Improving machine learning by improving image processing

Image processing becomes important for machines to learn from the environment in the real world. Nowadays, image processing by machine is still quite different from how human vision works. The differences may limit machine learning.

One difference is that human seeing the world in continuous time and a research shows that our brains process the visual information in a delay of one-tenth of a second. In order to track a moving object, our brains predict the path of the object. However, the image input for a machine is frame by frame; and the frames are sampled in a fixed time step, which may include noises and hence may confuse the machine.

The second difference is that human sees the world in controllable detail, while machine stores the image in pixels of an indexed 2d array. And we can track an object even if we are rotating our eyes, while machine only knows the indices of pixels of an object, and it requires many complex algorithms to track an object when moving the camera. Although we have the convolutional neuron network algorithms that can extract features from images, this method may not track the identity/reference of an object very well.

The third difference is that human can intentionally focus on an object or environment, while the machine has to scan whole image pixels in order to detect an object by detecting edges of the object, which consumes lots of computing power.

The proposal is to find a way to maximize simulation of human vision and apply it into machine learning and hence allows the robot to learn from the real-world environment in a possible way.

Approach: initially do some research about how our brains process the visual information and how computer simulates this process currently (studying computer vision). Experiments can be conducted to find out some evidence that supports the theory. An example hypothesis can be: human vision does not care about the colour but the difference of colours to identify an object. Then, models can be built to simulate how our brains process the visual information. A simple algorithm that simulates human paying attention can be:

Get input image, divide it into small parts;

Compute which part should be focused by using some algorithm simulating how eyes are attracted by objects;

For each part of the image:

If it is focused part:

Pay More attention to this part by getting more precise image input value only for this part;

Else if it is not focused part:

Ignore it, or take only a few sample of it;

Final Evaluation: we can build a dancing robot using the developed model and put it in a mall or train station and invite visitors to ‘teach’ the robot how to dance. Robots from 0 knowledge will ‘watch’ performers through the camera and make some moves. Performers then give feedback on how well the robot learns. Currently, robot Baxter learns to dance by ‘teacher’ moving its arms, not by vision.