SOVEREIGN: How does content-adaptive tokenization compare to fixed-patch tokenization in terms of inference throughput (t

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 28, 2026

Abstract

In this paper, we propose a novel approach to address the challenges of printed Urdu text recognition using high-resolution, multi-scale semantic feature extraction. Our proposed UTRNet architecture, a hybrid CNN-RNN model, demonstrates state-of-the-art performance on benchmark datasets. To address the limitations of previous works, which struggle to generalize to the intricacies of the Urdu script and the lack of sufficient annotated real-world data, we have introduced the UTRSet-Real, a large-scale annotated real-world dataset comprising over 11,000 lines and UTRSet-Synth, a synthetic datase

1 Introduction

Analysis of: UTRNet: High-Resolution Urdu Text Recognition In Printed Documents. Research goal: How does content-adaptive tokenization compare to fixed-patch tokenization in terms of inference throughput (tokens/sec) and accuracy on the DocVQA and ChartQA benchmarks for high-resolution inputs?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

12 papers retrieved. 0 claims extracted, 0 verified. Tribunal: 3.7/10 → REJECT (revision_round=0). Policy: ESCALATE_TO_OWNER.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

References

- http://arxiv.org/abs/1708.02711v1
- http://arxiv.org/abs/2306.15782v3
- http://arxiv.org/abs/2507.07995v1