DOI: https://doi.org/10.36094/sc.v92.2026.Benchmarking_Automatic_Scene_Word.Rahamatulla_Sengupta.692

ARTICLE

Sci. and Cult. 92 (9–10) : 692-696 (2026)

BENCHMARKING AUTOMATIC SCENE WORD TRANSCRIPTION ON ROMAN SCRIPT

RAHAMATULLA[1]* AND PAYEL SENGUPTA[1]

Scene Text Recognition (STR) has progressed rapidly in this era for Roman scripts, but robust recognition of Indic and other complex scripts like Hindi, Bengali, Telugu, and more in natural images remains underexplored. Existing work highlights the difficulty posed by combined characters, large words, different writing styles, and challenging capture conditions such as blur, low resolution, and complex backgrounds. To address this aperture, this work benchmarks the automatic scene word transcription on the SWord6k dataset, a collection of 6,661 cropped word-image components from Roman natural scenes. Performance is reported using Word Recognition Rate (WRR) and Character Recognition Rate (CRR), following recent Indic handwriting and STR competitions (Mondal & Jawahar, 2023; Mondal et al., 2024). The experiments show that both models achieve high CRR but noticeably lower WRR, especially for longer words and visually degraded sample images. The SWord6k benchmark and reported baselines provide a practical testbed for future work on Roman-script STR, transfer learning from larger word components datasets, and improved architectures for complex scripts. The paper below follows the structure and spirit of the AUTNT transcription benchmark while grounding design choices in existing scene text recognition.