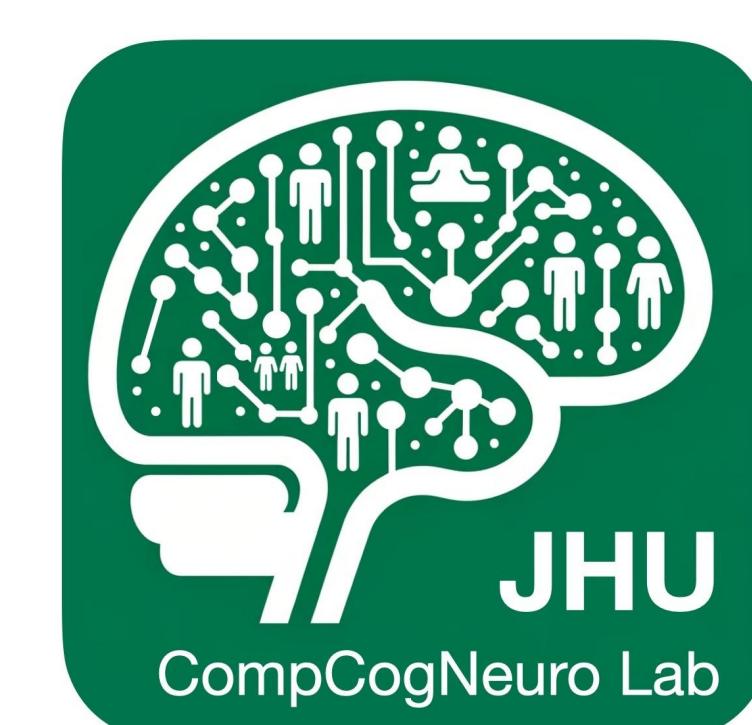




Relational Information Predicts Human Behavior and Neural Responses to Complex Social Scenes

Wenshuo Qin, Manasi Malik, Leyla Isik

Department of Cognitive Science at Johns Hopkins University



Introduction

Background: Relational information is critical for understanding social interactions and is represented in the human brain—particularly in the superior temporal sulcus (STS). However, most computational models overlook its importance.

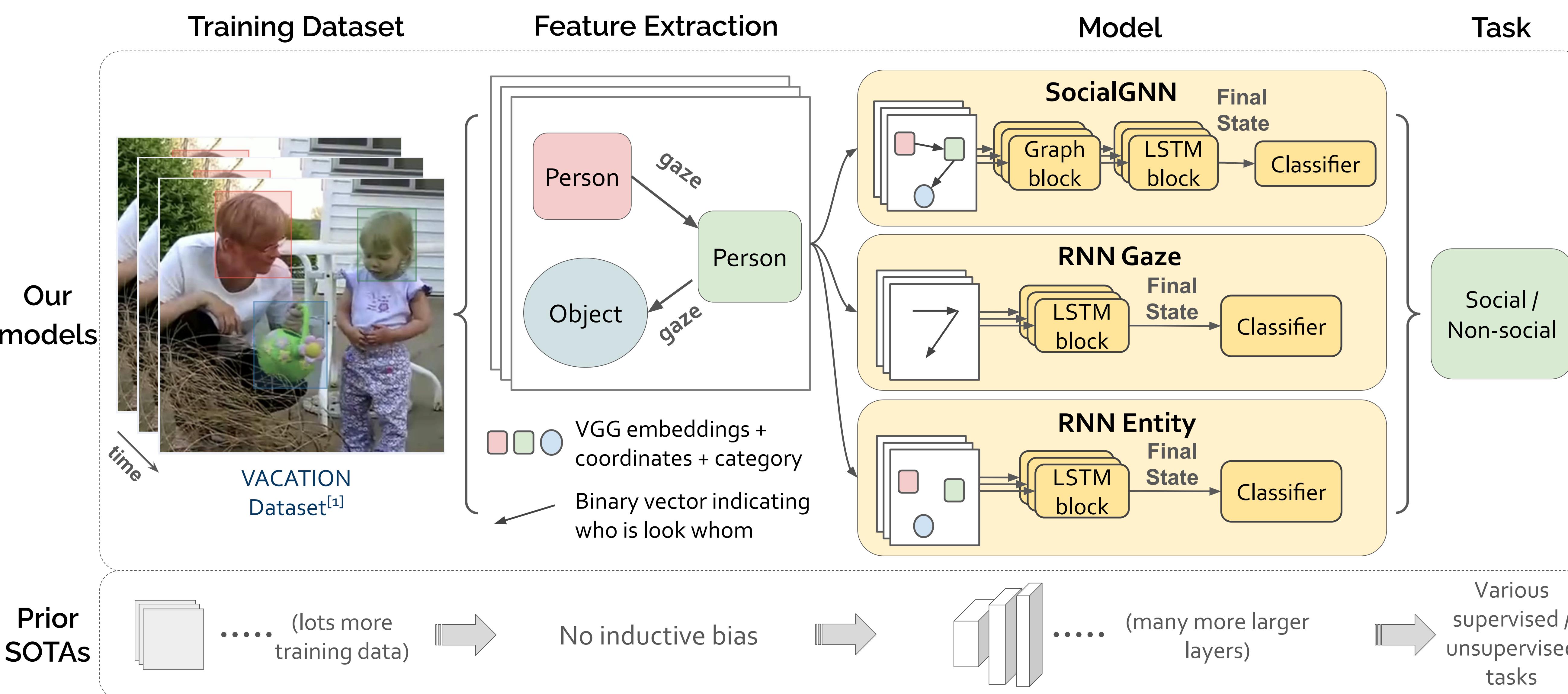
Goal: Compare relational models against a non-relational model and two prior SOTA models in predicting human behavioral and neural responses to social scenes.^[2]

Models:

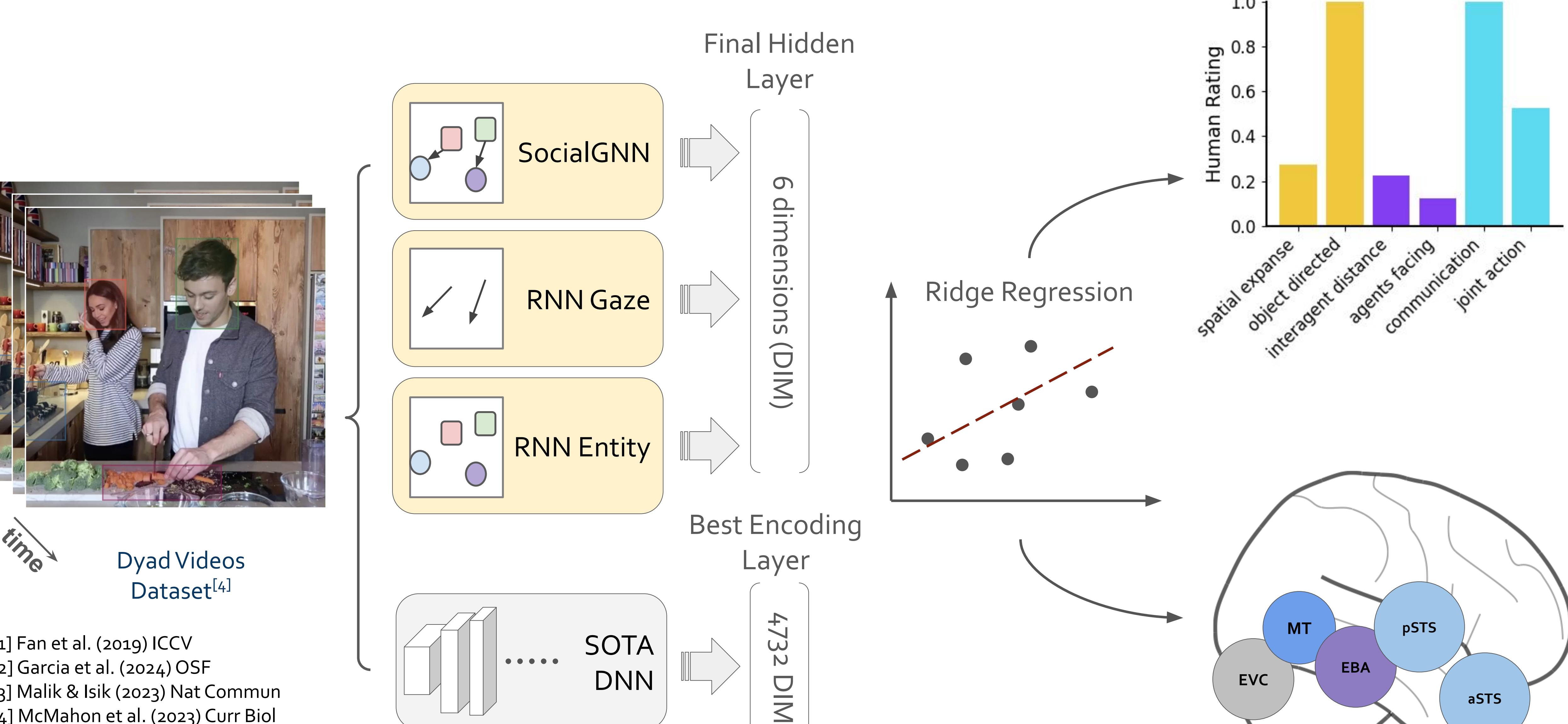
- Relational models: SocialGNN^[3] and RNN Gaze
- Non-relational model: RNN Entity
- 350 SOTA image models (CNNs and ViT) from prior benchmarking study^{[2][6]}

Method

TRAINING

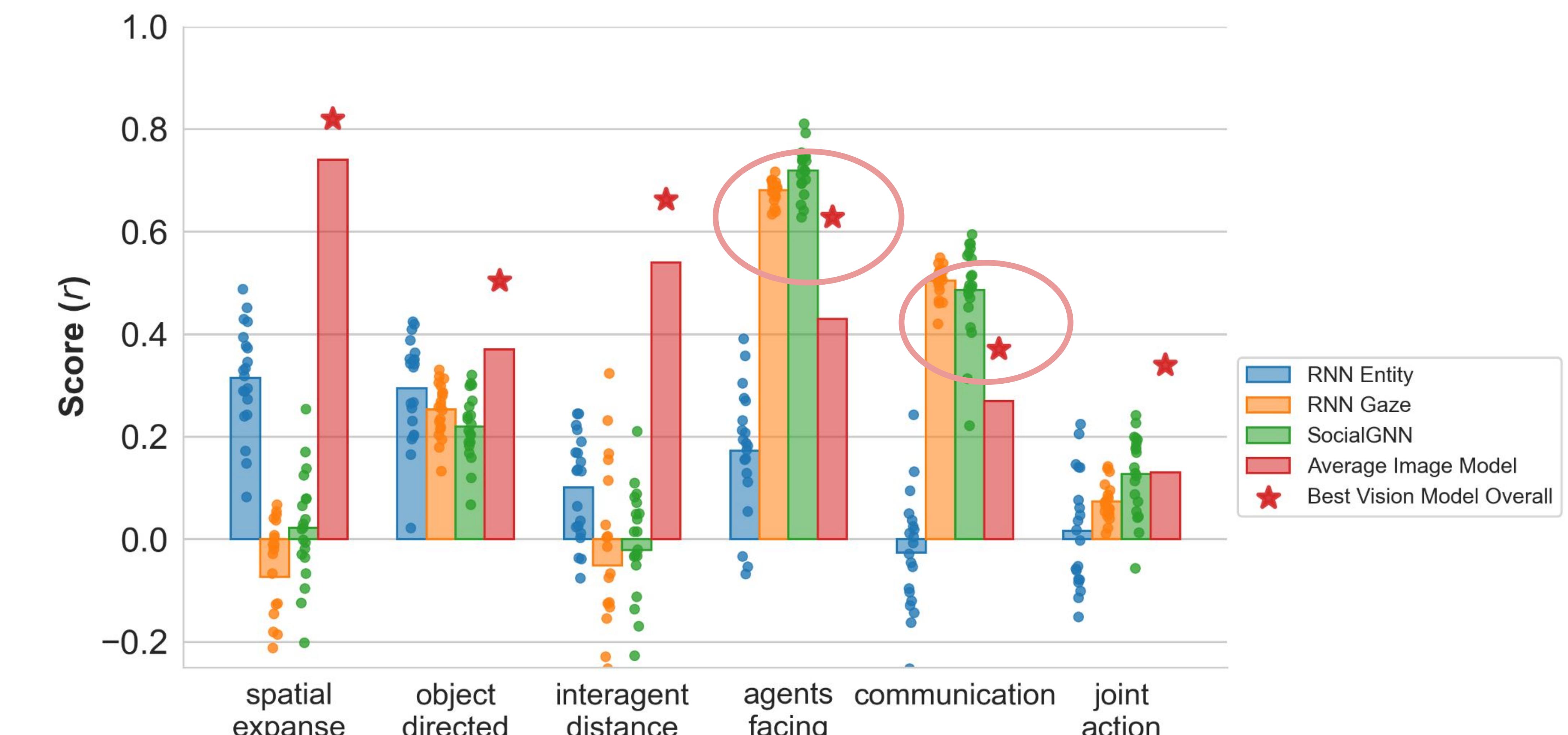


TESTING



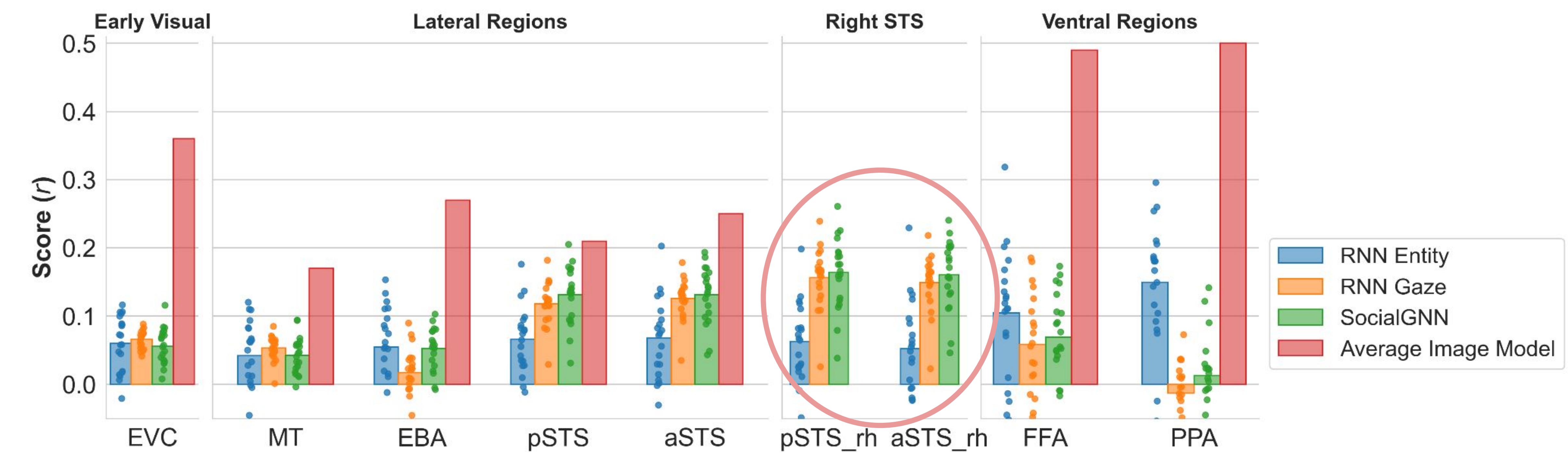
Results

Natural Videos Behavioral Encoding



- Relational models were good at identifying agent facing or communicating and did even better than the best SOTA AI vision model (DeiT)^{[2][6]}.
- Non-relational RNN Entity was better at capturing spatial and object-centric features.

Natural Videos Neural Encoding



- Relational models were equally better than RNN Entity at predicting the STS responses of people watching the social interaction clips^[4].
- RNN Entity was better at predicting the ventral regions activities

Discussion

TAKEAWAY: Simple neural networks trained on **relational cues** such as gaze outperform far more complex state-of-the-art vision models in predicting human judgments of **social interaction**, and also better account for neural responses in the **superior temporal sulcus (STS)**, a region implicated in social perception. These findings highlight the critical role of relational information and suggest future work should disentangle the contributions of relational and non-relational cues in shaping social perception.