

INTRODUCTION

- Natural history museums hold millions of preserved specimens whose collection dates and localities exist only on physical labels.
- These labels may be handwritten, machine-printed, faded, or written using historical orthography, making automated transcription difficult.
- The MuseumSCAT challenge asks models to extract the collection date and locality exactly as written from specimen photographs, or return MISSING when the information is absent. [1]

OBJECTIVES

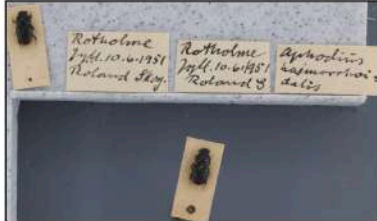
- Fine-tune HunyuanOCR-1.5 [2] to transcribe museum specimen collection dates and localities exactly as written.
- Return MISSING when a field is absent.
- Generate a confidence score between 0 and 1 for each prediction.
- Evaluate performance using area under the risk-coverage curve (AURC), where lower scores indicate better performance.

EXAMPLE TRAINING IMAGES

Training images contain handwritten, printed, faded, and historically formatted labels that must be transcribed verbatim.



Handwritten historical date: "Tebbestrøp Septbr '84" [1]



Handwritten date and locality: "Rotholme, Jyll. 10.6.1951" [1]



Close-up label example: "2-4-49, Røde Skov" [1]

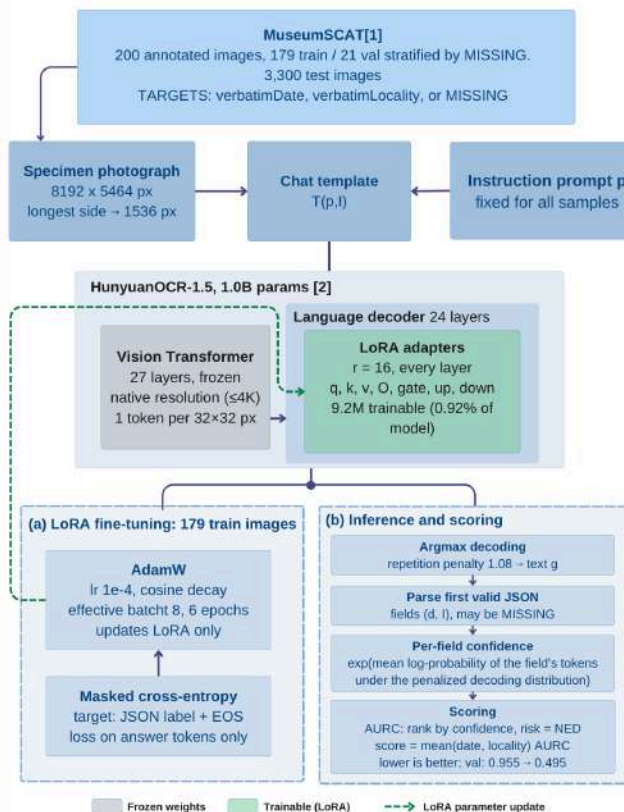
HUNYUANOCR-1.5

- 1.0B-parameter OCR vision-language model with open weights.
- Supports 4K native-resolution images and 331 languages.
- Preserves visually observed text, even when it conflicts with language priors.
- Supports structured JSON extraction for verbatim date and locality output.

PROPOSED APPROACH

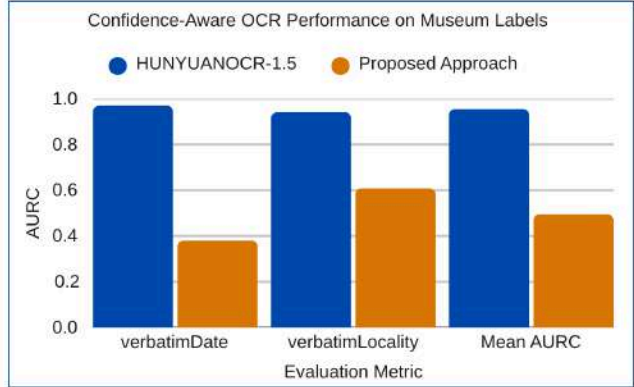
- LoRA fine-tuning: rank-16 adapters applied to all 7 decoder linear projections; vision tower frozen. [4]
- Trainable parameters: 9.2M (0.92% of the full model).
- Data split: 179 training / 21 validation images, stratified by MISSING pattern.
- Image preprocessing: resized to 1536 px on the longest side.
- Optimization: AdamW, learning rate of 1×10^{-4} , cosine decay, effective batch size 8, 6 epochs.

OVERVIEW OF THE PROPOSED APPROACH



RESULTS

- Mean AURC decreased from 0.955 to 0.495 after fine-tuning.
- Exact-match accuracy improved from 0% to 28.6% for dates and 0% to 14.3% for localities.
- Confidence scores were higher for MISSING predictions (0.45-0.75) than for transcribed values (0.24-0.58).



CONCLUSIONS

- LoRA fine-tuning substantially improved HunyuanOCR-1.5 performance on museum specimen label transcription [2].
- Training only 0.92% of the model's parameters on 179 images reduced mean AURC from 0.955 to 0.495.
- The fine-tuned model achieved verbatim recovery of 28.6% of dates and 14.3% of localities.
- Limitation: the model showed overfitting toward frequently represented training localities.

REFERENCES

- [1] Hansen, A. K., Pedersen, K. S., Hunt, R., and Natural History Museum of Denmark. (2026). "MuseumSCAT: Specimen Collection Annotation Task @ CVNH ECCV26" Kaggle, 2026. <https://kaggle.com/competitions/museumscat-specimen-collection-annotation-task>
- [2] Li, Gengluo, Xingyu Wan, Shangpin Peng, Weinong Wang, Hao Feng, Yongkun Du, Binghong Wu et al. "HunyuanOCR-1.5: Making Lightweight OCR VLMs Faster and Better." *arXiv preprint arXiv:2607.04884* (2026).
- [3] Team, Hunyuan Vision, Pengyuan Lyu, Xingyu Wan, Gengluo Li, Shangpin Peng, Weinong Wang, Liang Wu et al. "Hunyuanocr technical report." *arXiv preprint arXiv:2511.19575* (2025).
- [4] Hu, Edward J., Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. "Lora: Low-rank adaptation of large language models." *arXiv preprint arXiv:2106.09685* (2021).