

Transfer Learning for Scene Text Recognition in Indian Languages

Sanjana Gunna

IIIT Hyderabad

Rohit Saluja

IIIT Hyderabad

C. V. Jawahar

IIIT Hyderabad

<https://github.com/firesans/STRforIndicLanguages>

Centre for Vision Information Technology (CVIT)

International Institute of Information Technology, Hyderabad (IIIT-H), INDIA

Contents

- ❑ Introduction
- ❑ Motivation
- ❑ Data & Methodology
- ❑ Methods & Experiments
- ❑ Results
- ❑ Conclusion

Introduction

The recent algorithms proposed for Scene Text Recognition (STR) in English perform with an average Word Recognition Rate (WRR) of 92.9% on 4 benchmark datasets.

Model	Year	Train Data	IIIT	SVT	IC03	IC13
GCAM	2019	MJ + ST	93.9	91.3	95.3	95.7
GTC [Hu et al.]	2020	MJ + ST + SA + real (5.6M)	95.8	92.9	95.5	94.4
Luo et al.	2020	MJ + ST	95.4	92.7	96.3	94.8
Litman et al.	2020	MJ + ST+ SA	93.7	92.7	96.3	93.9

Table 1 - Performance comparison of recognition algorithms on benchmark datasets
(the values correspond to the “No lexicon” results)

- ❑ However, these models have not performed well on non-Latin datasets. The domain of STR in non-Latin languages has not been thoroughly explored as much as English.
- ❑ We set new benchmarks for 6 different Indian Languages (Next Slide).

Hindi / Devanagari

Bangla

Gujarati

Ground truth

Bounding Box

Baseline

Predictions

Our Model

Predictions

आप्रवासन
Immigration

आप्रवासन

आप्रवासन
आप्रवासन

K M GHOSH ROAD

के एम घोष रोड
के एम घोष रोड

के एम घोष रोड
के एम घोष रोड

સ્ટેશન સ્વરછ મેળવી
સ્ટેશન સ્વરછ મેળવી
આભાર આભાર

સ્ટેશન સ્વરછ મેળવી આભાર
સ્ટેશન સ્વરછ મેળવી આભાર

Telugu

Tamil

Malayalam

అనంతర

అనంతర
ఫైన్ డ్రైనింగ్ బ్యాంక్

ఫైన్ డ్రైనింగ్ బ్యాంక్

అనంతర బ్యాంక్ డ్రైనింగ్ ఫైన్
అనంతర బ్యాంక్ డ్రైనింగ్ ఫైన్

உதகமண்டலம்

உதகமண்டலம்
உதகமண்டலம்
UDAGAMANDALAM

பரசமண்டலம்
உதகமண்டலம்

എച്ച്.പി.ഓട്ടോ കെയർ
KOWDIAR

എച്ച്.പി.ഓട്ടോ കെയർ

കവടിയാർ

എച്ച്.പി.ഓട്ടോ കെയർ കവടിയാർ
എച്ച്.പി.ഓട്ടോ കെയർ കവടിയാർ

The correct predictions are in green, and the wrong predictions or missing characters are highlighted in red color.



ஆபரவாசன

ஆபரவாசன
ஆபரவாசன



கே எம் ஷோய ரோட
கே எம் ஷோய ரோட

Handwritten Sign Boards

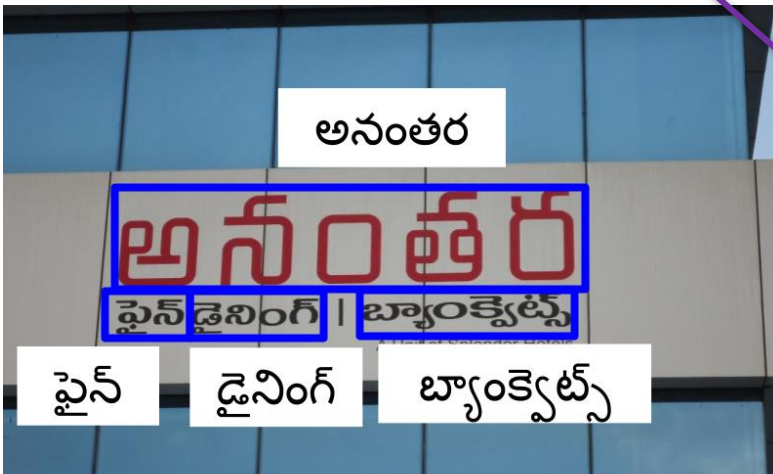
கே	எம்	ஷோய	ரோட
கே	எம்	ஷோய	ரோட



ஸ்டேஷன ஸ்வரஶ் மெலலி
ஸ்டேஷன ஸ்வரஶ் மெலலி
ஆலார

Printed Sign Boards

ஸ்டேஷன	ஸ்வரஶ்	மெலலி	ஆலார
ஸ்டேஷன	ஸ்வரஶ்	மெலலி	ஆலார



அனந்தர்

அனந்தர் டைனிங் டைனிங்

அனந்தர்	டெனிங்	டெனிங்	பைன்
அனந்தர்	டெனிங்	டெனிங்	பைன்



உதகமண்டலம்
English on sign boards

பரசமண்டலம்
உதகமண்டலம்



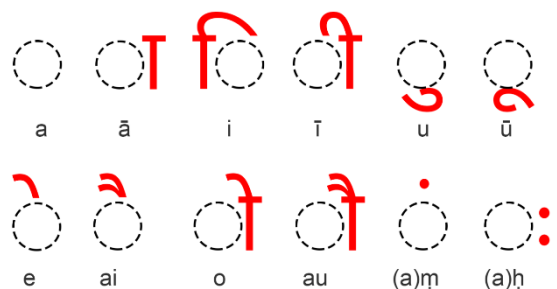
கவுசியார் கவுசியார்
கவுசியார்

கவுசியார்	கவுசியார்	கவுசியார்
கவுசியார்	கவுசியார்	கவுசியார்

Motivation

Motivation

- ❑ We compared n-grams of a corpus of 2 million words of the six languages in this study and found many striking similarities.
- ❑ The overall distribution curves of n-grams seen similar for all the languages. (Ex: Fig.1, Fig.2)
- ❑ The top 5 n-grams are usually not well-formed words in these languages unlike in English.



Diacritic characters in Hindi



"Halanta"

Fig.1: Top 5 n-gram Distribution Curve - Hindi

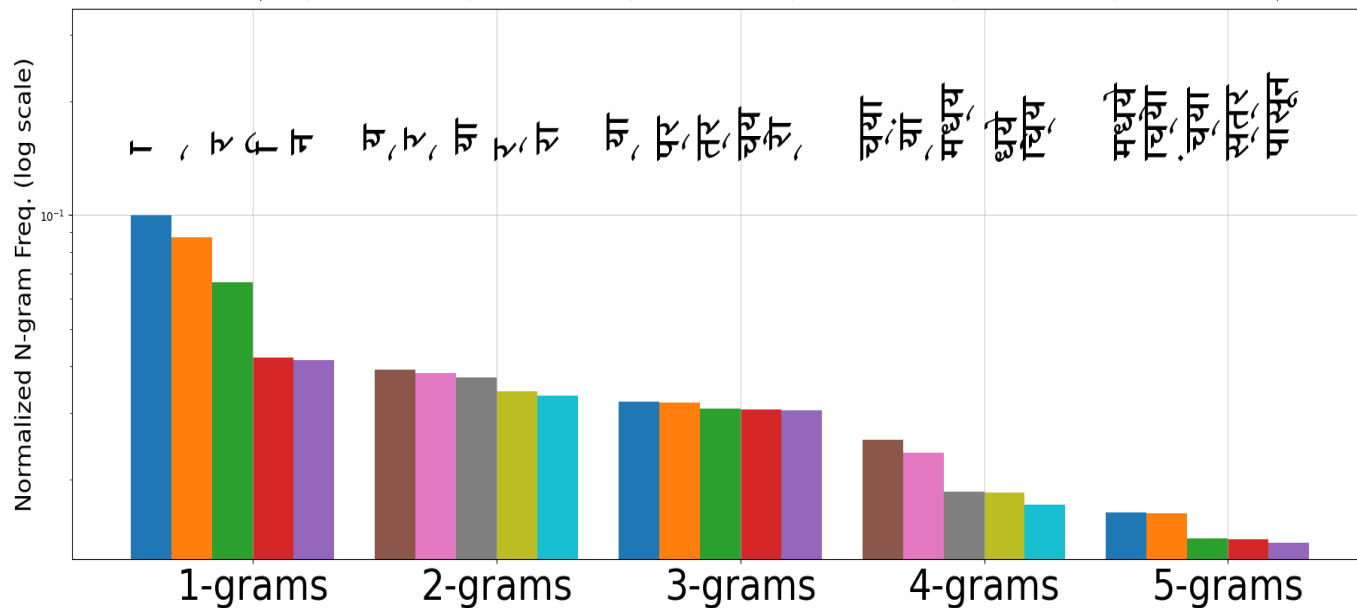
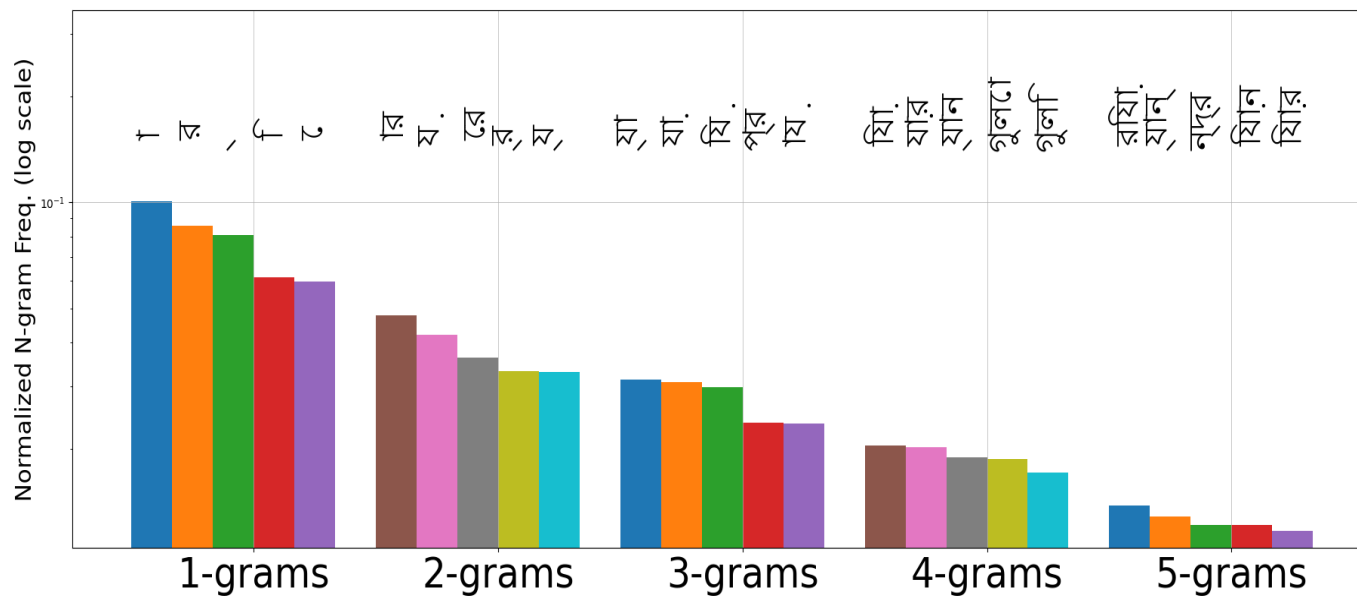


Fig.2: Top 5 n-gram Distribution Curve - Bangla



Motivation

“Use of transfer learning boosts the performance of the task amongst languages.”

□ Based on the word length statistics and the similarity between scripts, we choose the below mentioned order for transfer learning.

- Gujarati -> Hindi
- Hindi -> Bangla
- Bangla -> Tamil
- Tamil -> Telugu
- Telugu -> Malayalam

“Shirorekha”
(top-connector line)
Similar vowel
appearance

English
Gujarati
Hindi
Bangla
Tamil
Telugu
Malayalam



Data & Methodology

Languages & Data

- ❑ Our study focuses on 6 most popular Indian languages - Bengali, Gujarati , Hindi, Malayalam, Tamil, Telugu.
- ❑ We create >2.5 million synthetic data for training recognition algorithms using the methodology proposed by Mathew et al.

Language	#Images	Train	Test	(μ , σ) Word Length	#Fonts
English	17.5M	17M	0.5M	5.12, 2.99	>1200
Gujarati	2.5M	2M	0.5M	5.95, 1.85	12
Hindi	2.5M	2M	0.5M	8.73, 3.10	97
Bengali	2.5M	2M	0.5M	8.48, 2.98	68
Tamil	2.5M	2M	0.5M	10.92,3.75	158
Telugu	5M	5M	0.5M	9.75, 3.43	62
Malayalam	7.5M	7M	0.5M	12.29, 4.98	20



आप्रवासन



के एम घोष रोड
के एम घोष रोड



स्टेशन स्वरघ मेगपो
स्टेशन स्वरघ मेगपो
आभार
आभार

(#Actual S (s)

- Gu (words (105 scene
- Tamil - 2535 words (345 scenes)

के	एम	घोष	रोड
के	एम	घोष	रोड

रटशम	रवरघ	मवपो	अभार
स्टेशन	रवरघ	मोपो	आभार



अनंतर



உதகமண்டலம்



എച്ച്.പി.ഓട്ടോ കെയർ
എച്ച്.പി.ഓട്ടോ കെയർ
കവടിയാർ
കവടിയാർ

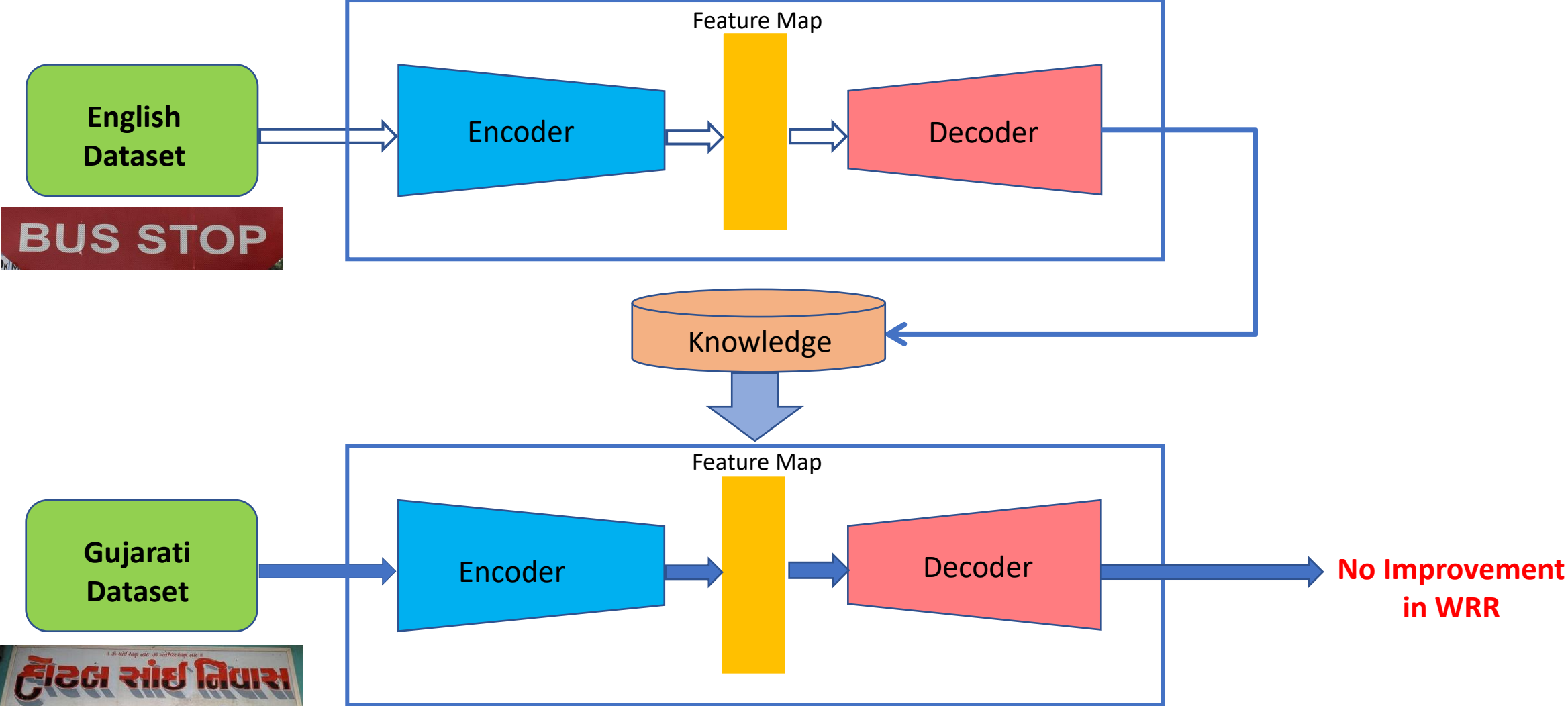
अनंतर
अनंतर

अनंतर	अनंतर	अनंतर	अनंतर
अनंतर	अनंतर	अनंतर	अनंतर

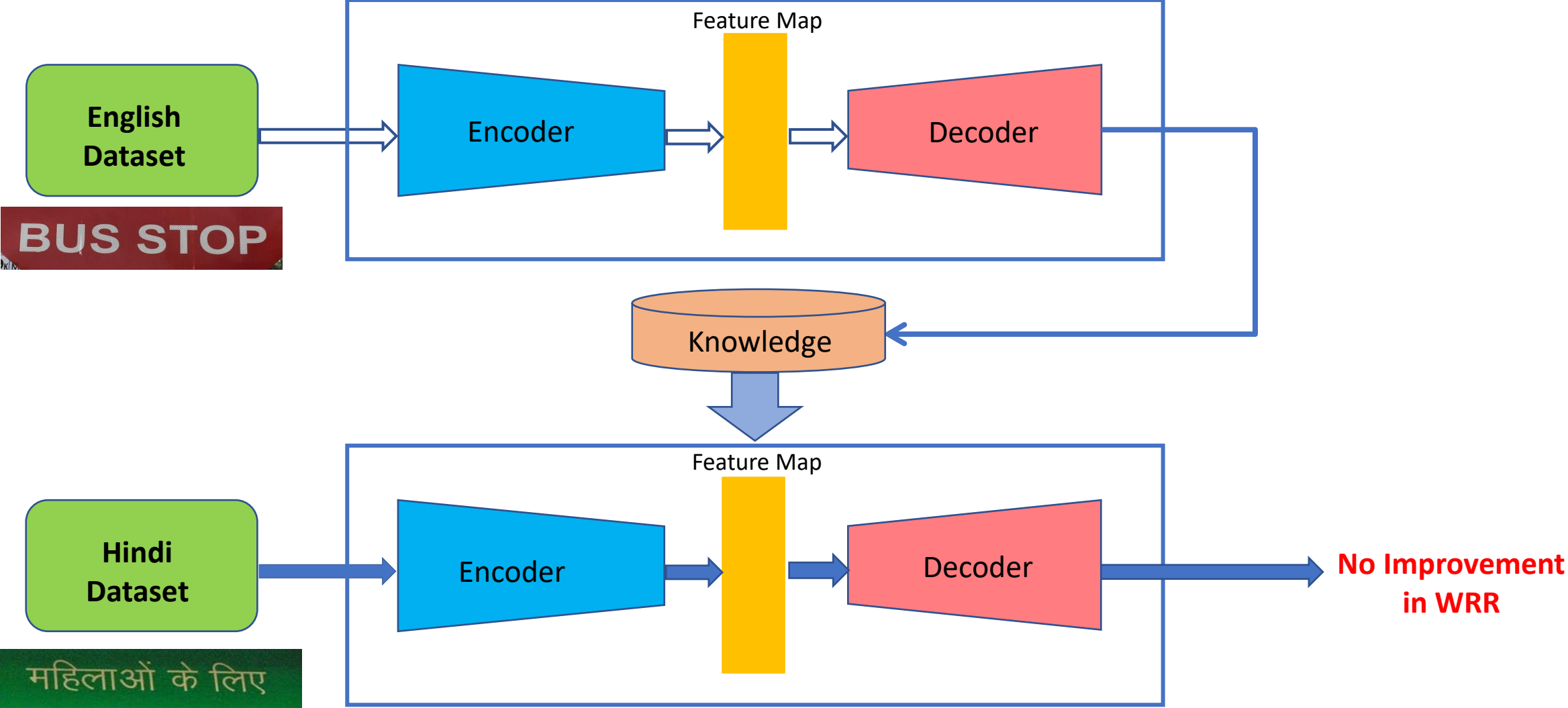
பரசமண்டலம்	உதகமண்டலம்
உதகமண்டலம்	உதகமண்டலம்

എച്ച്_പി_ഓട്ടോ	കെയർ	കവടിയാർ
എച്ച്_പി_ഓട്ടോ	കെയർ	കവടിയാർ

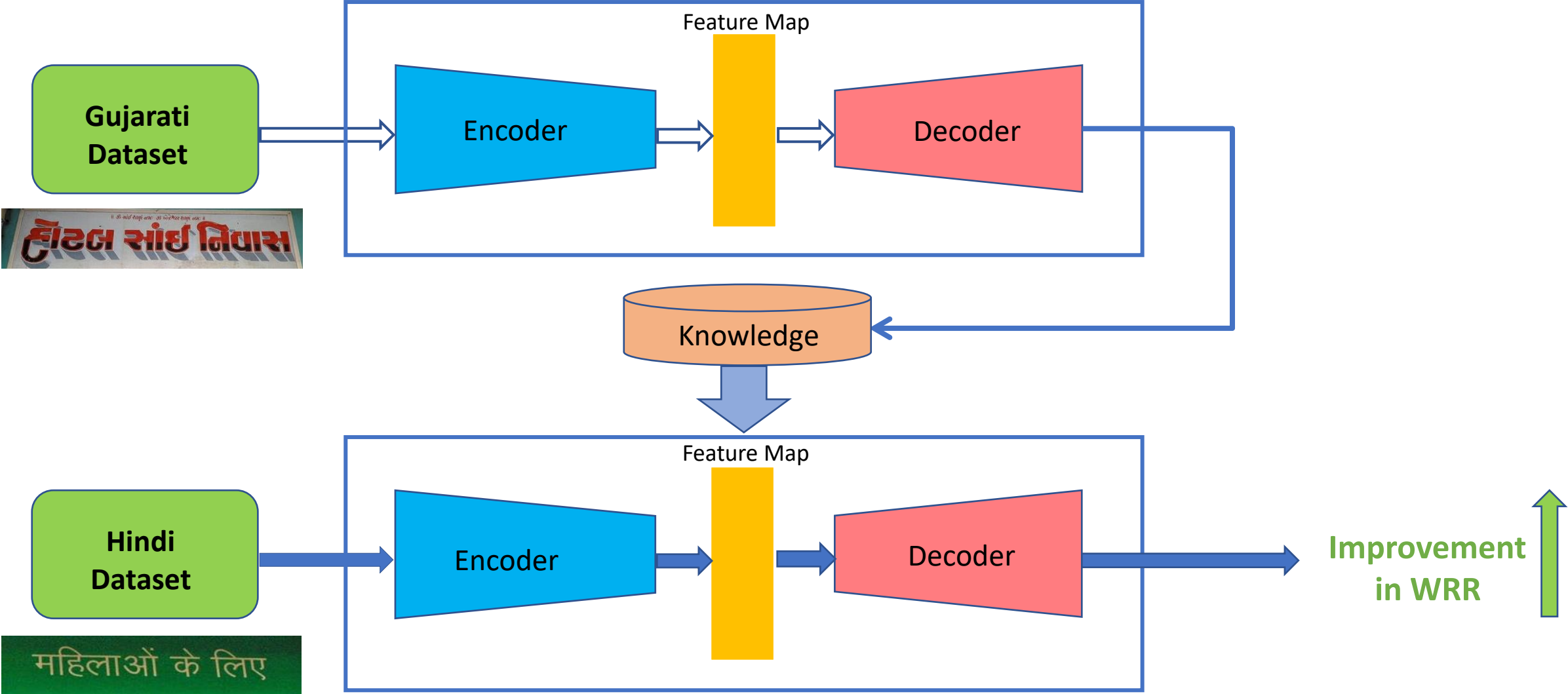
Our Contribution : Transfer Learning



Our Contribution : Transfer Learning



Our Contribution : Transfer Learning



Visualizations

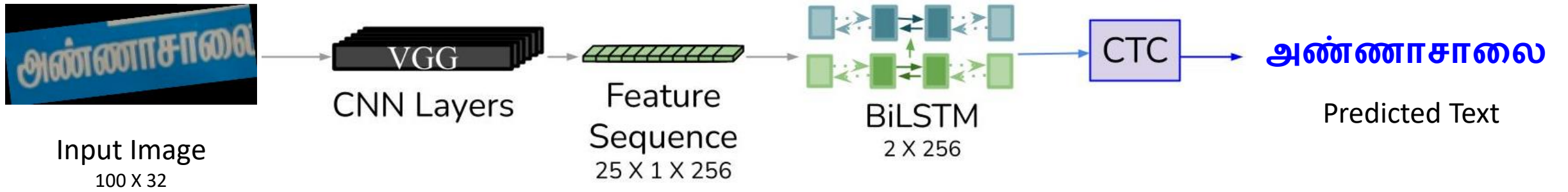
CRNN MODEL	WORD IMAGE	LAYER-1 FEATURES	LAYER-2 FEATURES	LAYER-3 FEATURES	LAYER-4 FEATURES
Hindi					
English -> Hindi					
Gujarati -> Hindi					
Hindi					
English -> Hindi					
Gujarati -> Hindi					
Hindi					
English -> Hindi					
Gujarati -> Hindi					

Activations pertaining to English transfer are shown in red boxes, which are not happening in better models, i.e. (Hindi and Gujarati → Hindi)

Methods & Experiments

Pipeline : Overview

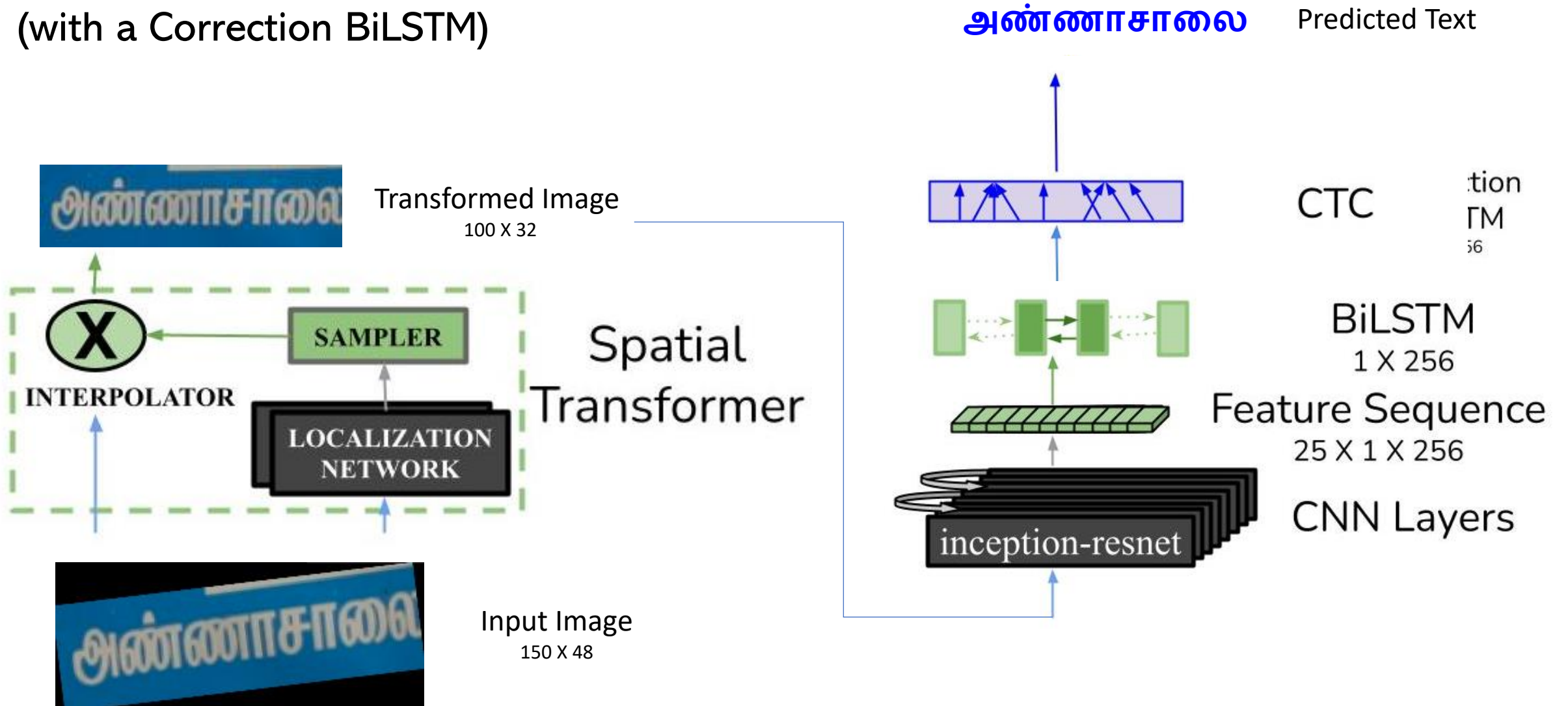
CRNN



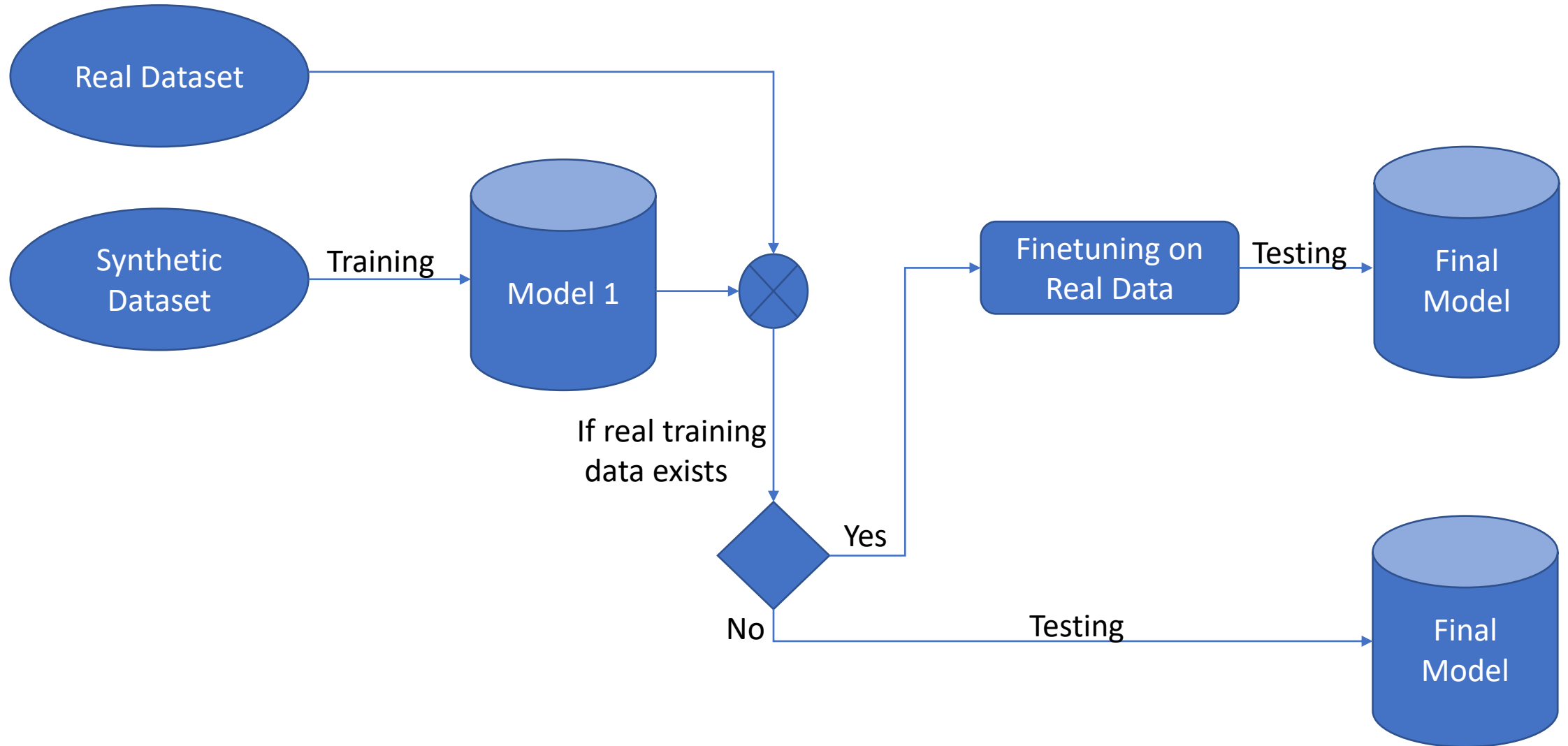
Pipeline : Overview

STAR-Net

(with a Correction BiLSTM)



Experiments



Results

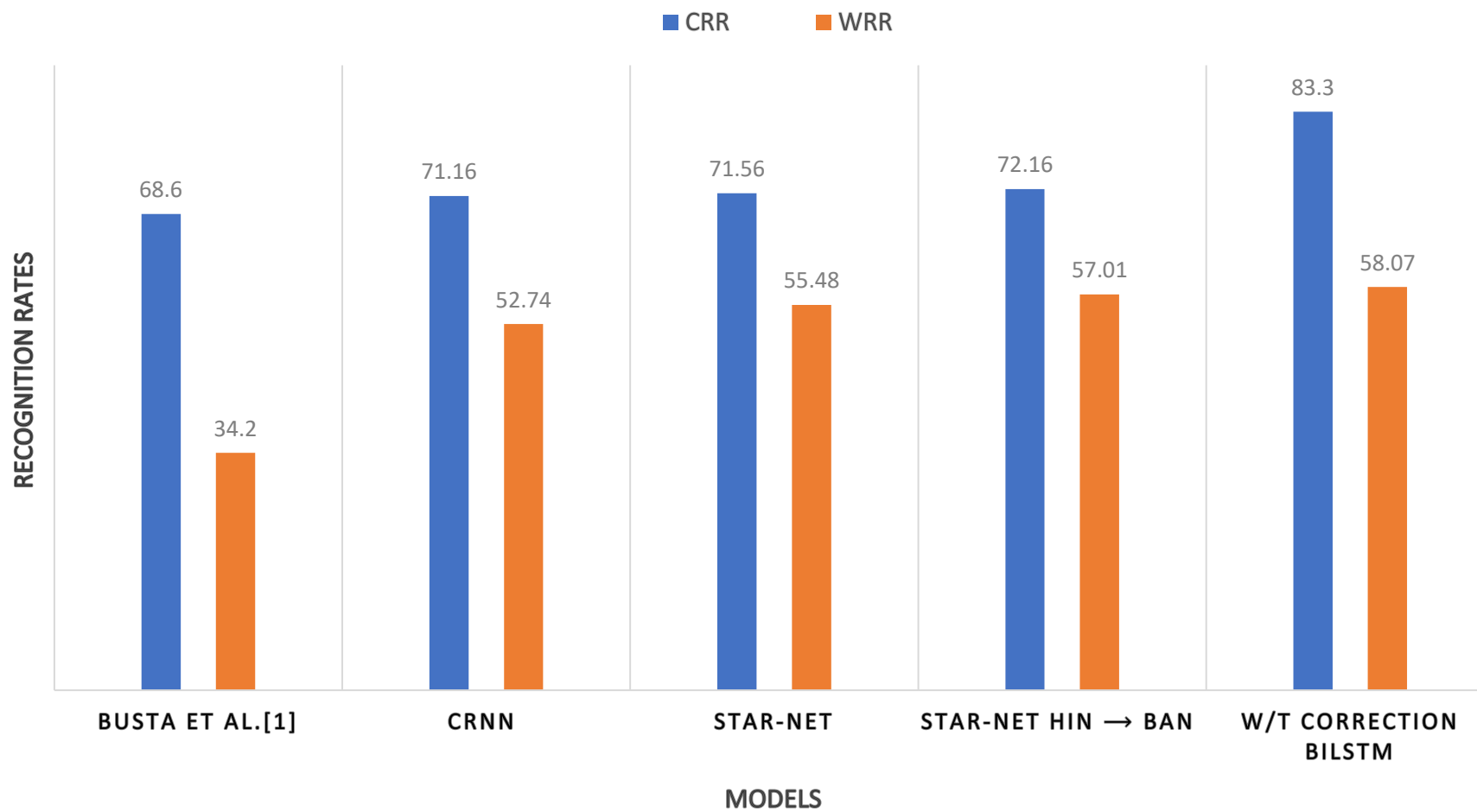
Results : Synthetic Data

Individual Language Model	CRNN (WRR)	STAR-Net (WRR)	Trained with Transfer Learning	CRNN (WRR)	STAR-Net (WRR)
Gujarati	81.85	91.40	English → Gujarati	77.06	90.90
			Hindi → Gujarati	84.21	92.81
Hindi	73.15	83.93	English → Hindi	70.12	80.90
			Gujarati → Hindi	73.12	84.32
Bangla	70.76	82.79	Hindi → Bangla	70.22	82.81
Tamil	48.19	79.90	Bangla → Tamil	44.74	81.73
Telugu	58.01	71.97	Tamil → Telugu	56.24	74.04
Malayalam	70.56	82.10	Telugu → Malayalam	65.78	77.97

The diagram illustrates the performance of CRNN and STAR-Net models on synthetic data for various language pairs. The table shows the Word Error Rate (WRR) for each model. Red ovals highlight the CRNN scores, and blue arrows indicate the transfer learning paths. Green and red arrows on the right indicate performance differences between the models.

Results : Real Data

PERFORMANCE (WRR) ON BANGLA DATASET – MLT-17

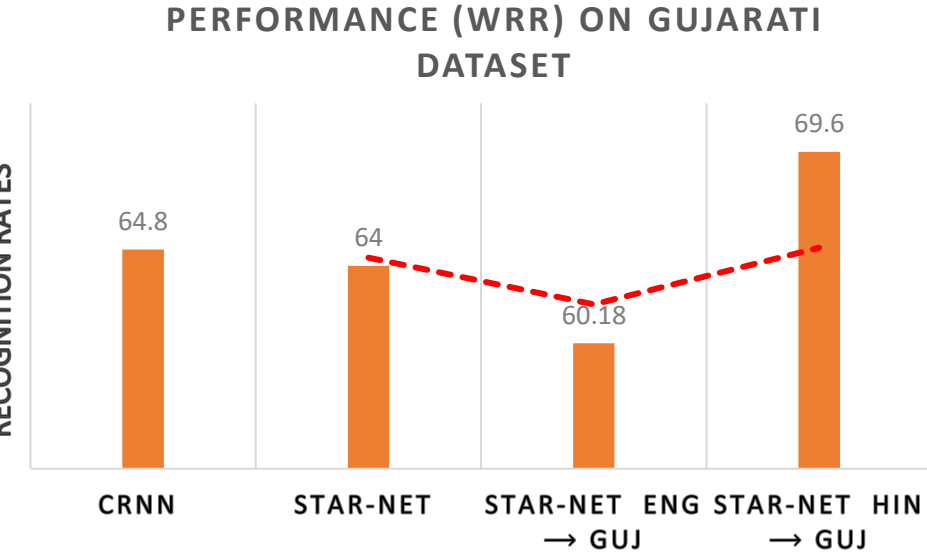
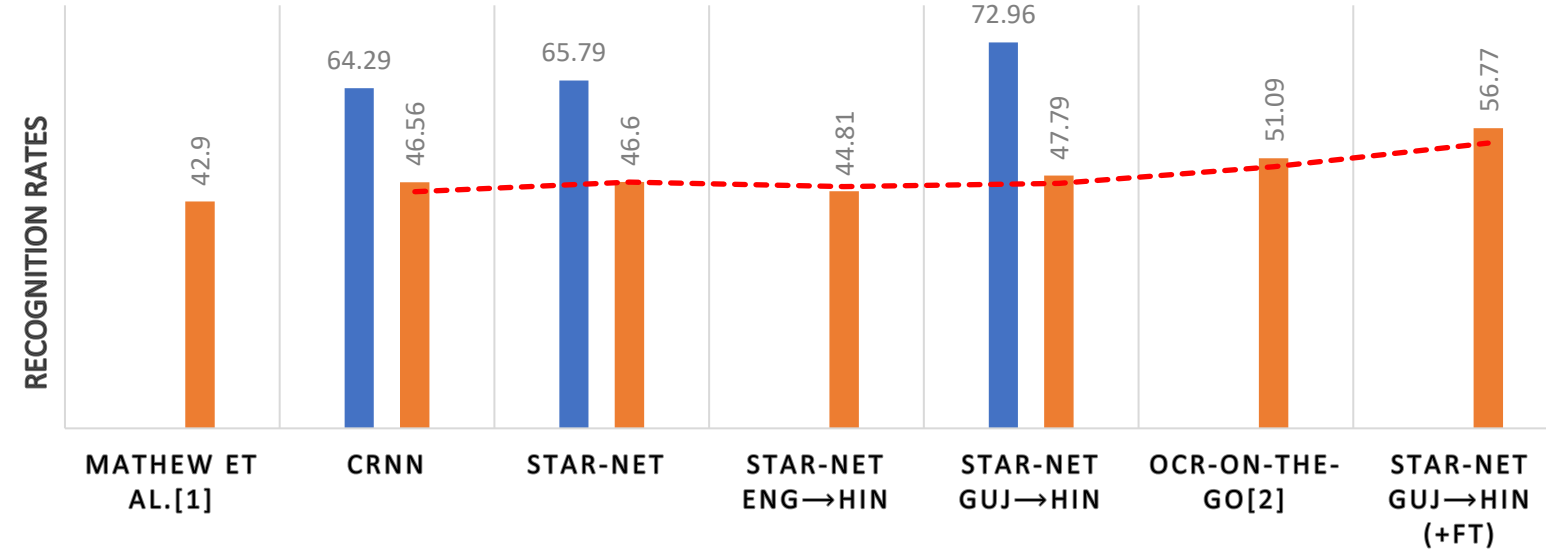


[[1]: Busta et al., "E2E-MLT-an Unconstrained End-to-End Method for Multi-Language Scene Text," ACCV, 2018]

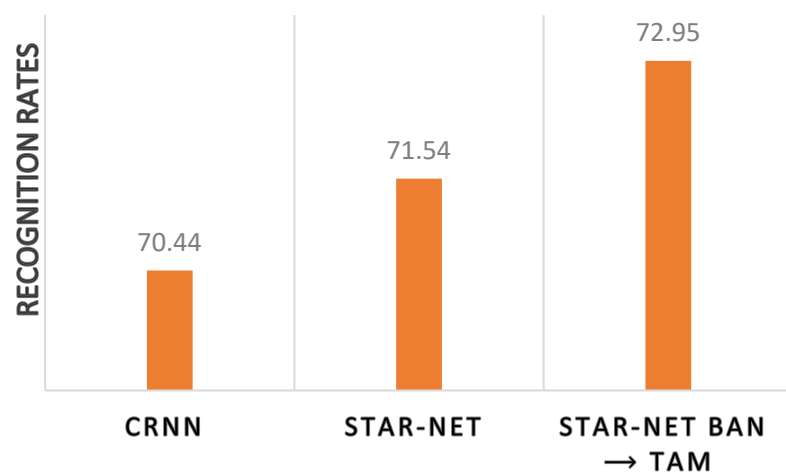
PERFORMANCE (WRR) ON HINDI DATASETS

Results on other languages and datasets follow the similar pattern

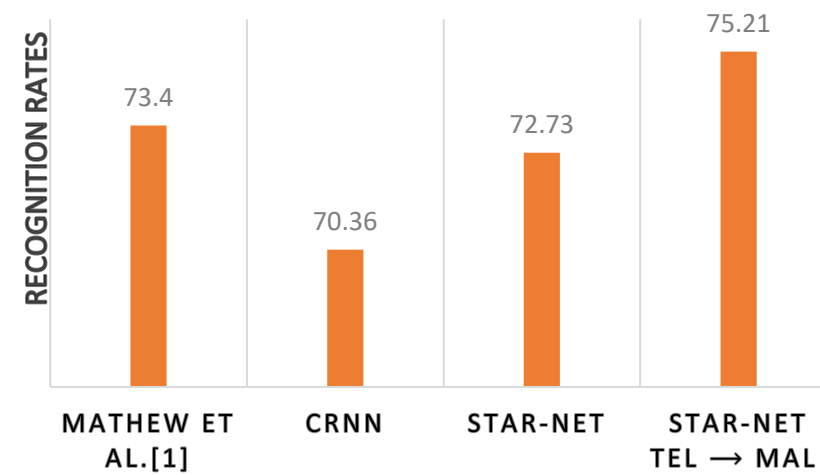
MLT-19
IIIT-ILST



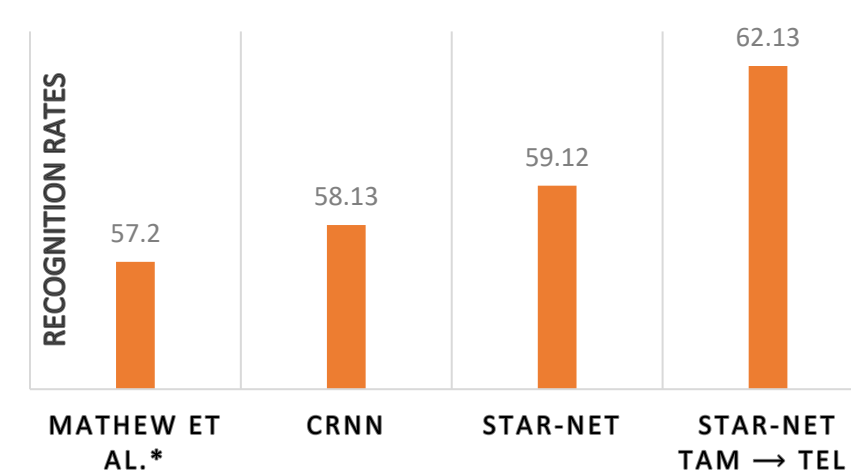
PERFORMANCE (WRR) ON TAMIL DATASET



PERFORMANCE (WRR & CRR) ON MALAYALAM DATASET – IIIT-ILST



PERFORMANCE (WRR) ON TELUGU DATASET –IIIT-ILST



[[1]: Mathew et al., "Benchmarking Scene Text Recognition in Devanagari, Telugu and Malayalam," MOCR,2017]

[[2]: Saluja et al., "OCR On-the-Go: Robust End-to-end Systems for Reading License Plates and Street Signs," ICDAR 2019]

Hindi / Devanagari



आप्रवासन
आप्रवासन

Bangla



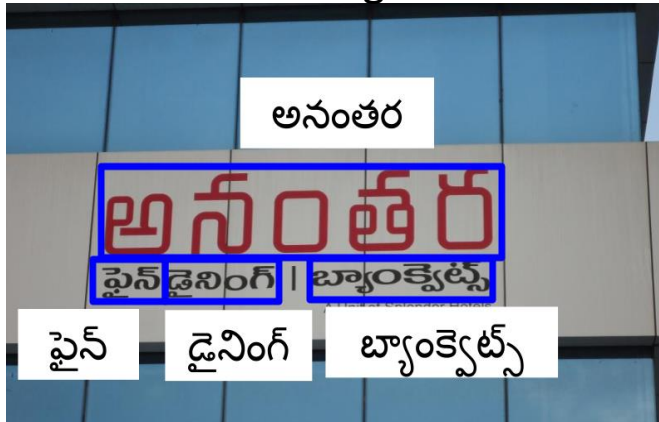
কে	এম	ঘোষ	রোড
কে	এম	ঘোষ	রোড

Gujarati



૨૮શમ	૨૯રછ	મવલો	અભાર
સ્ટેશન	૨૯રછ	મોલવો	આભાર

Telugu



అనంతరం	బ్యాంక్	డైనింగ్	ఊన్
అనంతరం	బ్యాంక్	డైనింగ్	ఊన్

Tamil



பரசமண்டவும்
உதகமண்டலம்

Malayalam



എച്ച്_പി_ഓട്ടോ	കയർ	കവടിയാർ
എച്ച്_പി_ഓട്ടോ	കയർ	കവടിയാർ

Ground truth

Correct Predictions

Incorrect or Missing predictions

Bounding Box

1st Row – CRNN Models (Baseline)

2nd Row – Best STAR-Net model (Transfer Learning)

Conclusion

- ❑ Transfer learning boosts performance over synthetic and real-world datasets, thereby setting new benchmarks for STR tasks in Indian languages.
- ❑ Sources of scene-text in Indian languages involve hand-painted signboards and wall paintings.
- ❑ There is a potential to utilize data across different modalities (ex. handwritten text) to augment recognition rates.
- ❑ This possibility of developing an all-in-one model for the Indian languages can be explored in the future.

THANK YOU

Please contact us @ [sanjana.gunna, rohit.saluja}@iiit.ac.in](mailto:{sanjana.gunna, rohit.saluja}@iiit.ac.in) for any further discussion or questions.