Unveiling the mystery of visual attributes of concrete and abstract concepts: Variability, nearest neighbors, and challenging categories

Tarun Tater, Sabine Schulte im Walde, Diego Frassinelli

Take Away

- In classification, **low-level features** outperform complex models → RQ 1
- **More images** improve performance only for concrete nouns → RQ 1
- Concrete and abstract targets show **significant visual variability** → RQ 2
- **Multiple reasons** for visual variability → RQ 3

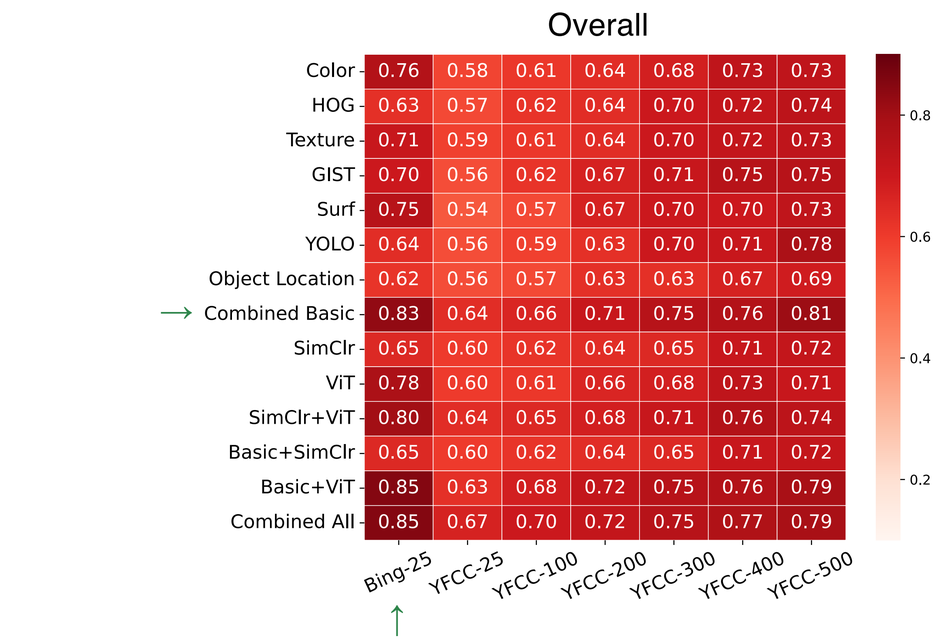
Motivation

- Language and vision play a crucial role for the understanding of the world surrounding us
- Multimodal studies → asymmetric contribution of vision and language
- Can visual information alone distinguish between concrete vs. abstract concepts?

Image Selection

- **1,000 target nouns** from Brysbaert Norms
  - 500 highly concrete
  - 500 highly abstract
- **Images:**
  - **Bing**
    - 25 per target
    - Search-based dataset → controlled data
  - **YFCC100M**
    - 25–500 per target → coverage issue!
    - User-tagged images from Flickr → diversity

RQ 1 – Can visual diversity differentiate between concrete & abstract concepts?

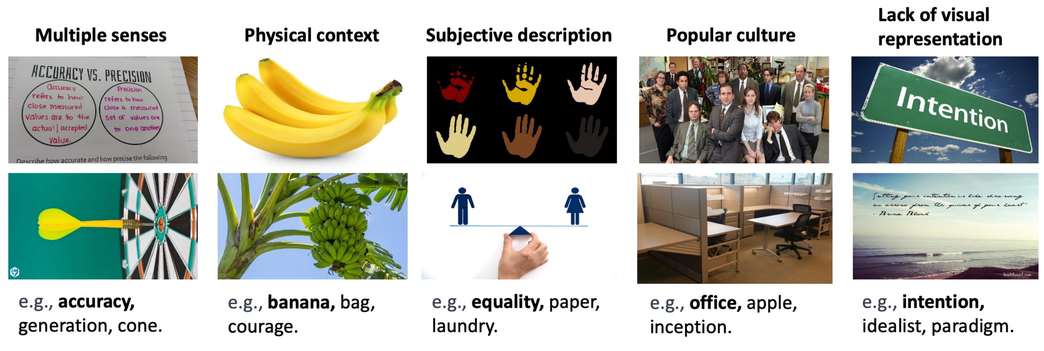


RQ 2 – How consistent are visual attributes across multiple images of the same concept?

Attribute	Bing-25		YFCC-25		YFCC-500	
	A	C	A	C	A	C
Color	0.68	0.96	1.70	0.95	0.81	0.65
HOG	0.48	1.44	0.68	0.58	0.36	0.44
Texture	0.29	0.33	0.35	0.26	0.28	0.27
GIST	0.55	1.88	1.03	0.76	0.52	0.56
SURF	0.64	1.70	0.93	0.62	0.40	0.38
YOLO	2.25	3.19	1.09	1.03	1.64	1.57
Object Loc.	0.18	0.39	0.15	0.18	0.24	0.27
Combined	0.64	2.14	1.40	0.99	0.69	0.75
Simclr	0.65	1.49	1.15	0.79	0.53	0.55
→ ViT	2.83	26.44	3.71	6.67	2.27	6.63

↑

RQ 3 – What are inherent yet plausible failure categories for visual representations?



Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN