

1. Which algorithms would you use to achieve this? Please provide at least 2 proposals

I would make use of semantic segmentation to label the different classifications of our interest, and some interesting algorithms might be skip connections by concatenation to combine predictions from features at different depths, or using the transposed convolution.

2. Which potential bottlenecks may this problem encounter? (i.e. non-static camera). Just list them (if any)

The confusion with the public, since it is a group of people that can be easily confused by the camera as a group of players.

3. Could your proposals be scaled to other team-based sports? (i.e. hockey, rugby, ...) Which would be worst-case scenarios?

Yes, for example in volleyball, lacrosse or waterpolo. By far waterpolo would be the worst-case scenario since all the players are mixed in the water and it is more confusing to differentiate who is who.

4. Extra: provide a sample code with some proposal running