Improving Word Recognition using Multiple Hypotheses and Deep Embeddings

Siddhant Bansal, Praveen Krishnan, and C.V. Jawahar
CVIT, IIIT Hyderabad

CVIT

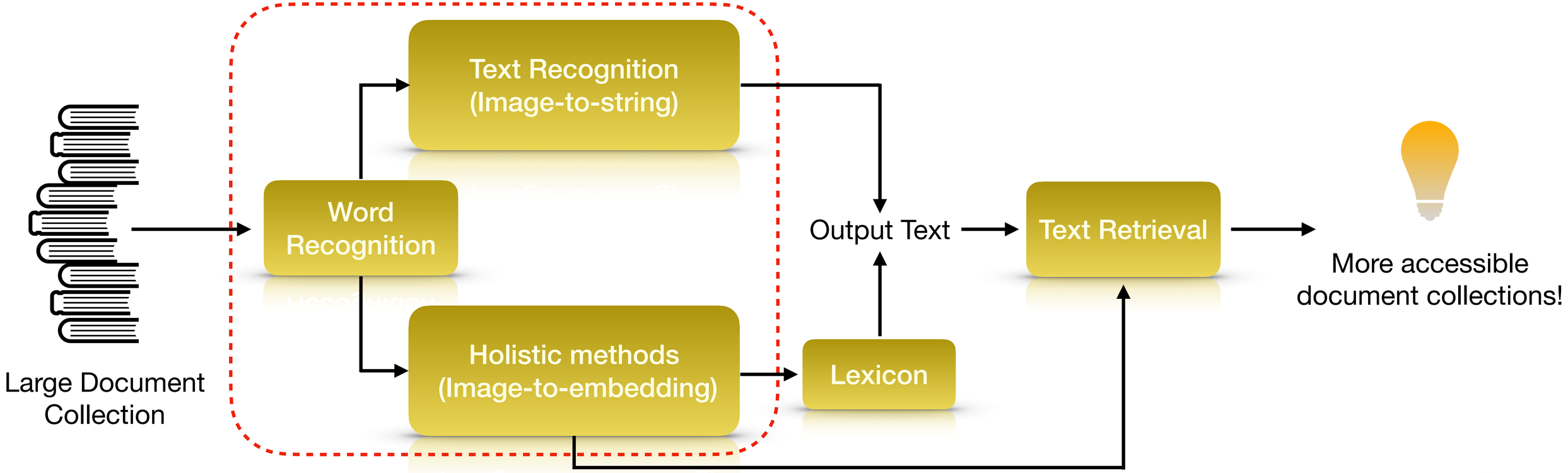
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD

Motivation

Motivation

Large Document Collections

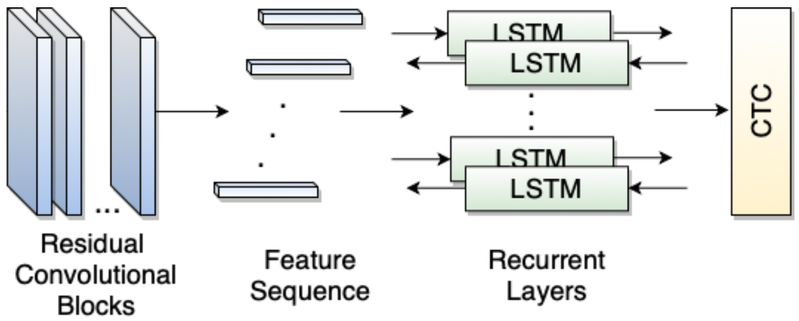
- Large Document Collections: Digital Library of India (DLI), Gutenberg Project.



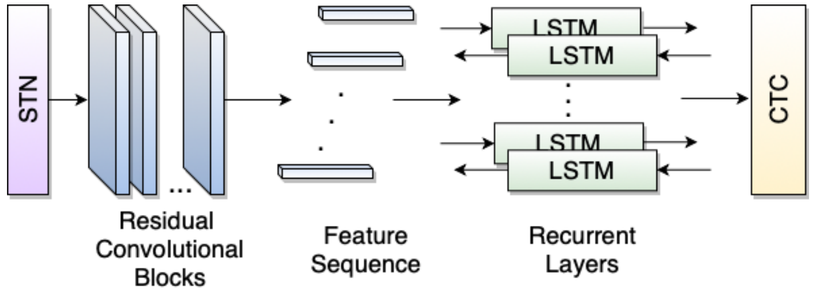
Text Recognition Methods

Word image-to-string

- Convolutional Recurrent Neural Network (CRNN) proposed by Shi et al.



- CRNN with Spatial Transformer Network (STN) proposed by Dutta et al.



Shi, B., Bai, X., Yao, C.: An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition. In PAMI 2017.
Dutta,K., Krishnan,P., Mathew,M., Jawahar,C.V.: Improving CNN-RNN Hybrid Networks for Handwriting Recognition. In ICHFR 2018.

Holistic Methods

Deep Embeddings

- Learn a holistic representation

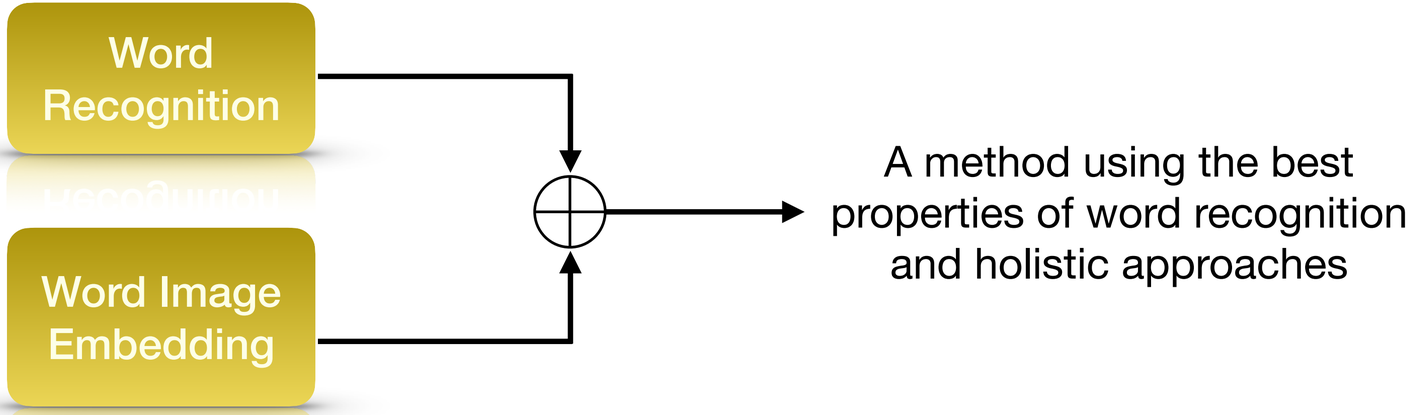
- **Deep learning** based methods for generating the representation
  - **PHOCNet** by Sudholt, S., Fink, G.A.
  - **HWNet** by Krishnan, P., Jawahar, C.V.

S. Sudholt, G. A. Fink: PHOCNet: A Deep Convolutional Neural Network for Word Spotting in Handwritten Documents. In ICFHR 2016.
Krishnan, P., Jawahar, C.V. HWNet v2: An efficient word image representation for handwritten documents. In *IJDAR 2019*

Creating a better system

Fusion

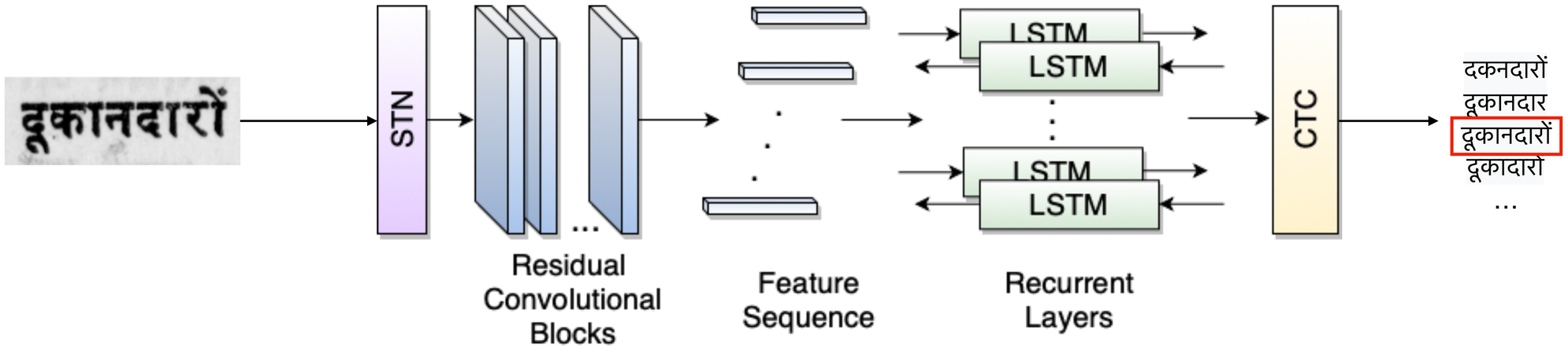
- Should we improve word recognition methods or holistic methods individually or should we fuse them?



Multiple Hypotheses

Effects of generating and using multiple hypotheses

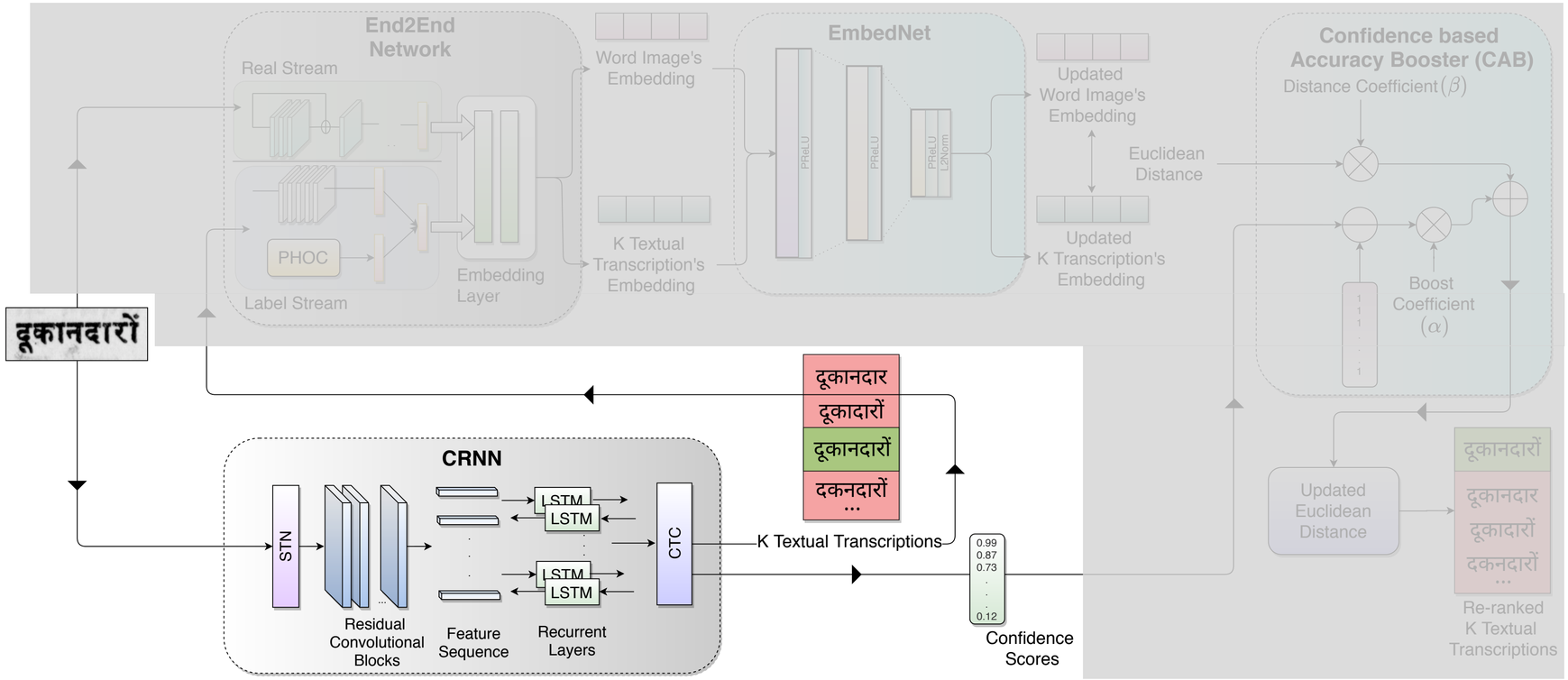
- Generating multiple (*K*) hypotheses using the beam-search decoding algorithm.
- Increasing the value of *K* leads to an increase in the word recognition accuracy.



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.

Baseline

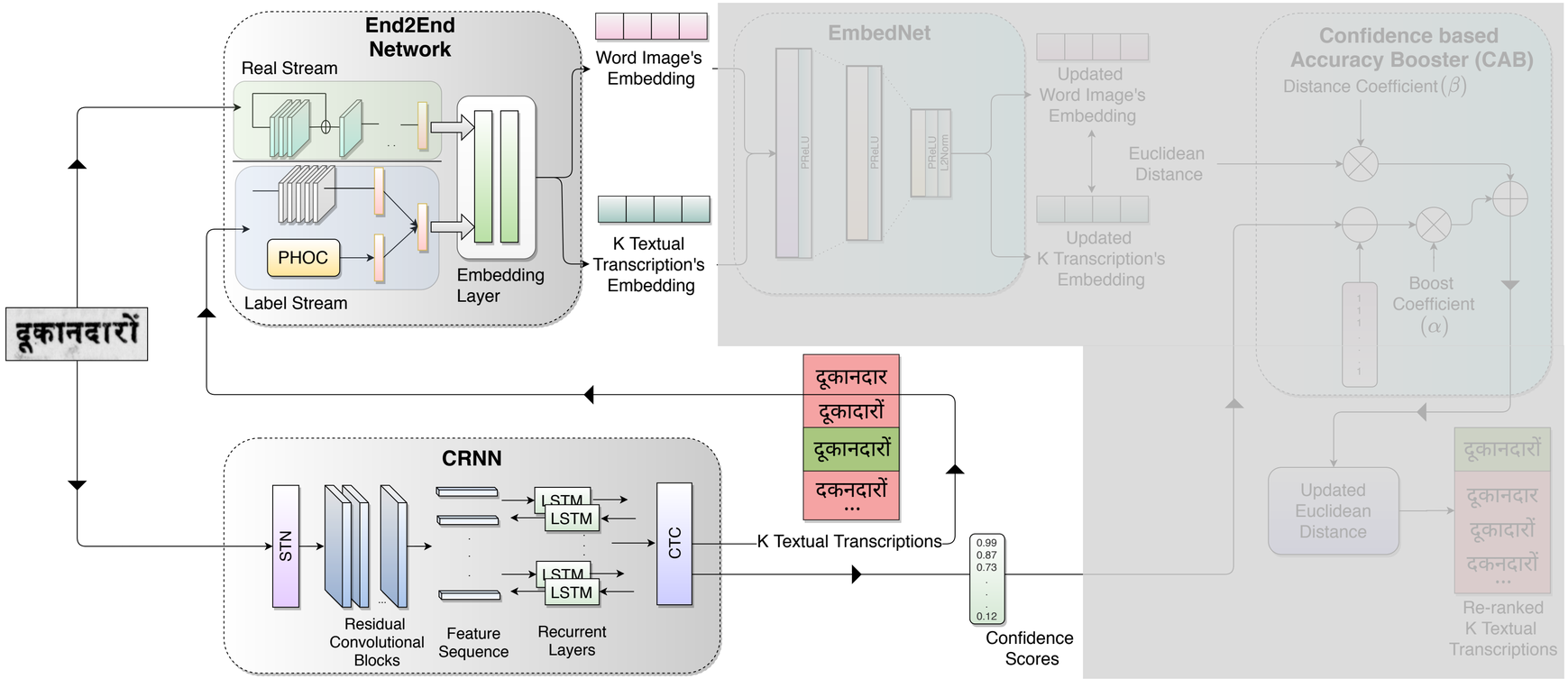
Text Recognition using CRNN



Dutta,K., Krishnan,P., Mathew,M., Jawahar,C.V.: Improving CNN-RNN Hybrid Networks for Handwriting Recognition. In ICHFR 2018.

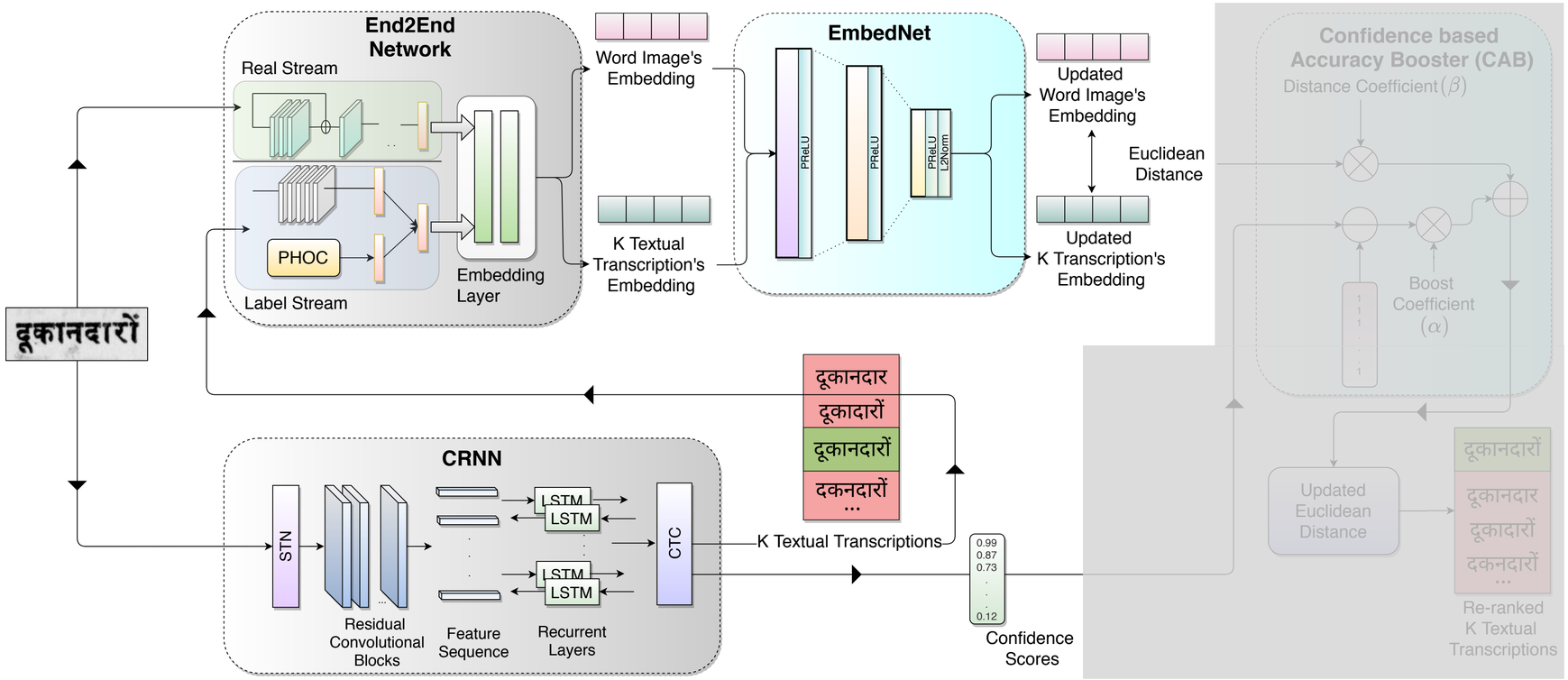
Baseline

End2End Network



Methods

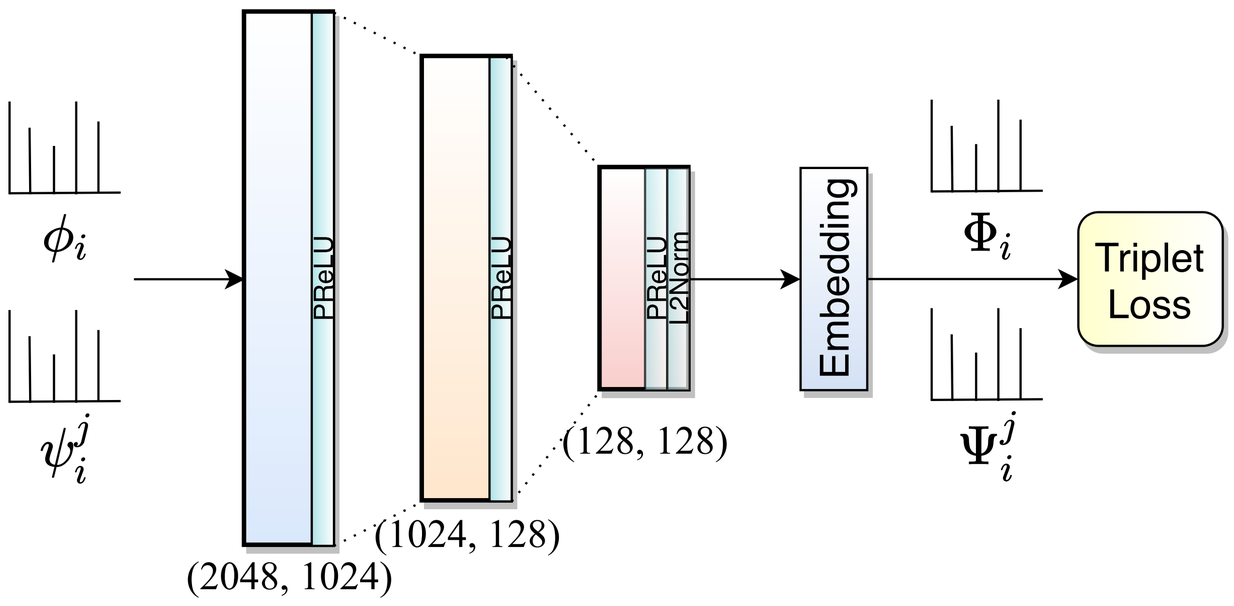
EmbedNet - Learning an updated embedding space



EmbedNet

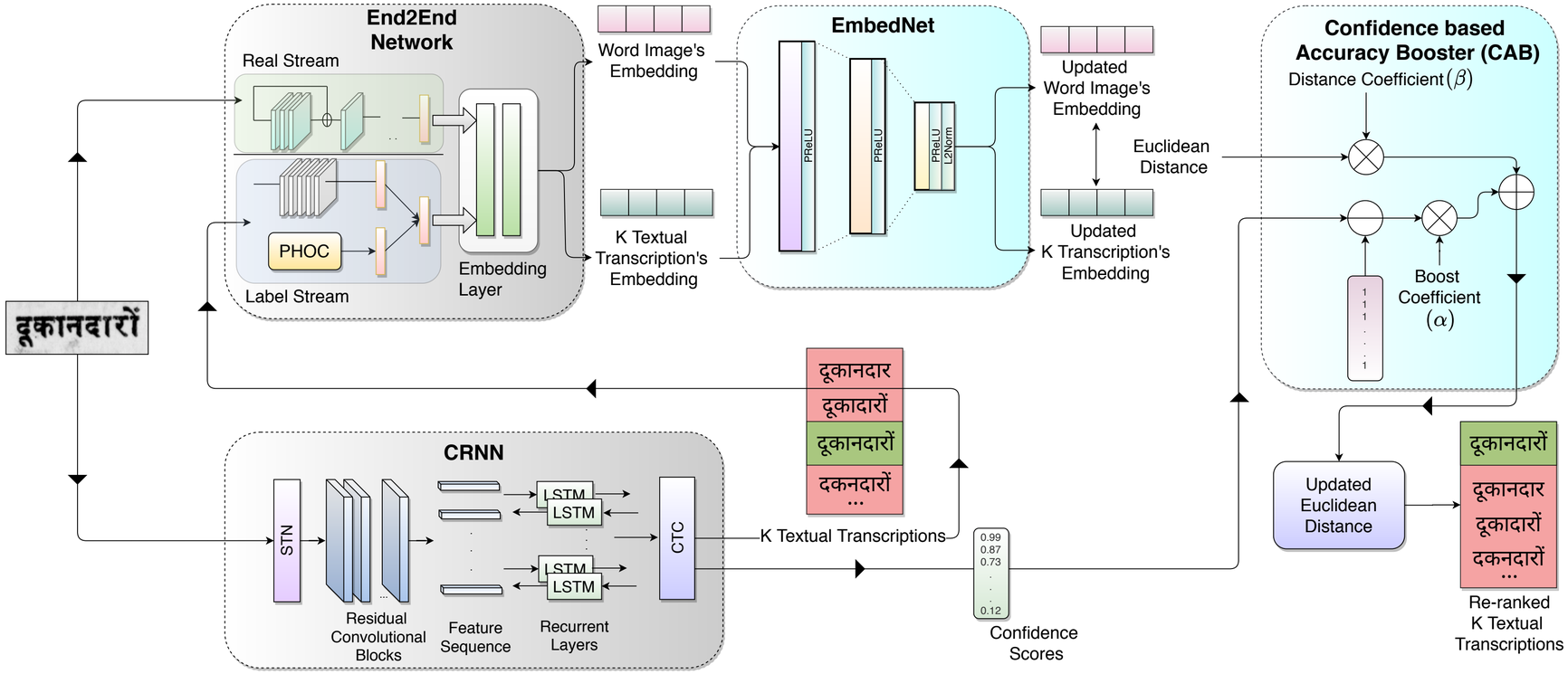
Learning an updated embedding space

- Projects embeddings to an updated embedding space.
- Trained using triplet loss.
  - Hard Negatives
  - Semi-hard Negatives
  - Easy Negatives
- Once trained, embeddings are updated.



Methods

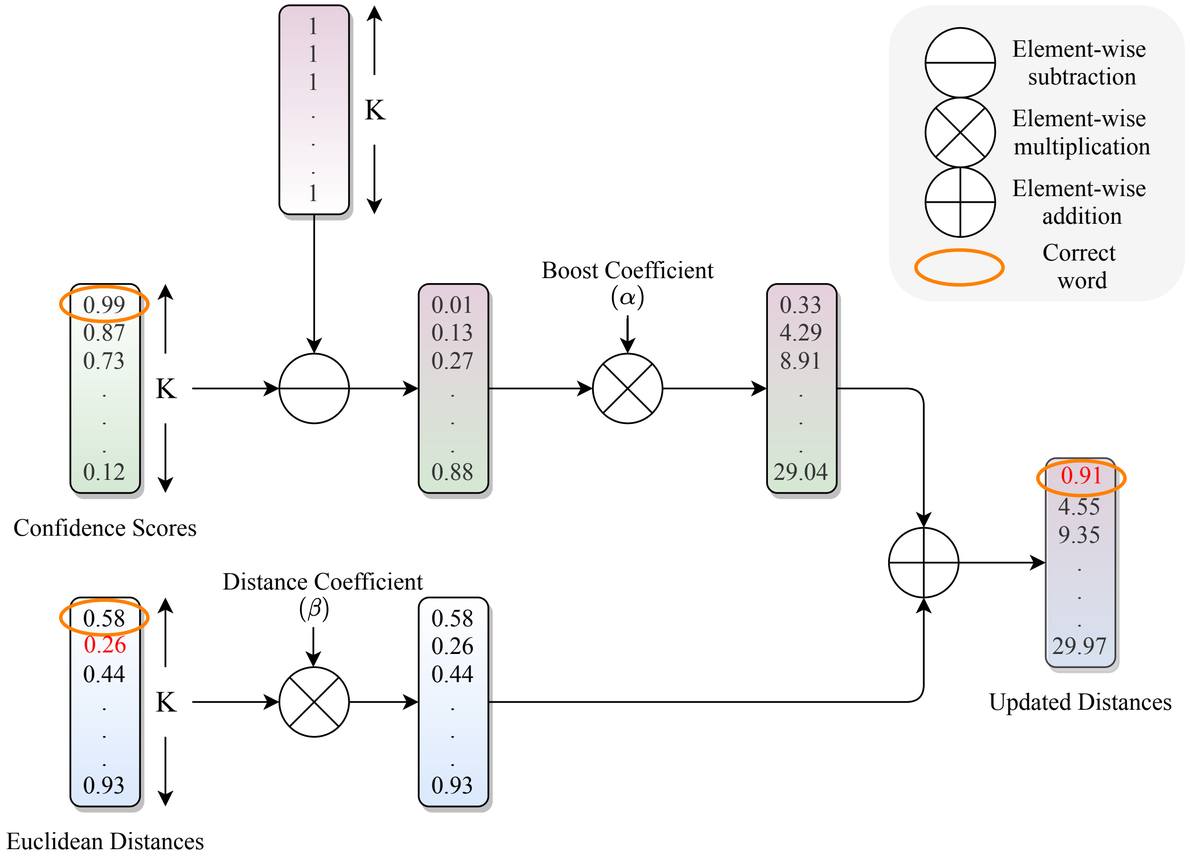
CAB - Further improving the word recognition accuracy



Confidence based Accuracy Booster (CAB)

Further improving the word recognition accuracy

- CAB uses confidence scores and Euclidean distance.
- It generates updated distances.
- Words with higher confidence scores are favoured.
- Helps in filtering out noisy predictions.



Results

Results

Dataset

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

Results

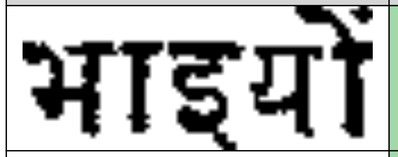
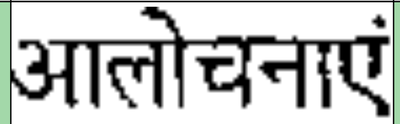
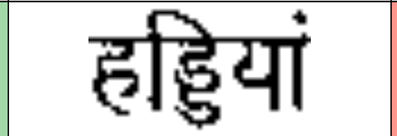
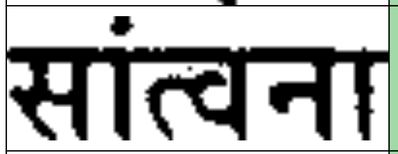
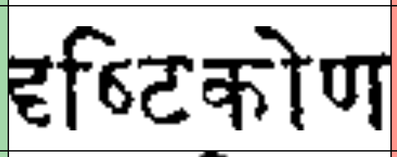
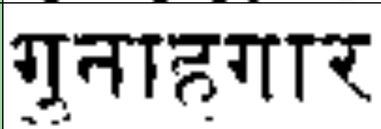
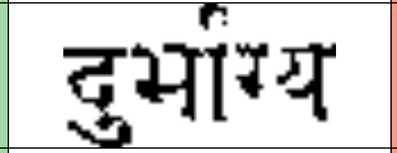
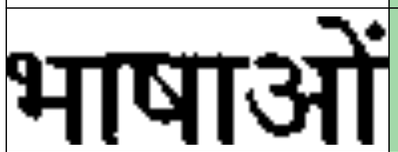
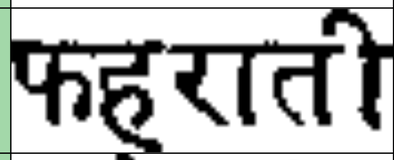
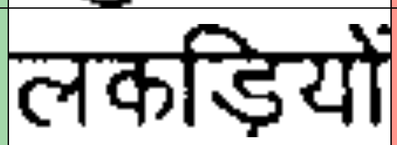
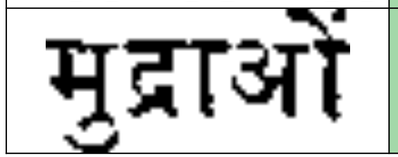
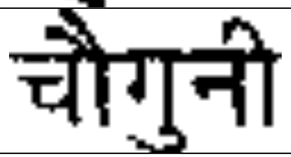
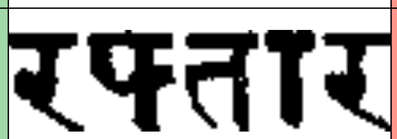
Quantitative

Sr. No.	Method	WRA	K_{high} (K)
1	Tesseract [1]	35.435	1 (1)
2	CRNN [2]	81.543	1 (1)
3	E2E + C [3]	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)**

[1] R. Smith, “An Overview of the Tesseract OCR Engine,” in International Conference on Document Analysis and Recognition (ICDAR), 2007.
[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] 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.

Results

Qualitative

Word Image	EmbedNet + CAB predictions	Word Image	EmbedNet + CAB predictions	Word Image	EmbedNet + CAB predictions
	भाइयों		आलोचनाएं		हड्डियां
	सांत्वना		चिंतन		दछिटिकोण
	त्यागना		गुनाहगार		दुर्भाग्य
	भाषाओं		फहराती		लकडियों
	मुद्राओं		चौगुनी		रफ्तार

Conclusion and Future Work

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.

Thank You!