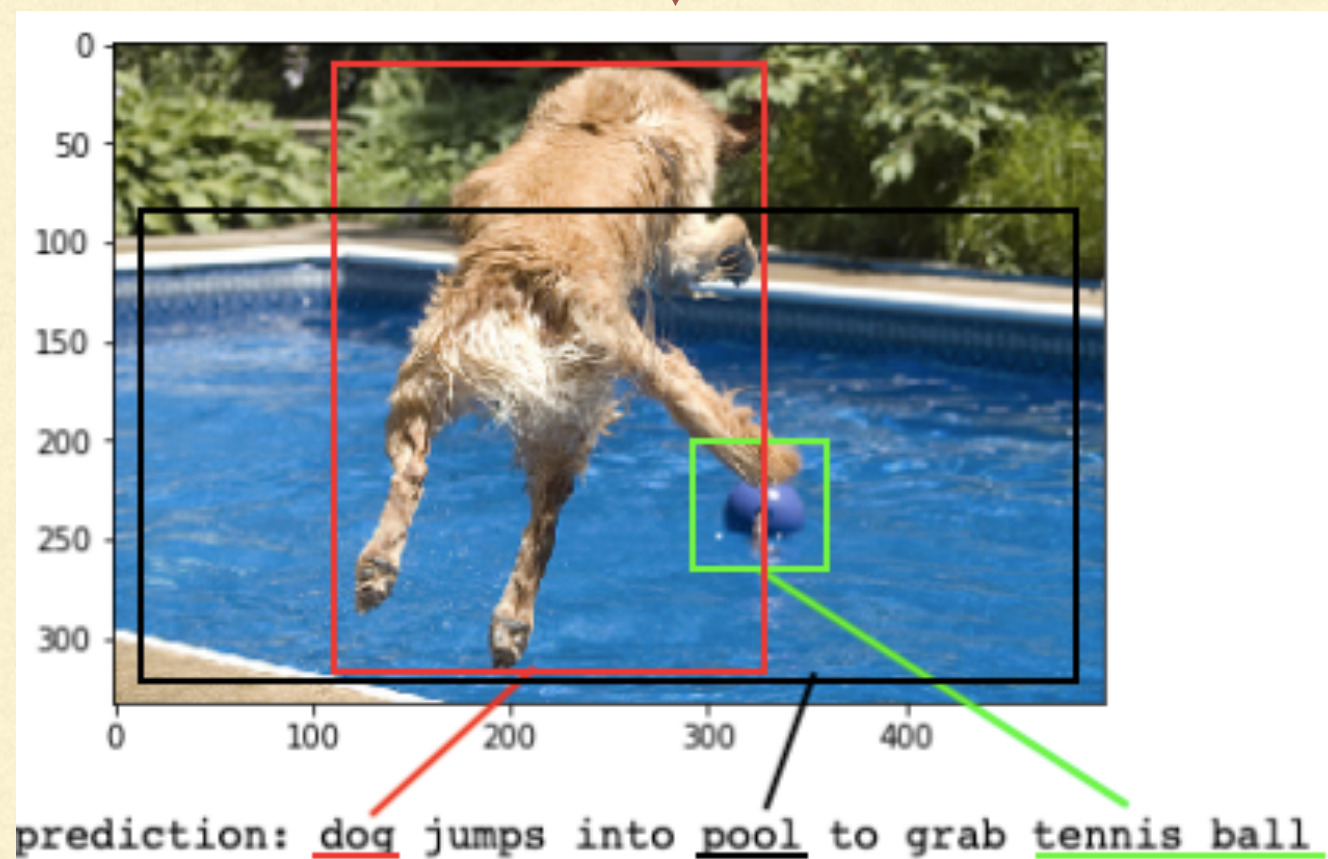
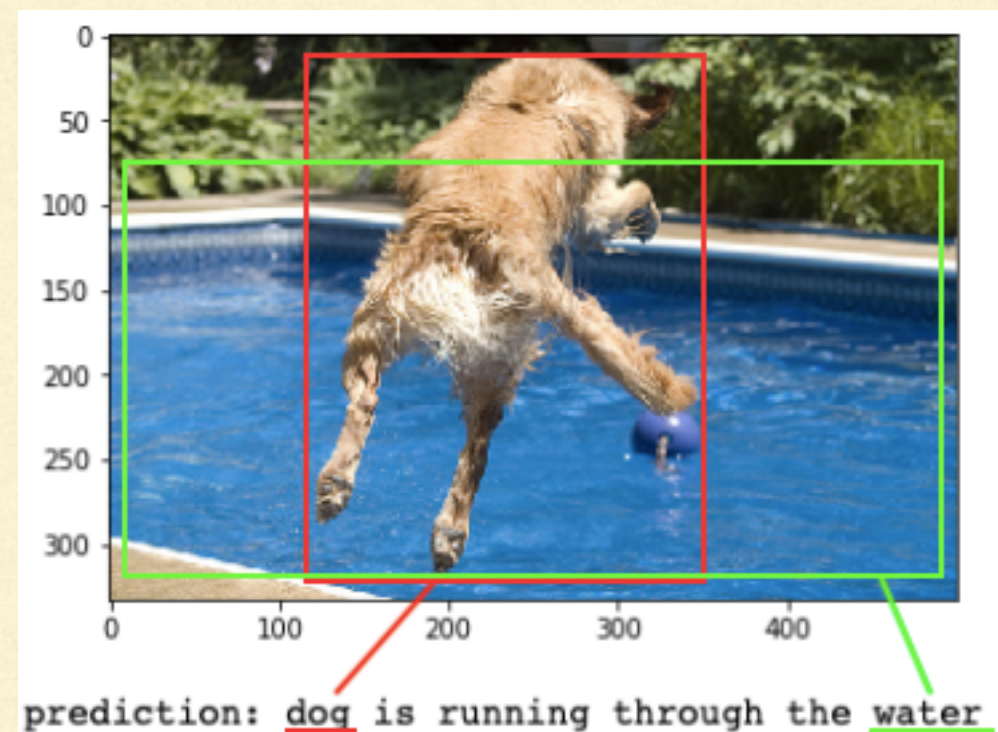
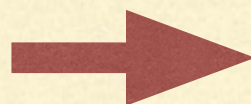
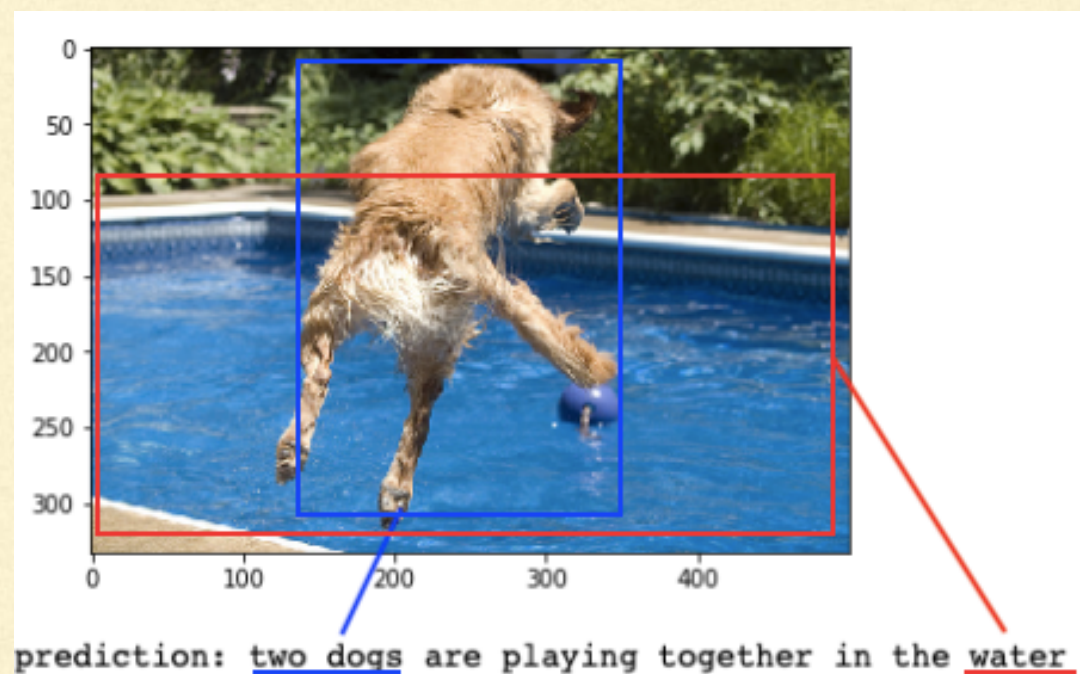
Image model | Language model

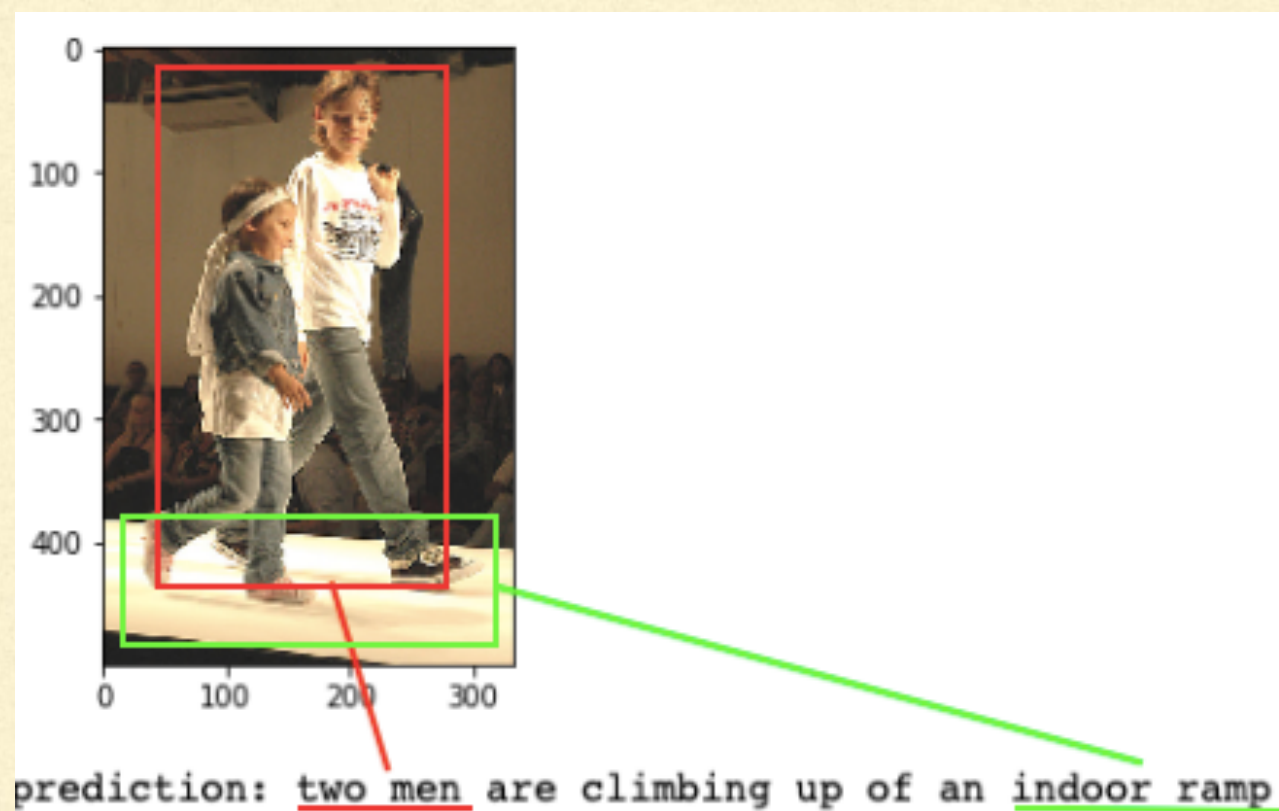
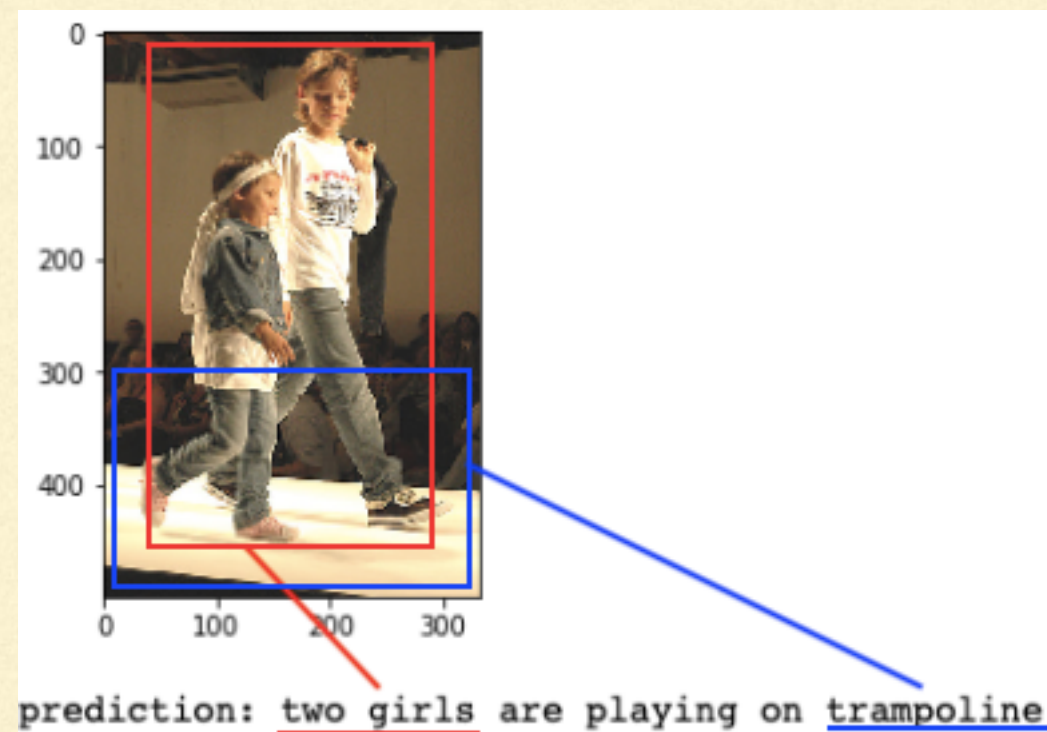
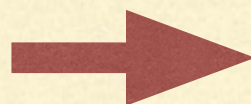
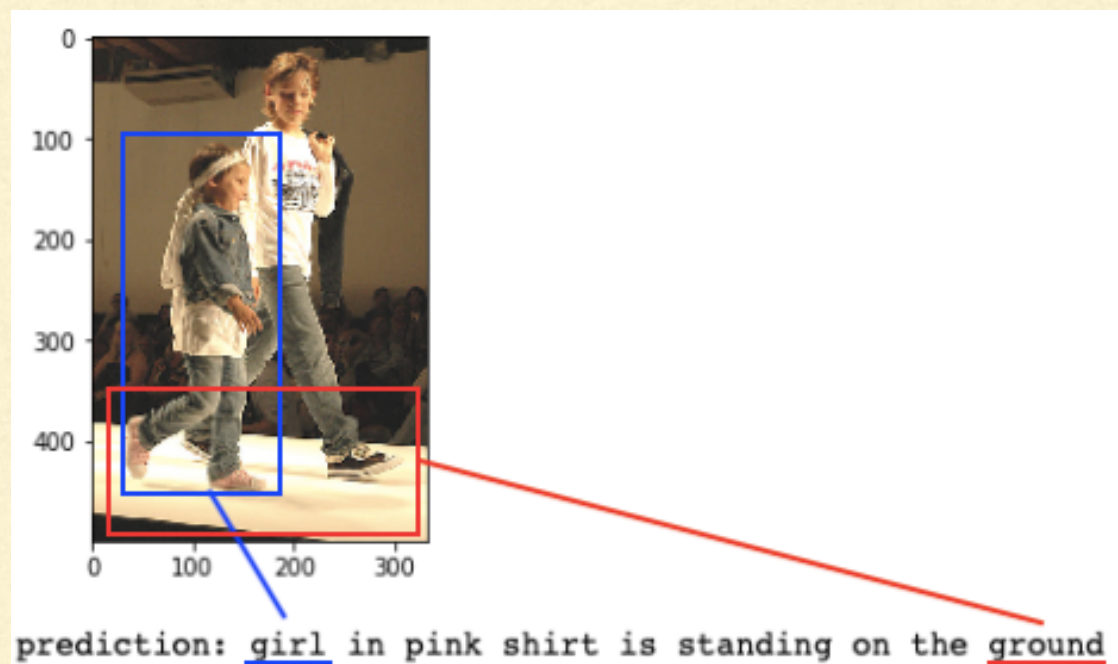
2048x1

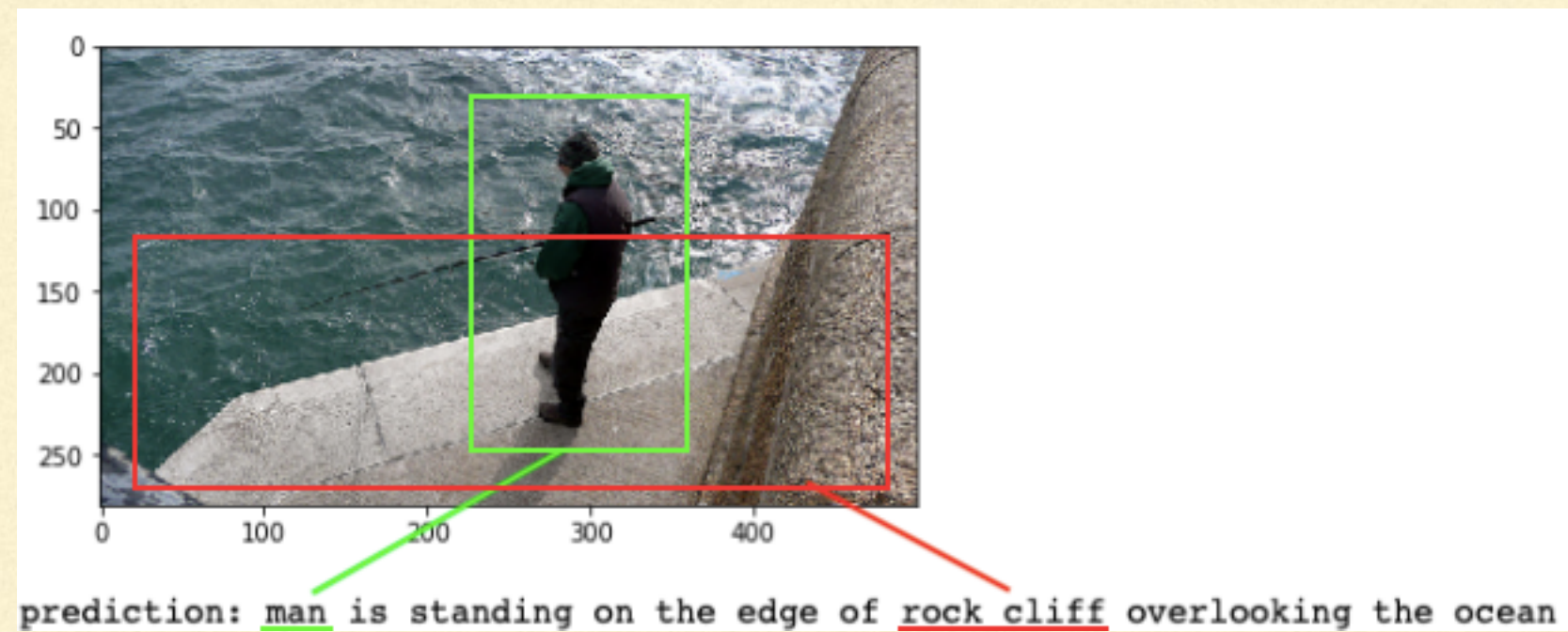
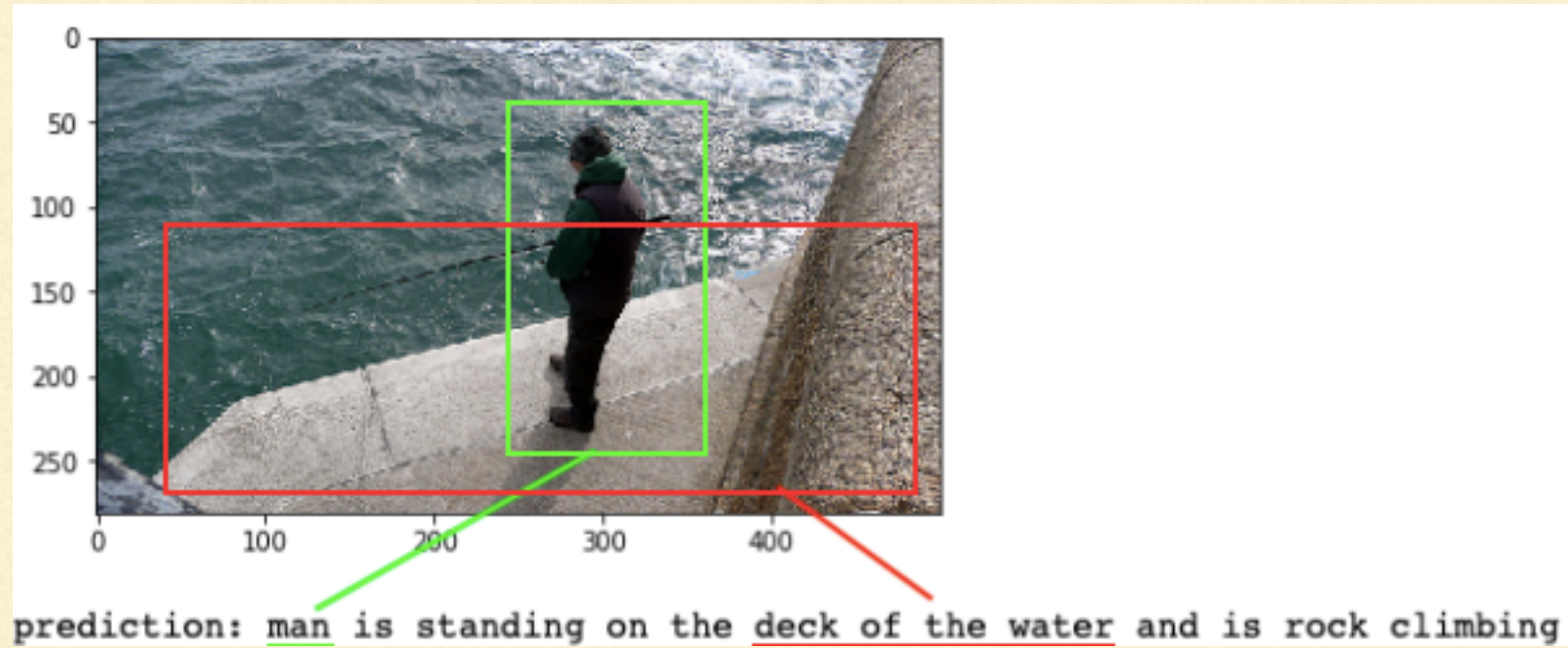
34x200

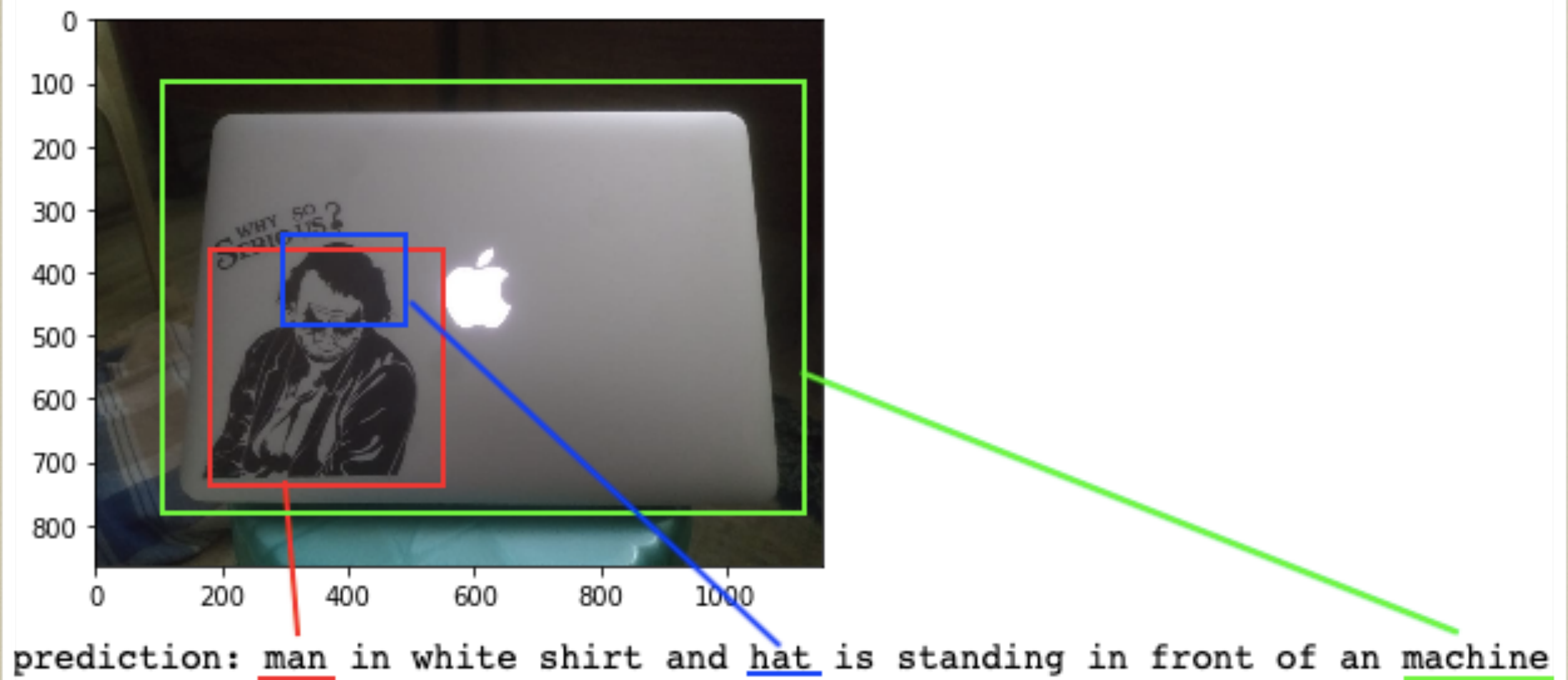
MODEL

Next word in the sequence









IMPROVEMENTS:

- Increase Dataset size
(Current Size is train/dev/test=6000/1000/1000)
- Increase the architecture
(Current Size is 164(InceptionResnetV2 model)+10(model) layers = 174 layers)
- Usage of BLEU, CIDEr, ROUGE, SPICE, Meteor
(Current Metrics is value of loss function)

THE END
