# RATM: Recurrent Attentive Tracking Model

mingzailao

*<2016-10-31 Mon>*

# Outline

# RATM: Recurrent Attentive Tracking Model

# IRNN

## Structure

Basic RNN

## Initialization of $W_rec$

A scaled version of the identity matrix

## Hidden activation function

ReLU

## Initialization of $h_0$

Zero vector

# RATM: Recurrent Attentive Tracking Model

# Attention Mechanism

## DRAW

Links :

# RATM: Recurrent Attentive Tracking Model

# Feature-extraction module

## Aims

The feature-extraction module computes a feature representation of a given input glimpse.

## Pre-train is useful

After pre-training, the feature extractor' s parameters can either be continued to be updated during end-to-end training, or kept fixed.

# Attention Module

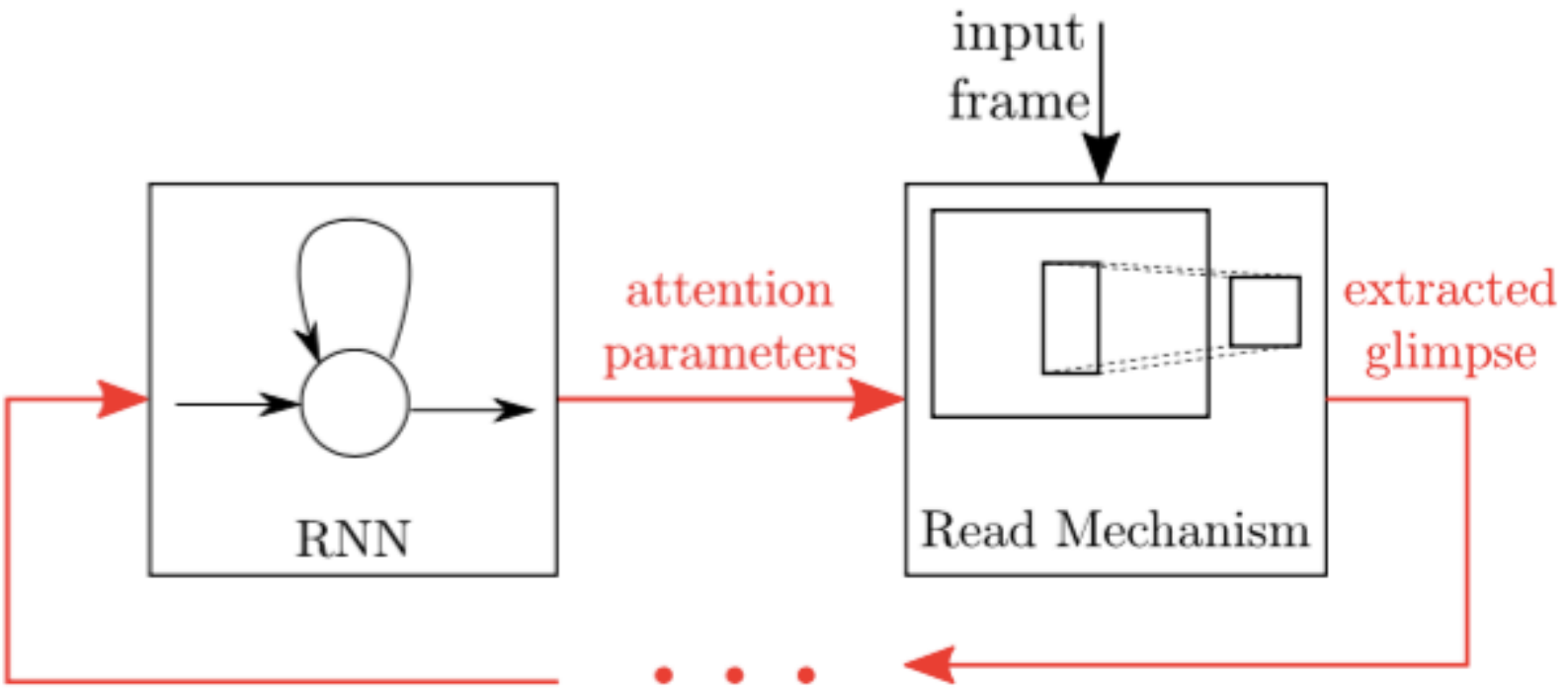


Fig. 4: The symbolic representation of a recurrent attention module, which is composed of an RNN and a read mechanism that extracts a glimpse from the input frame. The extracted glimpse is fed back to the RNN. The dots indicate, that the feed-back connection can involve intermediate processing steps, such as feature extraction.

# Objective Module

## Aims

An objective module guides the model to learn an attentional policy to solve a given task.

## Output

- Cost : function of its target and prediction inputs.

$$\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i=1}^{n} ||\mathbf{y}_{target} - \mathbf{y}_{pred}||_2^2 \quad (1)$$

# RATM: Recurrent Attentive Tracking Model

# Architecture

## Penalty Term

1. $$\mathcal{L}_{pixel} = ||\hat{p} - p||_2^2 \tag{2}$$

   where $\hat{p}$ : the glimpse extracted by the attention mechanism $p$ : ground truth image

2. $$\mathcal{L}_{feat} = ||f(\hat{p}) - f(p)||_2^2 \tag{3}$$

   where $f$ : feature extraction function

3. $$\mathcal{L}_{loc} = ||\hat{g} - g||_2^2 \tag{4}$$

   where $g$ : center of the ground truth.

# Evaluation of Tracking Performance

## IoU

$$IoU = \frac{B_{gt} \cap B_{pred}}{B_{gt} \cup B_{pred}} \tag{5}$$

# RATM: Recurrent Attentive Tracking Model