Detection Transformers Under the Knife – A Neuroscience-Inspired Approach to Ablations

Nils Hütten, Florian Hölken, Hasan Tercan, Tobias Meisen

BERGISCHE UNIVERSITÄT WUPPERTAL

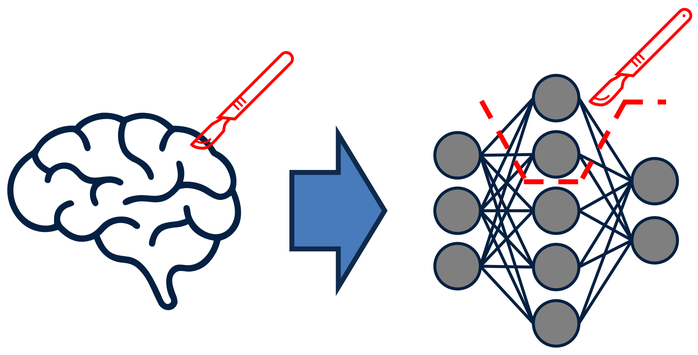
Institute for TMDT

BMVC 2025

Motivation

- Benchmark pursuit drives model complexity
- Internal knowledge organization remains opaque
- Real-world applications demand interpretability
- Neuroscience revealed brain organization by selective ablation

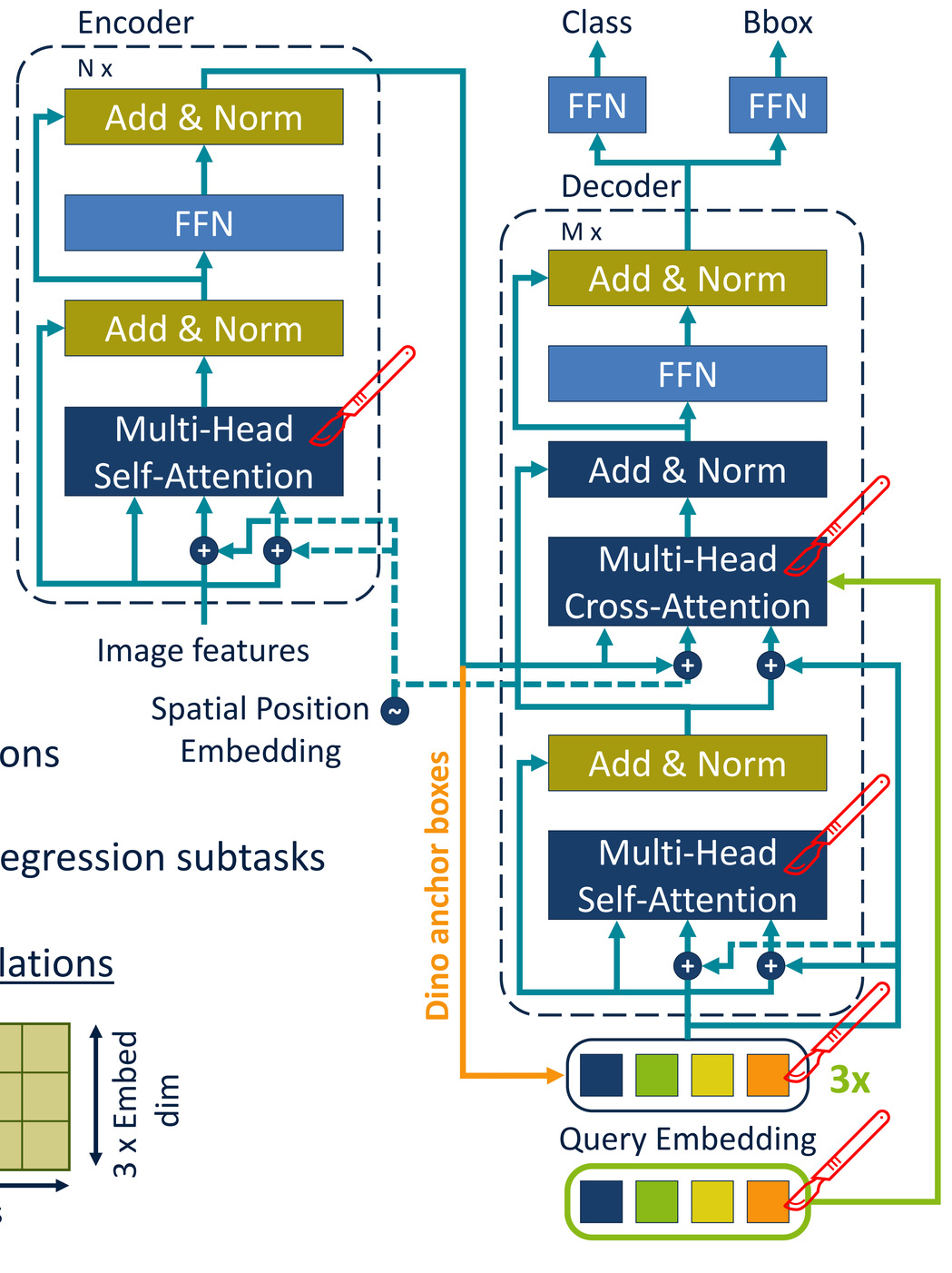
Application of neuroscience methodology to "*artificial nervous systems*"



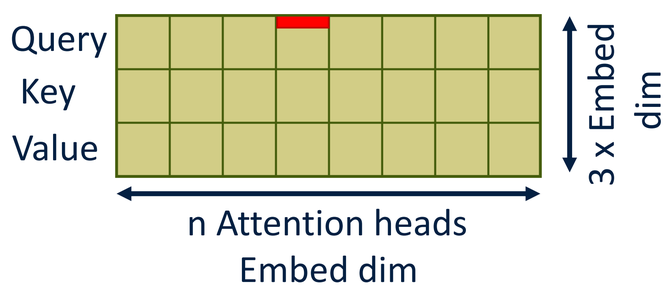
Methodology

Experimental Design

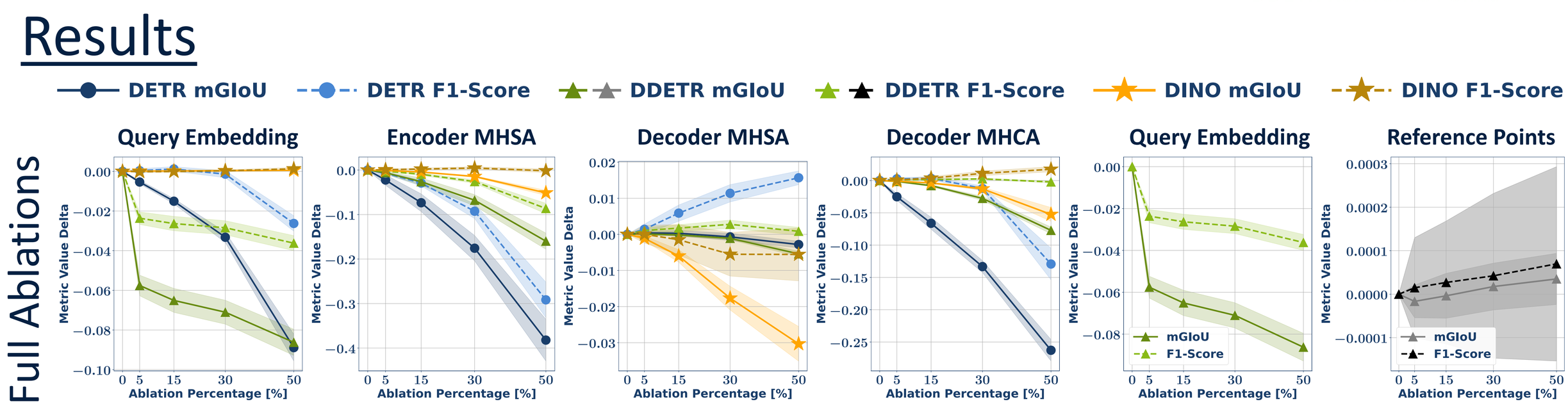
- 3 Models - **DETR**, **DDETR** & **DINO** trained on Coco
- Targets: DETR-specific components
- Full ablations: 5/15/30/50%
- Block-wise ablations: 30%
- 100 Randomly sampled iterations
- Evaluation of classification & regression subtasks

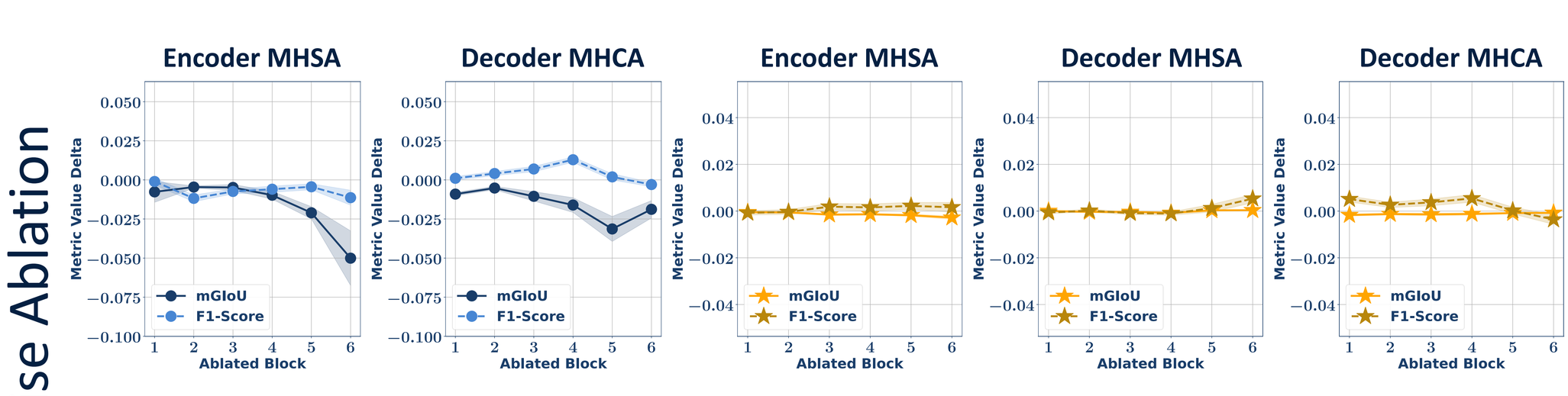


Projection Matrix Ablations



Results





Key Takeaways

- Newer models show increasing resilience and redundancy
- Regression subtask more susceptible than classification across models
- DETR /DDETR decoder MHSA shows very low contribution
- DDETR query embedding parameter increase promotes overfitting on frequent classes
- DINO's trained state renders static query embedding obsolete

Paper

Deep Dissect