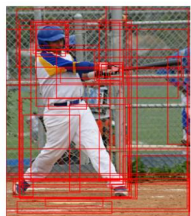


Q: What sport is this ?

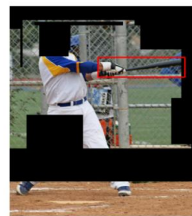


GT: baseball

Attention Layer 1



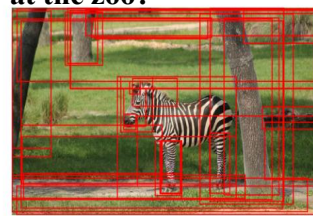
Attention Layer 2



Answer: baseball

(1)

Q: Can you see this animal at the zoo?

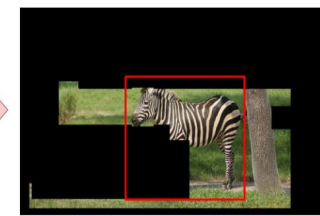


GT: yes

Attention Layer 1



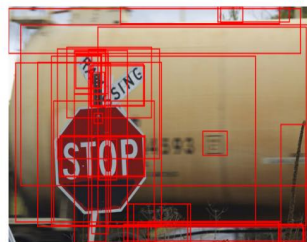
Attention Layer 2



Answer: yes

(2)

Q: What color is the stop sign?

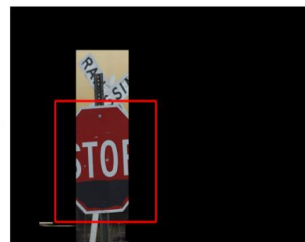


GT: red

Attention Layer 1



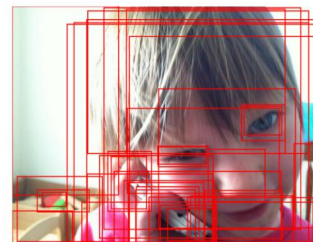
Attention Layer 2



Answer: red

(3)

Q: What color are the walls ?



GT: white

Attention Layer 1



Attention Layer 2



Answer: white

(4)

Q: What is the man holding?

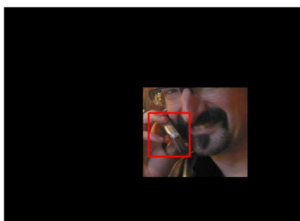


GT: cell phone

Attention Layer 1



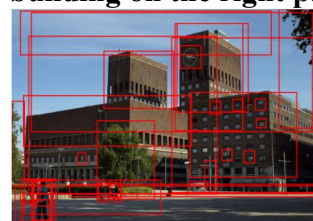
Attention Layer 2



Answer: cell phone

(5)

Q: How many colors is the building on the right painted?

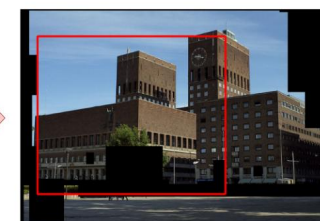


GT: one

Attention Layer 1



Attention Layer 2



Answer: one

(6)