Task 1: Controlled Trace Generation
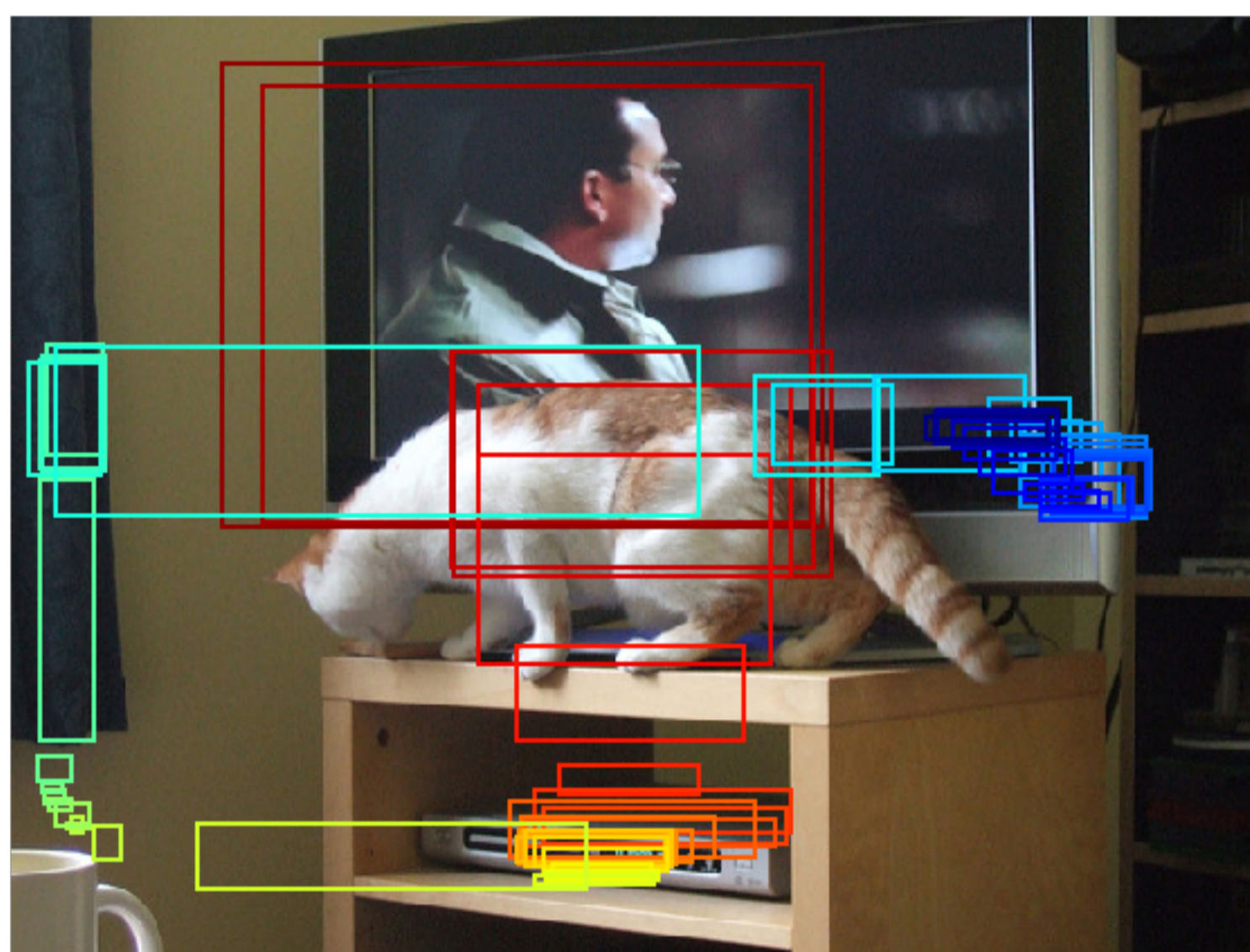
Ground-truth caption

Predicted trace

The image is taken on the road. In the center of the image there is a horse and there is a bus attached to the horse. We can see a man sitting in the bus. In the background there are many people. On the right there is a bench and people sitting on the bench. At the top there is sky.
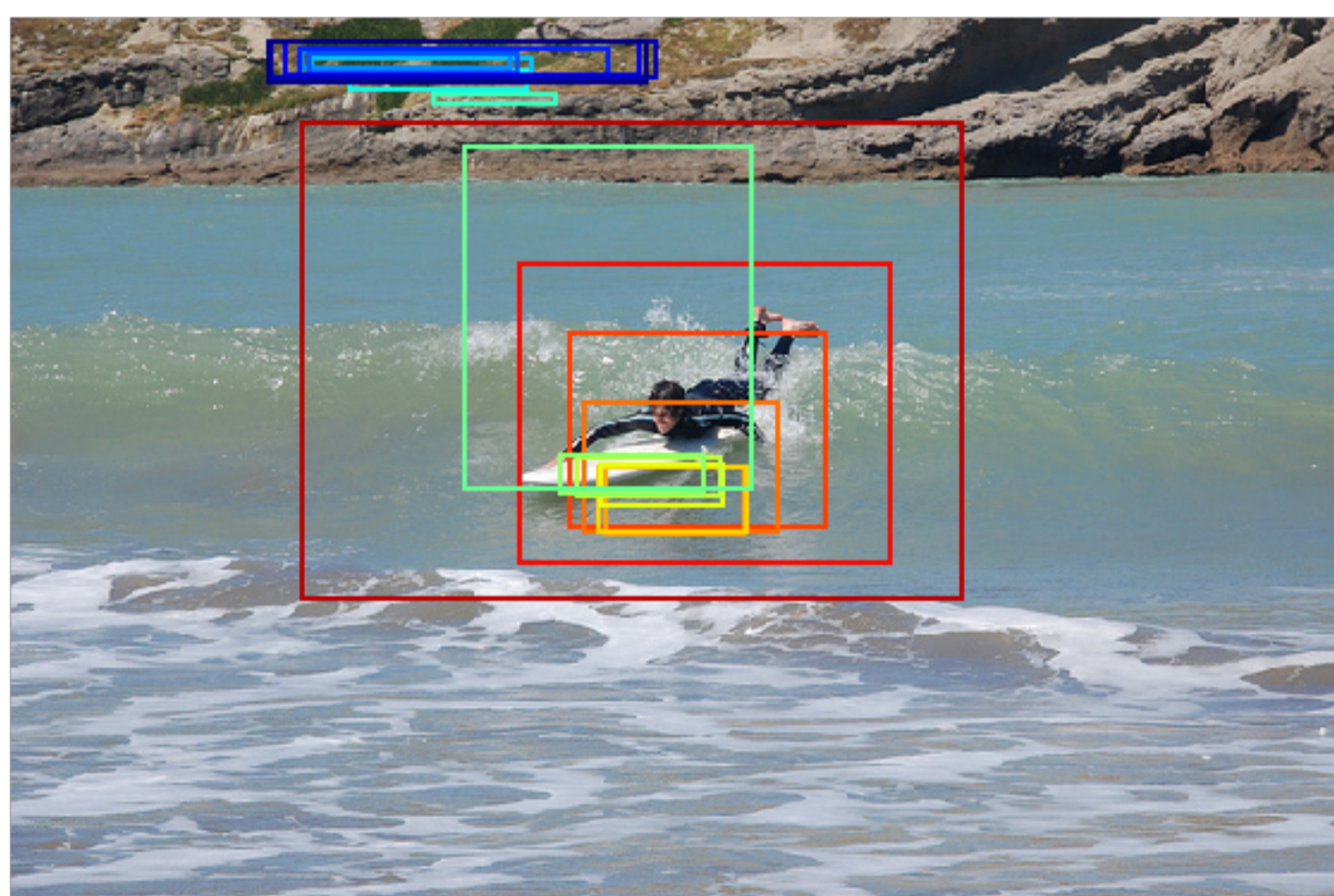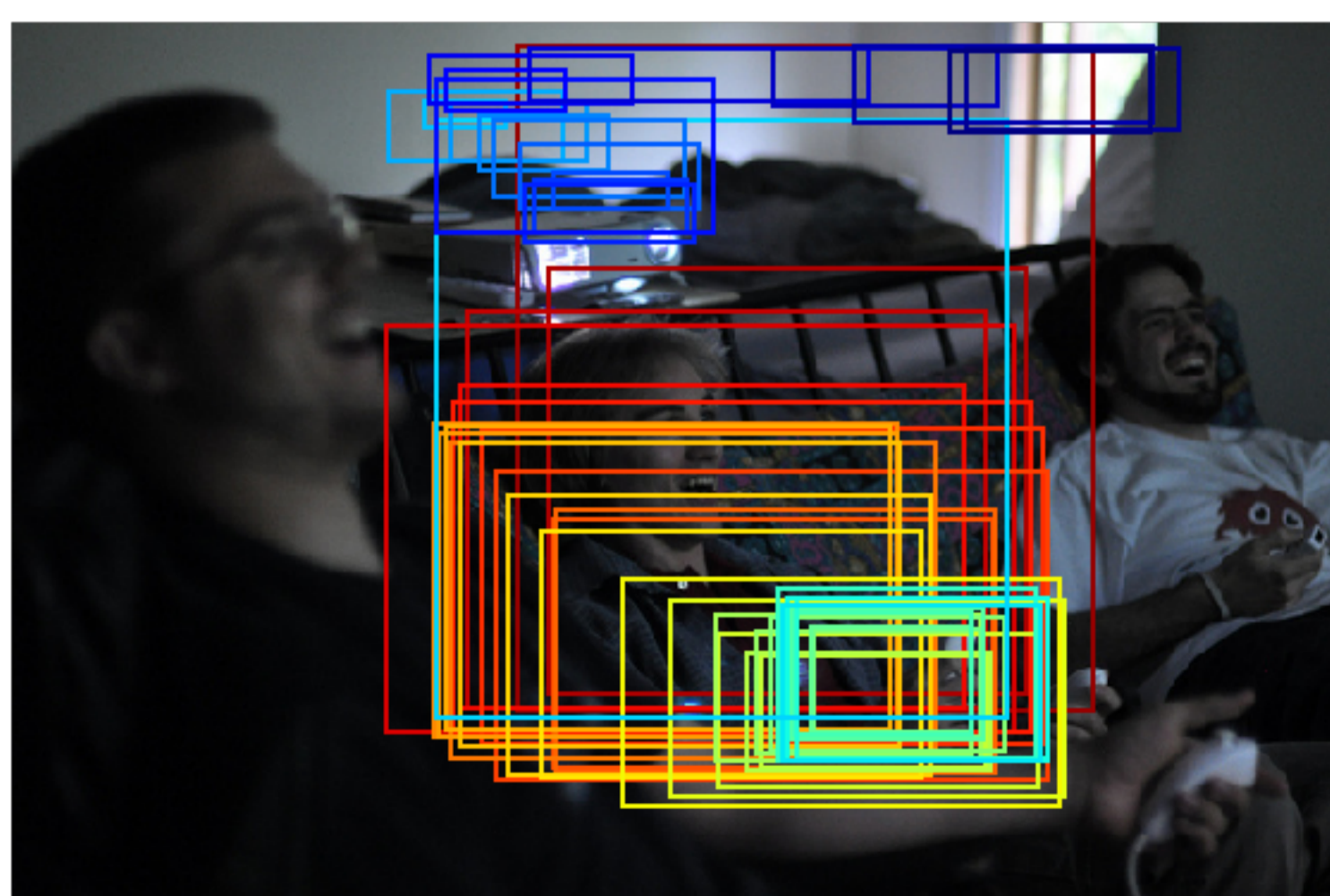
In this image, cat standing on the wooden table, there is a tv and player on the table. And left side, we can see a cup and curtain, wall. Right side, there is a big wooden shelf, some item items are placed on it. Here we can see few wires.

This person surfing on the water with surfboard, on the background we can see rock and grass.
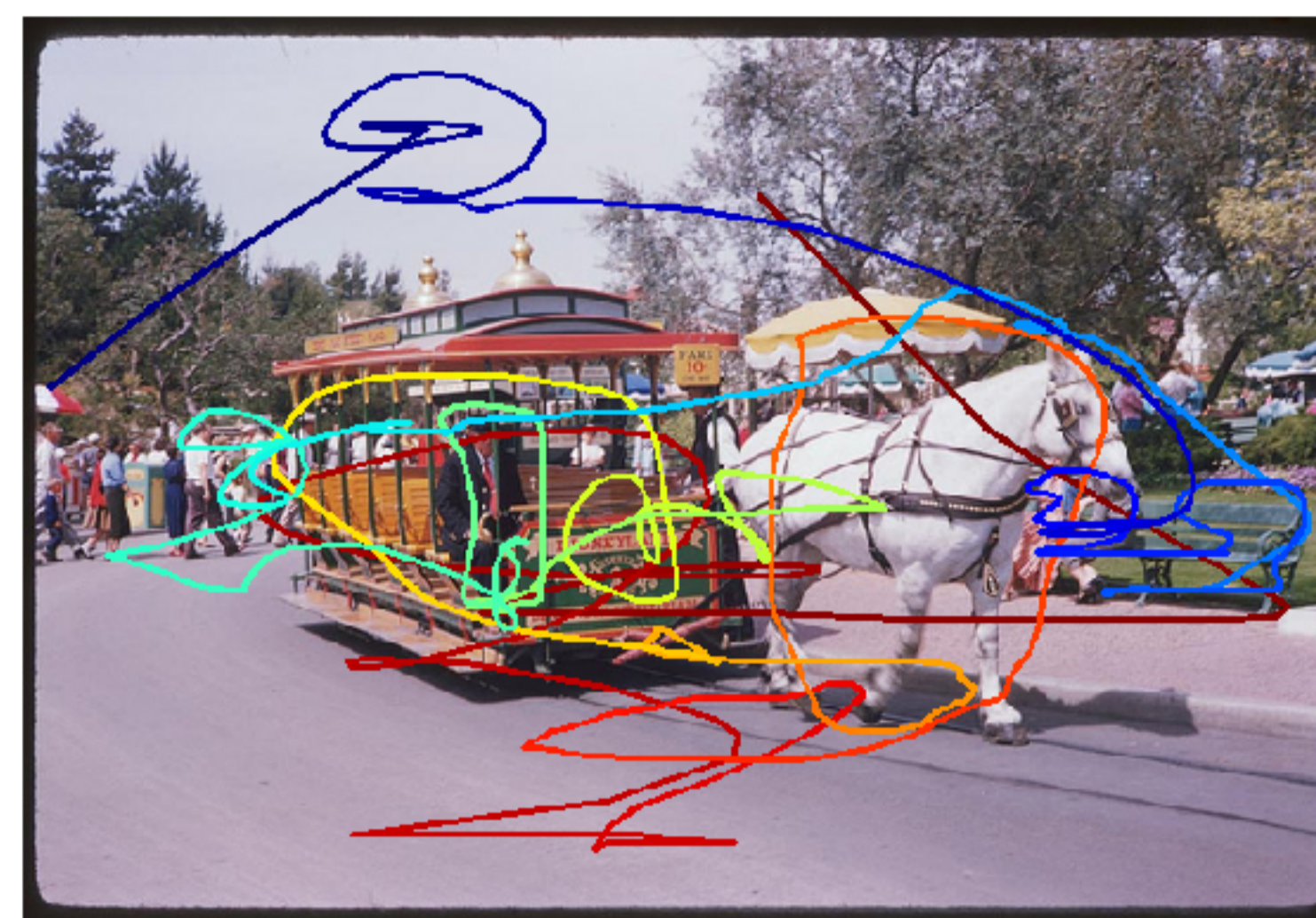
In this picture there are three people sitting on couch and they are smiling. They are holding joysticks in their hand and it seems to be they are playing video games. Behind them there is a projector on the table. In the background there wall and window.
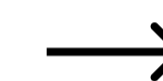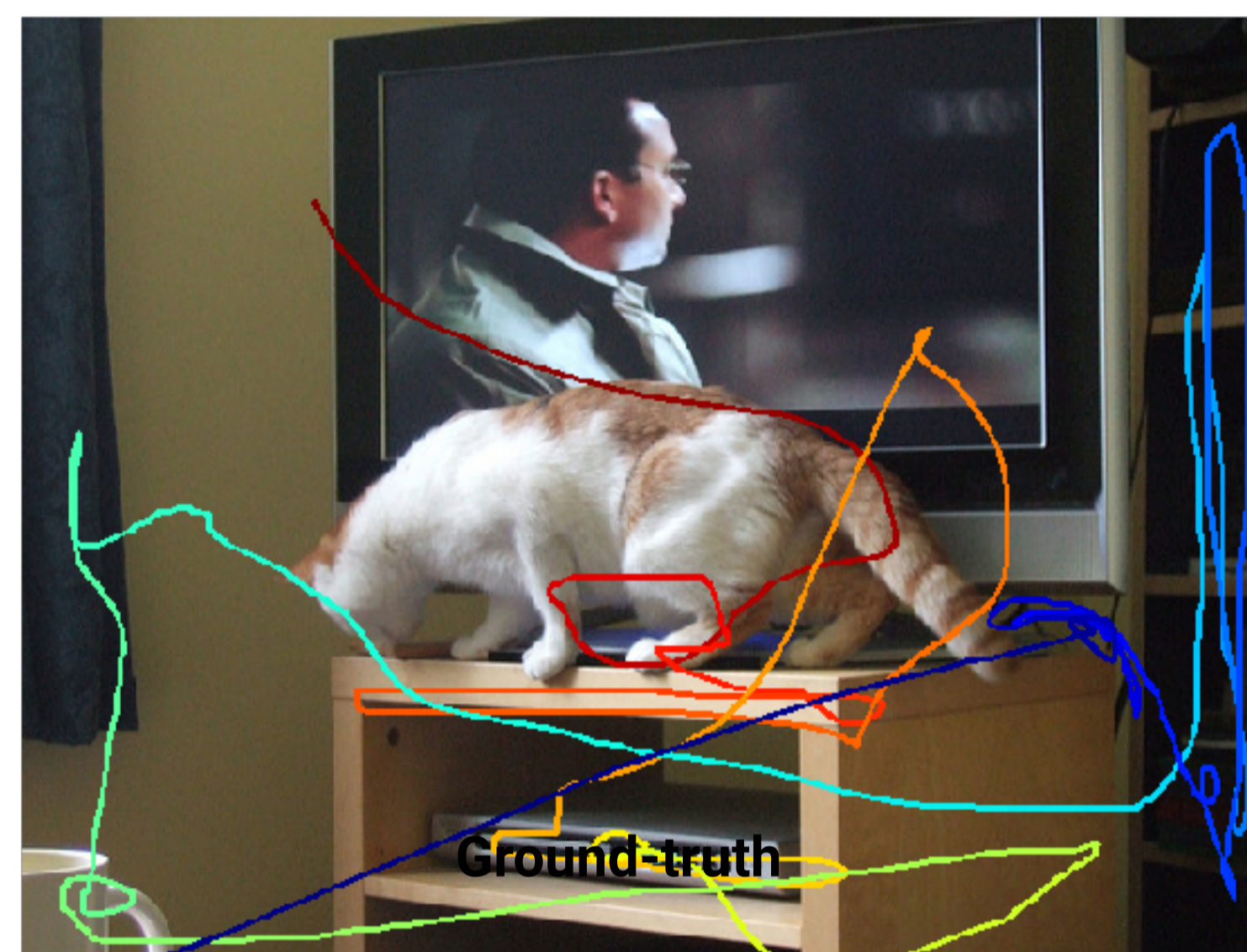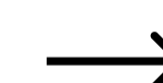
Task 2: Controlled Caption Generation

Ground-truth trace
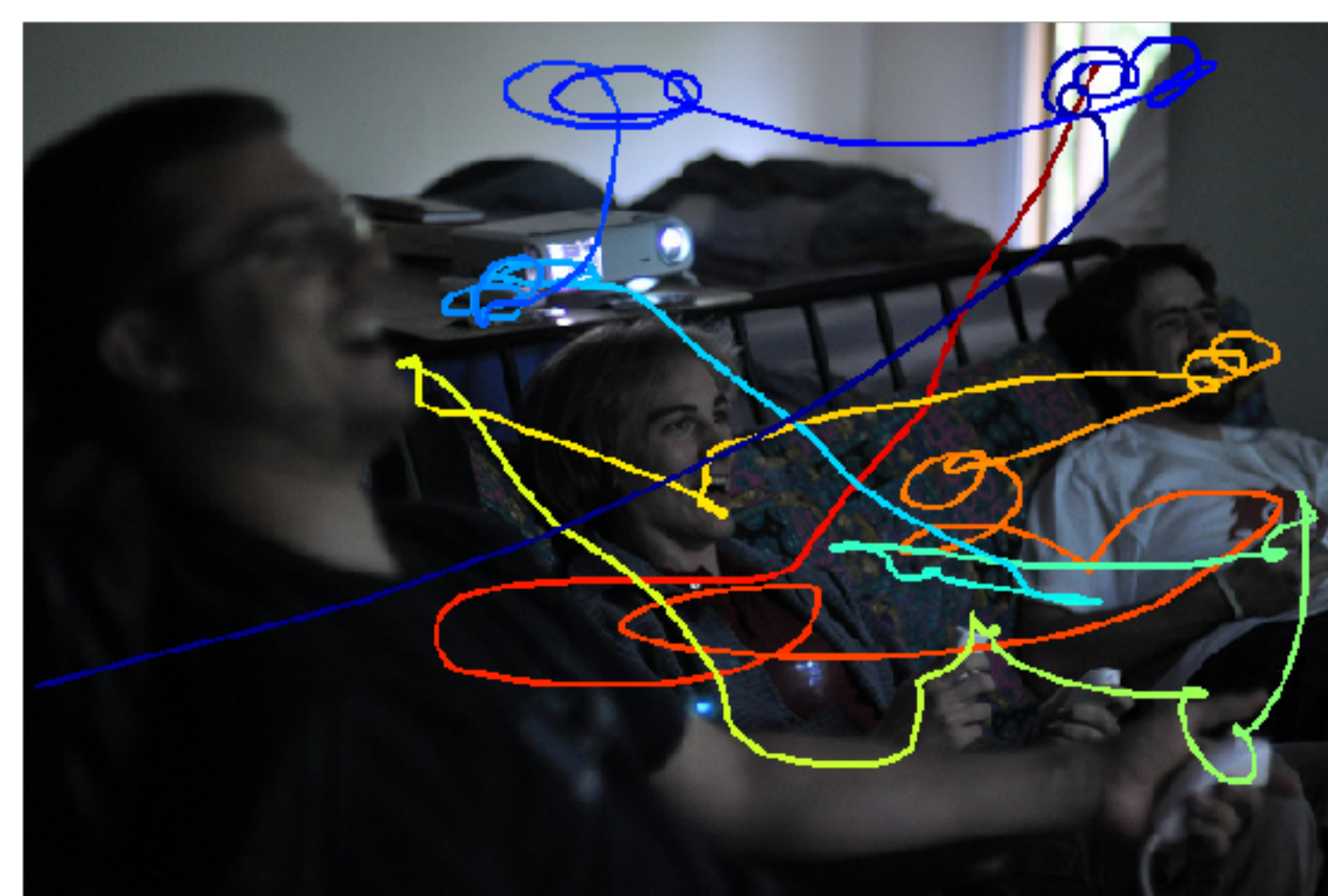
Predicted caption

This image is clicked on the road. In the center there is horse running on the ground and there is a cart and there are horses. On the cart there is a person sitting on the cart. There are group of persons. On the right there is bench. On the bench there are plants. In the background there are trees.

In this image, a cat is lay on the wooden table. At the bottom, we can see some machines, some containers. Left side, there is a cloth here. Left side, there is a wooden shelf, so many items we can see here some wooden piece. Right side, there is a wall.

Here there is a man surfing on the water. In the background there is a stone wall.

In this image there are few people sitting on couch and they are smiling. To the right there is a person sitting and holding a remote in his hand. There is a table on which there are few books placed on it. There is window behind them.