Improving machine learning by improving image processing

Image processing becomes important for learning machines to react the real world. Nowadays, machines process the image in digital form, different from how human would do. Although we have convolutional layer, there still are difference. One difference is that human seeing the world in continuous time which allows human to detect motions easily. While the image input for machines is frame by frame, although we can use the convolutional layer to eliminate the difference between similar frames, sometimes we want the machine to process sequence of frames in order to detect a motion. Nowadays we have markerless motion tracking techniques, but they are limited in detecting human’s motion. Another difference is that human sees the world in controllable detail, while machines stores the image in pixels of indexed 2d array. And we can track an object even if we move our eyes, while machine only knows the indexes of pixels of an object, and it requires complex algorithm to track an object when moving camera. The third difference is that human can intentionally focus on an object or environment while machines has to scan whole image pixels in order to detect objects, which consumes lots of computing power. The proposal is to find a way to maximise simulate human’s vision and how we process our vision in our brain. Firstly, human does not store the image frame by frame in their brain, instead, we are trying to understand what we see and keep the concept in our memories. (Mary Potter, MIT professor of brain and cognitive science) Secondly, human eyes process the light continuously rather than discretely, this allows human vision to track motions easily. Although we have markerless motion tracking techniques, they are limited in detecting human’s motion. Thirdly, human know to focus on something interested. (restructure to 2 points, human vision continuously, human focus interested) By simulating human’s image process method we can reduce the state space in machine learning. Improving image processing to allow motion definition for the agent. (So it can define a motion) to allow unsupervised learning more efficient. Sometimes human does not care about what colour is an object, but care more about if there is difference between environment.

Approach: do research in architecture of human eyes, and in analogue signal processing, computer vision. Simulate motion telling system and test it, test human as well. Do many experiment about human vision. E.g. does human care about lights colour, experiment can be show people picture and ask if they remember the colour, another experiment can be position people in light flashing room, then show then moving stuff?

Evaluation: build a dancing robot with this method and put it in the mall and invite customers to ‘teach’ the robot how to dance.

Precise system: vision focus and compute precise value if it is interested.

Change the way to store image. E.g. rather than using an array of pixels, we use structure of concept.

e.g. we build a 3d model in computer vision, by rather taking the picture and store it in 2d array with indexed each pixel