Human Perception of Navigational Affordances in Real-World Scenes
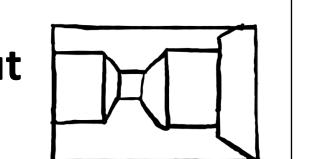
Clemens G. Bartnik, Iris I.A. Groen

Video & Image Sense Lab, Informatics Institute, University of Amsterdam, The Netherlands

UNIVERSITY OF AMSTERDAM

Aims

We move with ease through our environment. We can choose different paths and use a variety of different navigational actions, such as walking, swimming or climbing.

How do we represent navigational affordances of our immediate environment?

Past research focusing on indoor environments suggests that **scene layout** is an important determinant [1].
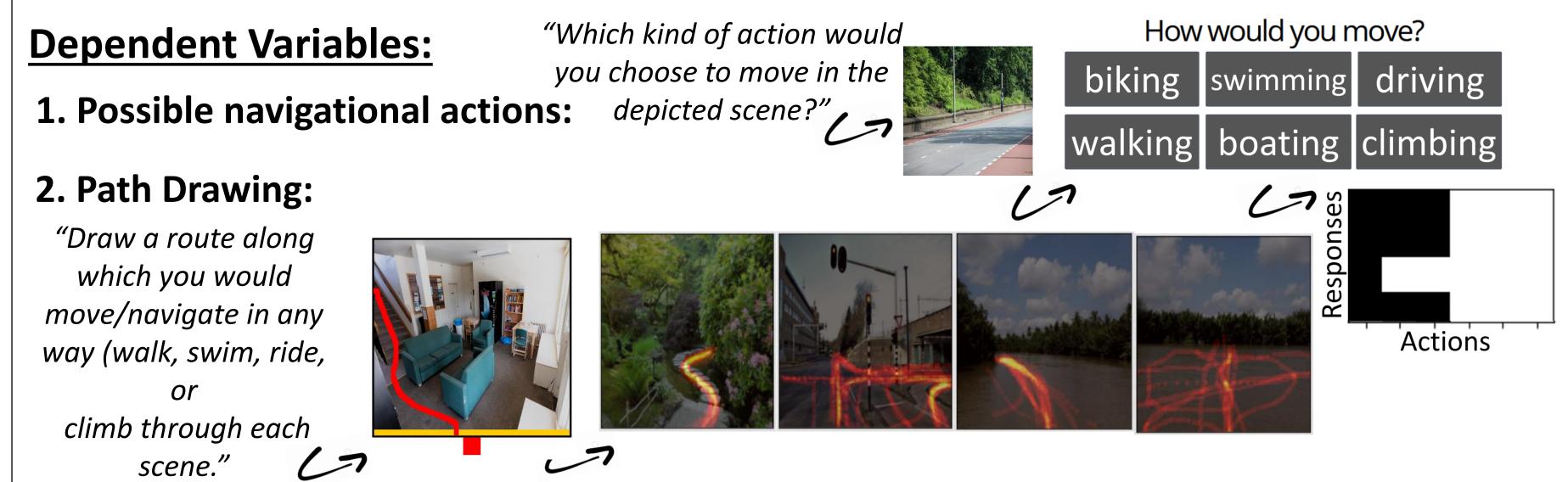
Here, we investigated which scene properties predict a larger range of navigational affordances in both indoor and outdoor environments.

Stimuli flickr

231 naturalistic real-world stimuli

(Number of images: indoor = outdoor natural = outdoor manmade)

t-SNE visualization of our stimuli covering the attribute space of large-scale dataset [2] (12k + Images) compared to previous research [1, 3].

Behavior

We collected human annotations (**N = 152**) for 7 tasks: 3 tasks involved navigation (DVs) and 4 tasks relied on readily identifiable qualitative scene properties (IVs)

Dependent Variables:

1. Possible navigational actions:

"Which kind of action would you choose to move in the depicted scene?"

How would you move?
biking | swimming | driving
walking | boating | climbing

2. Path Drawing:

"Draw a route along which you would move/navigate in any way (walk, swim, ride, or climb through each scene."

Responses: Actions

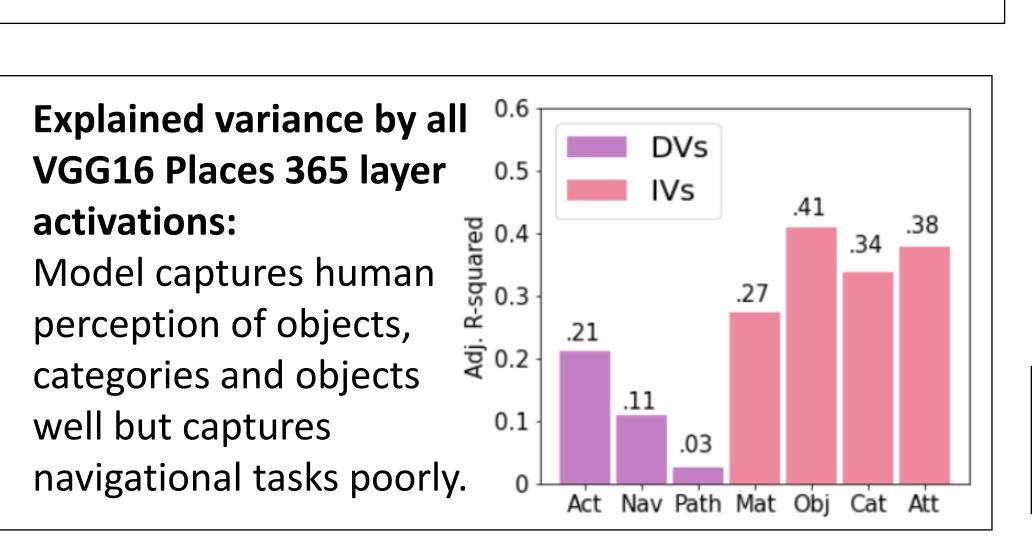**3. Navigability:** *"How easy would it be to move through the scene?"* Navigable ⟷ Not navigable
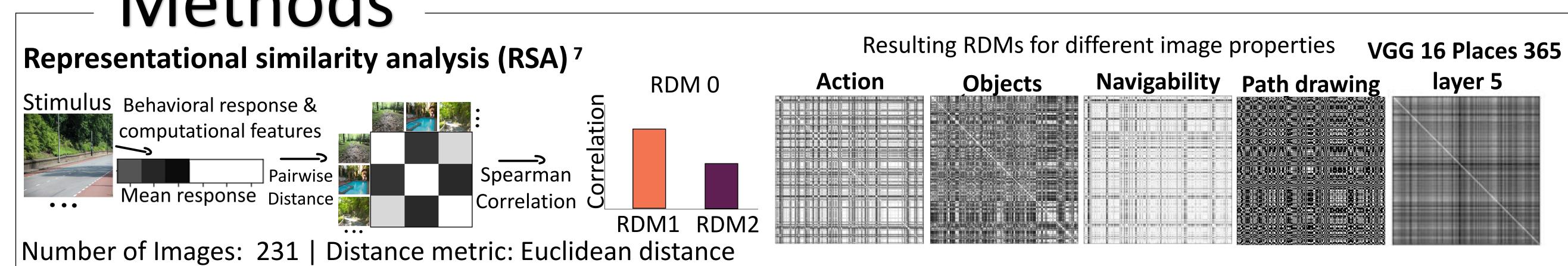
Independent Variables:

1. Material Properties: e.g.: water, wood (not tree), vegetation, dirt/soil, sand, stone/concrete [2]

2. Contained Objects: building/wall, tree/plant, road/street, furniture, body of water, rocks/stones [4]

3. Scene Category: room, inside city, hallway, forest, desert, open country, mountain, coast/river [5]

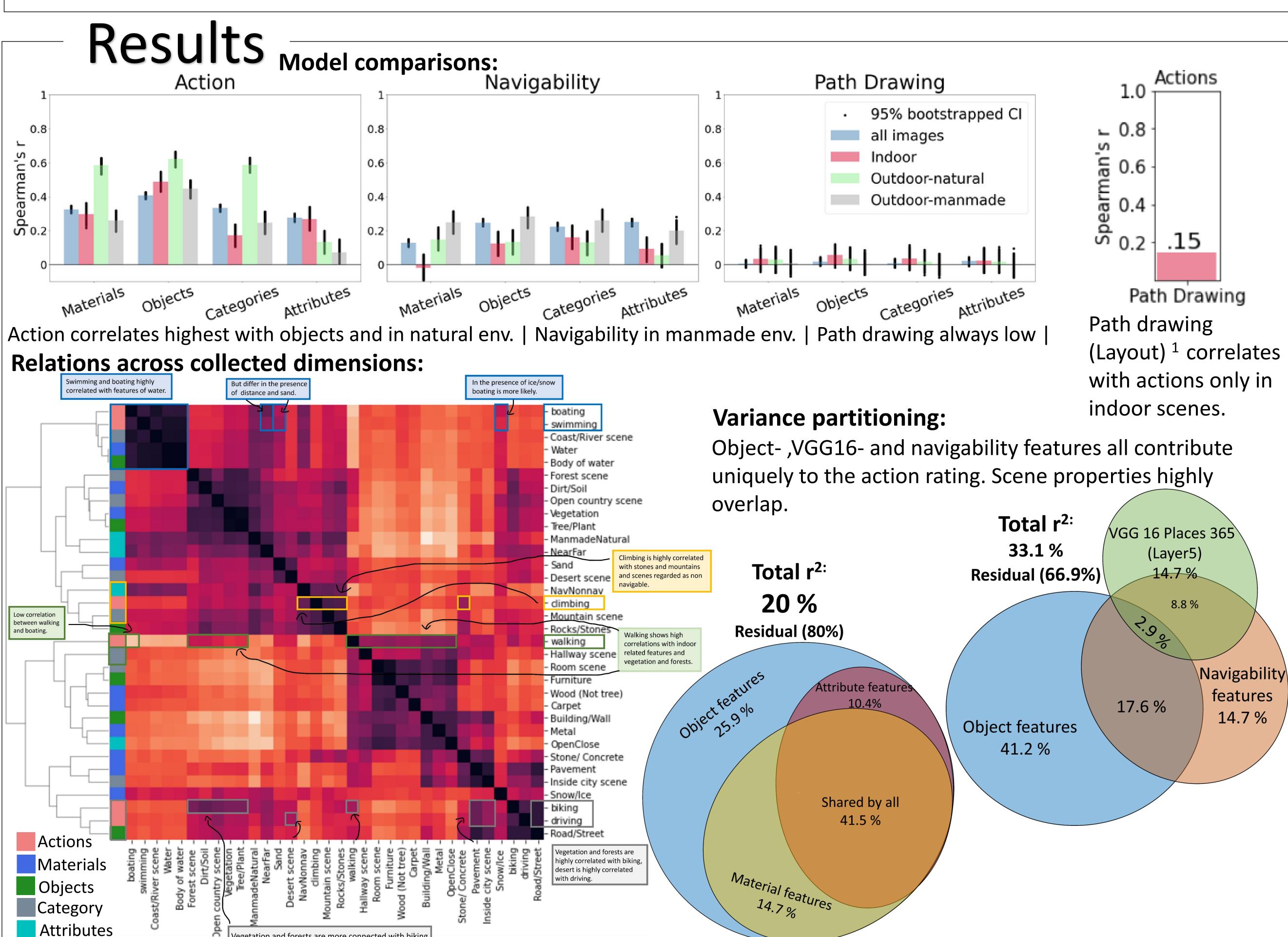**4. Scene Attributes:** open/close; manmade/natural; near/far [3]

Computation

For 231 stimuli, high-level visual features were collected via **conv layer 5 activations of VGG16 Places 365** [6]

Explained variance by all VGG16 Places 365 layer activations:

Model captures human perception of objects, categories and objects well but captures navigational tasks poorly.

Methods

Representational similarity analysis (RSA) [7]

Number of Images: 231 | Distance metric: Euclidean distance

Resulting RDMs for different image properties

Action | Objects | Navigability | Path drawing | VGG 16 Places 365 layer 5

Results

Model comparisons:

Action correlates highest with objects and in natural env. | Navigability in manmade env. | Path drawing always low |

95% bootstrapped CI — all images — Indoor — Outdoor-natural — Outdoor-manmade

Path drawing (Layout) [1] correlates with actions only in indoor scenes.

Relations across collected dimensions:

Swimming and boating highly correlated with features of water.

But differ in the presence of distance and sand.

In the presence of ice/snow boating is more likely.

Climbing is highly correlated with stones and mountains and scenes regarded as non navigable.

Low correlation between walking and boating.

Walking shows high correlations with indoor related features and vegetation and forests.

Vegetation and forests are highly correlated with biking, desert is highly correlated with driving.

Vegetation and forests are more connected with biking.

Actions | Materials | Objects | Category | Attributes

Variance partitioning:

Object-, VGG16- and navigability features all contribute uniquely to the action rating. Scene properties highly overlap.

Total r²: 20 % Residual (80%)

Object features 25.9 % | Attribute features 10.4% | Material features 14.7 % | Shared by all 41.5 %

Total r²: 33.1 % Residual (66.9%)

VGG 16 Places 365 (Layer5) 14.7 % | 8.8 % | 2.9 % | Object features 41.2 % | Navigability features 14.7 % | 17.6 %

Conclusions

- Navigational affordance are not adequately captured by readily identifiable scene properties. They only account for 20% and highly overlap.

- Layout captured by goal point localization drives indoor navigational affordances which has been shown by others [1] and was confirmed by us. But can not be generalized to all types of environments.

- Consequently, future models need to incorporate multiple relevant features that humans use to perceive the navigational affordances of natural scenes.

Total r²: 20 % Residual (80%)

Contact

✉ c.g.bartnik@uva.nl

@ Video & Image Sense Lab

cgbartnik

References

[1] Bonner & Epstein (2017), *PNAS* | [3] Kravitz et al. (2011), *J Neurosci* | [5] Oliva & Torralba (2001), *Int J Comput Vis* | [7] Kriegeskorte et al. (2008), *Front. Syst. Neurosci*

[2] Patterson & Hays (2012), CVPR | [4] Fei-Fei et al. (2007), *J Vis* | [6] Zhou et al. (2017), IEEE *Trans Pattern Anal Mach Intell* |