MAX PLANCK INSTITUTE FOR INFORMATICS

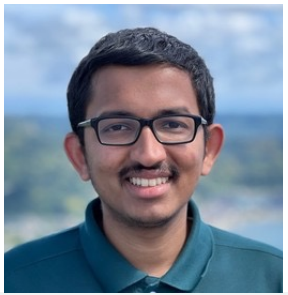
SIC Saarland Informatics Campus

ECCV 2026 | MALMÖ | SEPT 8–12

CFM: Language-aligned Concept Foundation Model for Vision



Kai Wittenmayer



Sukrut Rao



Amin Parchami

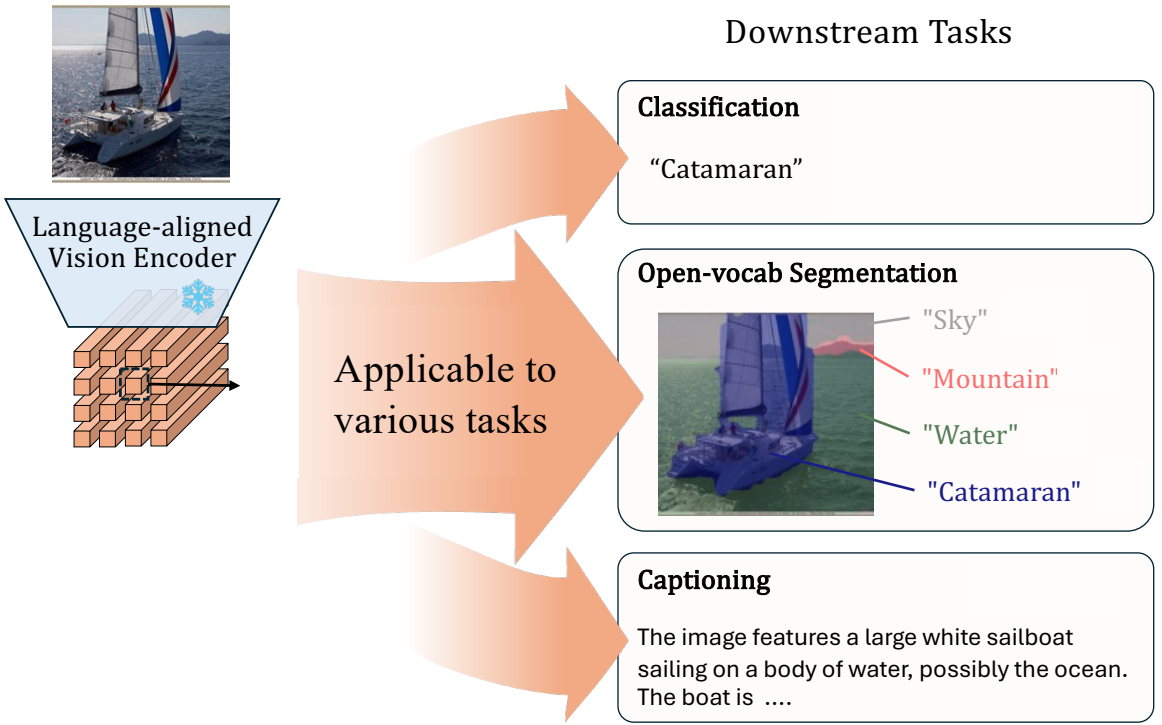


Bernt Schiele

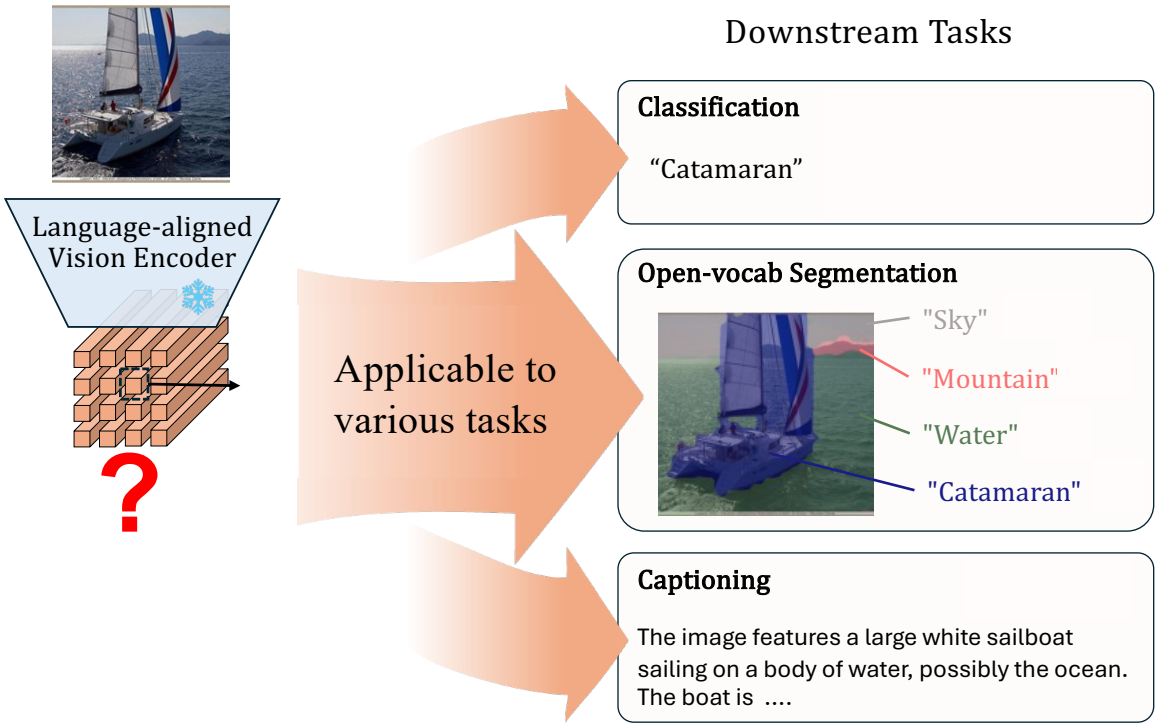


Jonas Fischer

Vision Foundation models are applicable to various downstream tasks

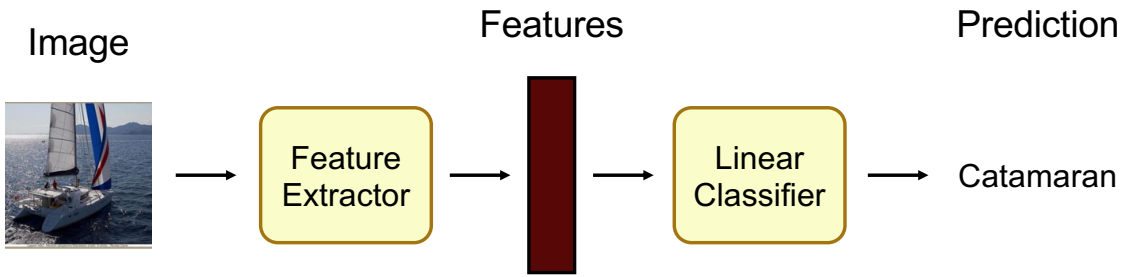


But they are opaque and hard to interpret

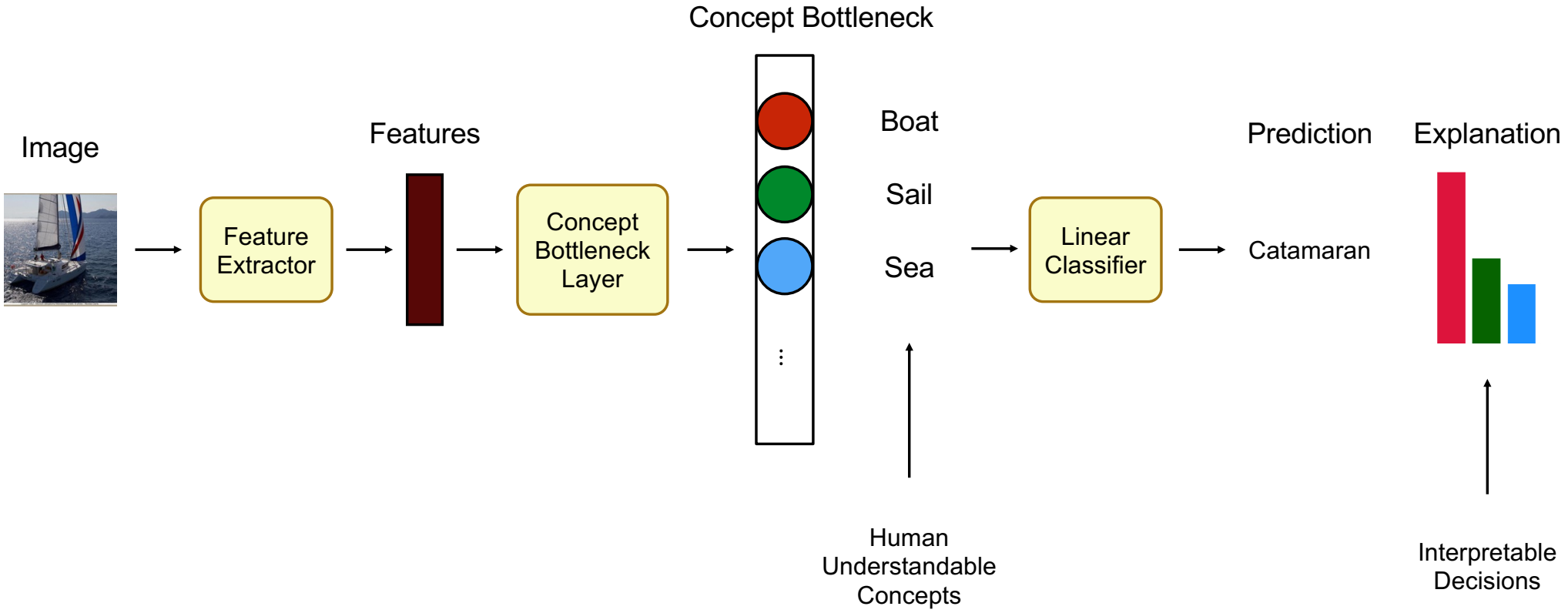


Interpretability can help **steer** models and make them **safer**

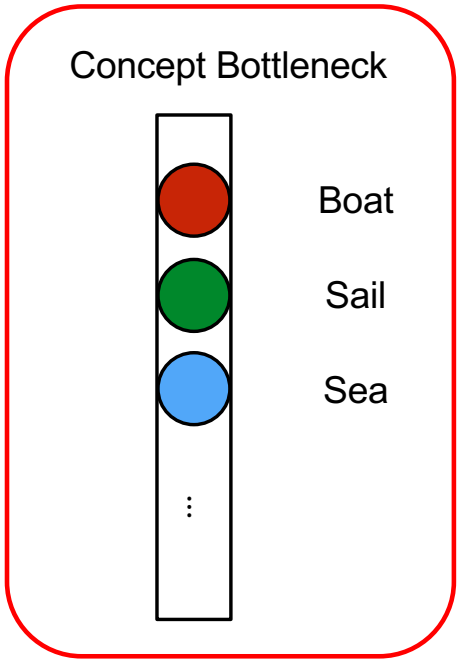
Concept Bottleneck Model (CBM)



Concept Bottleneck Model (CBM)



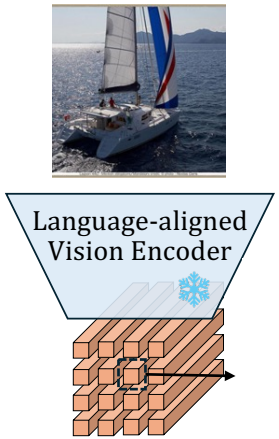
Concept Bottleneck Model (CBM)



✅ **human interpretable**

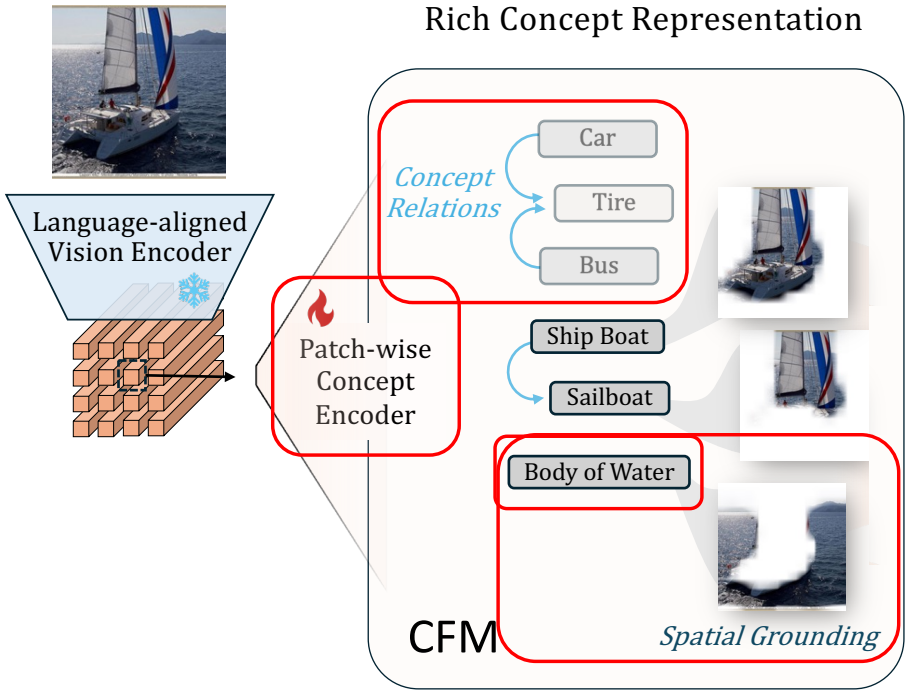
Typically: ❌ limited to classification ❌ no spatial grounding ❌ no hierarchies

Our Work: Concept Foundation Model (CFM)



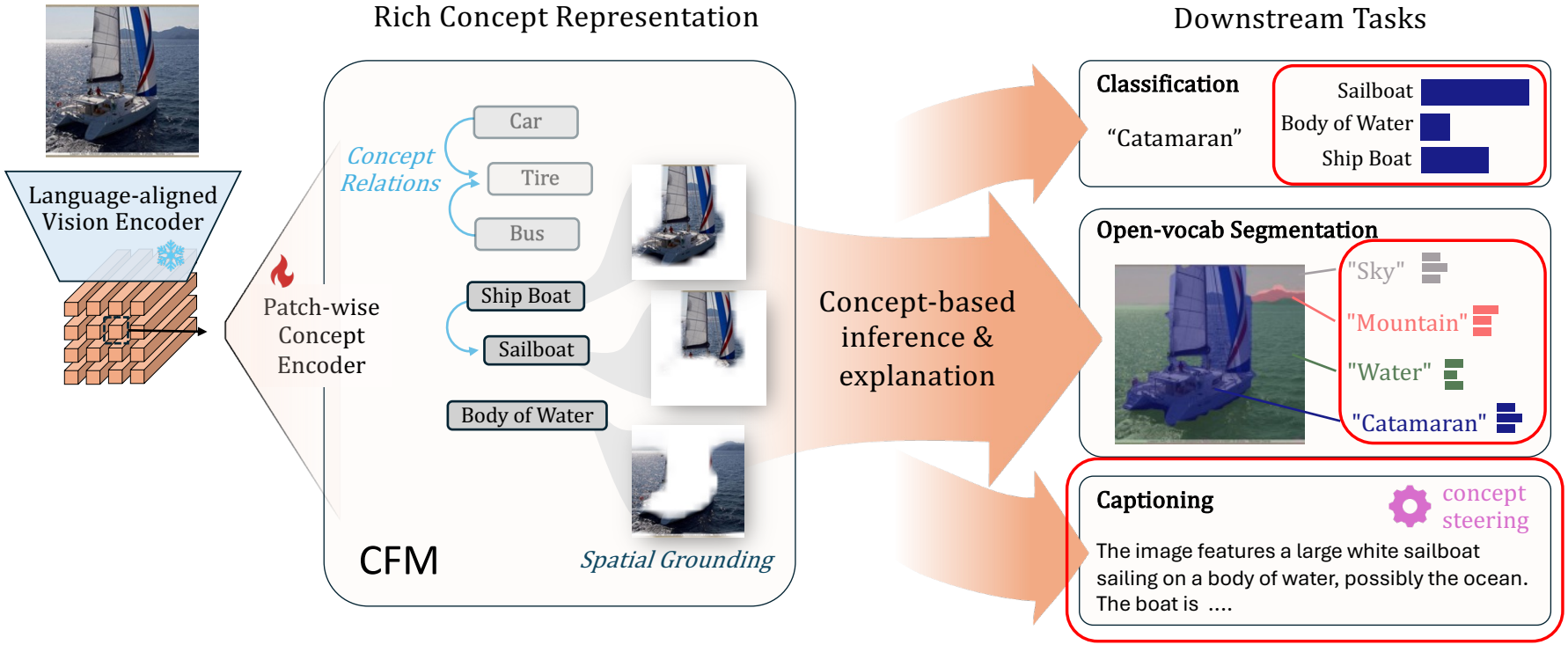
Concepts:

Our Work: Concept Foundation Model (CFM)



Concepts:

CFM preserves applicability accross all foundational model tasks



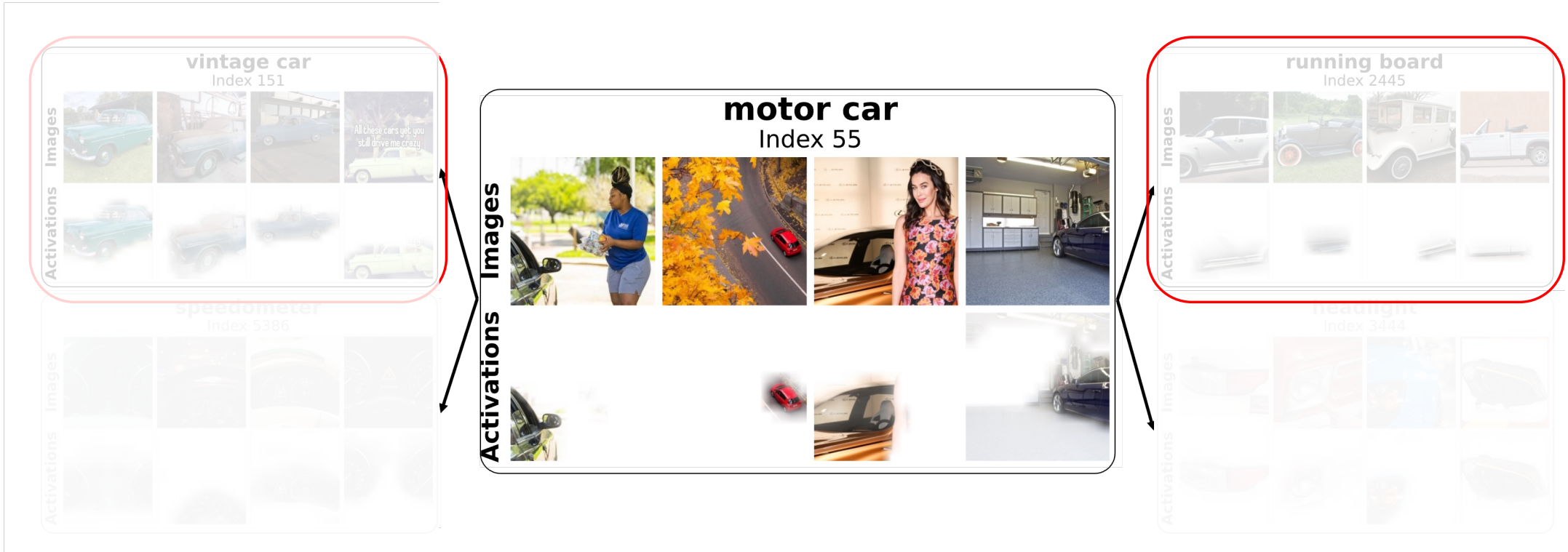
Concepts: ✅ language-aligned ✅ spatially grounded ✅ explicit hierarchies

Results

CFM provides consistent and well-grounded concepts



CFM provides consistent and well-grounded concepts



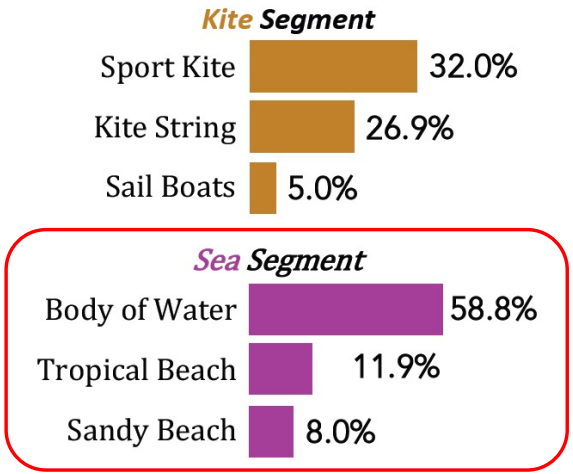
Meaningful relationships such as **specifications or subparts**!

CFM provides patch- or pixel-level segmentation explanations

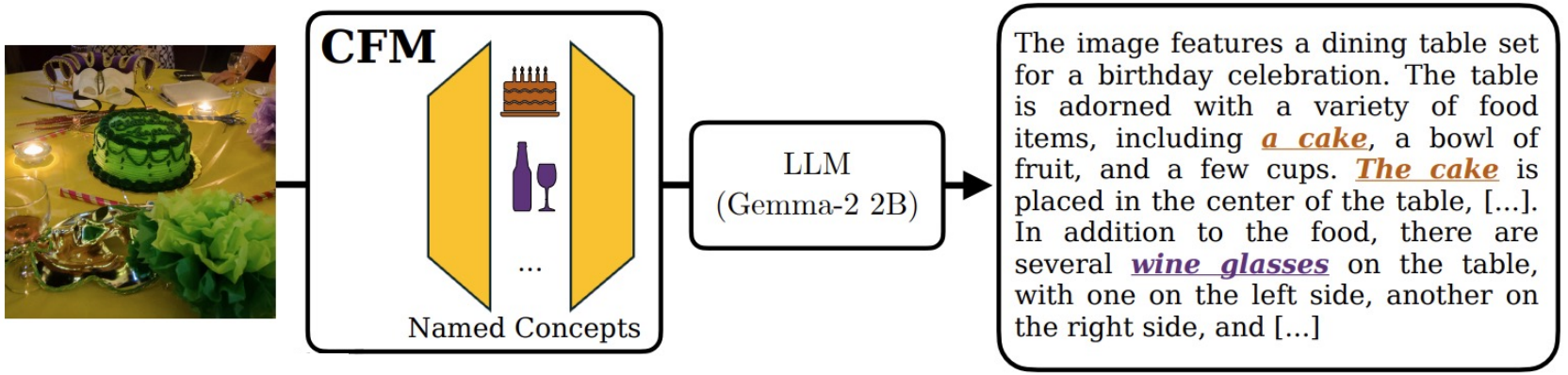


Open-vocabulary Segmentation

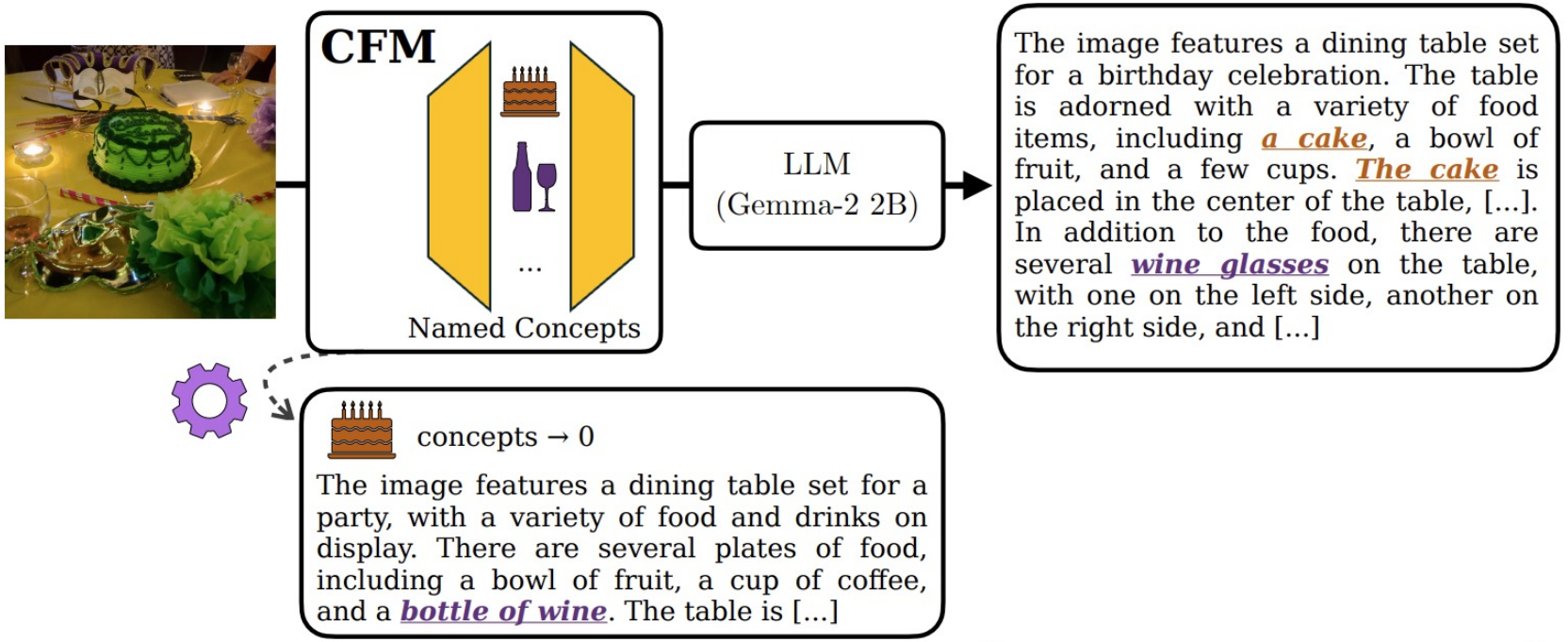
Top contributing concepts (per segment)



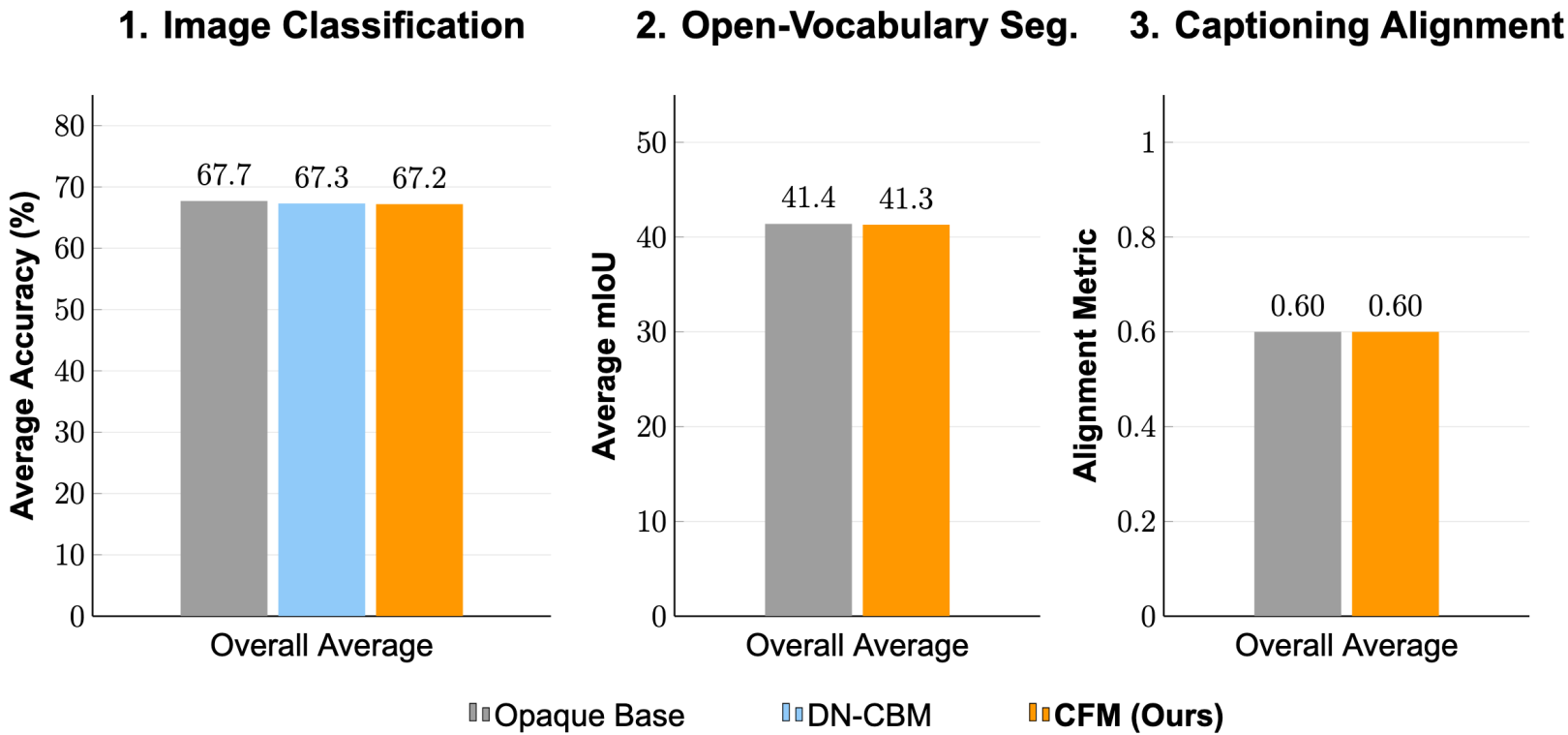
CFM yields steerable captioning



CFM yields steerable captioning

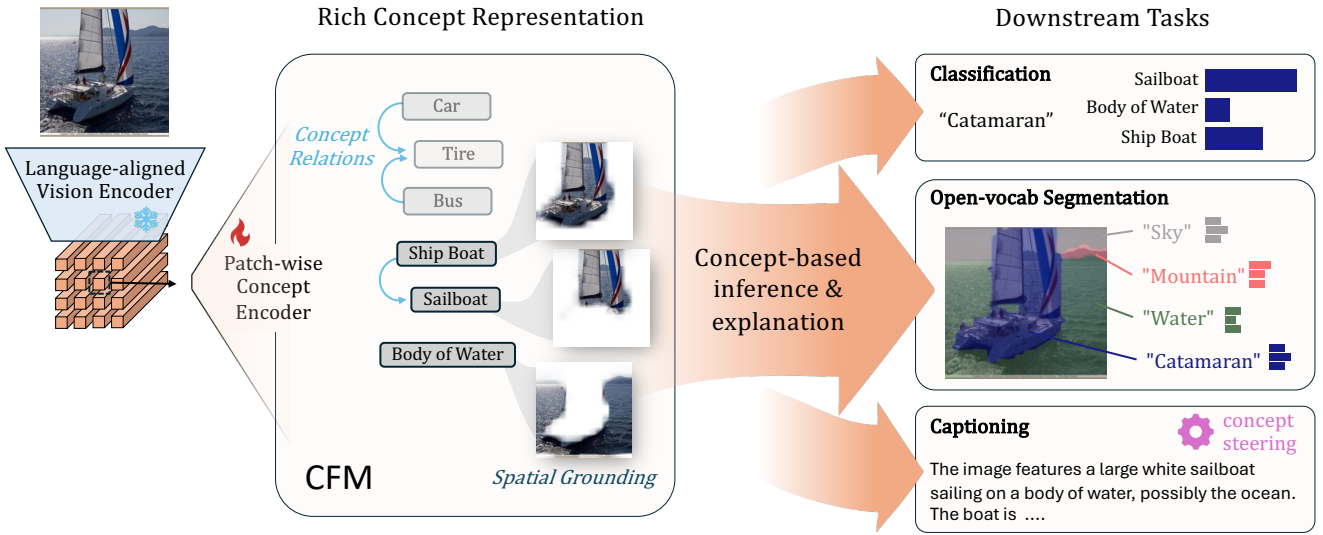


CFM performs on par with existing CBMs and opaque baselines



Conclusion

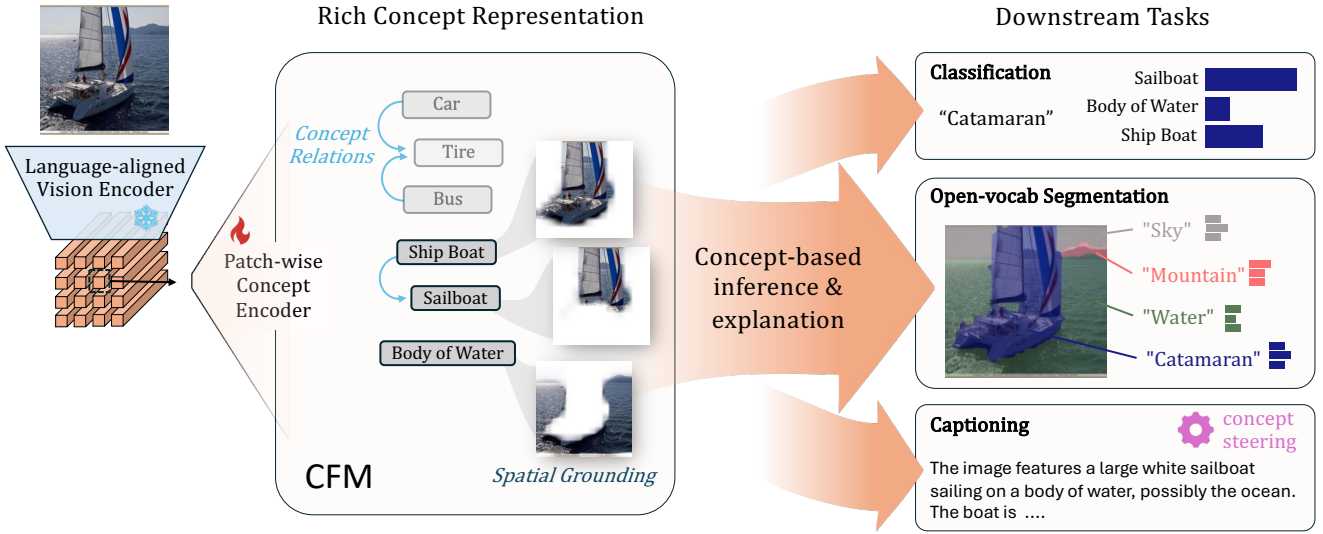
- ✅ **Spatially** grounded concepts
- ✅ Explicit **coarse-to-fine hierarchy**
- ✅ **Detailed naming**
- ✅ **Generality** of the foundation model



MAX PLANCK INSTITUTE FOR INFORMATICS

SIC Saarland Informatics Campus

Thank You



Poster ID: 323

Today 10:30 AM – 12:30 PM at ExHall

Paper

Code

Interactive Visualizations