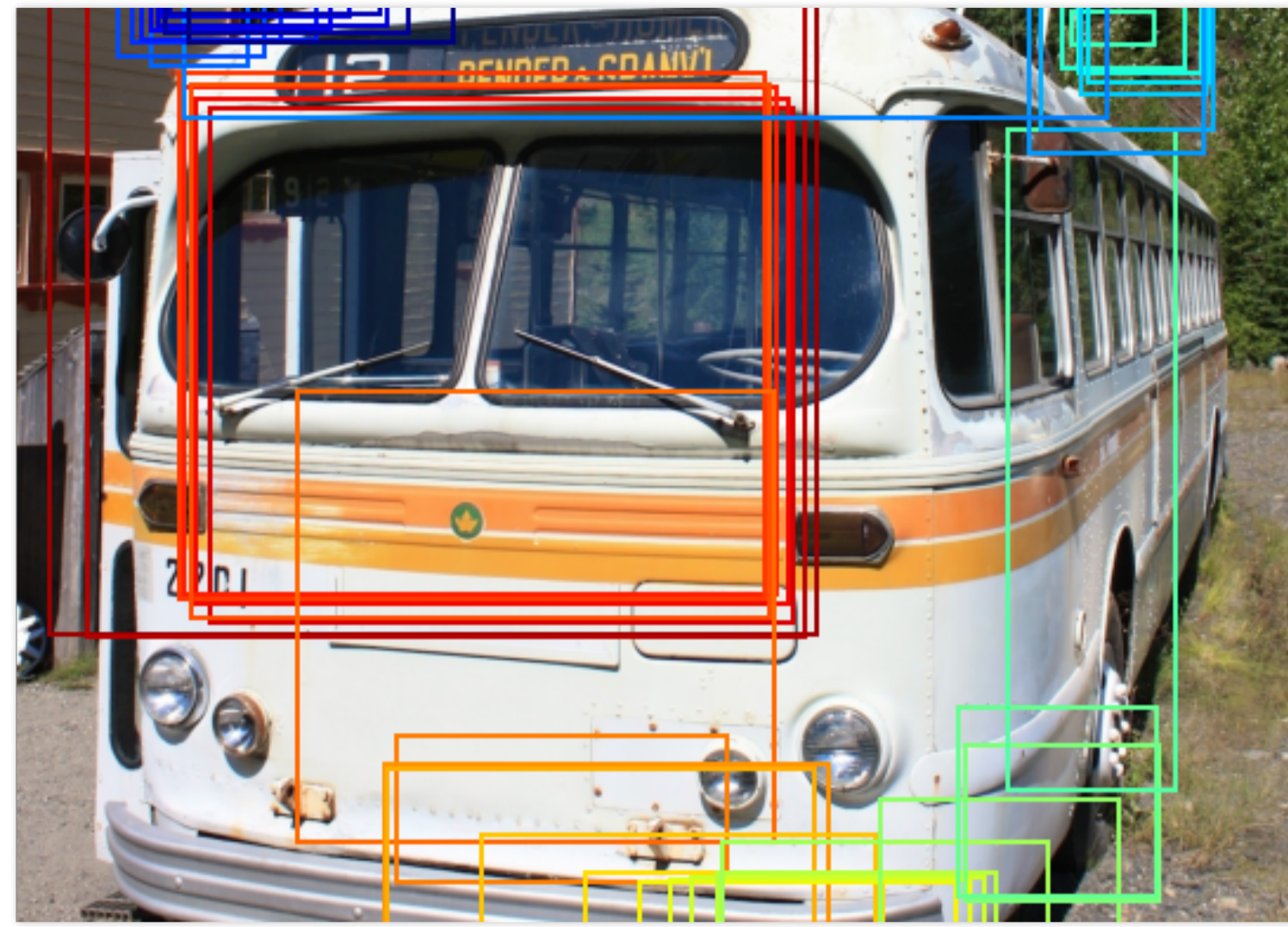
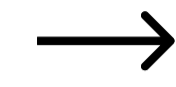


## Task 1: Controlled Trace Generation

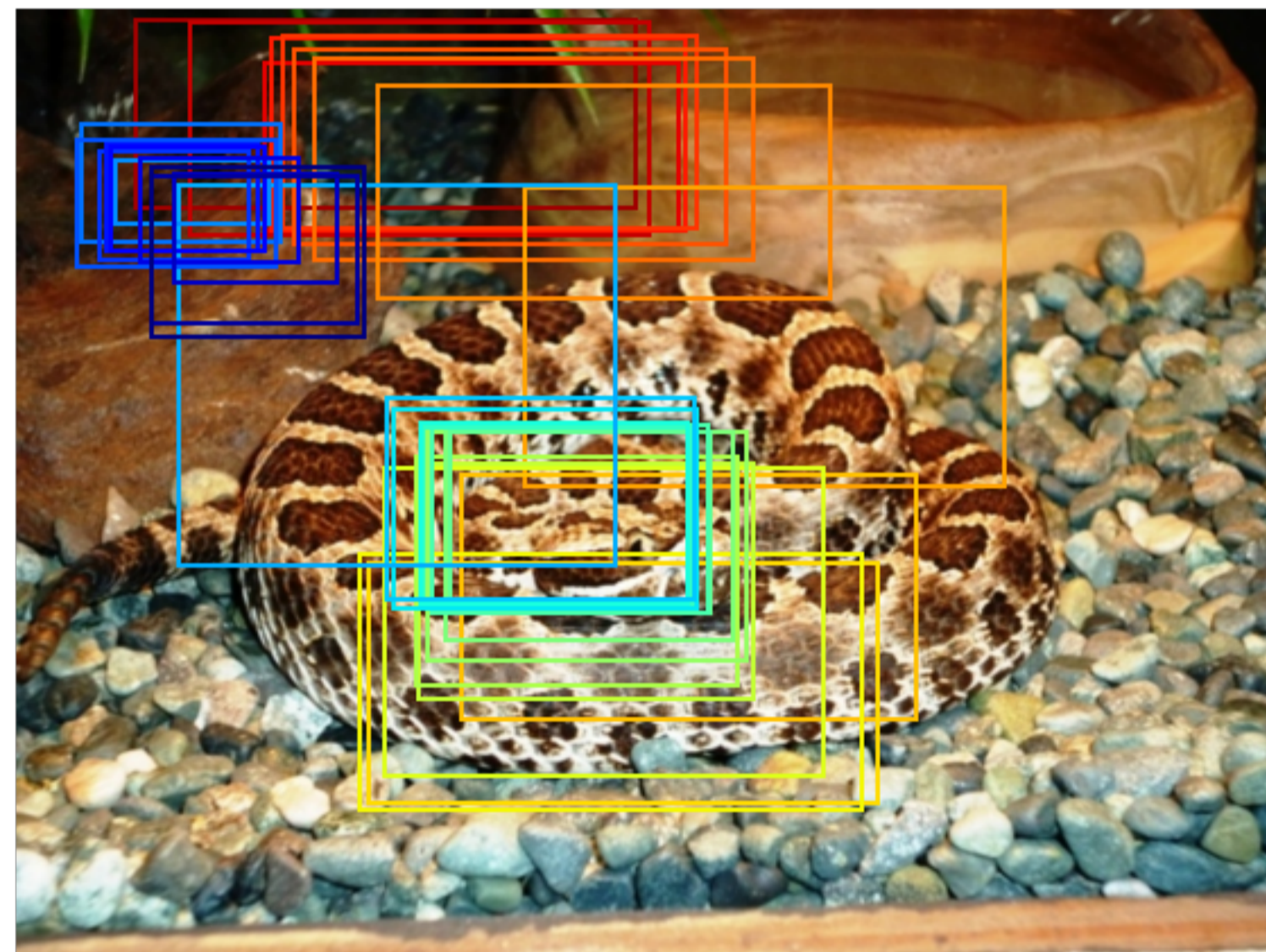
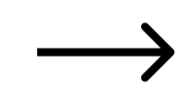
### Ground-truth caption

In this image we can see vehicles on the ground. On the ground we can see stones and grass. In the back we can see trees. On the left side we can see building with windows.

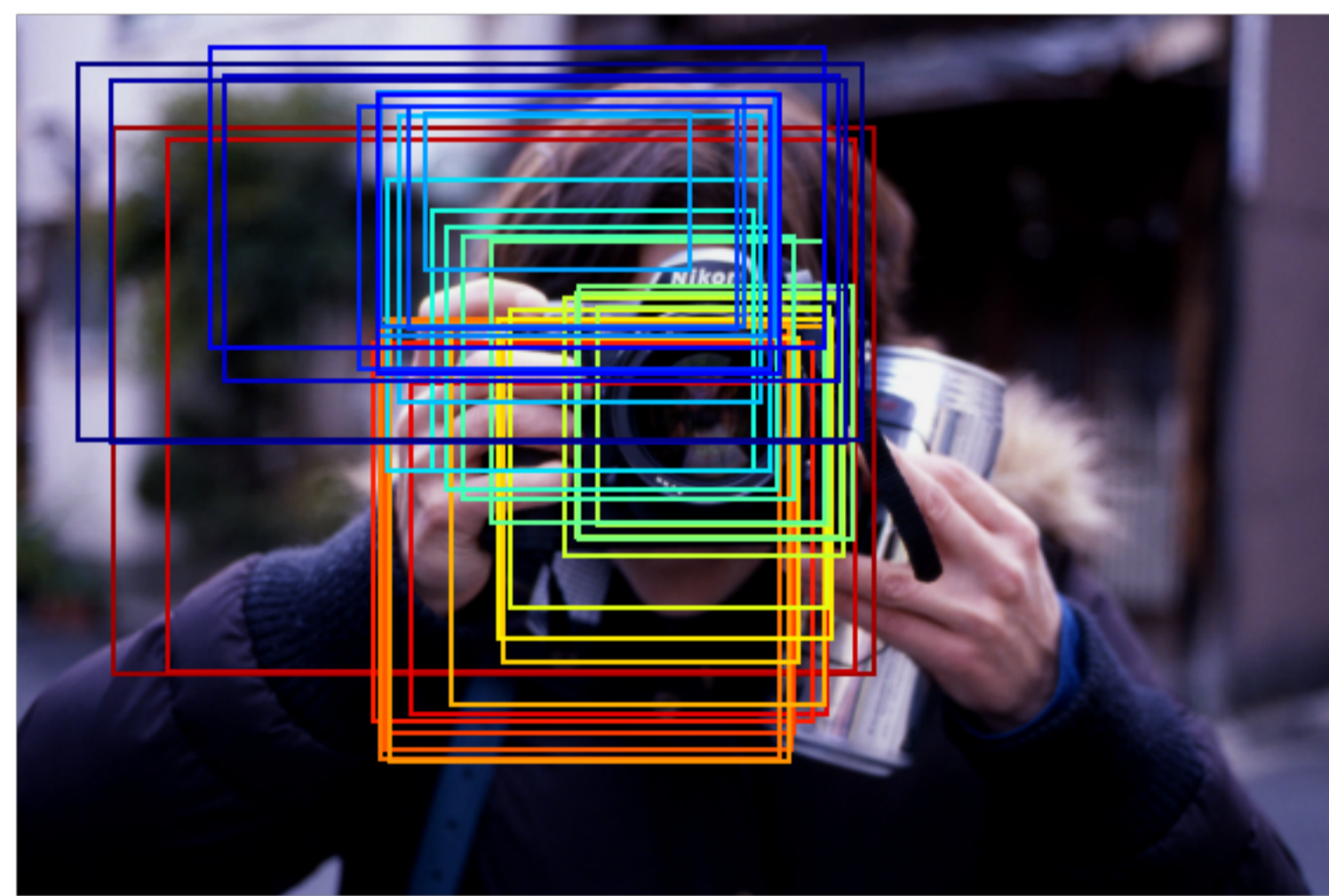
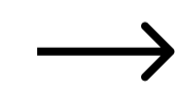


### Predicted trace

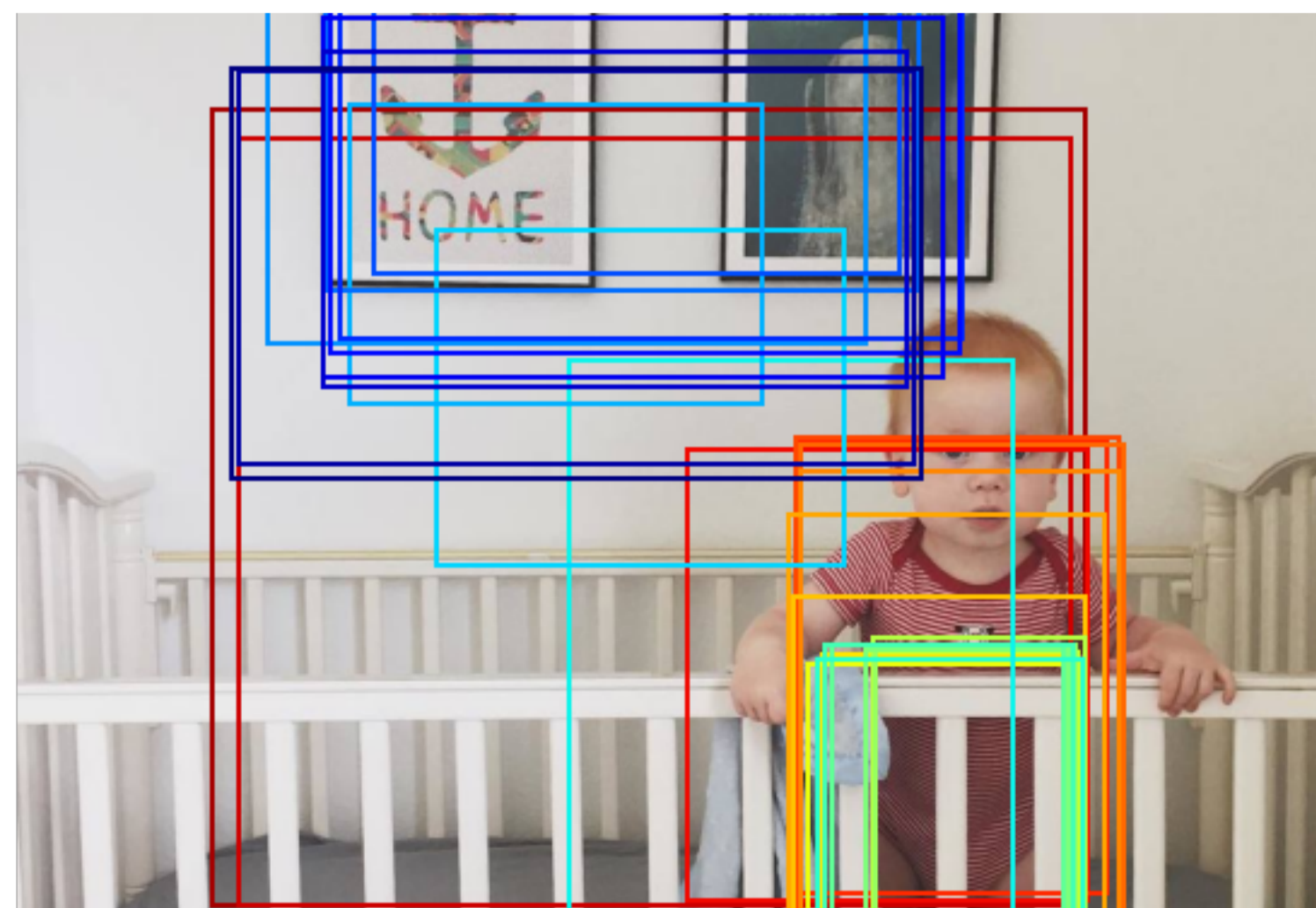
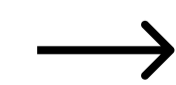
In this image, we can see wooden objects on stones. There is a snake in the middle of the image. There are leaves in the top left of the image.



In this picture we can see a person holding a steel can and a camera. We can see the greenery and the architecture in the background. Background is blurry.



In this picture we can see a baby holding a cloth and standing in a baby crib. We can see frames on the walls visible in the background.

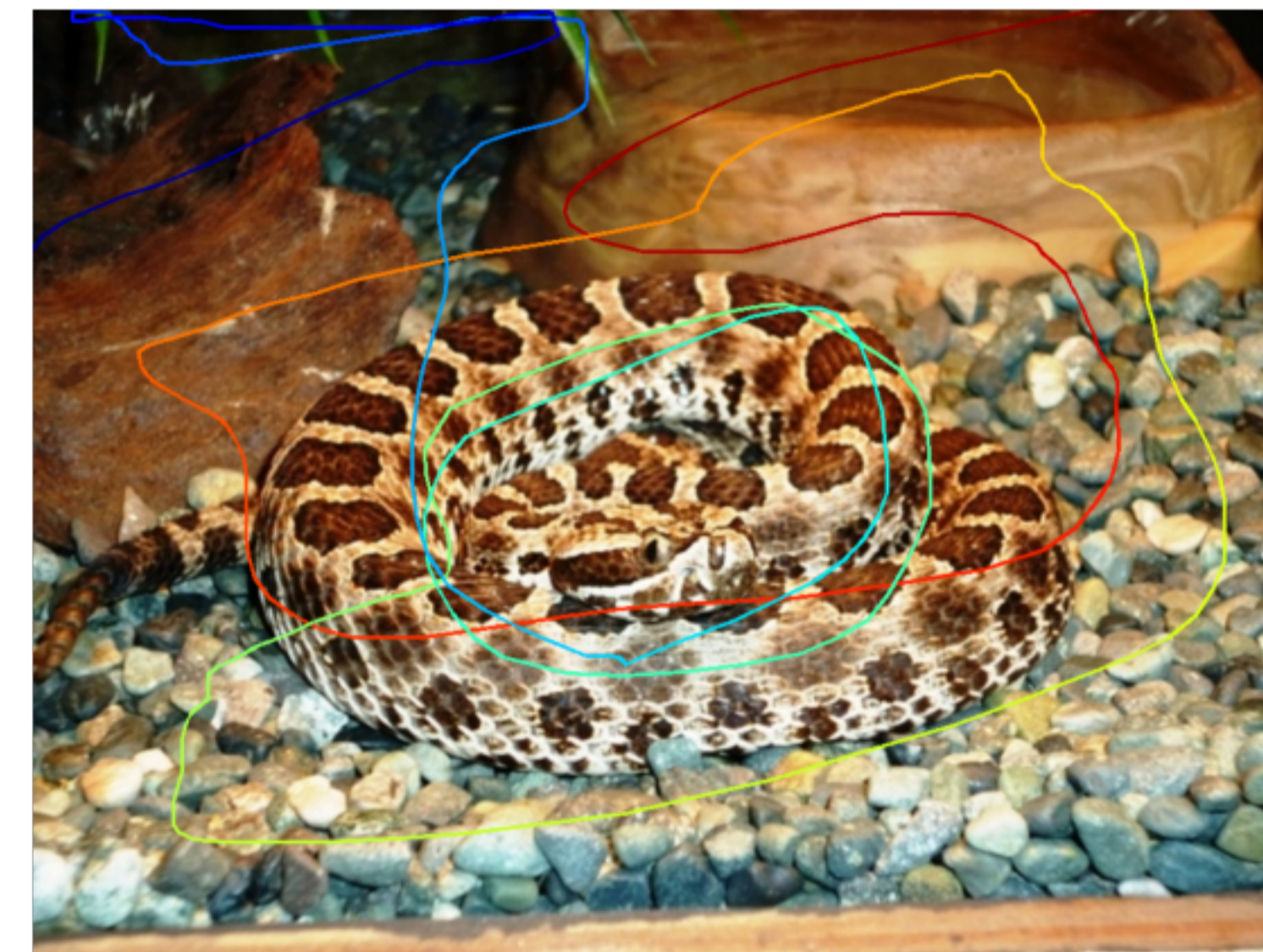


## Task 2: Controlled Caption Generation

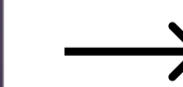
### Ground-truth trace



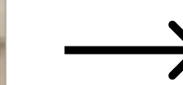
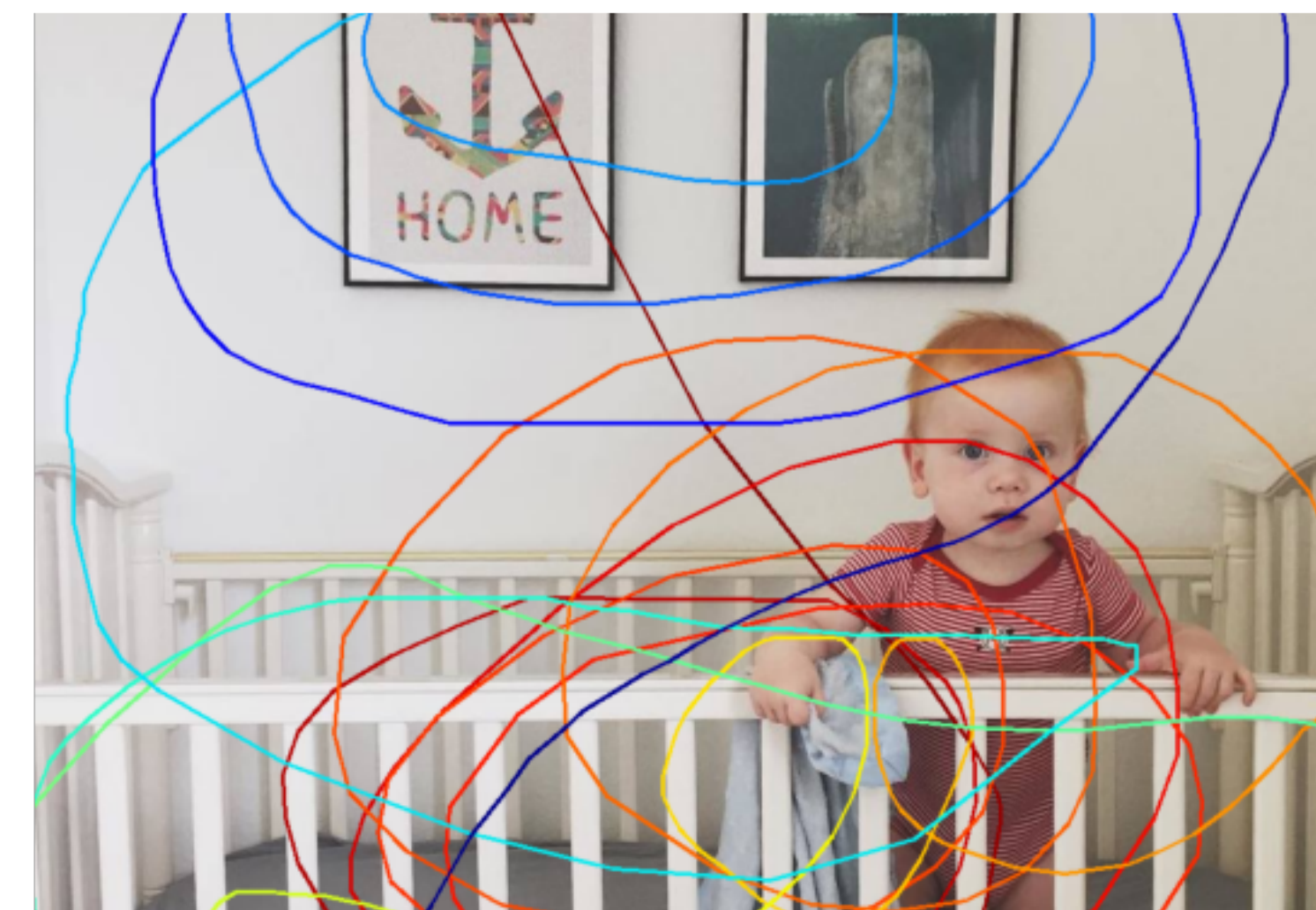
In this image we can see a vehicle parked on the ground. On the right side there is grass. In the background there is a building. On the left side there is a building with windows.



In this image, we can see a snake and stones. There is a snake in the middle of the image. There is a branch at the top of the image.



In this picture we can see a woman holding a camera in her hand and in front of her we can see a blurry background with some objects in the background.



In this image, we can see a kid and there is a railing on the railing. In the background, there are photo frames on the wall.