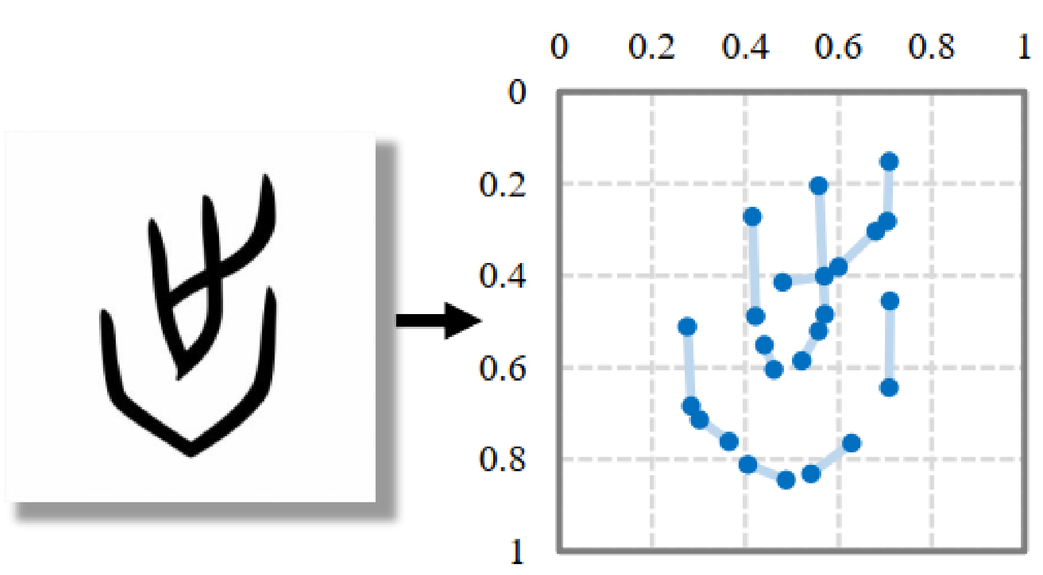
Enabling Stroke-Level Structural Analysis of Hieroglyphic Scripts without Language-Specific Priors

Fuwen Luo[1], Zihao Wan[1], Ziyue Wang[1], Yaluo Liu[3], Pau Tong Lin Xu[1], Xuanjia Qiao[4], Xiaolong Wang[1], Peng Li[2†], Yang Liu[1,2†]

[1] Dept. of Comp. Sci. & Tech., Institute for AI, Tsinghua University
[2] Institute for AI Industry Research (AIR), Tsinghua University
[3] Rixin College, Tsinghua University
[4] Department of Foreign Languages and Literatures, Tsinghua University

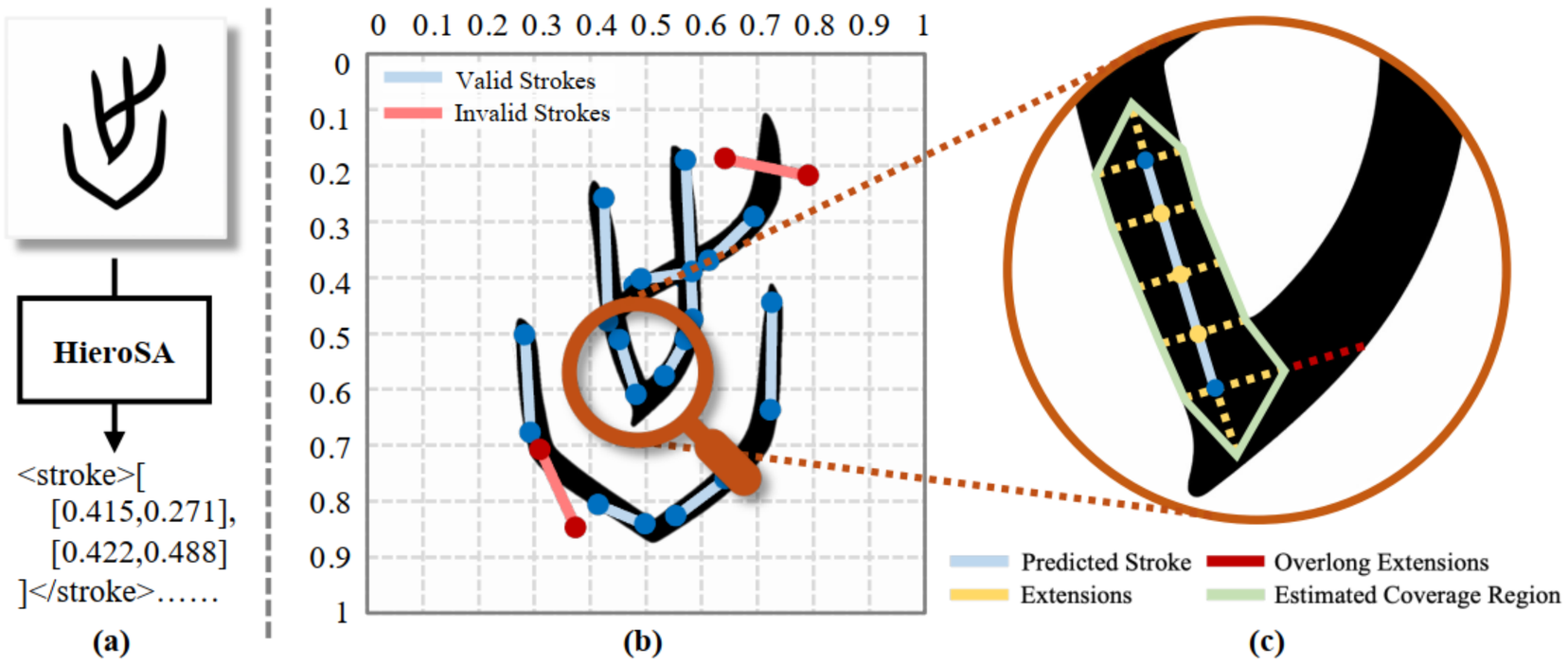
❖ Introduction

- Hieroglyphic and logographic scripts encode meaning inside character structure, especially in the configuration and combination of strokes.
- Current LLMs see characters as symbolic tokens; MLLMs see them as pixel grids. Both are weak at modeling stroke-level compositional logic.
- Existing structural analysis methods often rely on script-specific stroke inventories, decomposition rules, or manual annotation.
- Goal: infer explicit, interpretable stroke structures directly from bitmap images, without language-specific priors.



❖ HieroSA

- Representation: each stroke is a two-endpoint line segment in a normalized coordinate system.
- Learning signal: reward from bitmap coverage, invalid-stroke penalty, and structured output format.
- Training: GRPO fine-tuning of Qwen3-VL-4B-Instruct; no stroke annotations, no stroke order, no script-specific rules.



❖ Results

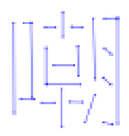
Model	Chinese (ZH)			Japanese (JA)			Oracle Bone Script (OBS)			AVG		
	RE↑	CO (%)↑	IS (%)↓	RE↑	CO (%)↑	IS (%)↓	RE↑	CO (%)↑	IS (%)↓	RE↑	CO (%)↑	IS (%)↓
GPT-5	0.133	3.6	88.2	0.129	1.5	92.8	0.139	3.8	92.0	0.134	3.0	91.0
Claude Sonnet 4	0.137	11.9	86.9	0.131	10.1	89.9	0.129	6.3	93.4	0.132	9.4	90.1
Qwen3-VL-4B	0.032	0.5	97.9	0.028	0.5	98.0	0.063	0.7	98.5	0.041	0.6	98.1
ELSED	0.026	0.4	83.8	0.014	0.3	80.3	0.100	0.8	73.9	0.047	0.5	79.3
DeepLSD	0.092	1.1	85.2	0.086	2.7	71.1	0.064	0.3	72.4	0.081	1.4	76.2
HieroSA (ZH)	**0.837**	**78.5**	**6.1**	**0.591**	**60.9**	19.7	**0.522**	50.5	**17.6**	**0.650**	**63.3**	**14.5**
HieroSA (JA)	0.756	72.2	10.2	0.584	59.4	**19.1**	0.455	45.0	23.9	0.598	58.9	17.7
HieroSA (OBS)	0.446	64.6	23.1	0.295	52.2	33.9	0.344	**53.3**	25.2	0.362	56.7	27.4

❖ Downstream Utility

- **OCR:** Adding HieroSA stroke representations improves average OCR accuracy from 82.4% to 83.5% after fine-tuning.

OCR Setting	Chinese	Japanese	AVG
Zero-shot	66.4	63.2	64.8
Trained w/o Stroke Rep	88.9	75.8	82.4
Trained w/ Stroke Rep	**89.7**	**77.3**	**83.5**

❖ Case Study

Model	Chinese (ZH)	Japanese (JA)	OBS
Input Bitmap Image			
HieroSA (ZH)			

Email: lfw23@mails.tsinghua.edu.cn

Page: https://github.com/THUNLP-MT/HieroSA/