Master Thesis Proposal

Name: Alisha Mund (Matriculation Nr: 23226214)

Supervisor: Mathias Zinnen (Pattern Recognition Lab, FAU)

December 1, 2025

Title: Multimodal Extraction of Lot-Level Metadata from Auction Catalogues using OCR and Vision Language Models

1. Description:

Historical auction catalogues represent a critical yet underutilized source of cultural heritage data. The German Sales collection, contain thousands of digitized catalogues from various time periods and publishers, present unique challenges for automated information extraction: handwritten annotations overlaying printed texts, irregular alignment of images and captions, diverse typographic styles and the scan quality is often inconsistent. These challenges make traditional OCR based extraction pipelines insufficient for recovering structured lot level metadata.

This thesis aims to design and evaluate a multimodal extraction pipeline that processes digitized catalogue pages and converts the extracted information into structured lot level metadata. Building upon a layered architecture consisting of document ingestion, perception and integration, the thesis will explore a combination of OCR, Object Detection Models, Vision Language Models (VLMs) and prompt based structured output generation using LLMs. The core objective is to investigate how these components, individually or jointly improve image extraction, isolate item illustrations and integrate them together with the corresponding lot descriptions.

The result of this work will support downstream tasks including multimodal search, retrieval systems and future integration with linked open data sources (e.g. Getty Provenance Index, Online art collections).

2. Primary Objectives:

2.1. Data Preparation and Schema Design

Establish a coherent data specification by assembling a representative collection of catalogues from various publishers and periods and specifying the desired target JSON schema. Create ground truth annotations (manual and semi-automated), including raw OCR content, detected image regions, lot number labels and complete structured output records for a selected subset of

catalogues. Construct finetuning datasets using page images with structured extraction instructions and expected JSON outputs.

2.2. Image Detection

Generation of high-quality image crops using a fine-tuned object detection model (e.g. YOLOv11n) on catalogue images and implement post processing to reduce false positives.

2.3. Systematic evaluation of three extraction strategies

- **Approach 1:** Classical OCR Text Extraction (Tesseract, PaddleOCR) + Rule based parsing using regular expression patterns (Baseline)
- **Approach 2:** OCR (PaddleOCR with layout analysis) + LLM structuring for the extracted text and images
- **Approach 3:** Finetuned VLMs for Direct Structured Output

2.4. Vision Language Model Finetuning for structured

- Explore various Parameter-Efficient Finetuning techniques such as LoRA, QLora or full finetuning.
- Analyse trade-offs between accuracy, training time, memory requirements, speed and identify optimal configuration for different model sizes

2.5. Implementation of Evaluation and Comparison metrics

- Text extraction evaluation using metrics such as Character Error Rate (CER), Word Error Rate (WER), Bag of Words F1 score, Named Entity Recognition (NER).
- Image detection evaluation using metrics such as mAP, recall, IoU.
- Additional VLM specific evaluation for structured output quality and confidence and approach level comparison.

2.6. End-to-End Integration and a structured JSON Output

Integrated pipeline that produces complete lot level records with correctly linked images and results in a clean and structured dataset for processed catalogues in a well-defined JSON Schema.

3. Optional Objectives:

3.1. Handwritten Text Recognition (HTR)

- Extension of the pipeline to detect and transcribe handwritten annotations.

3.2. Complete Catalogue Processing

- Scale the pipeline to process entire multi-year catalogues.

3.3. Dataset publication with FAIR Compliance

- Curate a representative dataset of annotated catalogue pages and their respective metadata and publish to Zenodo with licenses.

3.4. Visual Retrieval system integration and User Interface creation

- Enhance existing retrieval system with structured metadata.
- Create a web-based interface for uploading catalogue pages with visualization results.