Contextual Understanding of Visual Interactions

Mert Kilickaya

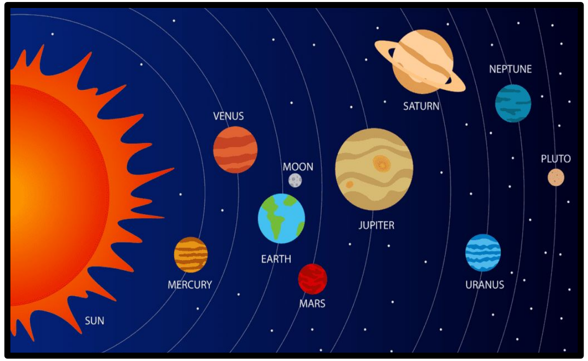
Supervisor: Prof. Arnold Smeulders,

Co-supervisor: Prof. Cees Snoek,

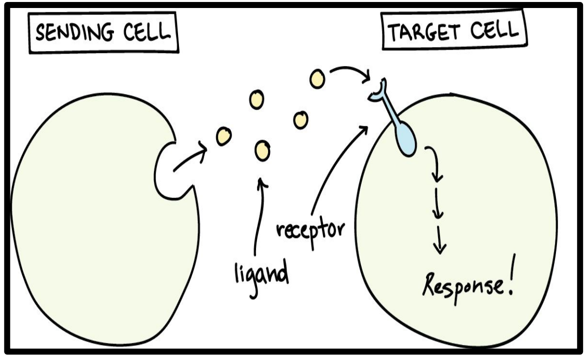
Collaborator: Assoc. Prof. Efstratios Gavves

Interactions are Fundamental to Life

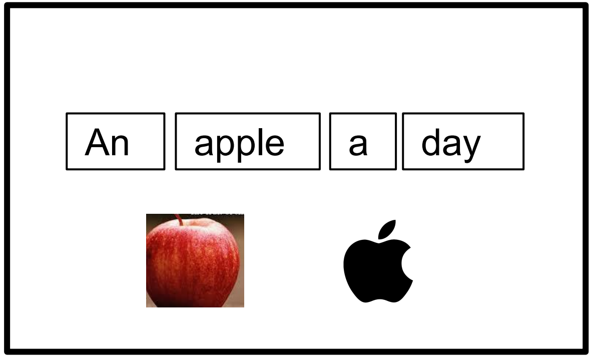
Gravitational Interactions



Biological Interactions



Lingual Interactions

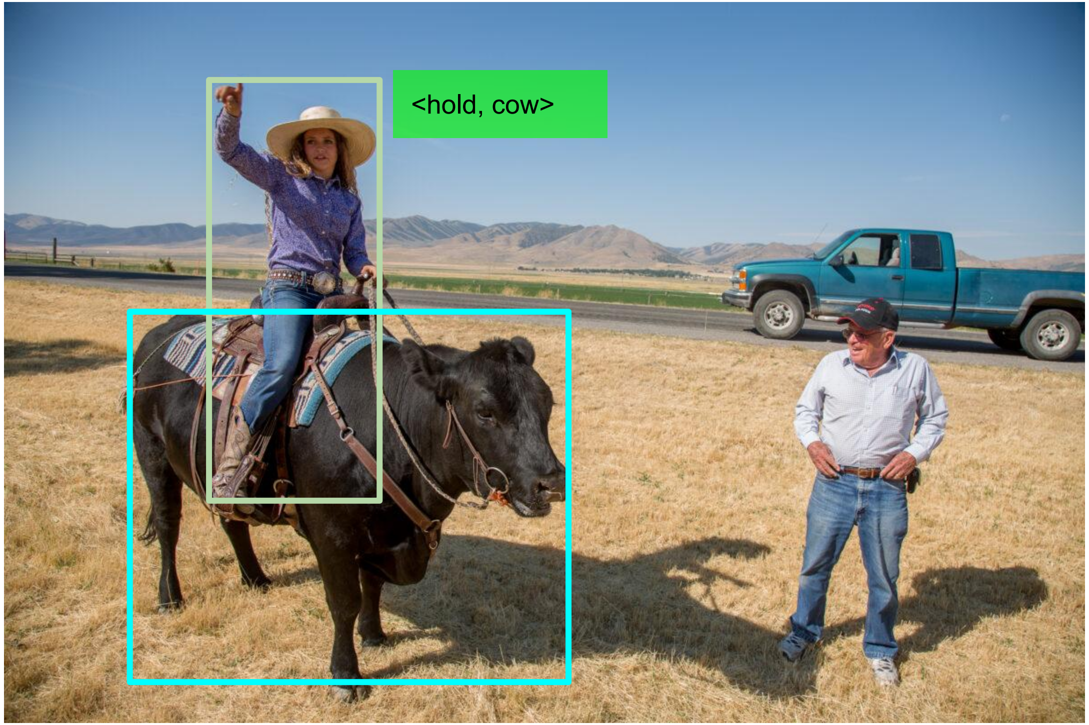


An apple a day

Visual Interactions

This thesis focuses on understanding visual interactions.

Understanding Visual Interactions: What



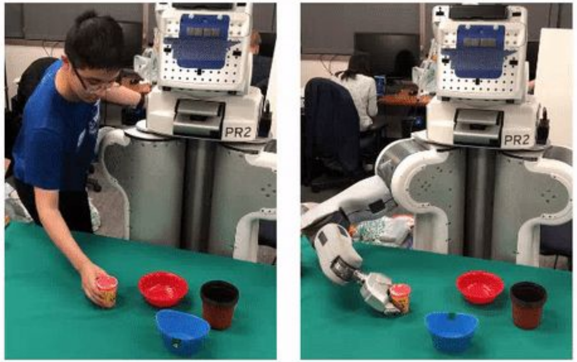
Understanding visual interactions entails: 1) Detecting human-objects, 2) Recognizing interactions

Understanding Visual Interactions: Why

“Cyclist Detection for Self-Driving”

“Affective Computing”

“Learning via Visual Imitation”



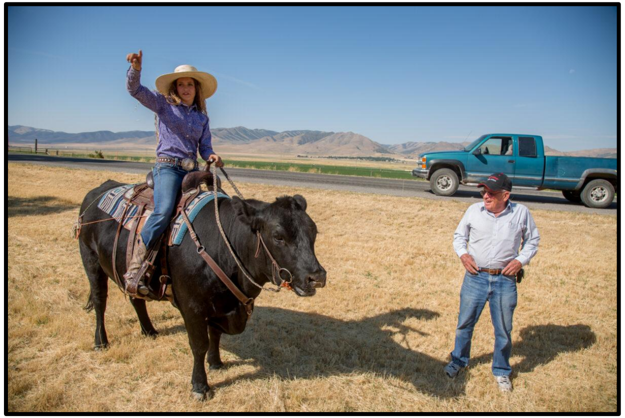
Understanding visual interactions is necessary to enable human-like abilities.

Understanding Visual Interactions: How

No-Context

- ✓ Scene Context (i.e. rural)
- ✓ Object Context (i.e. interactor, man and the car)

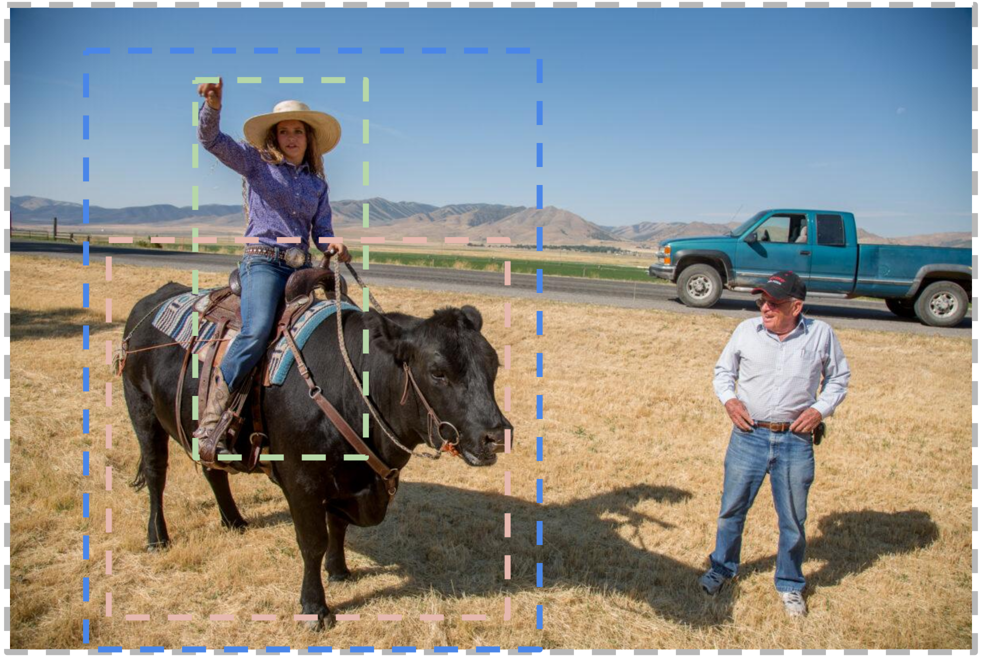
In-Context



- ✓ Spatial Context (i.e. on top)
- ✓ Appearance Context (i.e. pose, occlusion)

Visual context provides a multitude of information to understand interactions in the absence of time.

Contribution 1: The Context of Visual Interactions



Around human?

Around object?

Around human-object?

Everywhere?

Take-away: Interaction is everywhere, with a higher emphasis around the human-objects.

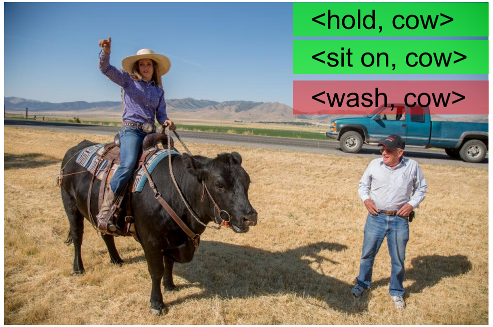
Contribution 2: Local Understanding of Visual Interactions

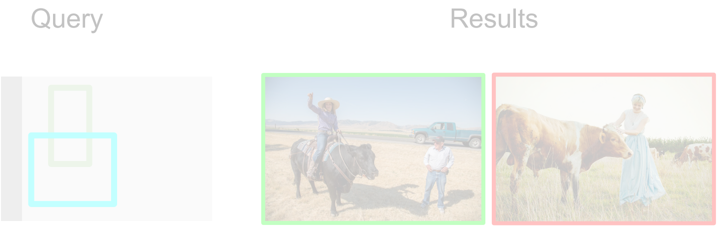
Interaction Recognition

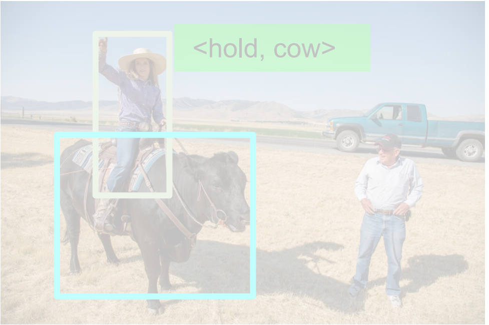
Interaction Search

Interaction Detection

What





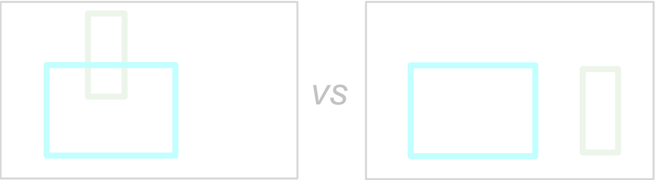


“Locality”

“Compositionality”

“Interactivity”

How



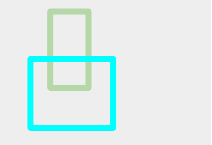
Take-away: Local context such as pose and deformation are useful signals for interaction recognition.

Contribution 3: Compositional Understanding of Visual Interactions

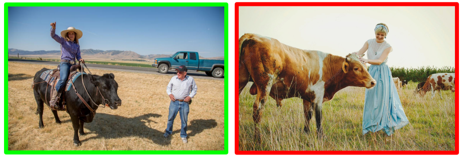
Interaction Recognition | Interaction Search | Interaction Detection

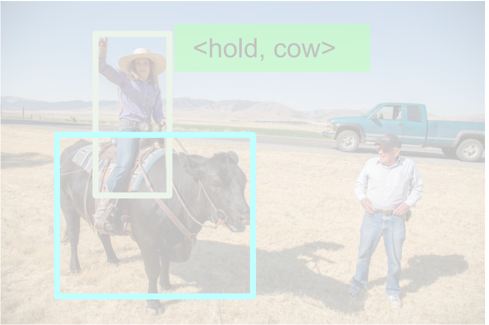
What

Query



Results



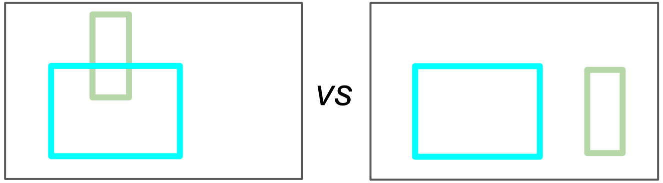


"Locality"

"Compositionality"

"Interactivity"

How



Take-away: Spatial context is useful in searching for visual interactions over large databases.

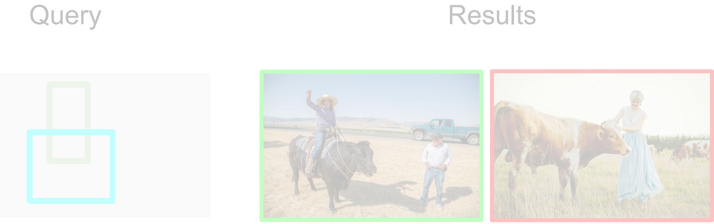
Contribution 4: Interactivity Understanding of Visual Interactions

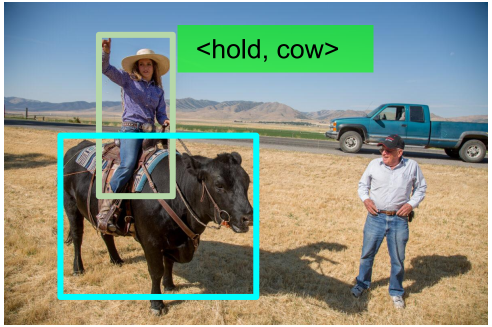
Interaction Recognition

Interaction Search

Interaction Detection

What



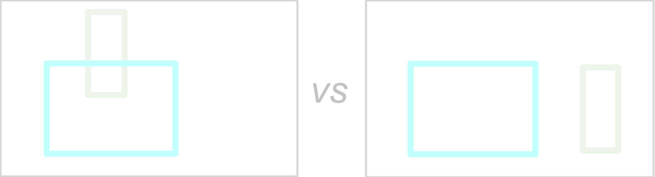


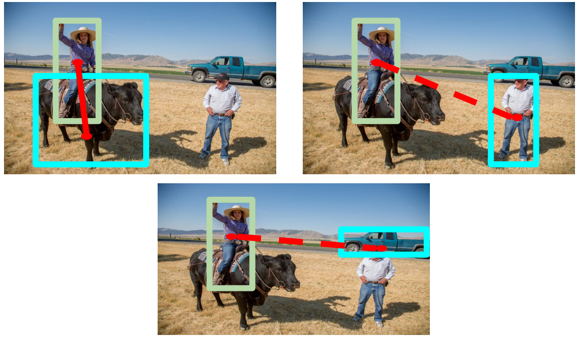
"Locality"

"Compositionality"

"Interactivity"

How





Take-away: Sparse interactivity context is crucial in finding the real human-object interactors.

Conclusion

C1: The thesis proposes a contextual understanding of visual human-object interactions.

C2: Interaction is the *absence* of isolation, within the context of others.

C3: The context of visual interactions is in *detail*: Locality, Compositionality, Interactivity.

C4: Representing visual details can help us to distinguish across interactions and interactors.

Thanks a lot for listening!

References

Kilickaya, M., & Gavves, E., Smeulders, A. “*Where is the Interaction? An Empirical Study*”. ArXiv, 2019.

Kilickaya, M., Hussein, N., Gavves, E., & Smeulders, A. “*Self-selective context for interaction recognition*”. ICPR, 2020.

Kilickaya, M., & Smeulders, A. “*Structured Visual Search Via Composition-Aware Learning*”. WACV, 2021.

Kilickaya, M., & Smeulders, A. “*Human-Object Interaction Detection via Weak Supervision*”. BMVC, 2021.