



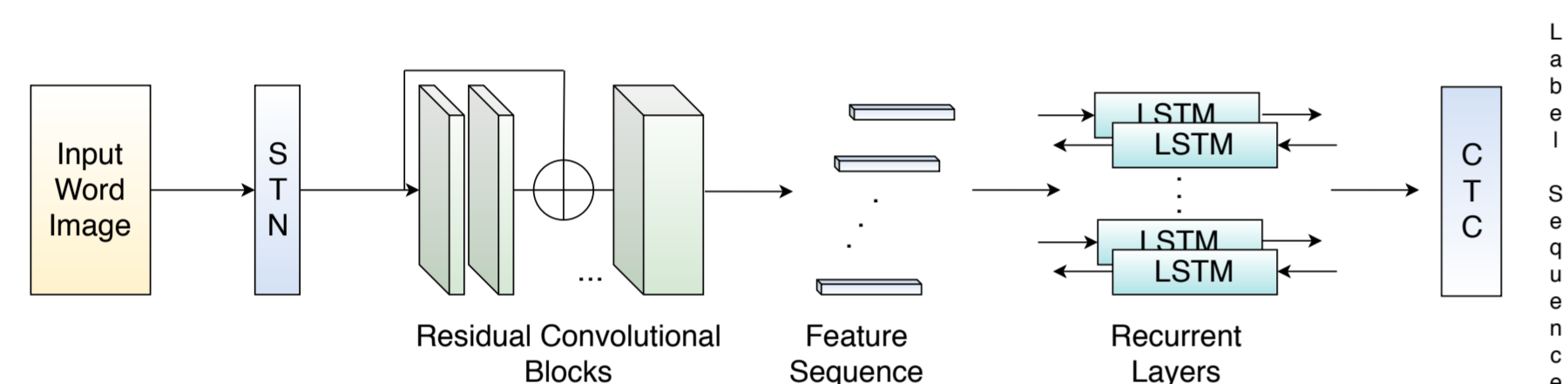
Improving Word Recognition and Retrieval

Motivation

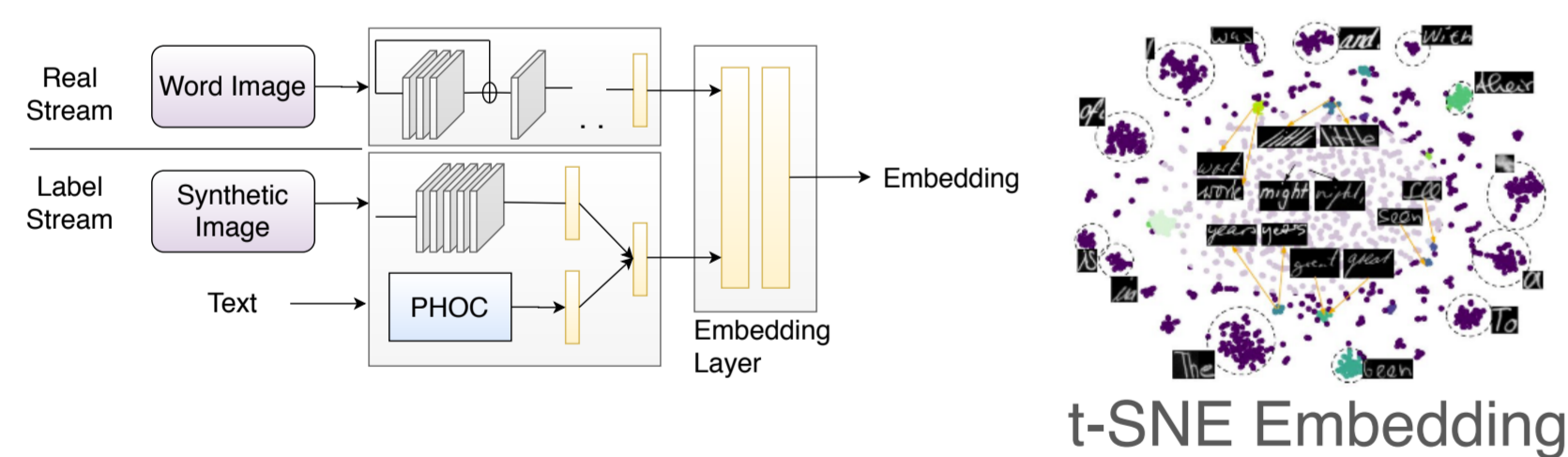
Enabling content level searching in scanned documents.

Text Recogniser and Deep Embeddings

- To convert handwritten human text to machine encoded text.

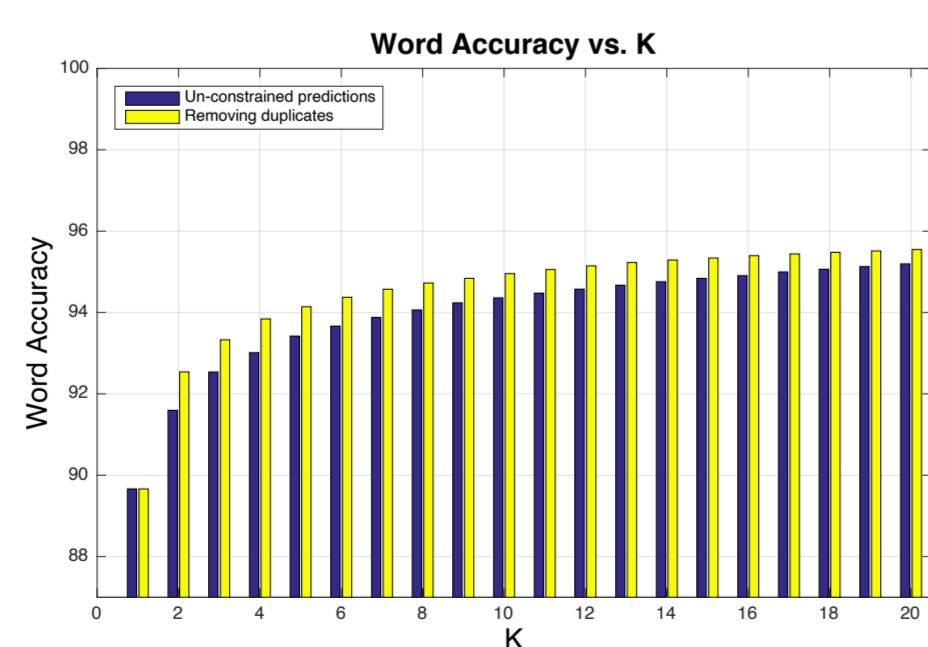


- End2End network learns a representation space which respects lexical similarity both in image and text domain.



Analysis

- Multiple hypotheses
- $K \propto$ word accuracy
- Deep embedding \rightarrow higher recall
- Text recogniser \rightarrow higher precision

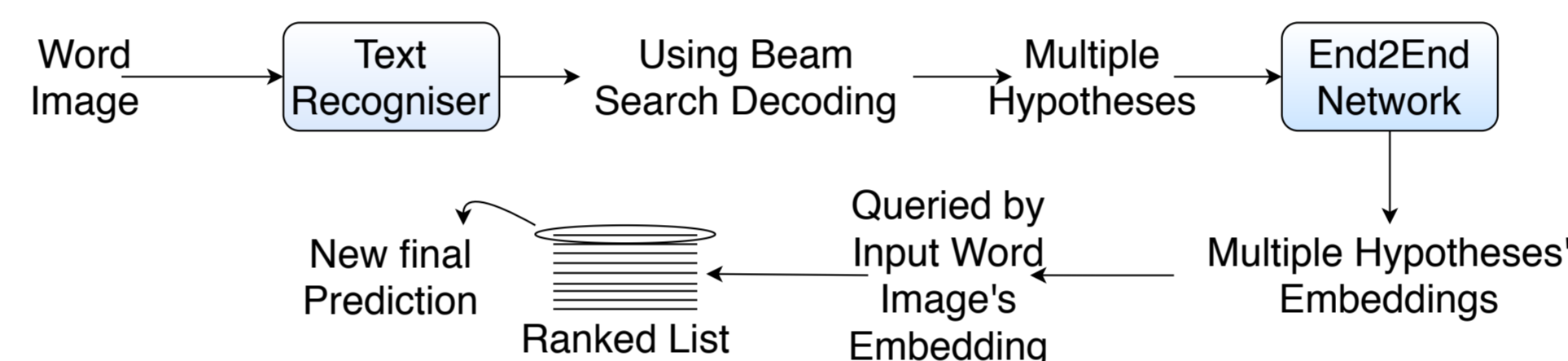


TopK Hypotheses Analysis

Fusing Text Recogniser and Deep Embeddings

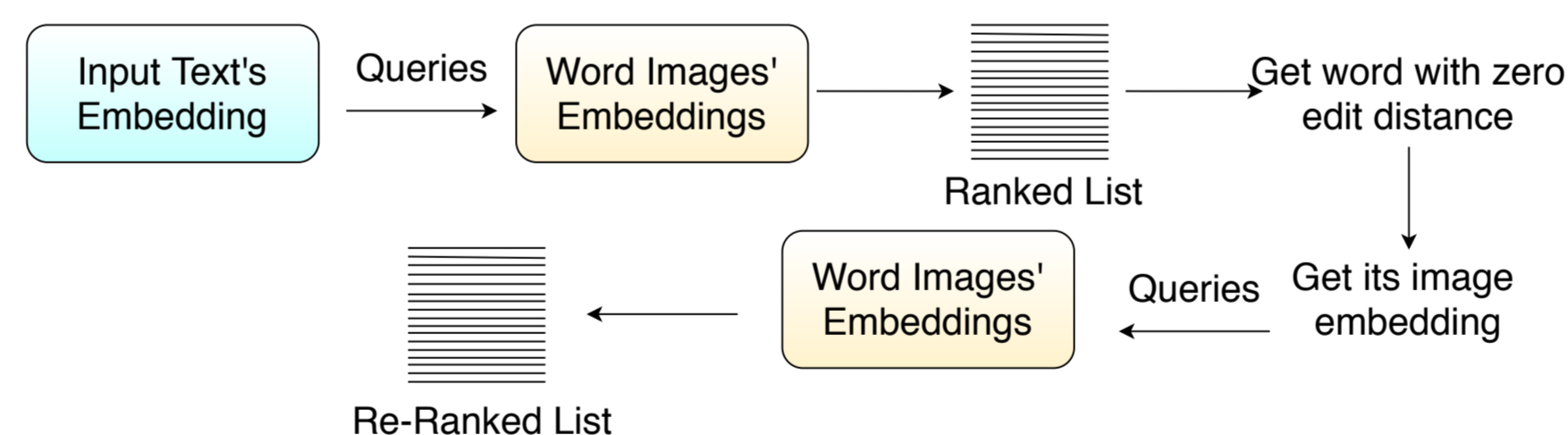
- Improving Word Recognition using Multiple Hypotheses

Multiple hypotheses' embeddings generated by the text recogniser for an input word image which are queried by the input word image's embedding to get a new final prediction.



- Improving Word Retrieval using Deep Embeddings

Ranked list is used to obtain image embedding of the word with zero edit distance w.r.t. the input text which is re-queried to generate a re-ranked list.



Applications

- Querying large documents' collections prepared by scanning books using Optical Character Recognition (OCR) techniques.
- Creating reliable text recogniser systems: An improvement of 1.4% in word accuracy for the Hindi language is observed using the methods.
- Automatic Annotations Tools.
- Universal methods: The methods proposed can be used with a variety of text recogniser and word spotting techniques.

Results

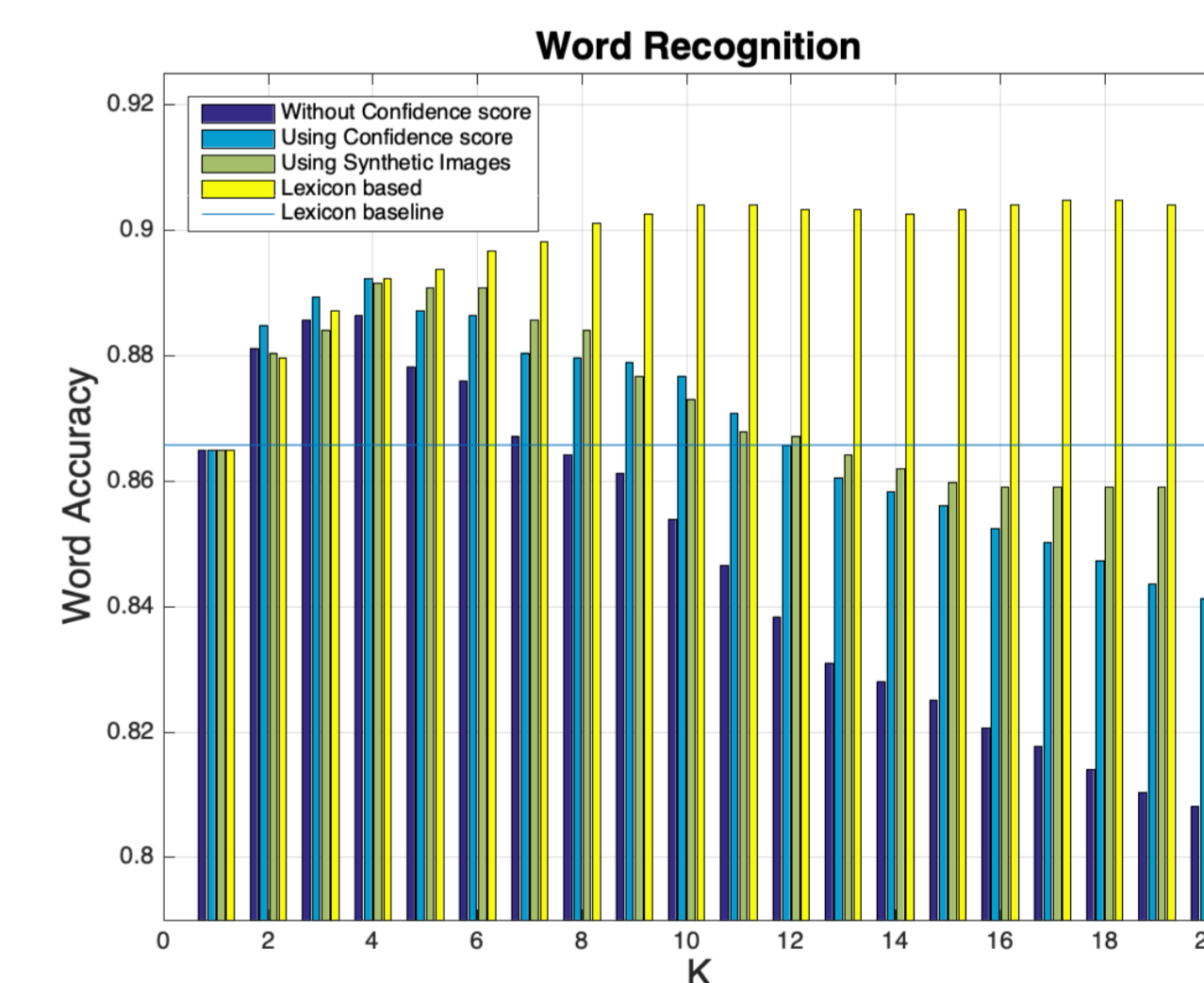
Word retrieval results

Method	Baseline	Naive Merge	Query Expansion	Avg. Fusion	Max Fusion
mAP	90.18	92.18	93.18	93.07	92.79



Retrieval results for the word 'Vatavaran'

Word recognition results



	Input Image	Baseline Prediction	Prediction using Confidence Score	Lexicon based Prediction
(a)	आलोचनाएं	आलोचनाएं	आलोचनाएं	आलोचनाएं
(b)	बेचैनी	बेचनी	बेचैनी	बेचैनी
(c)	बर्हिमुखी	बरहिमुखी	बरहिमुखी	बर्हिमुखी
(d)	टेलीफोन	टेलीफोन	टेलीफोन	टेलीफोन