Master Thesis Proposal

Title: Unsupervised Image Retrieval for Auction Catalogues

Name: Rahul Ramraje

Supervisor : Mathias Zinnen

Description: This thesis builds upon the foundational research presented in "Aiding Provenance Research: Computer-Assisted Image Retrieval in Auction Catalogs". The thesis aims to systematically explore how various unsupervised and self-supervised pretraining schemes, along with feature extraction from multiple network stages, influence the effectiveness of image retrieval within auction catalogues. The catalogues are gathered from GermanSales, a platform of more than 10,000 digitized auction catalogues.

Mandatory Goals:

- Evaluate multiple unsupervised and self-supervised pretraining schemes (e.g., DINO, SwAV, CLIP).
- Compare retrieval performance against ImageNet-pretrained models.
- Analyze features from different layers and scales for their impact on retrieval.
- Benchmark classical feature descriptors (e.g., SIFT, HOG) against deep learning features.
- Scale evaluation from a subset to the full set of auction catalogues.

Optional Goals:

- Plan to fine-tune top unsupervised models on catalogue data to improve retrieval scores,exact training objective and process still to be decided.
- Test different feature extraction backbones and sizes, including ResNet-18, to investigate the effect of architecture and model size on the results.
- Explore indexing and clustering methods (e.g., IVF, HNSW, PQ) for speed and efficiency.