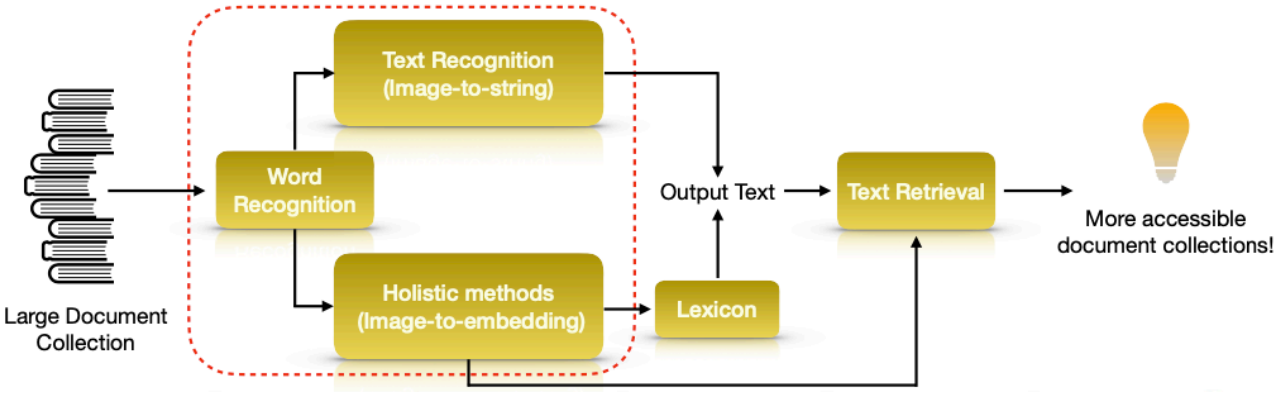
Improving Word Recognition using Multiple Hypotheses and Deep Embeddings

Siddhant Bansal, Praveen Krishnan, C.V. Jawahar
CVIT, IIIT, Hyderabad, India
ICPR 2020

Motivation

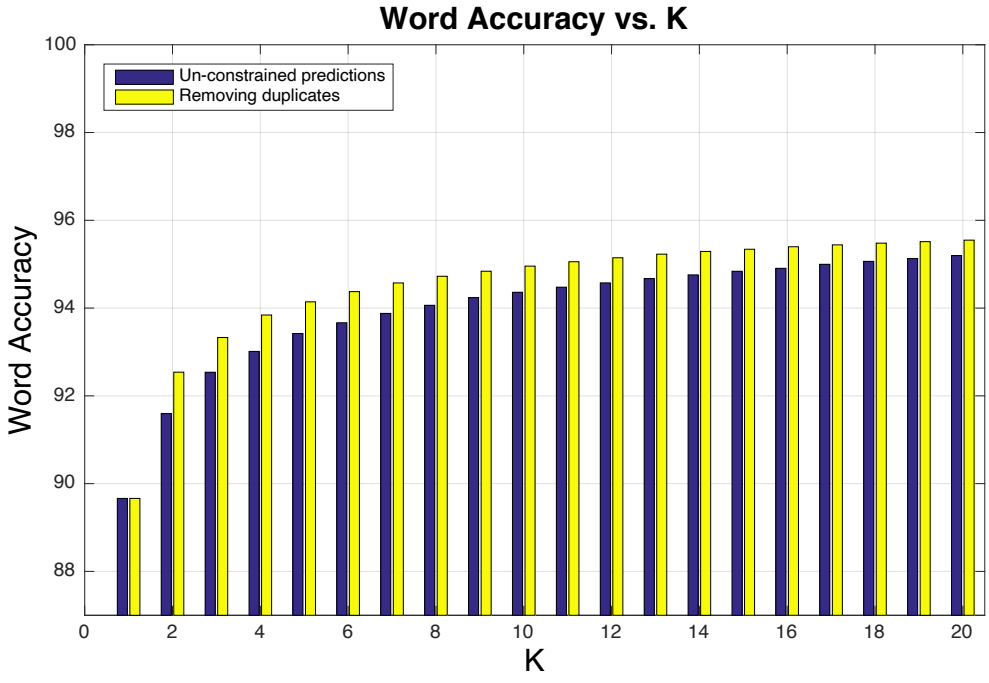
Enabling content level search in scanned documents.



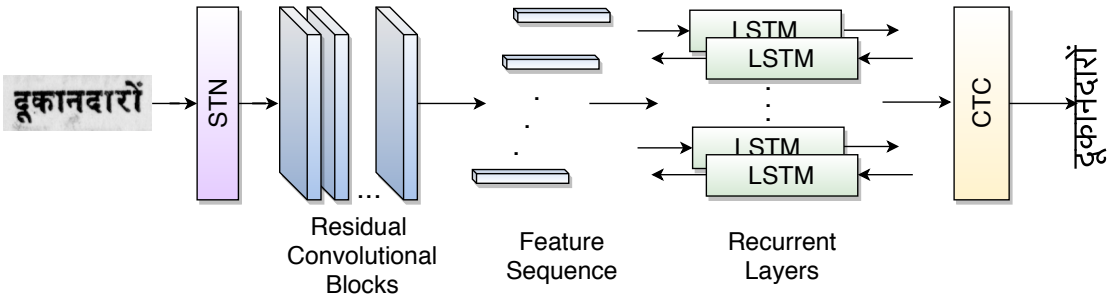
- We aim to make document collections more accessible by fusing text recognition and holistic approaches.

Analysis

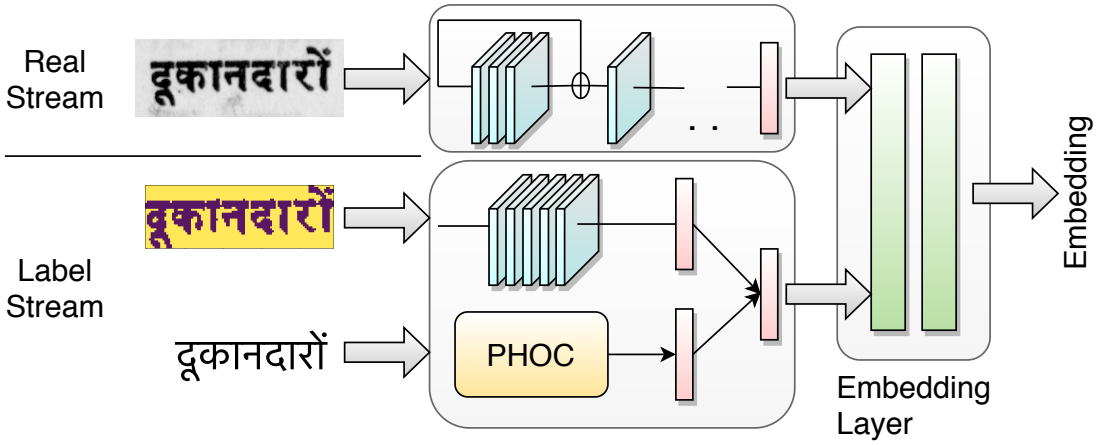
- We generate multiple hypotheses (K) using the beam search decoding algorithm. Increasing the value of K results in improvement in the word recognition accuracy.



Text Recogniser and Deep Embeddings



CRNN Architecture [2] takes in a word image and passes it through the Spatial Transform Layer (STN), followed by residual convolutional blocks which learn features maps. These feature maps are further given as an input to the BLSTM layer.

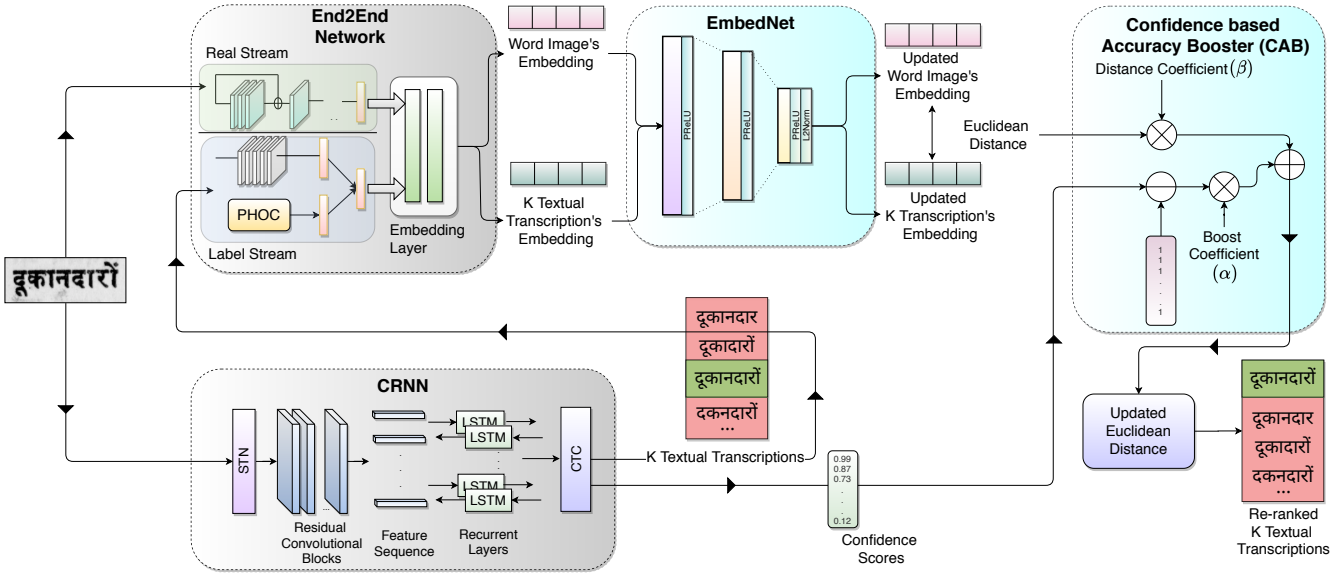


End2End network [3] for learning both textual and image embedding using a multi-task loss function.

Fusing Text Recogniser and Deep Embeddings

- We use EmbedNet for projecting deep embeddings to an updated embedding space.
- Once trained, the deep embeddings used for recognition purpose are first passed through the EmbedNet.

- To further improve the word recognition accuracy, we propose a novel plug-and-play module called Confidence based Accuracy Booster (CAB).
- CAB uses confidence scores and Euclidean distance.
- CAB favours words with higher confidence scores.
- Therefore, it helps in filtering out the noisy predictions.



- We pass the input word image through the CRNN [2] and the End2End network [3] and get multiple (K) hypotheses and their corresponding embeddings, respectively.
- The embeddings are then passed through the EmbedNet for generating an updated set of embeddings.
- The Euclidean distances between the embeddings and the confidence scores from the CRNN are then passed through the CAB module for generating updated distances. This helps in selecting the correct prediction.

Results

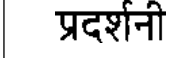
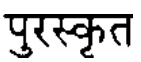
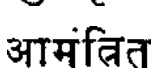
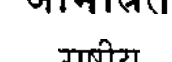
- Dataset

Language	Annotated	Number of Pages	Number of Word Images		
			Train	Validation	Test
Hindi	Yes	402	72,000	8,000	25,475

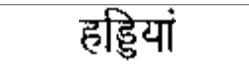
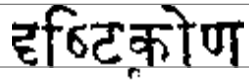
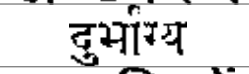
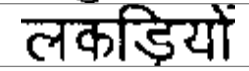
- Quantitative Results

Sr. No.	Method	WRA	K_{high} (K)
1	Tesseract	35.435	1 (1)
2	CRNN	81.543	1 (1)
3	E2E + C	83.062	2 (20)
4	E2E + C + CAB	84.358	11 (20)
5	MLP	83.259	3 (20)
6	EmbedNet	83.216	2 (20)
7	MLP + CAB	84.782	20 (20)
8	**EmbedNet + CAB**	**85.364**	**20 (20)**

- Qualitative Results

	Input Image	MLP	MLP + CAB	EmbedNet	EmbedNet + CAB
(a)		प्रदर्शनी	प्रदर्शनी	प्रदर्शनी	प्रदर्शनी
(b)		पुरस्कृत	पुरस्कृत	पुरस्कृत	पुरस्कृत
(c)		आमंत्रत	आमत्रित	आमंत्ित	आमंत्रित
(d)		राष्टय	राष्टीय	राष्टीय	राष्ट्रीय

- Corner Cases

Word Image	EmbedNet + CAB Predictions
	हड्डियां
	दछिटिकोण
	दुर्भीग्य
	लकडियों

Conclusion and Future Work

- We propose a new direction of improving word recognition by fusing word recognition and image embedding techniques.
- As future work, we aim to create an end-to-end architecture for fusion.

References

1. S. Bansal, P. Krishnan, and C. V. Jawahar, “Fused Text Recogniser and Deep Embeddings Improve Word Recognition and Retrieval,” in Document Analysis Systems (DAS), 2020.
2. K.Dutta, P.Krishnan, M.Mathew, and C.V. Jawahar, “ImprovingCNN-RNN Hybrid Networks for Handwriting Recognition,” in International Conference on Frontiers in Handwriting Recognition (ICFHR), 2018.
3. P. Krishnan, K. Dutta, and C. V. Jawahar, “Word Spotting and Recognition Using Deep Embedding,” in IAPR International Workshop on Document Analysis Systems (DAS), 2018.