

Task 3: Joint Caption + Trace Generation

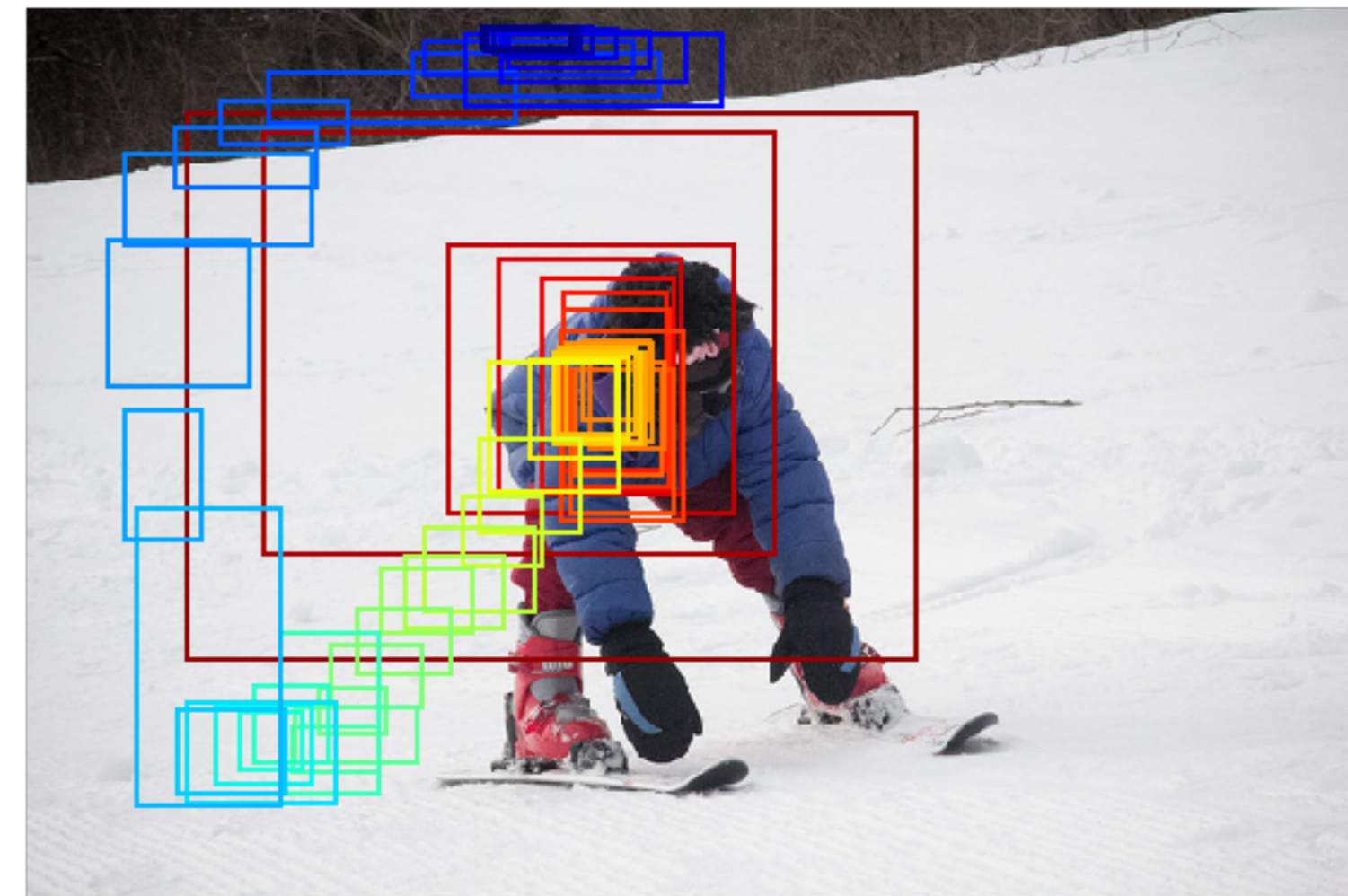
Image



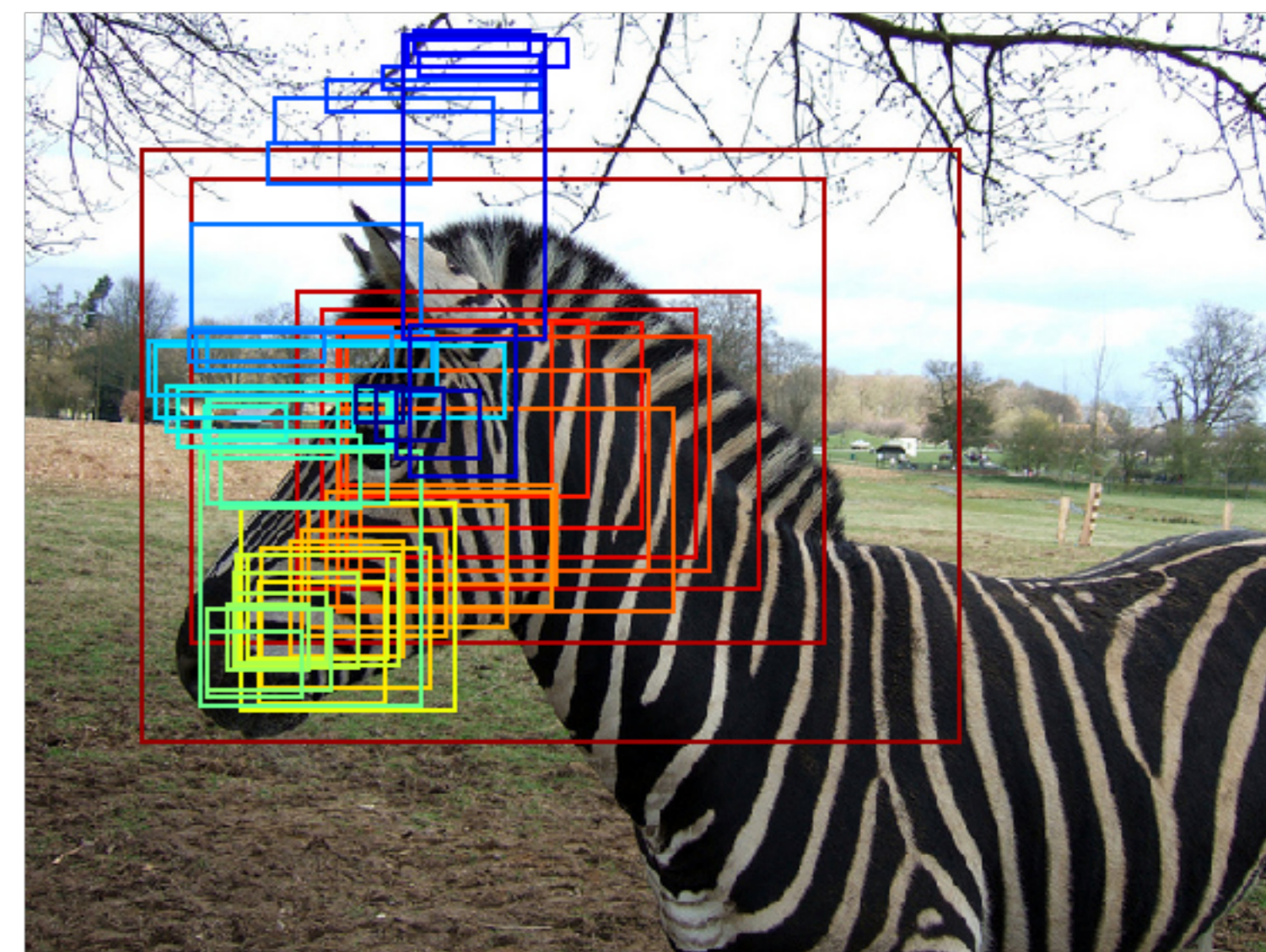
Predicted caption

→ In this image i can see a person wearing blue jacket, black gloves, black pant is standing on the ski boards and skiing on the snow. I can see the snow. In the background i can see the trees.

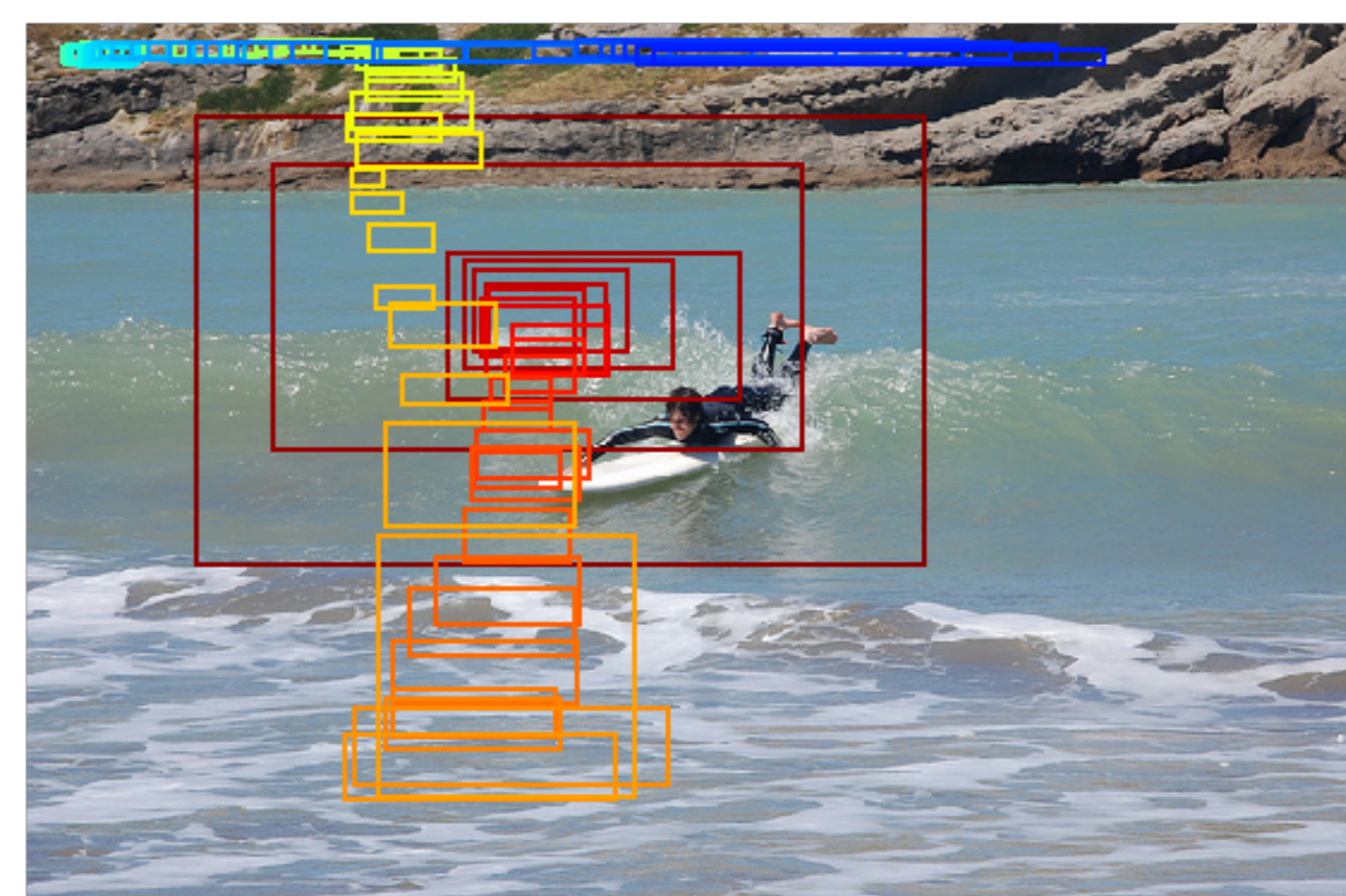
Predicted trace



→ In this image we can see a zebra. On the ground there is a grass. On the ground there is a grass. In the back there are houses, trees and a vehicle. On the top there is a sky.



→ In this image, we can see a person surfing a surfboard on the water. At the bottom, there is a water. Top of the image, we can see stones.



→ In this image we can see a blue color bus which is of blue color and something written on it. On the right side of the image we can see a pole and a board. In the background, we can see a pole, a light, a building, a tree, a cloudy sky, we can see a pole.

