

# Rerouting Connection: Hybrid Computer Vision Analysis Reveals Visual Similarity Between Indus and Tibetan-Yi Corridor Writing Systems

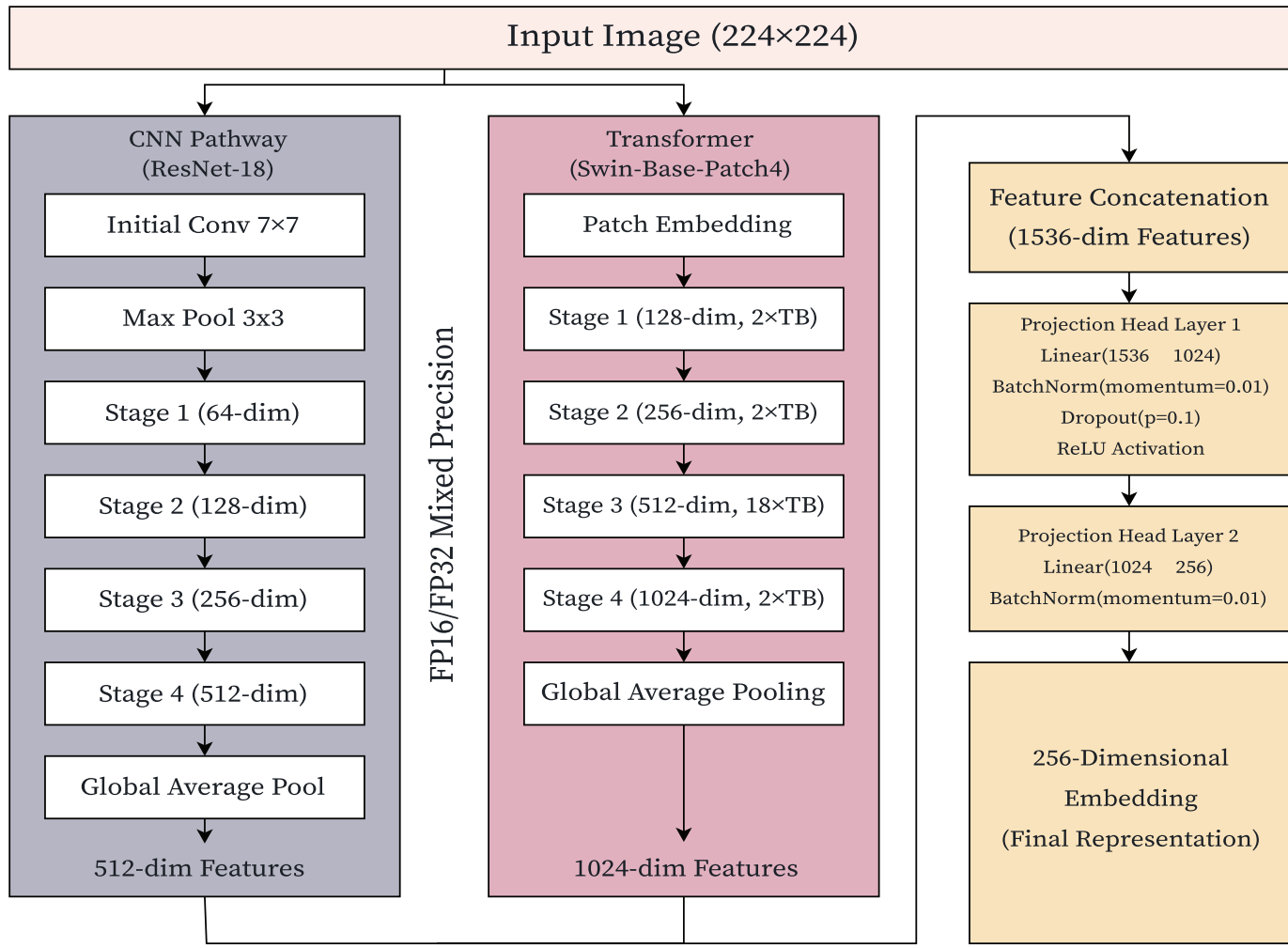
Ooha Lakkadi Reddy (Computation and Design/Digital Media), Class of 2025  
Mentors: Prof. Lincoln Rathnam and Prof. Mustafa Misir

## Summary

CNN-Transformer hybrid ensemble analysis reveals ancient Tibetan-Yi scripts share 61.7%-63.5% visual similarity with Indus Valley signs versus 7.6%-10.9% with West Asian scripts. Contrary to our current understanding of the networks of the Indus Valley Civilization, the Indus script unexpectedly maps closer to Tibetan-Yi Corridor scripts than to the aforementioned contemporaneous West Asian signaries, despite their geographic proximity and known trade relations.

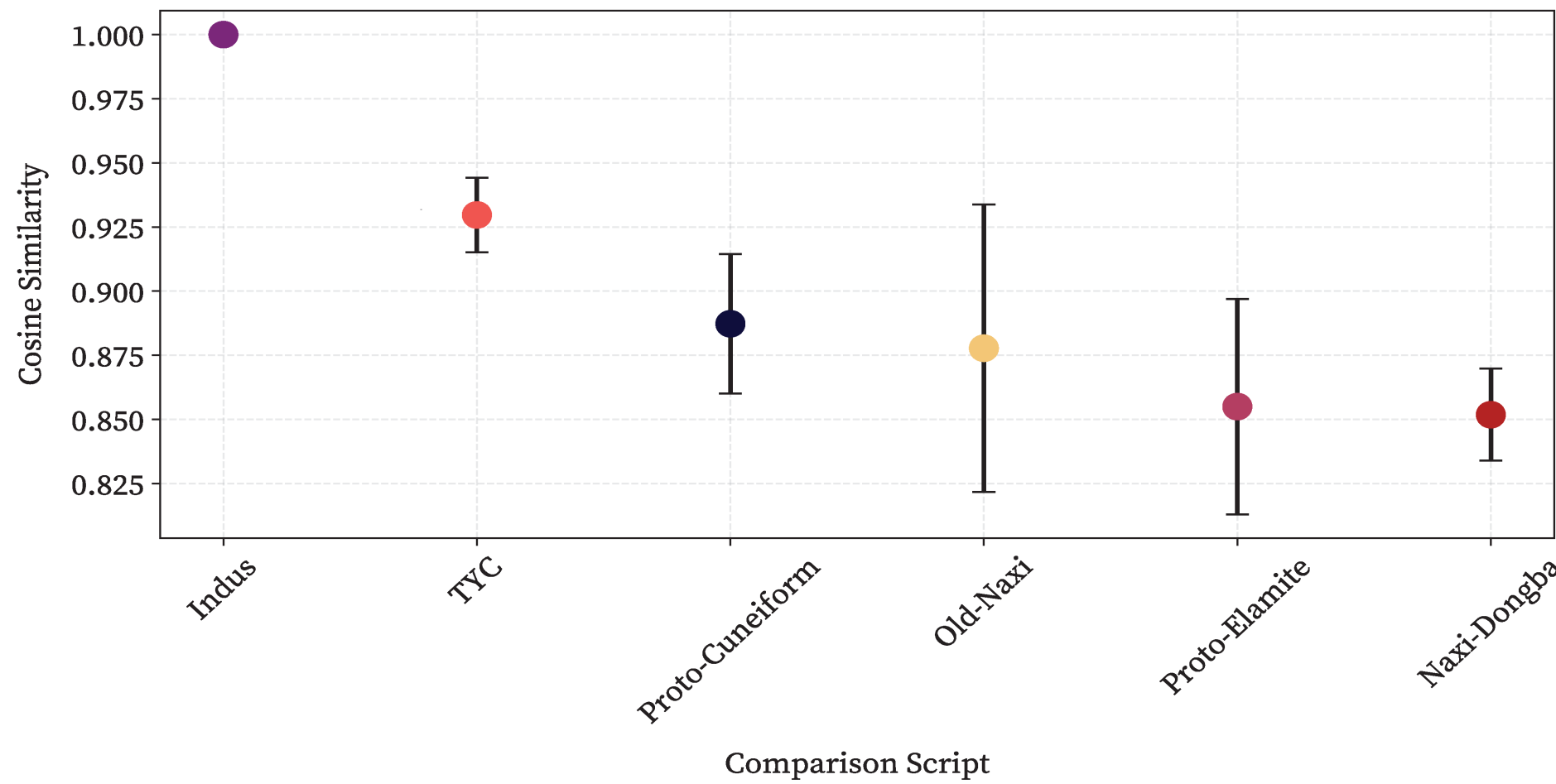
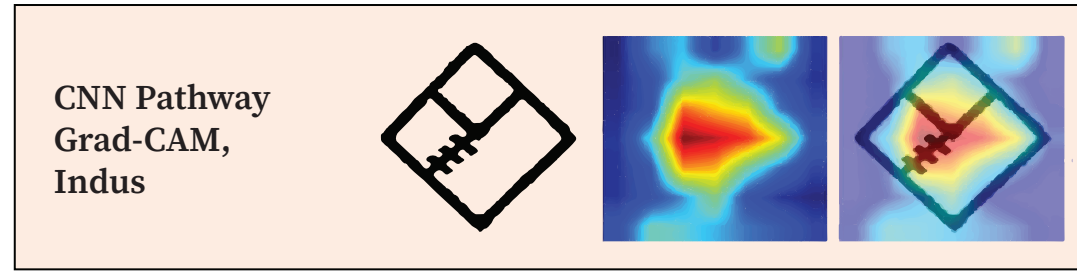
## Background

The Indus Valley Civilization (IVC, 2600–1900 BCE) was among the world's earliest urban societies, famous for its undeciphered pictographic script. Centuries later, nomadic peoples migrated through the Tibetan-Yi Corridor (TYC), a mountainous crossroads formed by the Hengduan Mountains. Within the TYC, the Naxi developed the Dongba script—the world's only surviving pictographic tradition. Although previous data mining studies clustered the Indus script with Proto-Cuneiform and Proto-Elamite, our qualitative observations, corroborated with computational results, suggest closer kinship with ancient TYC scripts. We consider the evidence for early iconographic transmission along the ancient Shu-Shendu route linking South Asia to Southwest China.



## Architecture

- Combined ResNet-18 & Swin Transformer for local/global feature extraction
- Features fused into 1536D embeddings with 2-layer projection head
- Trained via self-supervised contrastive learning on A100 GPU
- Used 5-model ensembles per script to improve stability
- Validated with Grad-CAM for interpretability



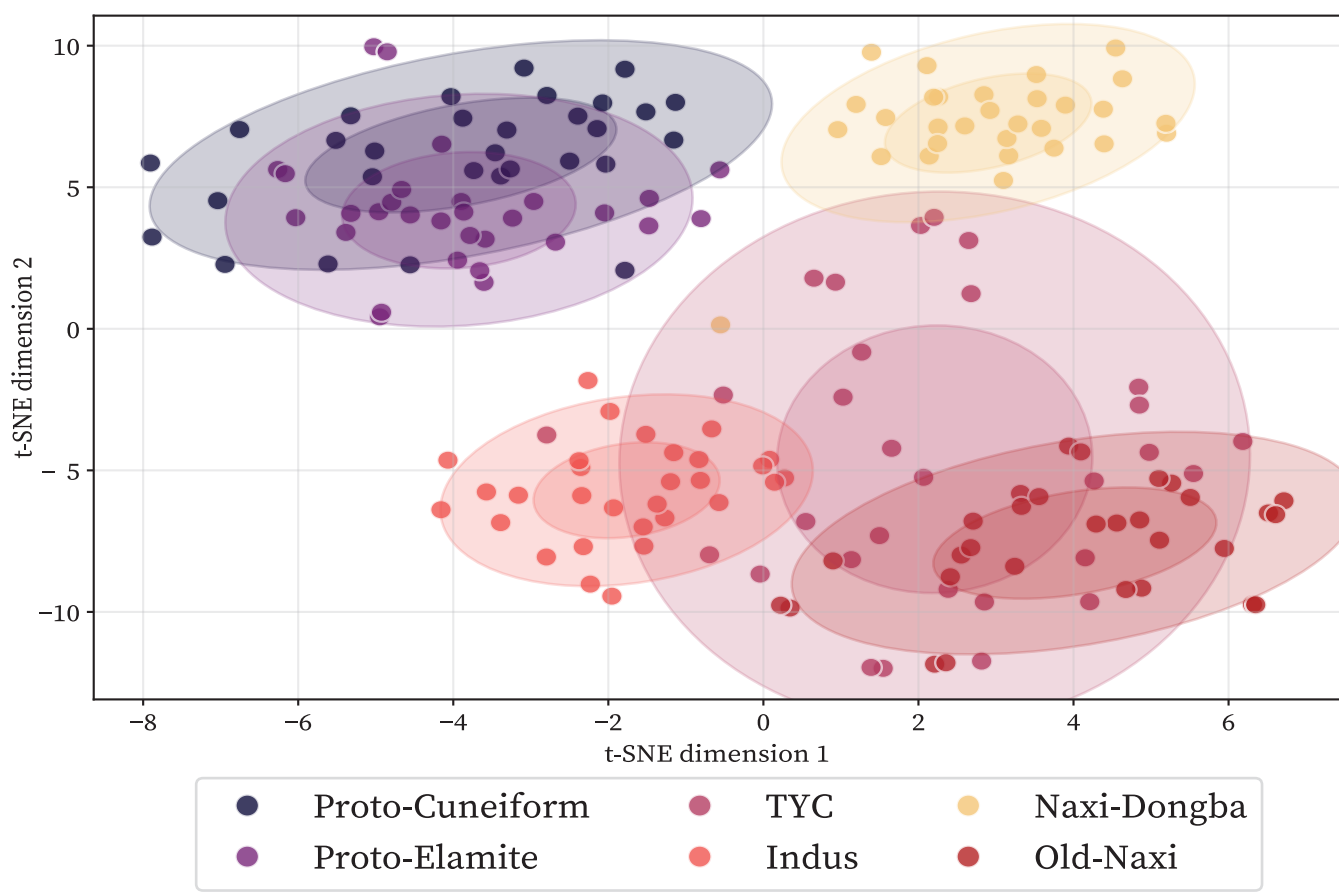
## Cosine Similarity Analysis & Validation

- TYC script embeddings were closest to Indus embeddings, showing substantially higher similarity to Indus (0.629) than to Proto-Cuneiform (0.102) or Proto-Elamite (0.078)—**approximately six-fold greater**.
- Indus–TYC consistently exhibited higher cosine similarity (mean similarity = 0.629) compared to Indus–Proto-Cuneiform (0.104) and Indus–Proto-Elamite (0.080).
- Welch's t-tests confirmed these differences as statistically significant ( $p < 0.001$  after Bonferroni correction).
- Cohen's d values exceeded 7.0 for Indus–TYC vs. Indus–West Asian scripts, indicating strong practical significance.
- All 5 independently trained ensemble models agreed on this ordering (Indus closer to TYC scripts), underscoring robustness.

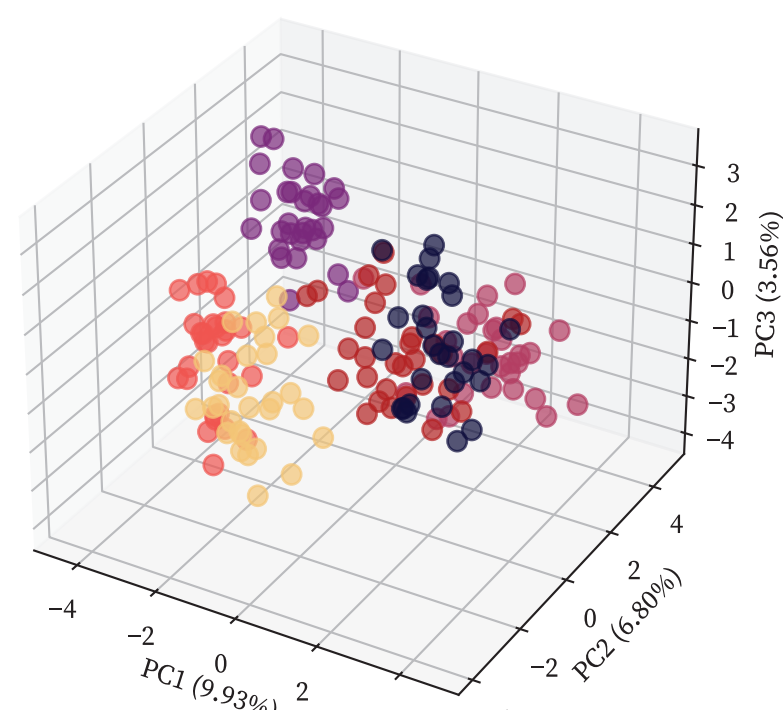
## Unsupervised Clustering

- Indus scripts consistently clustered with older TYC scripts, distinctly separate from Proto-Cuneiform and Proto-Elamite via hierarchical agglomerative clustering.
- Ensemble t-SNE showed clear separation into three clusters:
  - Proto-Cuneiform/Proto-Elamite
  - Indus/older TYC
  - newer Dongba (Sinicized)
- Ensemble PCA showed a West Asian/TYC/Indus cluster separate from newer Dongba.
- Clusters remained consistent across all linkage methods (single, complete, average, Ward's), demonstrating stable method-independence.

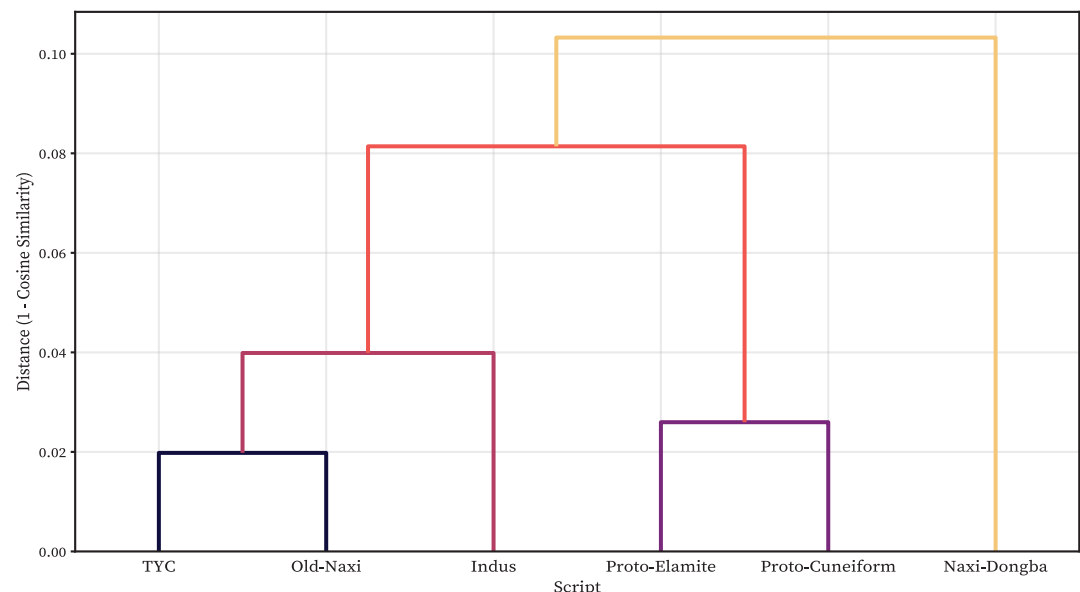
Proto-Cuneiform Ensemble t-SNE Visualization



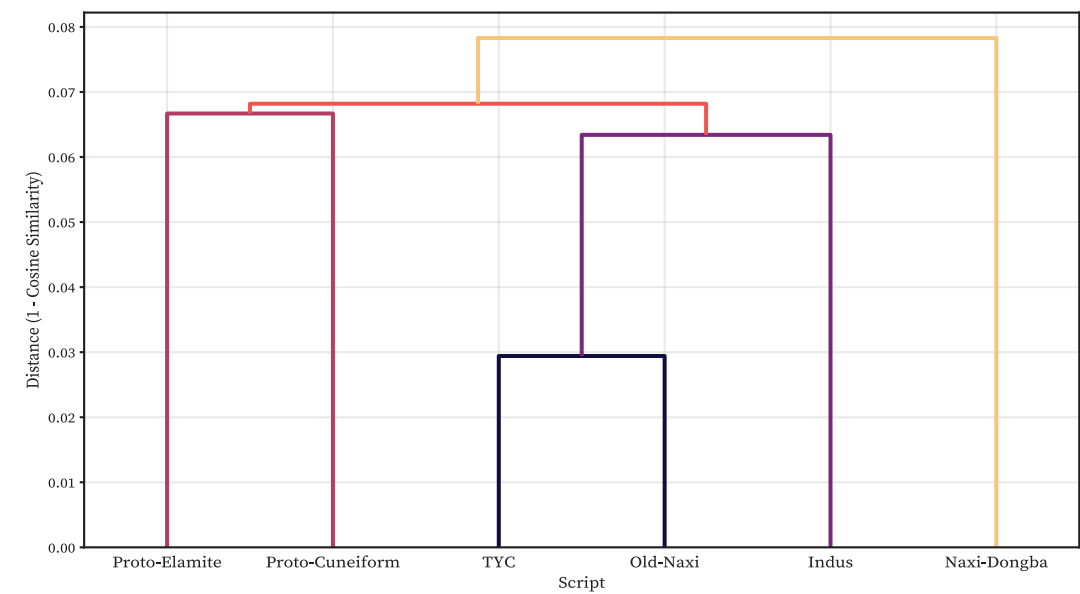
Proto-Elamite - Ensemble 3D PCA Visualization  
Total Explained Variance: 20.29%



Proto-Elamite - Ensemble Hierarchical Clustering (Average Linkage)



Indus - Ensemble Hierarchical Clustering (Single Linkage)



Proto-Elamite - Ensemble Clustered Similarity Matrix (Single Linkage)



## Future Implications

Our multidisciplinary analysis opens promising research pathways across a plethora of fields and marks a step forwards in solving the long-standing questions surrounding the IVC. These findings challenge the current state of scholarship regarding IVC relations, suggesting more extensive trans-Himalayan exchange than previously recognized. Future studies should prioritize collaborative decipherment efforts, targeted archaeological/geological surveys along the proposed Shu-Shendu route, and further application of our trained hybrid model to investigate and compare potential connections with late IVC contemporaneous Tibetan rock glyphs, 甲骨文 (oracle bone script), as well as other ancient and modern scripts as control groups to evaluate our comparisons. While these findings are unprecedented, their robust alignment with the presented anthropological evidence could represent an initial step toward substantial breakthroughs in understanding ancient Asian script encoding—and, farsightedly, the civilizations that created them along with those that inherited them.

