InpaintCOCO: A retrieval dataset **with minimum differences**. CL w/ hard negatives help retrieval.

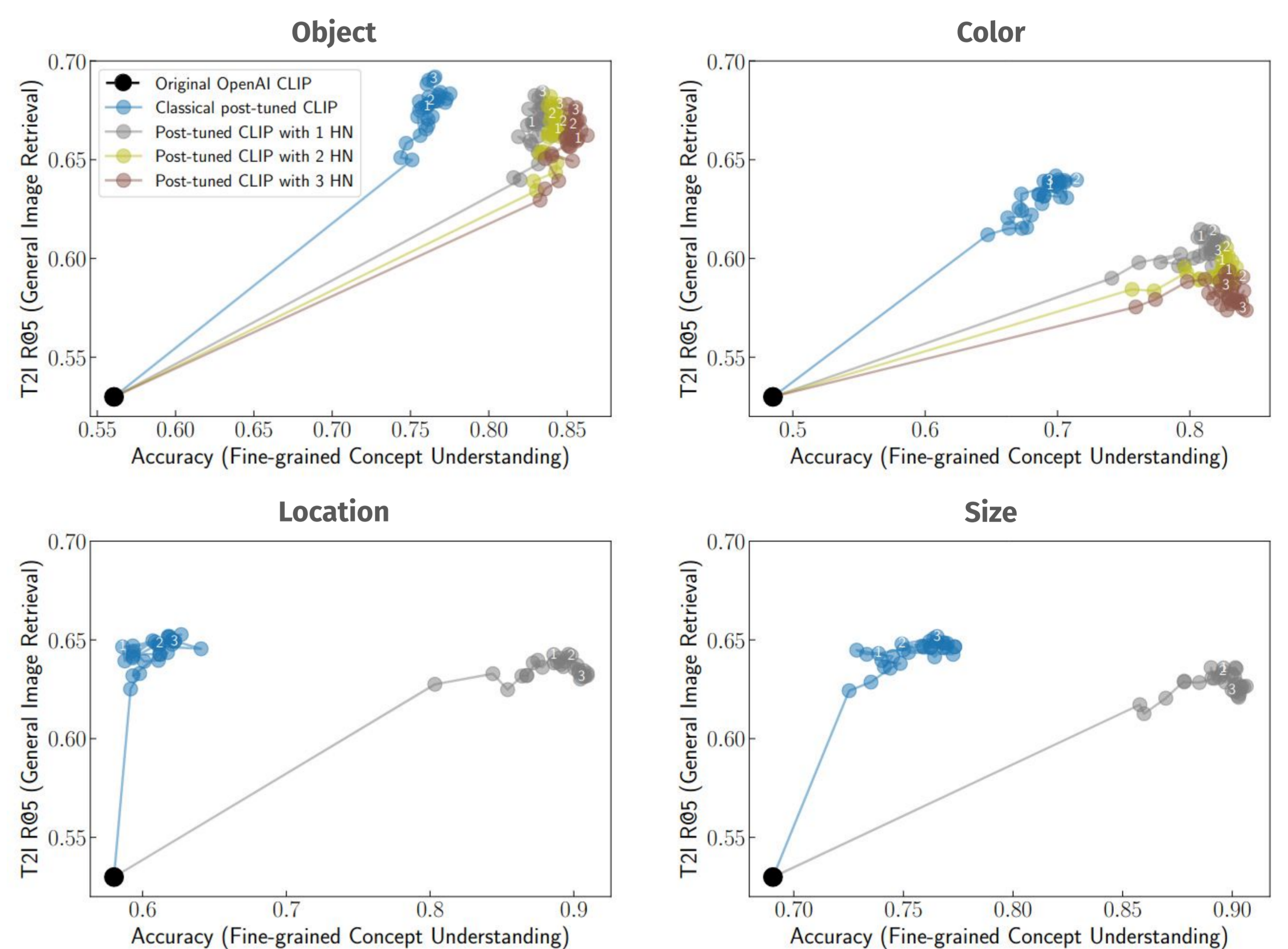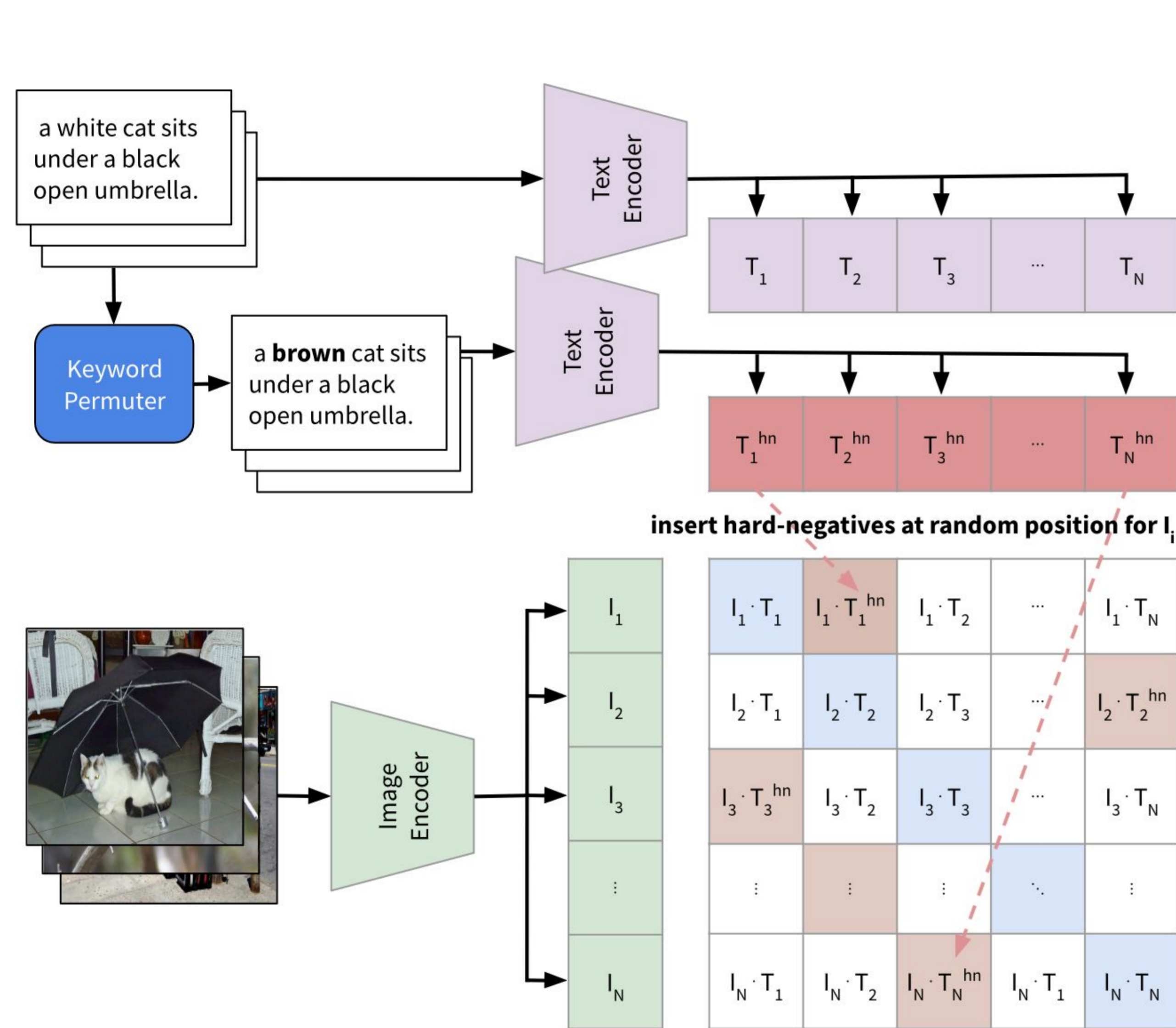## Enhancing Conceptual Understanding in Multimodal Contrastive Learning through Hard Negative Samples

Philipp J. Rösch   *philipp.roesch@unibw.de*
Norbert Oswald
Michaela Geierhos
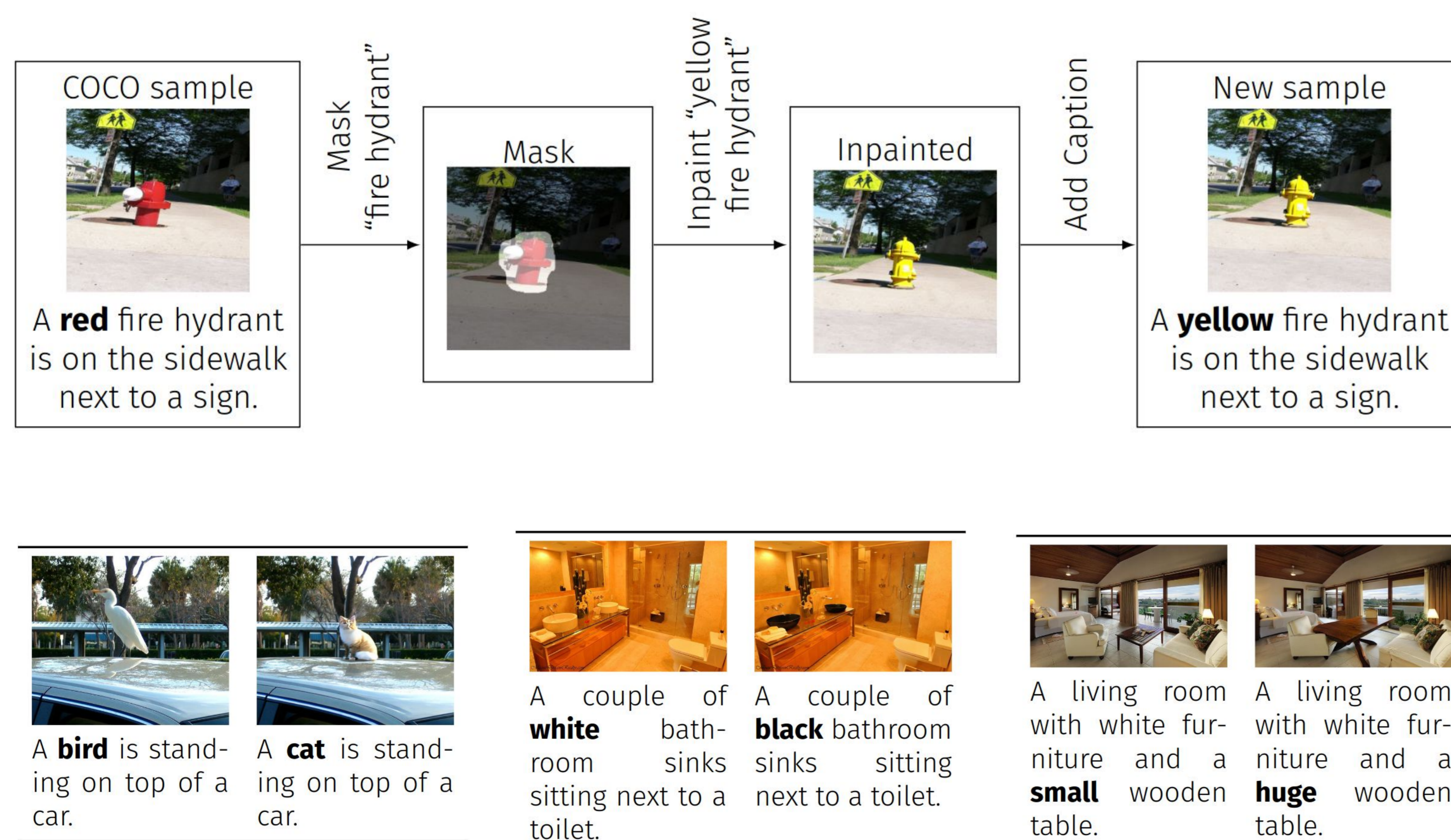**Jindřich Libovický**   *libovicky@ufal.mff.cuni.cz*

Universität der Bundeswehr München
Charles University

Hard negatives for specific concepts

InpaintCOCO Dataset with single-concept differences

A **red** fire hydrant is on the sidewalk next to a sign.
A **yellow** fire hydrant is on the sidewalk next to a sign.

A **bird** is standing on top of a car.
A **cat** is standing on top of a car.

A couple of **white** bathroom sinks sitting next to a toilet.
A couple of **black** bathroom sinks sitting next to a toilet.

A living room with white furniture and a **small** wooden table.
A living room with white furniture and a **huge** wooden table.

Hard negative sampling on InpaintCOCO

Object

Epoch	Orig.	Clas.	HN1	HN2	HN3
1	0.78	0.85	0.88	0.88	0.88
2	0.78	0.84	0.86	0.88	0.88
3	0.78	0.87	0.88	0.89	0.90

Color

Epoch	Orig.	Clas.	HN1	HN2	HN3
1	0.62	0.83	0.89	0.89	0.89
2	0.62	0.83	0.90	0.90	0.91
3	0.62	0.83	0.90	0.91	0.92

Size

Epoch	Orig.	Clas.	HN1
1	0.26	0.27	0.55
2	0.26	0.28	0.57
3	0.26	0.30	0.60