Iterative Visual Reasoning Beyond Convolutions

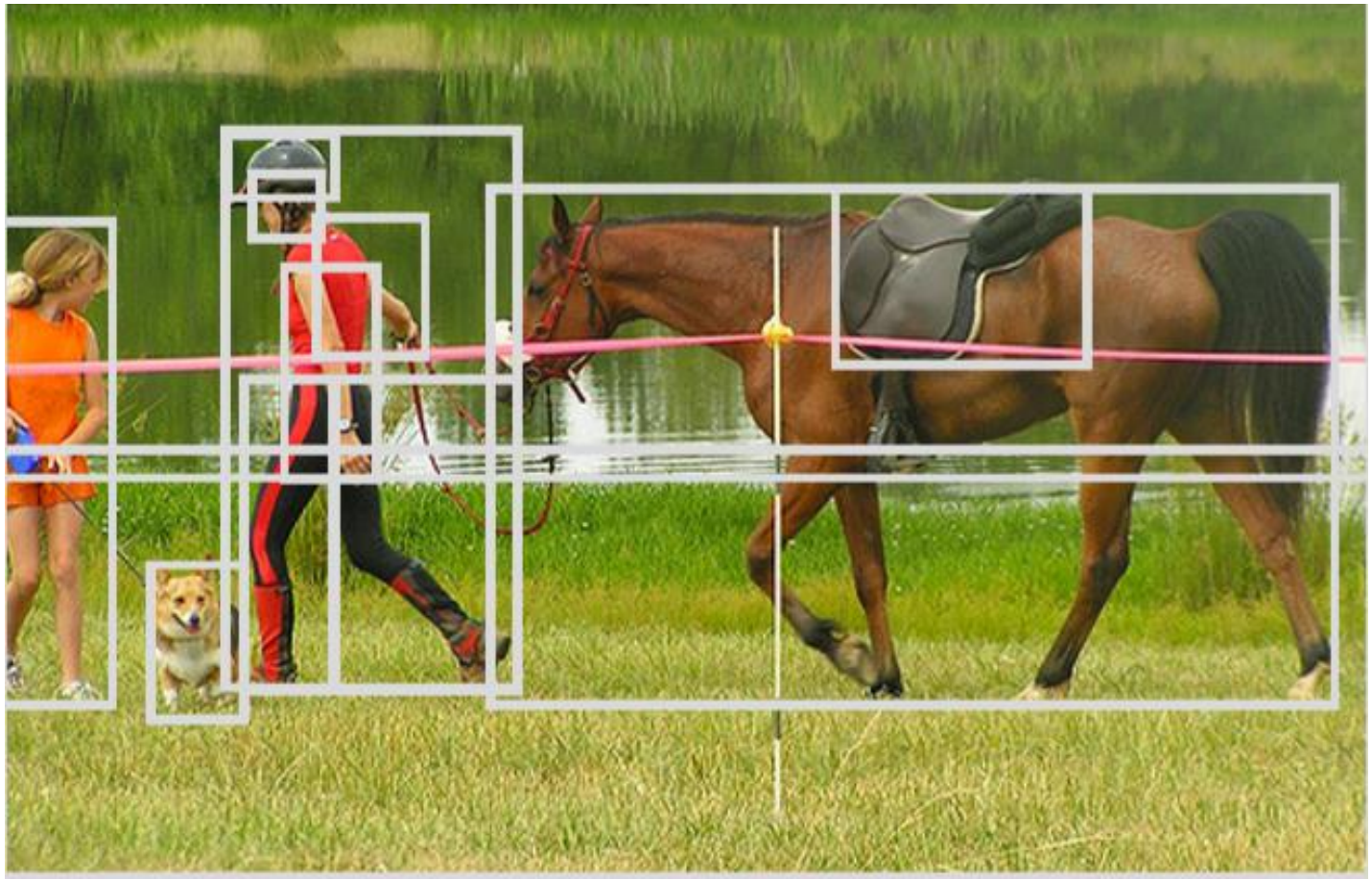
Xinlei Chen, Li-Jia Li, Li Fei-Fei and Abhinav Gupta.

CVPR 2018 (https://arxiv.org/abs/1803.11189)

Jung Hun Oh

Seoul National University

Motivation



Spatial relationship

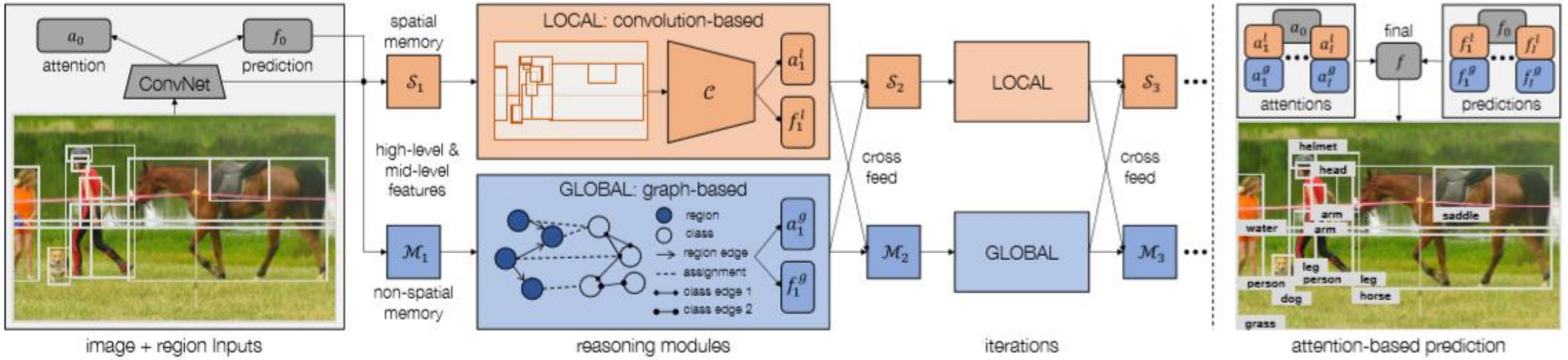
+

Semantic relationship

↓

Global level reasoning

Model



- Local module: performs pixel-level reasoning using ConvNet.
- Global module: performs global-level reasoning using graphs

Model

- **Graphs in Global module**

- Region to Region graphs: Spatial relationship

(e.g. left-right, top-down …)

- Region to Class (class to Region): Assignment

- Class to Class: Semantic relationship

(e.g. is-kind-of (cake-food…), is-part-of (wheel-car, nose-face…) …)

Model

- **Feature embeding**

- Spatial path: processing spatial features about all regions(M_r)

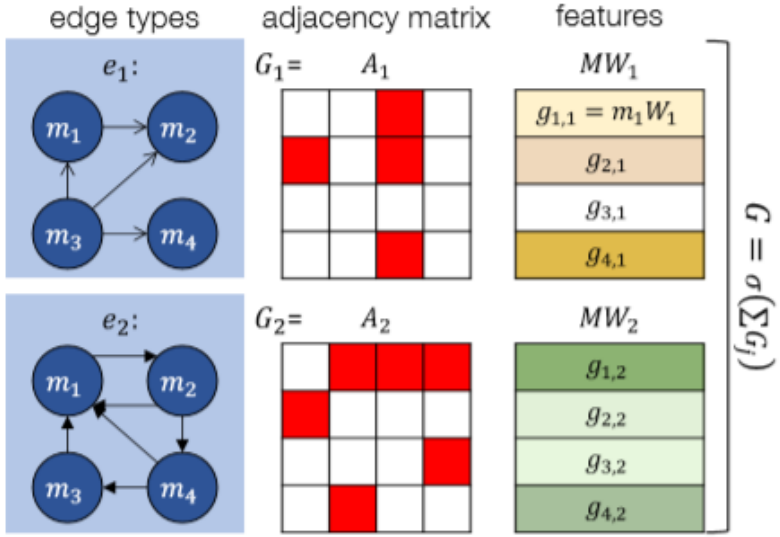
$$G_r^{spatial} = \sum_{e \in \mathcal{E}_{r \to r}} A_e M_r W_e,$$

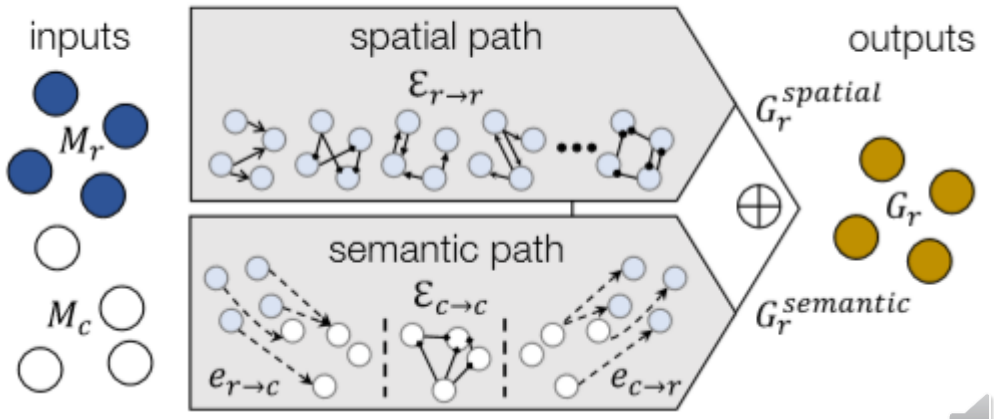
- Semantic path: processing semantic features(M_c)

$$G_c^{semantic} = \sum_{e \in \mathcal{E}_{c \to c}} A_e \sigma(A_{e_{r \to c}} M_r W_{e_{r \to c}} + M_c W_c) W_e,$$

- Final output

$$G_r = \sigma(G_r^{spatial} + \sigma(A_{e_{c \to r}} G_c^{semantic} W_{e_{c \to r}})),$$





Results