Our model is characterized by:
1. Image-dependent search trajectories
2. Can adjust speed-accuracy trade-off at test time
3. Context is aggregated as search proceeds
4. Search policy and detector parameters jointly learned with deep RL (policy gradients)
5. Automatically terminates search

State representation $S_t$
- High-level feature map
- Objectness and box offset norms
- Class-specific history of earlier detections

Action representation
- Policy
- GRU
- Exploration penalty $\beta$
- Softmax $\text{softmax}(i)$
- Fixation probability map $F_t$
- Fixation $\ell(i, j)$
- Not done $\text{nd}(i, j)$
- State of fixation $S_{fix}(i, j)$

Sequential object detection system

Overview of proposed drl-RPN object detector

1) Action: Fixate next location
   i) Classify RoIs in fixation area
   ii) Locally detected: person
   iii) Update class-specific history
   iv) Fixate action reward (training):
      \[
      \hat{r}_t' = -\beta \sum_i (g_i/\text{IoU}) > \text{th} \geq 1 \frac{\text{vol}_{max}}{\text{vol}}
      \]

2) Action: Fixate next location
   i) Classify RoIs in fixation area
   ii) Locally detected: moto
   iii) Update class-specific history
   iv) Fixate action reward (training):
      \[
      \hat{r}_t' = -\beta \sum_i (g_i/\text{IoU}) > \text{th} \geq 1 \frac{\text{vol}_{max}}{\text{vol}}
      \]

3) Action: Done - terminate search
   i) Posterior class-probability adjustments
   ii) Final detections: person, moto
   iii) Done action reward (training):
      \[
      \hat{r}_t' = \sum_i (g_i/\text{IoU})_{\text{max}} \geq 1 \frac{\text{vol}_{max}}{\text{vol}}
      \]

Experimental results
- Baseline: VGG-16 based Faster R-CNN with standard RPN
- drl-RPN more accurate on MS COCO, PASCAL VOC (1-2% mAP)
- drl-RPN not as fast, but possible to set speed-accuracy trade-off