As you may know, biological systems, for example humans, have very poor sensory information. Rods in the eye respond seemingly random to photons hitting them and more importantly clarity in the human eye exists in one single location called the fovea. The areas surrounding the fovea decrease in precision the further they are.

Given such poor sensory information, it is assumed that the brain may rely on Bayesian inference in order to determine the true state of the world. As visual information is collected through ocular fixations our belief about the true state of the world becomes more and more precise. This is important for robotics, artificial systems have to deal with sensor uncertainty no matter how good your hardware is.

The focus of the project was on visual perception and teaching an artificial agent to behave similar to how humans would.

Our main assumption is that humans direct their sensory input to the locations that will increase the success of their current activity. We know from literature that humans direct gaze to the locations that aid their current tasks. We take a step further assuming that it is not just the current task, but to the most important activity which gives the highest reward.

The experimental setup is similar to a visual perception task as found in some brain training websites online. Imagine a computer screen that flashes two objects for short duration, a few milliseconds. The subject is then required to click on the exact center of the objects and he receives a reward if he succeeds or a punishment otherwise.

Everything makes sense so far? Good.

Now imagine the two objects are not yet shown on the screen. Then our belief of where the objects will be, is spread across the entire screen. They can be anywhere.

When the screen first lights up and shows the objects, no matter our current fixation we get some visual information about their location. But the information is not precise, because we perceive the objects with our peripheral vision instead of the fovea. Unless we get lucky and they happen to appear where we are looking.

When we make the first ocular fixation to what is believed to be true object location, we get another set of visual stimulation and our belief about their location becomes more precise.

Our agent encodes this belief state changes through something called a particle filter. Which is a common algorithm in robotics, and may be known in psychology as Metropolis Haystings.

<<nod and short pause to observe if audience is still on track>>

The agent has to learn two things. The objective of the task, and where to optimally look. Both are challenges in our model, although the main objective is to test a training strategy for optimal gaze. Namely directing visual input to the direction that increases reward.

We use an actor critic model for reinforcement learning. The agent takes some actions in the world, and the environment either rewards him or punishes him. The difference eventually trains the behavior of the agent to do more of the actions that reward him.

Actor-critic is the overall training strategy. What we really train is a set of radial basis function networks, that given the spread of uncertainty with the value of each object, and returns a probability for each action it can take. Namely, to gaze at object one or object two and to attempt to grasp an object or not to. If the spread of uncertainty is high, it might be unwise to grasp an object because missing its center will result in a large punishment.

It is a bit hard to explain what the value function is and represents. In short, it is used for training. It represents an expected value of reward given our current belief. For example, making a grasp action for a certain belief, should have a large expected reward compared to making a grasp for very uncertain belief. This difference in expected reward over time steps is used to train the probabilities of taking different actions.

So we train our grasp, and we train our gaze, and we can see, that as certainty about the world increases, the grasp action probability increases and we start directing our gaze to the object that gives us higher reward.

<<wait for visual confirmation that they understood>>

From the result we can see the result of training. Our assumption that directed gaze is more optimal than randomly fixating between object is quite clear and significant. This small reward function on top is what brings that difference and was the objective of the placement.

In laymen terms what the formula represents is:

<<said slow>>

reward for gaze is the sum of expected improvement over grasping reward. The reward of reaching picking up either object.