

Point-by-Point Response to the Editorial Comments

Submission ID: 5e3eebcf-7254-4032-8499-a733072c8fe7

Manuscript title: Unsupervised Domain Adaptation for Vertical Cursive Script Recognition with Geometric Stroke Enhancement and Pseudo-Label Refinement

Journal: The Visual Computer

Dear Professor Sheng and Editorial Team,

We sincerely thank you for the careful editorial assessment of our manuscript and for giving us the opportunity to submit a substantially revised version. We have revised the manuscript to strengthen its relevance to computer graphics and visual computing, improve presentation quality and discoverability, clarify the reusable research contribution, and enhance transparency and reproducibility through open-source materials and benchmark protocols.

Below we provide a point-by-point response and a detailed list of revisions.

Required Questions

Q1. Does the manuscript clearly articulate a reusable problem formulation, benchmark, dataset, codebase, taxonomy, or methodological insight that other researchers can cite and build upon?

Yes. In the revised manuscript, we explicitly formulate the task as strict unsupervised domain adaptation for isolated vertical cursive script recognition. The revised problem formulation defines the labeled synthetic source domain, the unlabeled authentic target domain, the target-label exclusion rule during training and model selection, and the final evaluation protocol.

We also strengthened the manuscript as a reusable benchmark by adding clear source/target split rules, page-level partitioning to prevent cross-column leakage, dataset statistics, baseline comparisons, ablation settings, and evaluation metrics. The methodological insights that other researchers can build upon are now stated more clearly: the vertical-spine-aware geometric prior, the SASE physical degradation synthesis strategy, sequence-aware adversarial alignment, and valid-frame-normalized CTC entropy for pseudo-label refinement.

Revision made: The Introduction now includes a concise contribution statement. The Methodology section clarifies the strict UDA formulation, and the Experiments section provides benchmark statistics, baselines, ablations, and leakage-prevention rules.

Q2. Are the title, abstract, keywords, contribution statement, figures, and experimental comparisons sufficiently discoverable by researchers outside the immediate subfield?

Yes. We revised the title, abstract, keywords, and contribution statement to make the paper more discoverable to readers in visual computing, document image analysis, cultural heritage digitization, and low-resource recognition. The revised title follows the editorial suggestion and highlights the broader task of vertical cursive script recognition, rather than only naming traditional Mongolian OCR.

We also revised the abstract to reduce excessive technical density, clarify the cultural-heritage motivation, summarize the core method, and highlight the quantitative result. The keywords now include broader terms such as "vertical cursive script recognition," "document image analysis," and "cultural heritage digitization." We further strengthened the literature discussion by citing recent related articles from The Visual Computer and connecting them to structure-preserving visual restoration and structure-aware recognition.

Revision made: The title, abstract, keywords, Introduction, Related Work, and Data Availability sections were revised. We also clarified how the figures support the visual computing contribution, especially the comparison between synthetic degradation, authentic archival samples, and the vertical-spine-aware enhancement rationale.

Q3. Are code, data, pretrained models, demos, or benchmark protocols made available, when ethically and technically feasible?

Yes, to the extent ethically and technically feasible. We will provide a permanent public repository and Zenodo archive for the code, pretrained models, benchmark protocol, environment requirements, inference examples, and a releasable desensitized dataset subset.

The complete authentic archival dataset is subject to institutional and cultural-heritage access restrictions, so it cannot be fully released publicly. To support reproducibility, we will release a desensitized subset of isolated column patches, split files, annotation format examples, and scripts that reproduce the reported evaluation protocol when authorized users have access to the restricted data. The repository will explicitly state that the code and artifacts are associated with the manuscript submitted to The Visual Computer and will include a citation request for this manuscript.

Revision made: The abstract and the Data Availability and Reproducibility Statement now include explicit repository and DOI placeholders. These should be replaced by the final GitHub/Zenodo links before resubmission.

Detailed Revision List

1. Title revision

Comment addressed: The original title was long and less discoverable outside the immediate OCR subfield.

Revision: We changed the title to:

Unsupervised Domain Adaptation for Vertical Cursive Script Recognition with Geometric Stroke Enhancement and Pseudo-Label Refinement

Rationale: The new title foregrounds the reusable problem setting, geometric visual-computing contribution, and pseudo-label refinement, while still accurately describing the method.

2. Abstract optimization

Comment addressed: The abstract should better highlight the cultural-heritage context, core innovation, quantitative results, and broader impact.

Revision: We rewrote the abstract to clarify the problem, method, result, benchmark contribution, and reproducibility resources. The revised abstract reports the main result of $7.8 \pm 0.2\%$ CER and states that code, trained models, benchmark protocols, and a releasable dataset subset will be made available.

3. Keywords revision

Comment addressed: The manuscript should be more discoverable by readers outside the narrow subfield.

Revision: We revised the keywords to include broader and more searchable terms:

vertical cursive script recognition; traditional Mongolian script; historical archives; unsupervised domain adaptation; document image analysis; pseudo-label refinement; cultural heritage digitization.

4. Contribution statement added

Comment addressed: The manuscript should clearly articulate reusable contributions that other researchers can cite and build upon.

Revision: We added a contribution list in the Introduction. It now explicitly states the reusable UDA benchmark, the VS-MSSE geometric stroke enhancement module, the sequence-aware adversarial and valid-frame CTC entropy pseudo-labeling strategy, and the open-source release plan.

5. Relevance to The Visual Computer strengthened

Comment addressed: The manuscript should better fit the computer graphics and visual computing community.

Revision: We revised the Related Work section to connect the proposed method to structure-preserving image restoration and structure-aware visual recognition. We added two recent The Visual Computer references suggested by the editorial office:

- Detail-aware image denoising via structure preserved network and residual diffusion model. The Visual Computer, 2025, 41(1): 639-658.
- SATD: syntax-aware handwritten mathematical expression recognition based on tree-structured transformer decoder. The Visual Computer, 2025, 41(2): 883-900.

Rationale: These citations are used selectively to strengthen the visual-computing context without excessive citation.

6. Methodological novelty clarified

Comment addressed: The novelty and theoretical depth should be clarified relative to recent UDA and text recognition methods.

Revision: We clarified that standard UDA methods mainly align global feature distributions, while our method introduces a script-geometry-aware visual prior and a CTC-specific pseudo-label confidence criterion. The manuscript now more clearly distinguishes VS-MSSE from symmetrical multi-scale modules such as ASPP and distinguishes valid-frame-normalized CTC entropy from raw softmax confidence.

7. Experimental comparisons and ablations clarified

Comment addressed: Comparative experiments should be expanded and explained more clearly.

Revision: The revised manuscript retains comparisons against CRNN, TrOCR, PARSeq, CycleGAN+CRNN, DANN, CDAN, and a fully supervised upper-bound reference. We also present ablation studies for SASE, VS-MSSE, DANN, and pseudo-label refinement, plus architectural ablation for the vertical asymmetric kernels.

8. Figure and visualization explanation strengthened

Comment addressed: Add intuitive visualizations to validate the vertical-spine-aware module.

Revision: We clarified the role of synthetic and authentic sample figures and the Grad-CAM-based explanation in the Error Analysis and Limitations section. The revised text emphasizes that VS-MSSE activates along the central vertical spine and that remaining failure modes are tied to broken spines, visually similar vowels, and stains/noise.

Recommended next step before final resubmission: add or improve a dedicated visual figure showing VS-MSSE activation maps or before/after enhancement examples, if space permits.

9. Open-source and reproducibility statement expanded

Comment addressed: Code should be permanently hosted with comprehensive documentation and, preferably, DOI.

Revision: We rewrote the Data Availability and Reproducibility Statement. It now states that the repository will include environment requirements, training/evaluation scripts, key algorithm implementations, pretrained model weights, inference examples, benchmark protocols, and a releasable desensitized dataset subset. It also states that the repository will identify the connection to this The Visual Computer manuscript and request citation.

Important action before submission: replace the placeholders below with final permanent links:

- GitHub repository: <https://github.com/REPLACE-WITH-REPOSITORY>
- Zenodo DOI: <https://doi.org/REPLACE-WITH-ZENODO-DOI>

10. Ethical and technical data-release constraints clarified

Comment addressed: Data should be released when ethically and technically feasible.

Revision: We clarified that the complete authentic archival dataset is restricted by institutional and cultural-heritage access conditions. To balance reproducibility and ethical constraints, the revised manuscript commits to releasing a desensitized subset,

split files, annotation examples, and evaluation scripts.

Closing Statement

We appreciate the editorial guidance and have revised the manuscript substantially to improve its clarity, reproducibility, visual-computing relevance, and discoverability. We hope the revised manuscript and the accompanying itemized response demonstrate that the key concerns have been addressed carefully and completely.

Sincerely,

Lanying Liang and Yuefeng Liu