



universität
uulm

Funded by
DFG Deutsche
Forschungsgemeinschaft
German Research Foundation



Invariance properties of predicted eye movements in visual search models

Atif Khurshid*, Matthias Kohn*, Heiko Neumann

Institute for Neural Information Processing, Ulm University

*Equal contribution

Invariance in visual search

- **Visual search:** Looking for a target object among distractors
- Relies on saccadic eye movements in complex scenarios
- Fairly **robust to variations** in pose, scale and viewpoint
- Computational models can help uncover mechanisms behind this ability
- **Invariant Visual Search Network**
 - Biologically-inspired model of target-guided visual search
 - First-order approximation to predict human behavior
 - Claims invariance to changes in target object appearance



Image: Hollingworth et al. Eye Tracking in Visual Search Experiments. *Neuromethods*. (2020)

Examples of object transformations

Invariance in visual search

- **Visual search:** Looking for a target object among distractors
- Relies on saccadic eye movements in complex scenarios



▪ Fail

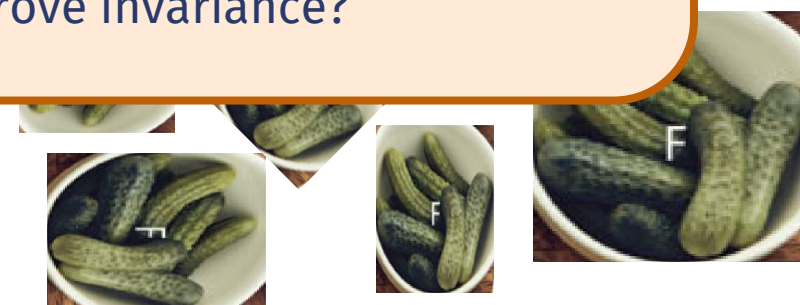
▪ Can be

▪ Inv

Research Questions

1. How invariant is the IVSN actually?
2. Where does invariance arise in the visual hierarchy?
3. Does a Gabor filterbank frontend ($\text{IVSN}_{\text{Gabor}}$) improve invariance?

- Biologically-inspired model of target-guided visual search
- First-order approximation to predict human behavior
- Claims invariance to changes in target object appearance

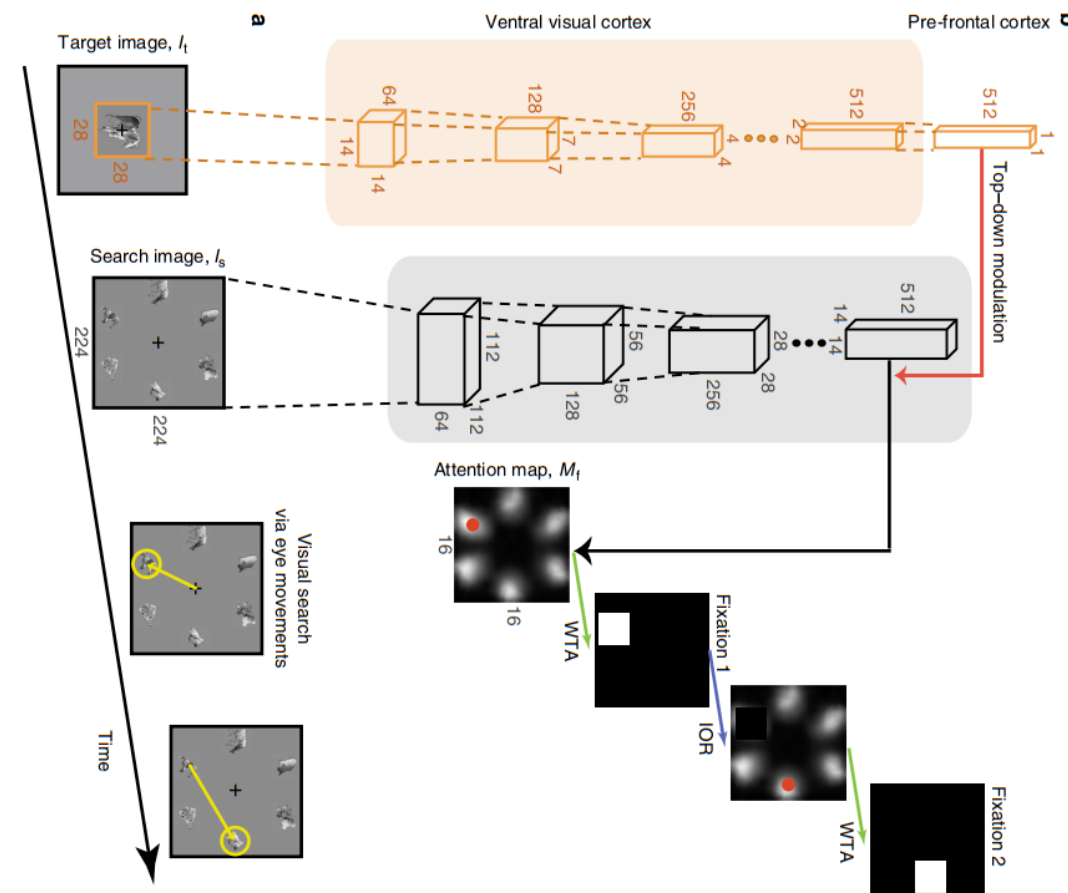


IVSN model architecture

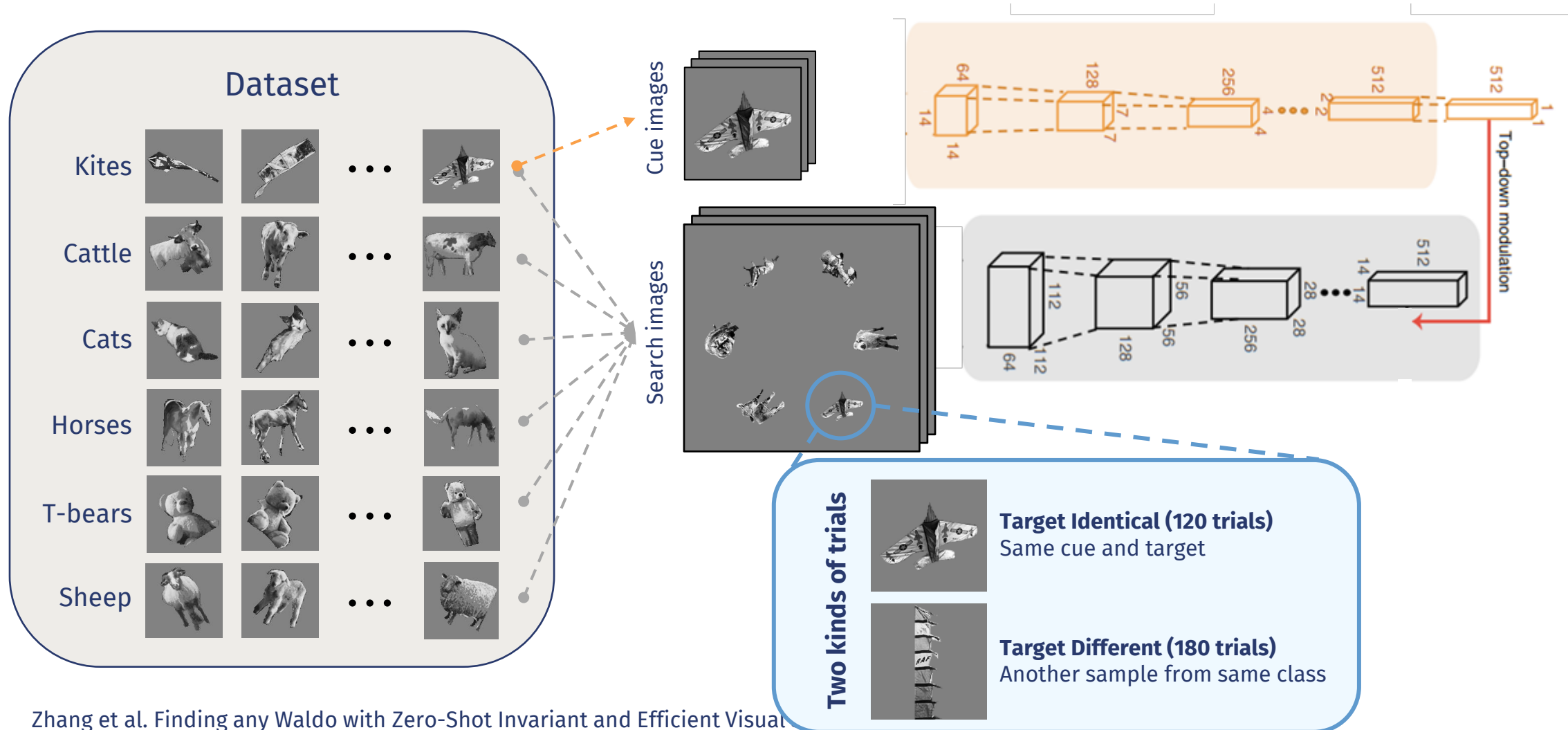
Core Component: ImageNet-trained CNN
approximating processing in ventral visual stream

Workflow:

1. Cue and search image processed by CNN
2. Top-layer cue features act as target template in PFC
3. Cue features convolved with search image features to generate priority map
4. Sequential selection of fixation points from priority map until target location fixated

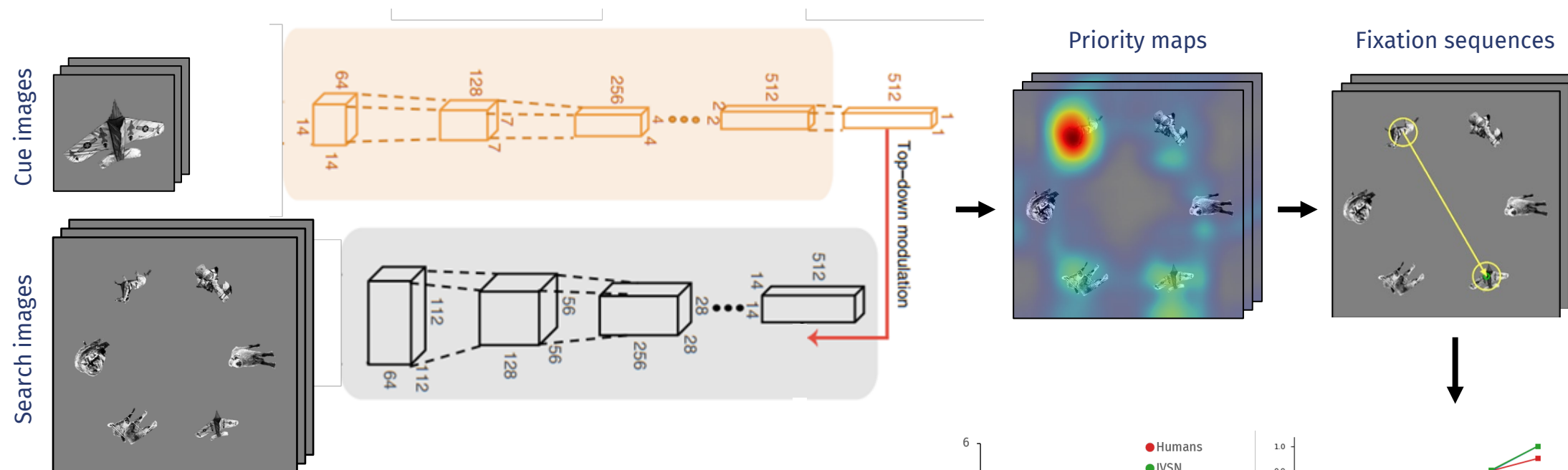


Original experimental setup of IVSN



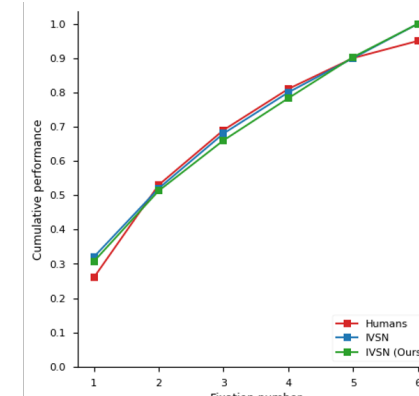
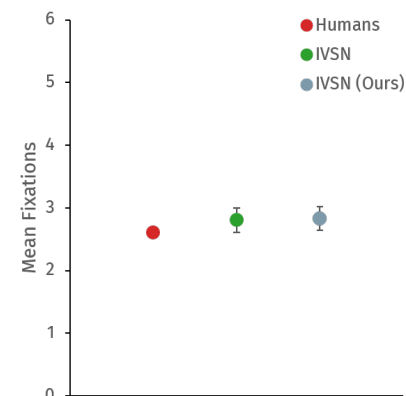
Zhang et al. Finding any Waldo with Zero-Shot Invariant and Efficient Visual

Original experimental setup of IVSN



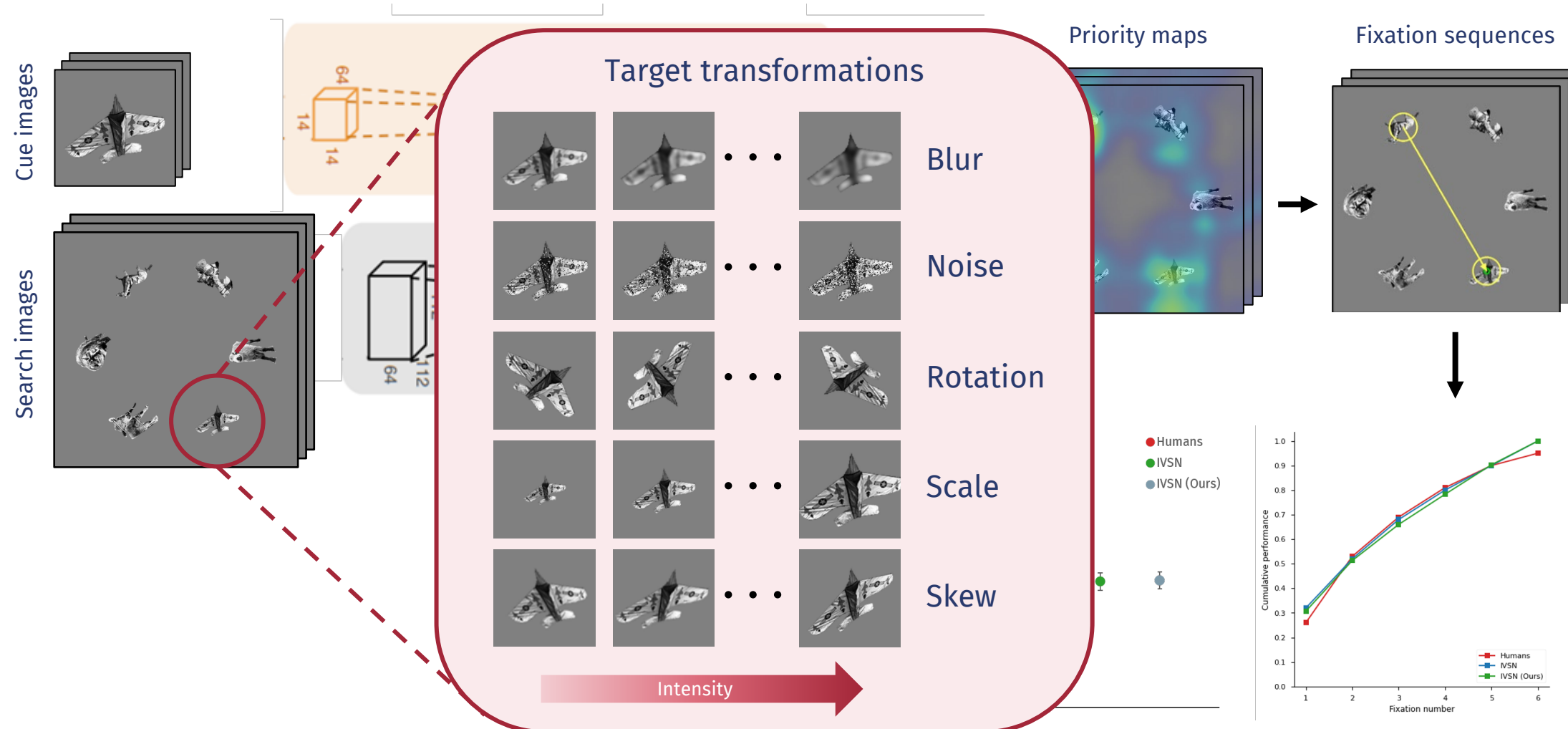
Evaluation Metrics

- ❖ Mean fixations across trials
- ❖ Found within n fixations (%)



Zhang et al. Finding any Waldo with Zero-Shot Invariant and Efficient Visual Search. *Nat. Commun.* (2018)

Our experimental setup for evaluating invariance



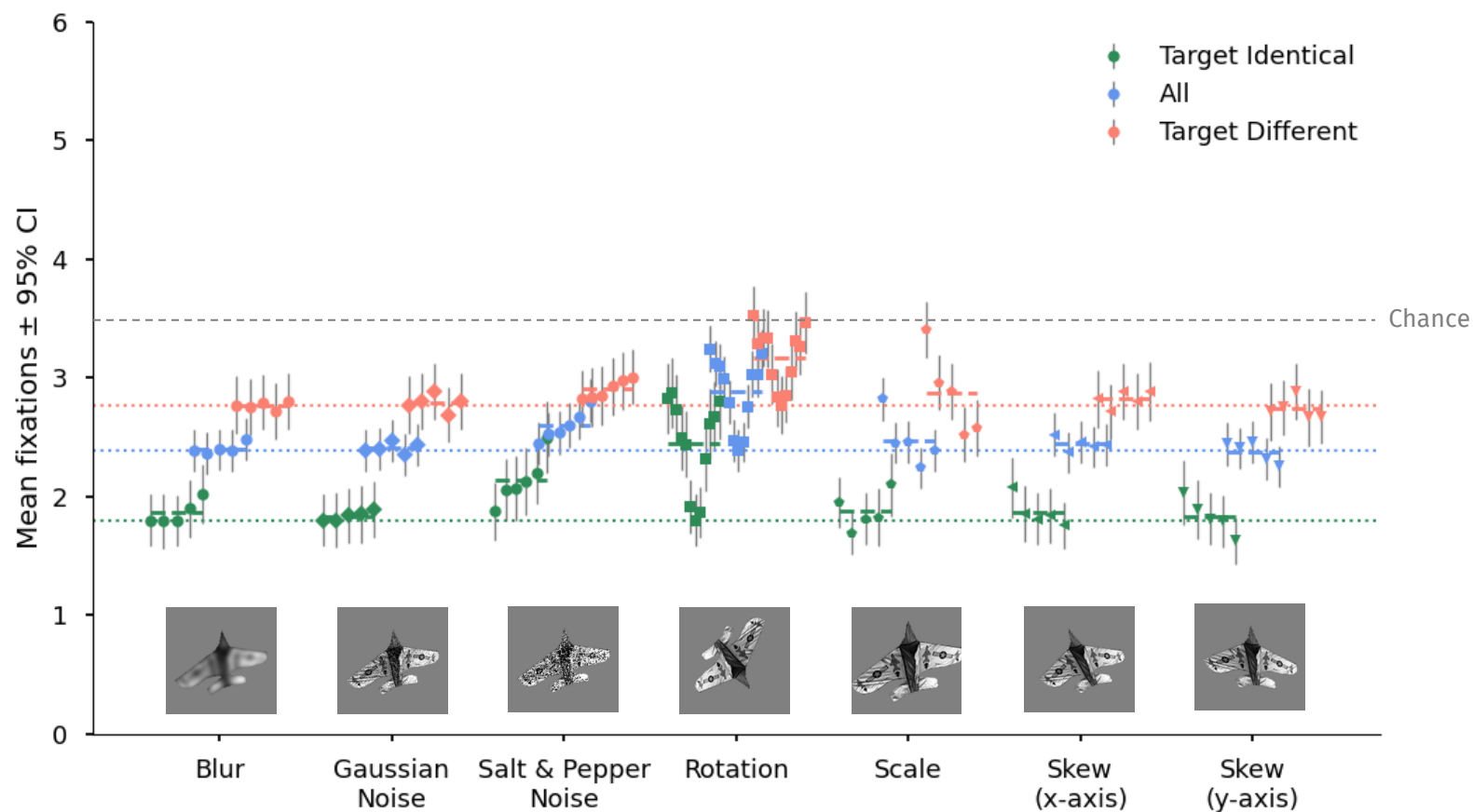
Zhang et al. Finding any Waldo with Zero-Shot Invariant and Efficient Visual Search. *Nat. Commun.* (2018)

How invariant is the IVSN actually?

Analyzing number of mean fixations to find target across different target object transformations

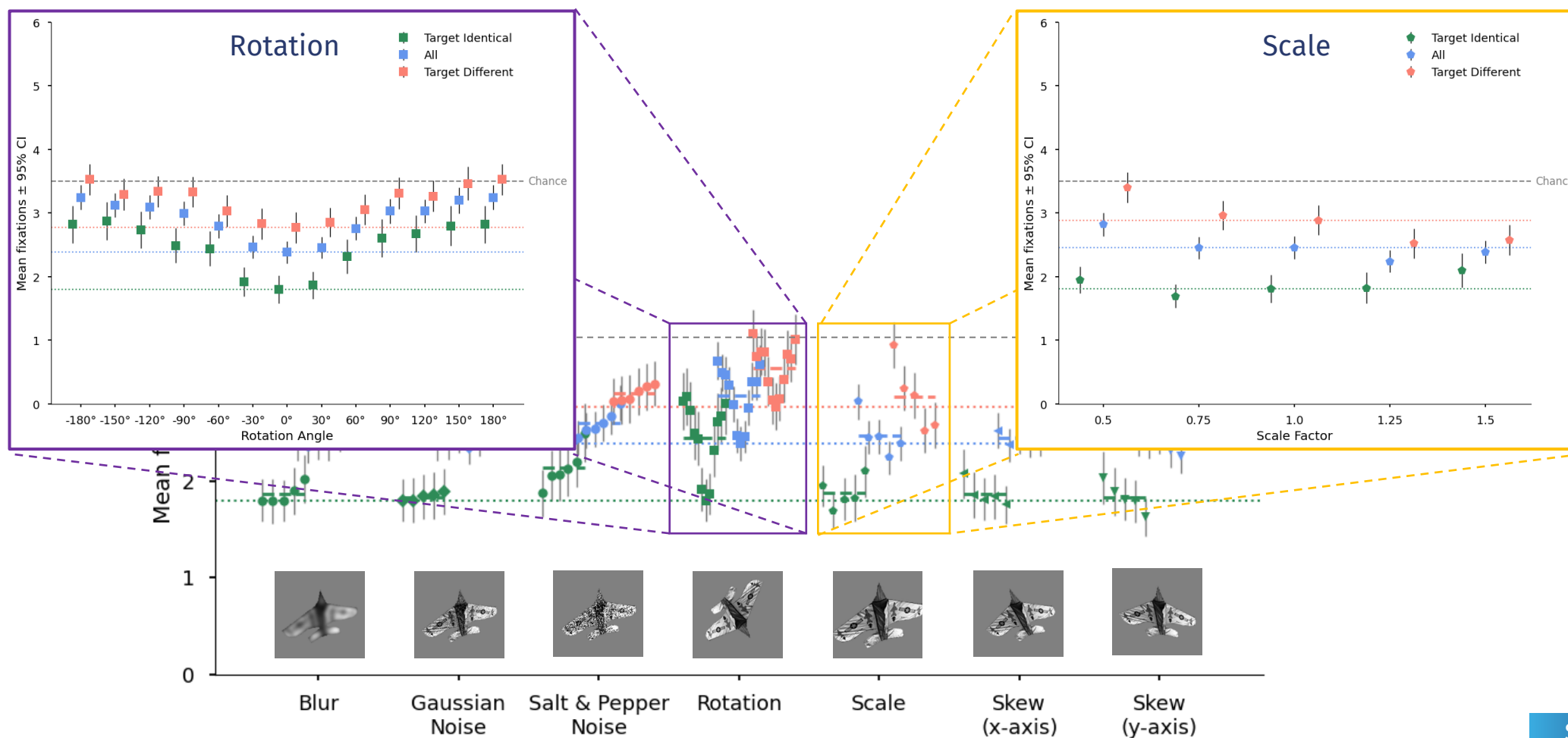
Evaluating invariance on 6-object circular array

- Aggregate results across all transformation types
- Each point represents mean of 300 trials at a single intensity level



Evaluating invariance on 6-object circular array

- Rotation and scale transformations have the strongest effect on performance



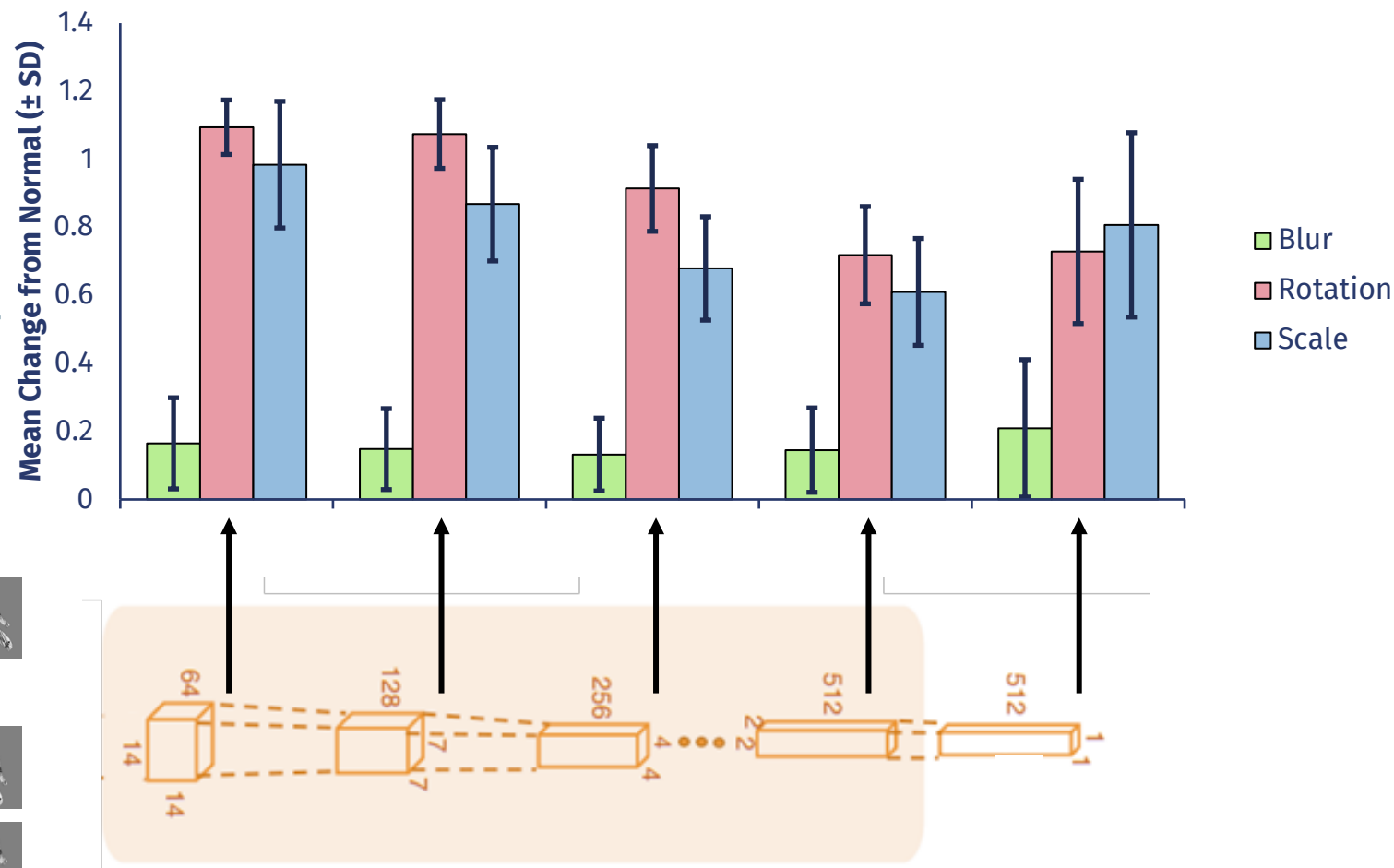
Where does invariance arise in the visual hierarchy?

Comparing internal representations of the model at different levels of processing hierarchy

Change in activations at different layers due to transformations

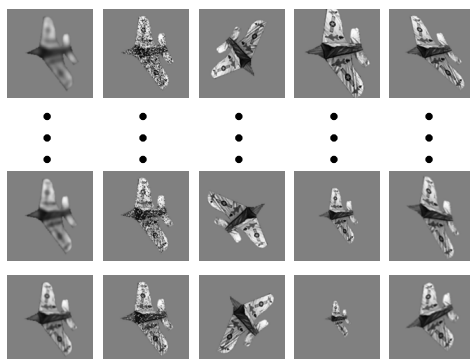
$$\frac{\|A_{orig} - A_{trans}\|_2}{\|A_{orig}\|_2}$$

Higher layer activations
are less affected by
transformations



Normal

vs



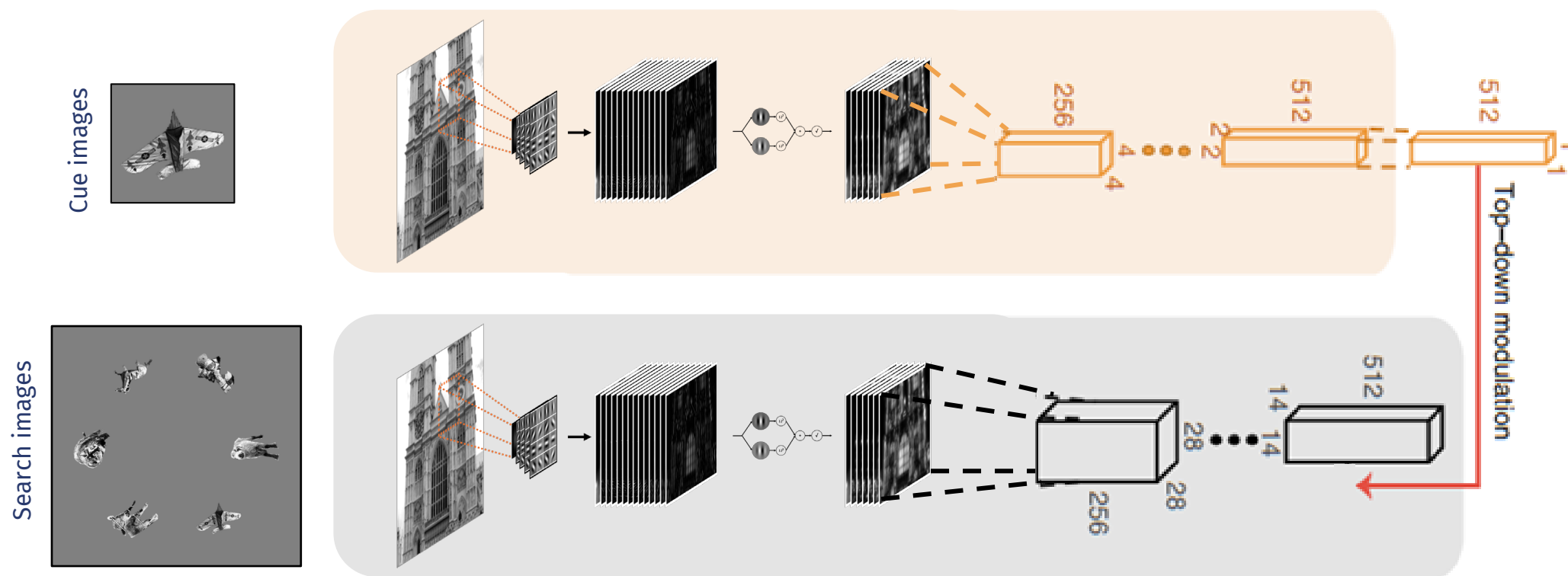
Transformed

Does a Gabor filterbank frontend improve invariance?

Comparing IVSN against a variant ($\text{IVSN}_{\text{Gabor}}$) with a Gabor filterbank frontend

IVSN_{Gabor} model architecture

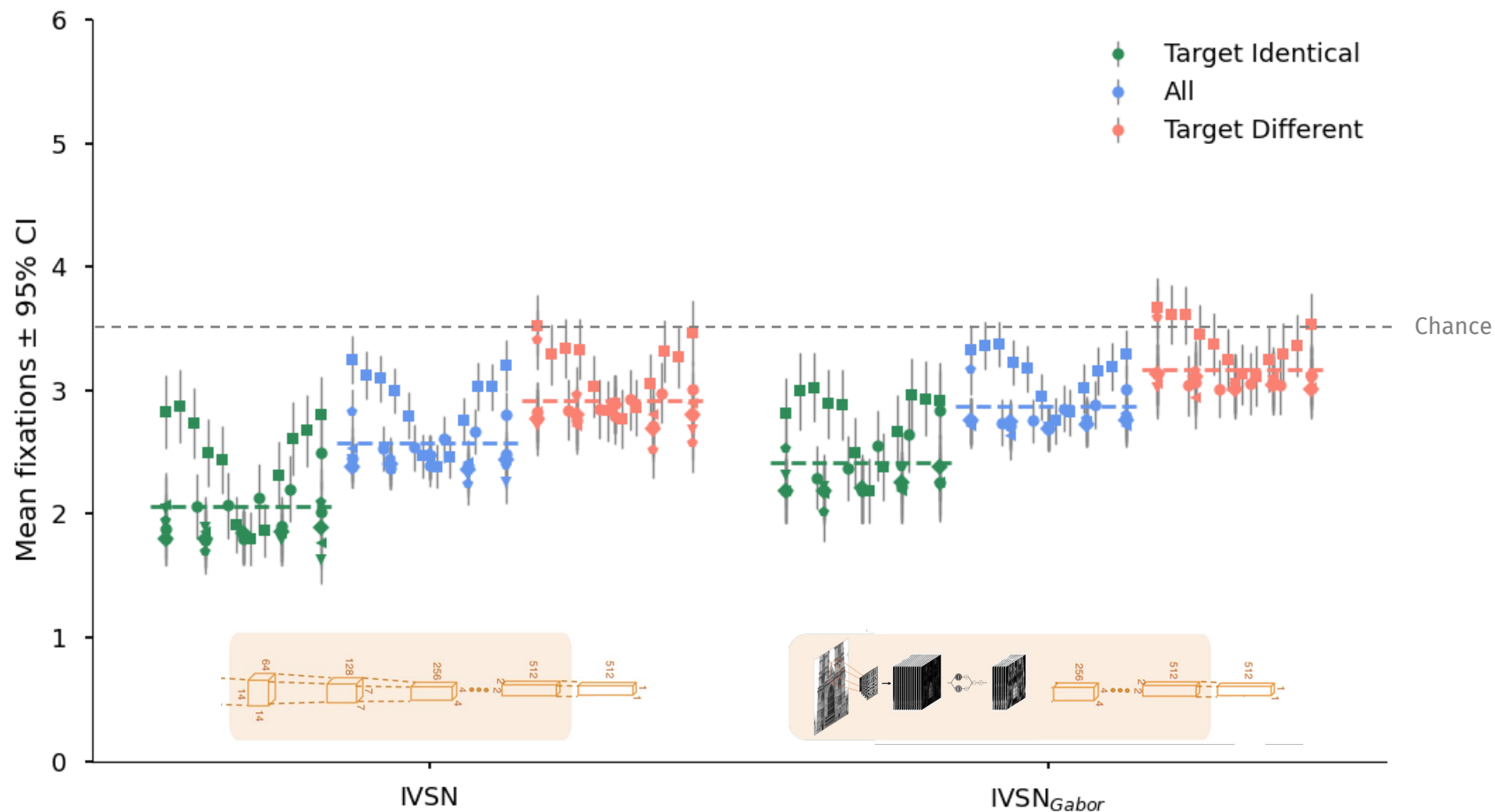
- First two blocks of CNN replaced by a Gabor filterbank – approximates V1 processing
- Known to improve robustness to noise and image corruptions¹



¹ Dapello et al. Simulating a Primary Visual Cortex at the Front of CNNs Improves Robustness to Image Perturbations. *NeurIPS*. (2020)

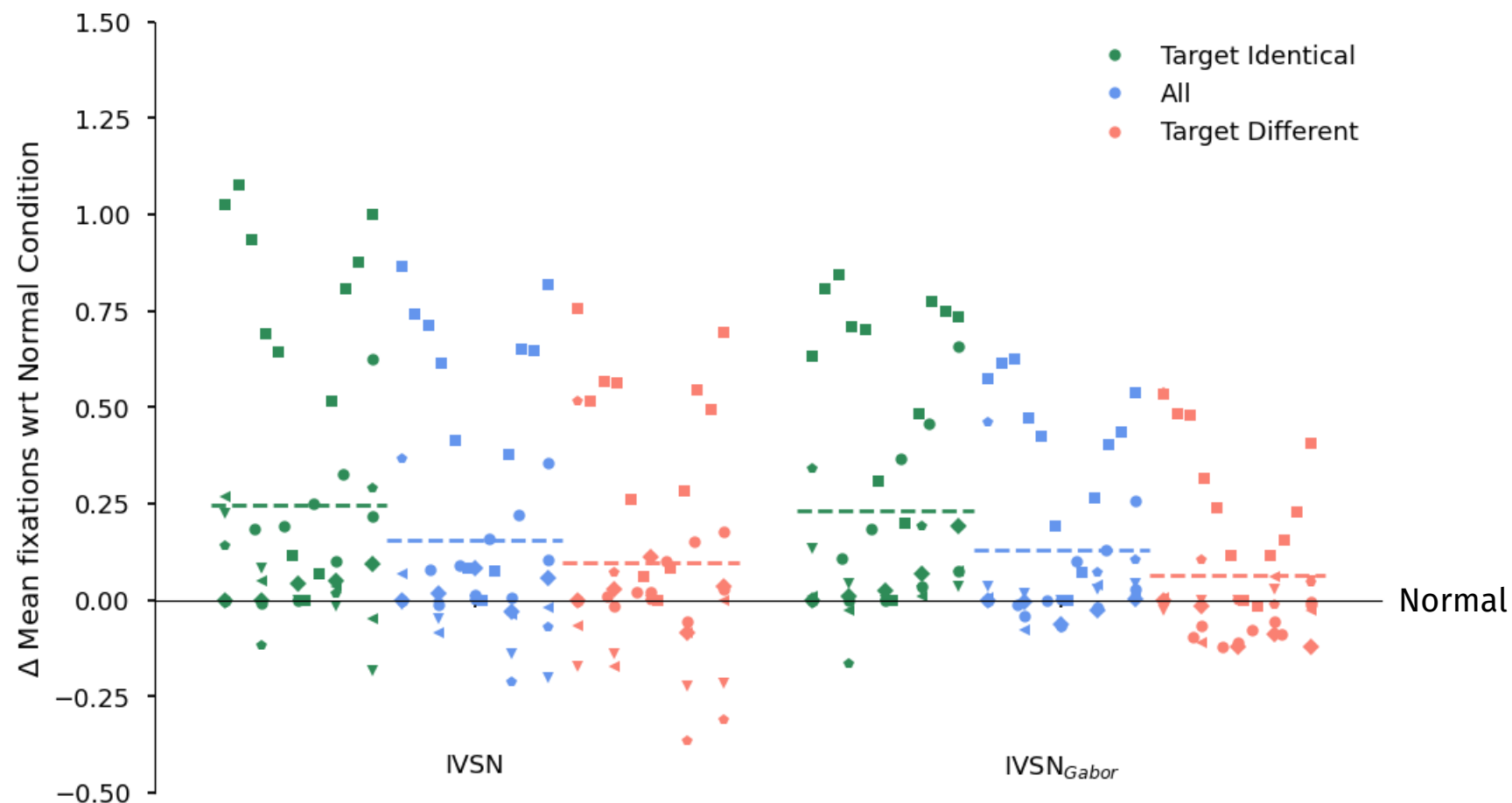
Comparing IVSN and IVSN_{Gabor} on 6-object circular array

- IVSN_{Gabor} has worse baseline performance ...



Comparing IVSN and IVSN_{Gabor} on 6-object circular array

- IVSN_{Gabor} has worse baseline performance ... but slightly better invariance properties



Conclusion

1. How invariant is the IVSN actually?

Not all transformations are equal ...

- IVSN is invariant to blur, (some types of) noise and skew transformations
- Rotation and scale are more difficult to overcome

2. Where does invariance arise in the visual hierarchy?

In the later stages of the feed-forward processing hierarchy

3. Does a Gabor filterbank frontend ($\text{IVSN}_{\text{Gabor}}$) improve invariance?

Slightly reduces variability in performance due to transformations

Open Questions

Why is the model invariant to some transformations and not others?

Is invariance solely a property of the visual representation, or do eye movements contribute?

Why is $IVSN_{\text{Gabor}}$ more invariant? How do the internal representations differ?



universität
uulm

Funded by
DFG Deutsche
Forschungsgemeinschaft
German Research Foundation



International Graduate School
in Molecular Medicine Ulm



KEMAI



Code



Thank you for your attention.
Questions?

Special thanks to

Julian Zeitler · Nadine von Hohenzollern-Emden

DFG Research Training Group 3012/1