

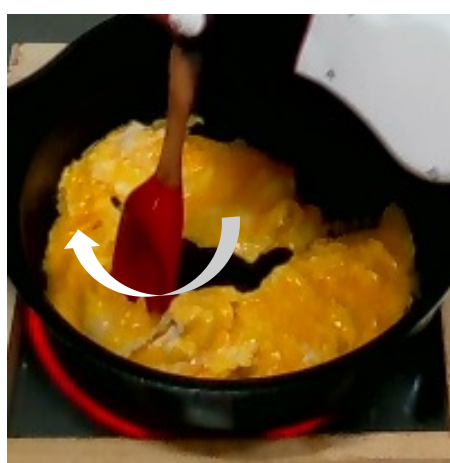




(a) Collecting training data with teleoperation





Need to adjust the motion depends on states of the egg



Raw → Hard


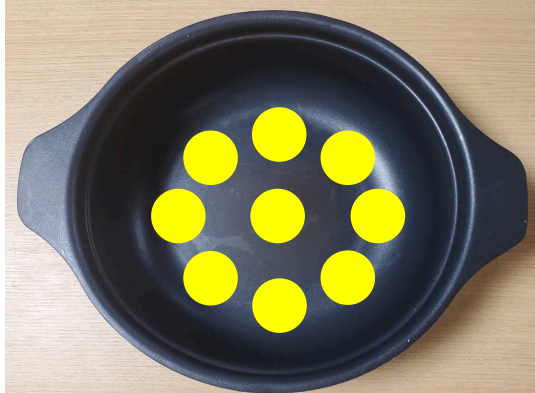
Otherwise the egg will

- Burned
- Undercooked
- Become large blocks



Bad example With our model


Teleoperation using key motions



Key-points Flipping key-motions


Training data variation in

Vision




Plain

Touch




Seaweed

Vision & Touch




Corn




Sausage


(b) Learning from demonstration



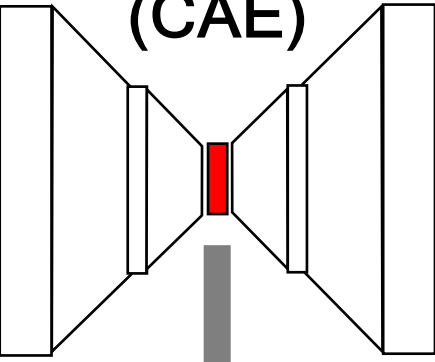
Whole image (t)
720000 dim
(500 × 480 × 3)




Trimmed image(t)
120000 dim
(200 × 200 × 3)




Convolutional Auto Encoder (CAE)



Reconstruct whole image



Reconstruct trimmed image



Whole image feature (t) 20 dim
Trimmed image feature (t) 30 dim

Motor angle(t) 7 dim
Torque sensor (t) 7 dim
Tactile sensor (t) 4 dim

Motor angle (t+1) 7 dim

Multiple Timescales Recurrent Neural Network (MTRNN)

Slow Context (Cs)


Fast Context (Cf)

Attention

Input-Output (IO) node

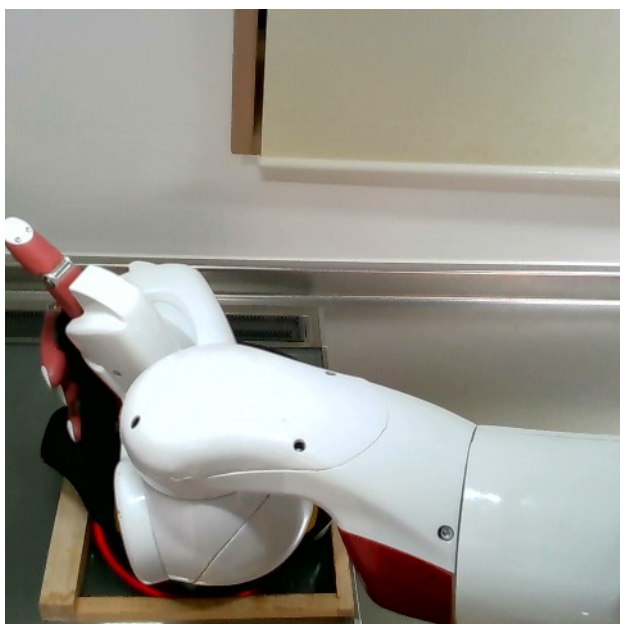
(c) Evaluate motion generation with untrained ingredients

Efficient perception with attention mechanism



Turner does not touch the pot

Focus on Image




The arm occludes the pot


Focus on Torque and Tactile

Variety in

Vision




Soy sauce




Red food coloring

Touch




Bamboo shoots




Cheese

Vision & Touch



Minced meat



Spinach