IIT Bombay

22M2119

Home Logout IITB Website

Roll no	22M2119	Name	Badri Vishal Kasuba
Department	Computer Science and Engineering	Program	Master of Science By Research

Payment | Performance Summary | New Entrants | Graduation Requirements | Personal Information | Forms/Requests

Academic Performance Summary

Year	Sem	SPI	CPI	Sem Credits Used for SPI	Completed Semester Credits	Cumulative Credits Used for CPI	Completed Cumulative Credits
2023	Spring	8.0	9.03	12.0	12.0	58.0	58.0
2023	Autumn	0.0	9.3	0.0	0.0	46.0	46.0
2022	Spring	9.25	9.3	24.0	24.0	46.0	46.0
2022	Autumn	9.36	9.36	22.0	22.0	22.0	22.0

Semester-wise Details

*This registration is subject to approval(s) from faculty advisor/Course Instructor/Academic office.

Report Problem

Year/Semester: 2023-24/Spring

Course Code	Course Name	Credits	Tag	Grade	Credit/Audit
CS 753	Automatic Speech Recognition	6.0	Department elective	BB	C
CS 769	Optimization in Machine Learning	6.0	Department elective	BB	C

Year/Semester: 2023-24/Autumn

Course Code	Course Name	Credits	Tag	Grade	Credit/Audit
CS 663	Fundamentals of Digital Image Processing	6.0	Additional Learning	BC	C

Year/Semester: 2022-23/Spring

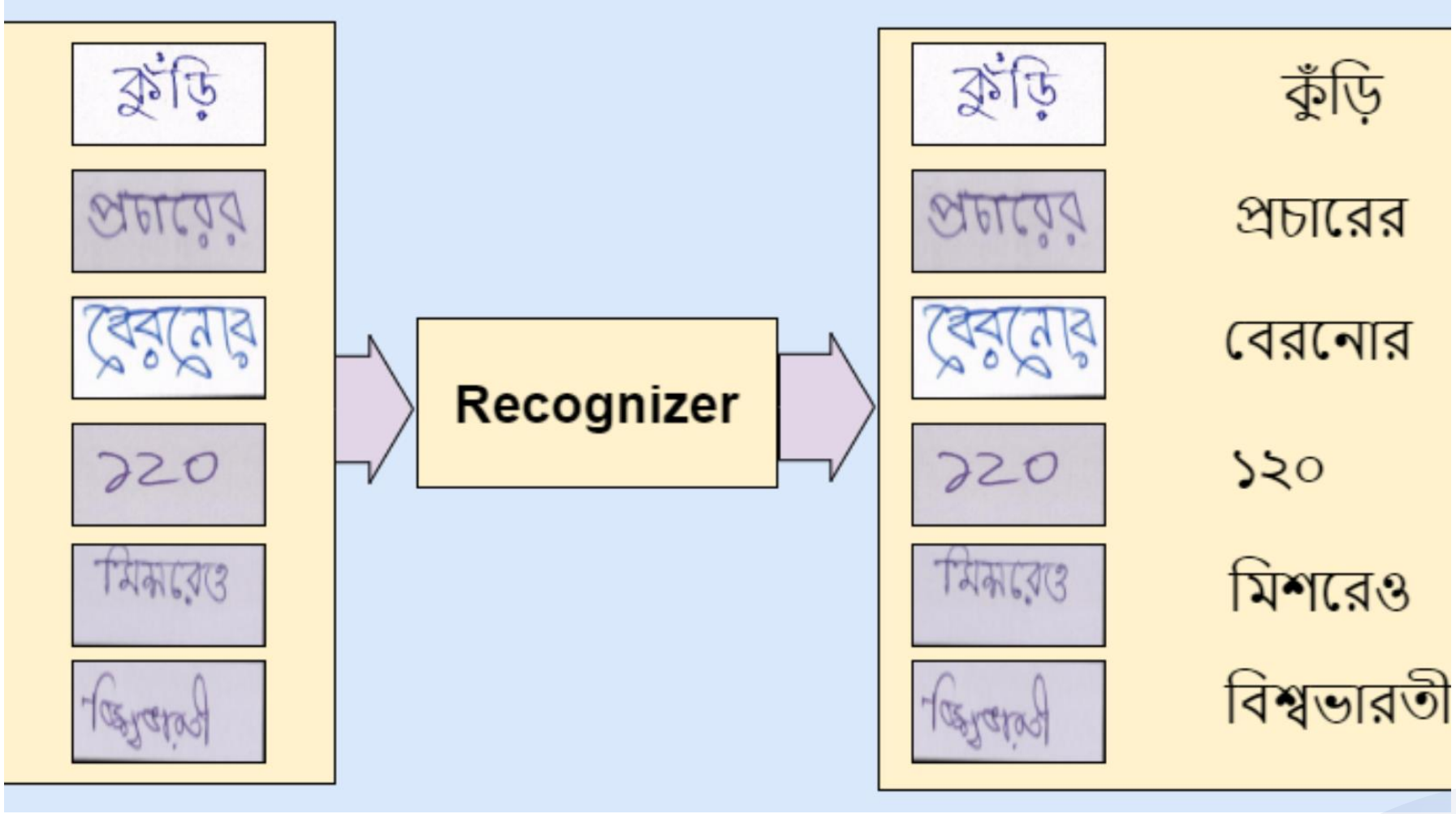
Course Code	Course Name	Credits	Tag	Grade	Credit/Audit
CS 694	Seminar	4.0	Core course	AA	C
CS 763	Computer Vision	6.0	Department elective	BB	C
CS 772	Deep Learning for Natural Language Processing	6.0	Department elective	AB	C
CS 778	M.S. R&D 2	8.0	Core course	AA	C
CS 899	Communication Skills	6.0	Core course	PP	N
TA 101	Teaching Assistant Skill Enhancement & Training (TASET)	0.0	Core course	PP	N

Year/Semester: 2022-23/Autumn

DocToSpeech
OCR-Summarization-MT-TTS
Final Project Discussion

Yashwant Singh Parihar, 22M0798
Abhishek Kumar Singh, 22M2104
Badri Vishal Kasuba, 22M2119

30th April, 2023



1

Attractive and Complex Problem

Problem Statement , Significance and Challenges

Problem Statement (Introduction)

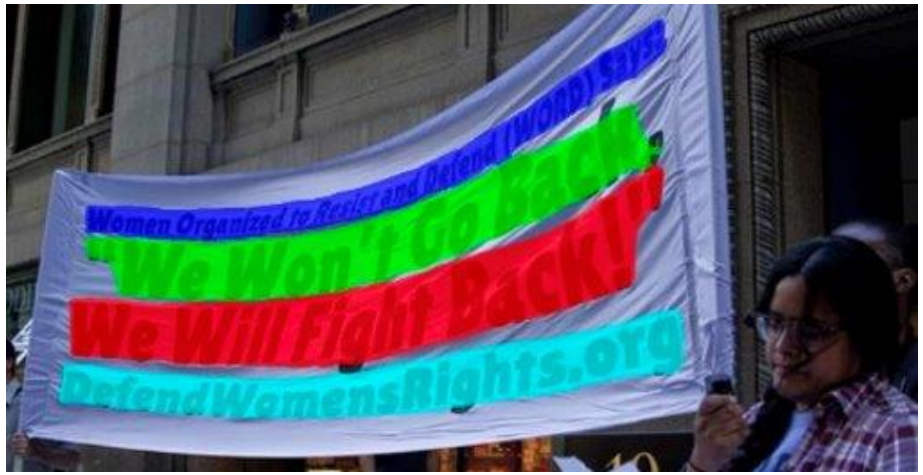
- Creation of end-to-end pipeline which
  - OCR's text from documents *(Machine could read and understand it)*
  - Summarizes the OCRed data (Condensing information is handy)
  - Translates summarization to Hindi (Useful for crossing Language barrier)
  - Hindi Text to Hindi Speech Creation (Beneficial for Emotional Exchange)

OCR - Text Summarization-Machine Translation-TTS

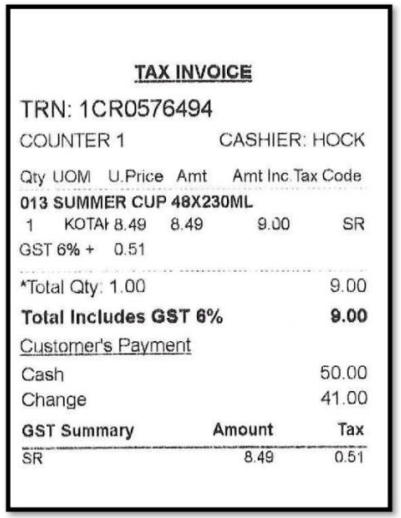
1. Text Detection
2. Text Recognition

Problem Statement (Significance)

The Task



The Significance



TAX INVOICE

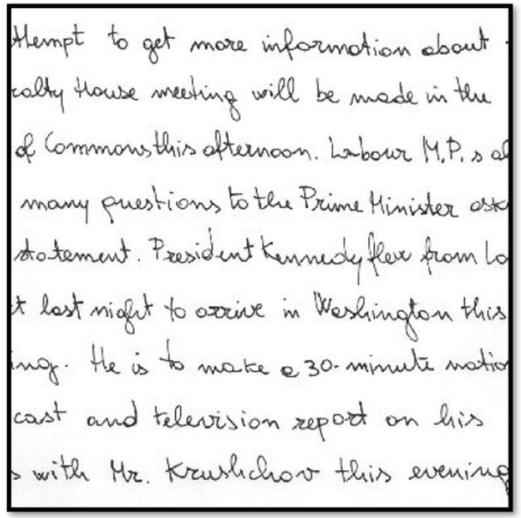
TRN: 1CR0576494

COUNTER 1 CASHIER: HOCK

Qty	UOM	U.Price	Amt	Amt Inc	Tax Code
013 SUMMER CUP 48X230ML					
1	KOTAI	8.49	8.49	9.00	SR
GST 6% +		0.51			
*Total Qty: 1.00				9.00	
Total Includes GST 6%				**9.00**	
Customer's Payment					
Cash				50.00	
Change				41.00	

GST Summary	Amount	Tax
SR	8.49	0.51

Text in Printed Documents



Attempt to get more information about
ealth House meeting will be made in the
of Commons this afternoon. Labour M.P.s a
many questions to the Prime Minister ab
statement. President Kennedy flew from Lo
t last night to arrive in Washington this
ng. He is to make a 30-minute natio
cast and television report on his
with Mr. Krushchov this evening

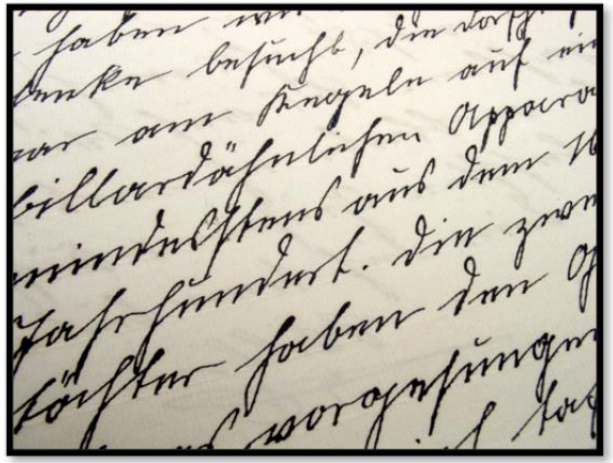
Text in Handwritten Documents

Text captured in the Wild

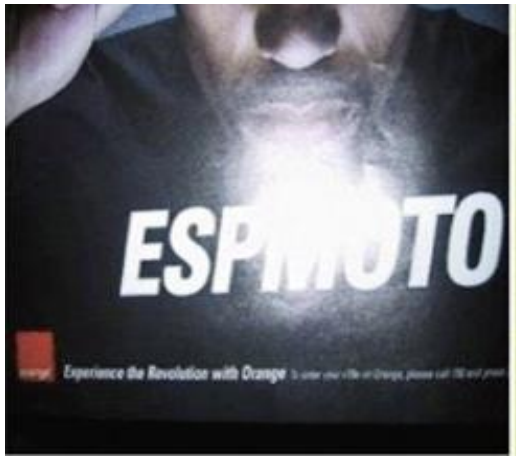
Problem Statement (Challenges)

Curved Text

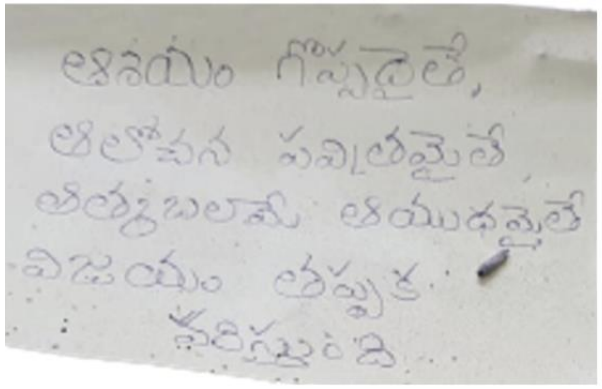
Multi-Oriented Text



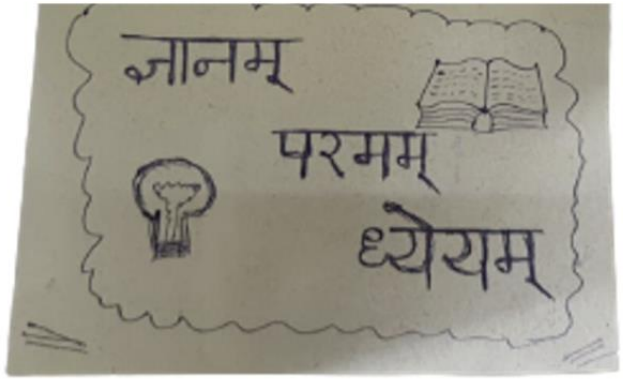
Illumination



Font Variations



Blurred Text



Handwritten Text

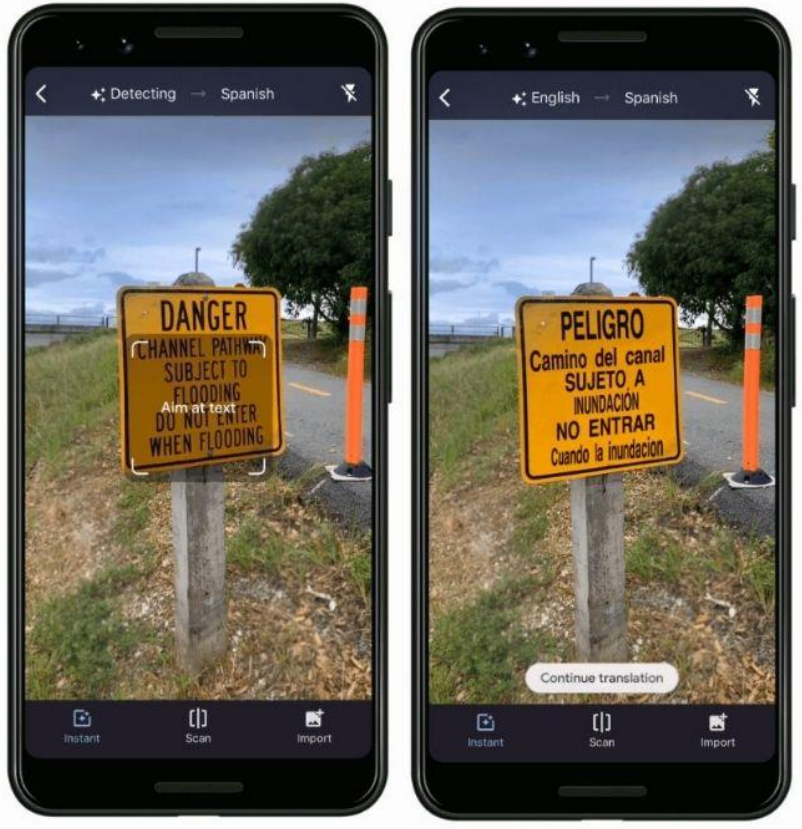
Scene-Text

Multi-Lingual Text

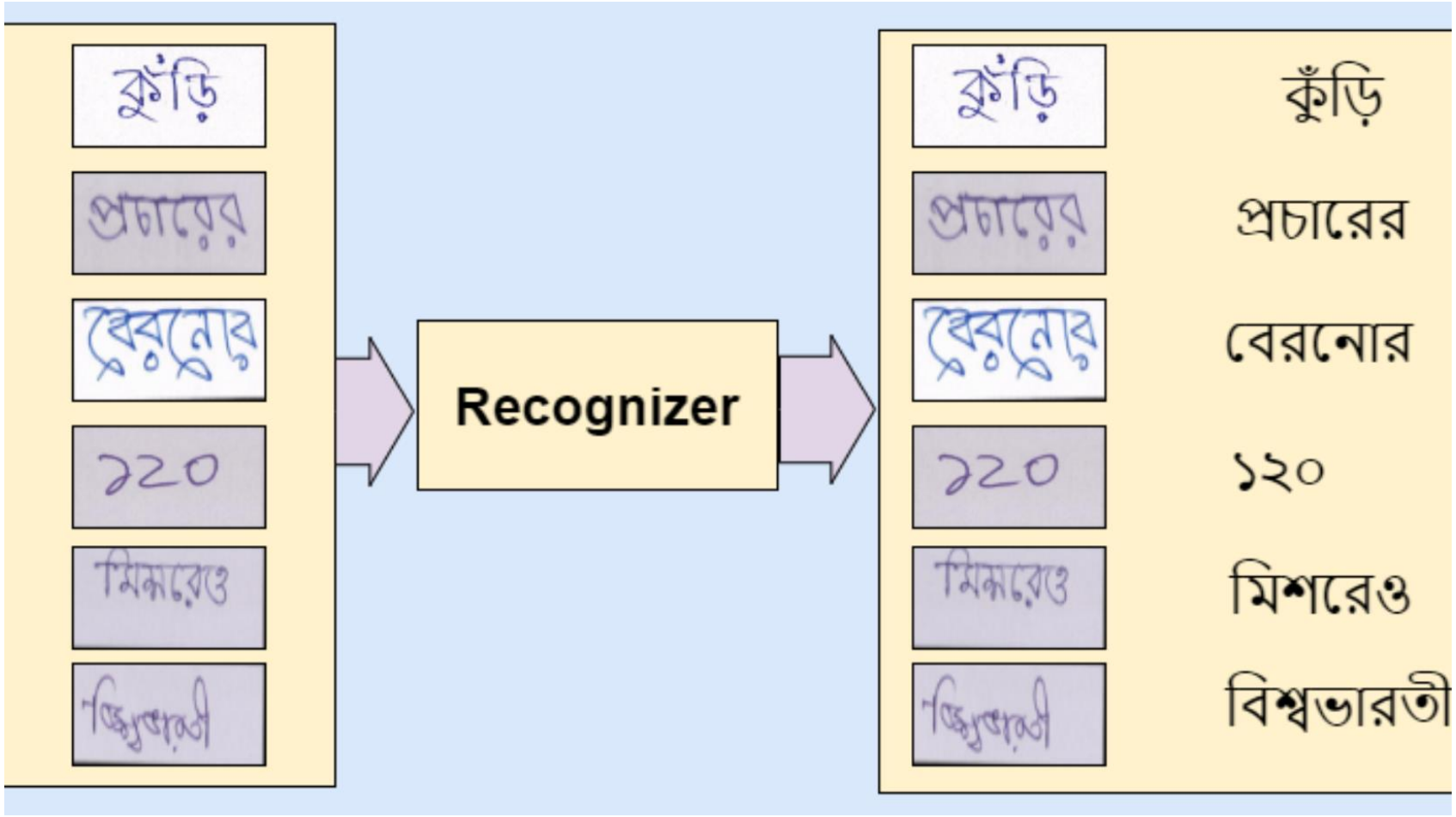
- Text Representation
- Environment based

Problem Statement (Attractive)

HOW TO TRANSLATE TEXT ON YOUR ANDROID PHONE USING GOOGLE LENS



Challenging Google Lens!!!!!



2

Clarity in Task and I/O Behaviour

Problem Statement Proposal and Related Works

Problem Statement (Proposal)

Pipeline:

1. Creation of end-to-end pipeline for all the mentioned approaches of OCR, summarization, MT, TTS.
2. Creation of efficient **product** which could be used in various downstream tasks

OCR:

1. Implementation of different pretrained sota models to perform OCR on benchmark DocBank dataset
2. Pre-processing Datasets as required to perform seamless OCR (**figures, equations and tables** removal)
3. Compare different OCR models through proper evaluation metrics
4. Analyse the performance of the models and make analytical study based on results
5. Training and Evaluating OCR for Indian Languages **(Hindi, Telugu and Tamil)**
6. Training end-to-end OCR model by fine-tuning pre-trained model on DocBank dataset to improve performance of the models
7. Using Trained Model of OCR in the production stage level

Repository Link - https://github.com/dl-in-nlp-gabru-geeks/ocr-mt-tts

Problem Statement (Proposal)

Text Summarization:

1. Summarize the OCRed text through summarization models after processing the OCRed predictions
2. For Text Summarization, we use pretrained models for getting predictions

Machine Translation:

1. Using pretrained model for Machine translation from English-to-Hindi
2. Training MT model from scratch on Hindi Dataset of CFILT
3. Comparative study of the models implemented and analysis
4. If possible, Implement MT models of other languages
5. Evaluating the model performance which was trained from scratch and using it in the product phase

Text-to-Speech:

1. Convert Hindi translated text into Hindi speech using pretrained models
2. For Text to Speech conversion, we use pretrained models for getting predictions

Repository Link - https://github.com/dl-in-nlp-gabru-geeks/ocr-mt-tts

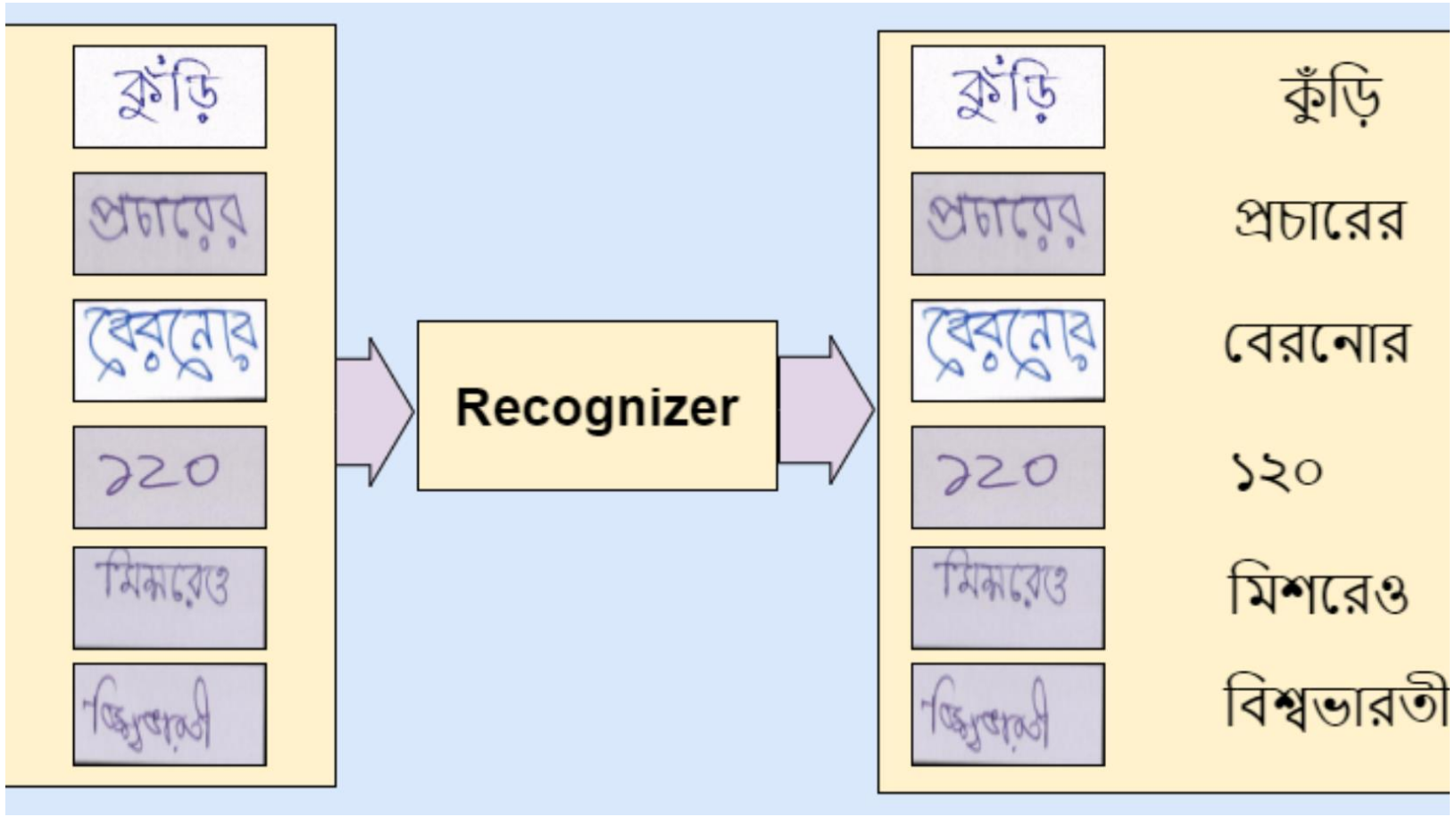
Related Works and Literature Review

1. DocBank: A Benchmark Dataset for Document Layout Analysis
   - [COLING 2020, 84 citations]
   - [Dataset for OCR Model predictions and training]
2. Real-time Scene Text Detection with Differentiable Binarization
   - [AAAI 2020, 371 citations]
   - [Text Detection Model]
3. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition
   - [CVPR 2015, 2227 citations]
   - [Text Recognition Model]
4. An Overview of the Tesseract OCR Engine
   - [2387 citations]
   - [Text Detection and Text Recognition]
5. The IIT Bombay English-Hindi Parallel Corpus.
   - [LREC 2018, 206 citations]
   - [Dataset for MT Model Training]
6. DocTR Library
   - Opensource utilized in implementing pretrained models for OCR

Related Works and Literature Review

1. IIIT-INDIC-HW-WORDS: A Dataset for Indic Handwritten Text Recognition
    1. 31 citations, accepted in CVPR 2021 conference
    2. Indic handwritten word-level images for 10 languages
2. Vakyansh – For Text to Speech
    1. Open-source library for various ASR tasks
3. Pegasus – For Text Summarization
    1. Sota model for Summarization by Google
    2. **A Self-Supervised Objective for Summarization**

3



Dataset Effort – Collection and Annotation

Datasets for OCR and MT tasks, Preprocessing and analysis

Dataset (s)

DocBank

1. DocBank is a large-scale dataset that is constructed using a weak supervision approach.
2. DocBank enables models to integrate both the textual and layout information for downstream tasks.
3. The current DocBank dataset totally includes **500K** document pages, where 400K for training, 50K for validation and 50K for testing.
4. We use DocBank's sample set of labelled images for making comparison of our detected text and recognized text of word-level data
5. Link - https://github.com/doc-analysis/DocBank

The IIT Bombay English-Hindi Parallel Corpus:-

1. The IIT Bombay English-Hindi corpus contains parallel corpus for English-Hindi as well as monolingual Hindi corpus collected from a variety of existing sources and corpora developed at the Center for Indian Language Technology, IIT Bombay over the years.
2. The Dataset contains **1561841 sentence pairs** to the parallel corpus
3. Link - https://www.cfilt.iitb.ac.in/iitb_parallel/

Dataset (s)

IIIT-Indic Words

1. A Benchmark dataset published in 2021 for Recognition of Handwritten Word images for 10 Indian languages
2. More than 100k word level images are present for training and we have chosen to work on Devanagari, Telugu and Tamil datasets for our project

3. Link - http://cvit.iiit.ac.in/research/projects/cvit-projects/iiit-indic-hw-words

Docbank data

IIIT-Indic words

Dataset for OCR Task - Docbank

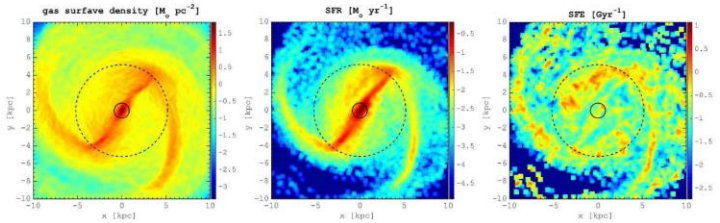


Fig. 13. Maps of the gas surface density (left), surface star formation rate (center) and star formation efficiency (right) for barred galaxy model at a single time, 1.2 Gyr. Black circles correspond to the positions of inner Lindblad resonance (solid line) and corotation radius (dashed line) for the bar.

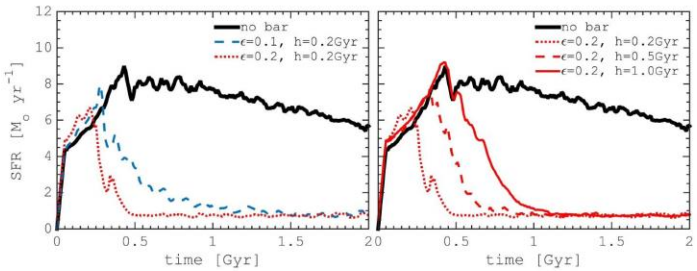


Fig. 14. Star-formation history in the different models. Unbarred galaxy SFR is shown by black line in all panels. *Left:* Comparison of models with two different bar strengths, ε_b = 0.1 and 0.2; *right:* Models with three different bar formation time-scales, h = 0.2, 0.5, 1 Gyr.

decreases only due to the conversion of gas into stars, with a slower decay than in our basic no-bar simulation.

In barred galaxy simulations, star formation decreases not as rapidly after the bar formation in comparison to our standard simulations. The SFR decreases because the gas is driven to central regions and converted to stars. In the standard model star formation is suppressed very rapidly right after the bar formation while in a model where star formation does not depend on the gas random motions star formation continues after the bar formation, but at lower level decreasing only due to gas consumption and depletion. Globally, in the case of no converging flow, bar formation reduces the star-formation rate by a factor of few which is much slower than in our fiducial runs. At the end of the simulation (after 2 Gyr), SFR is still high: ≈ 3 M$_\odot$ yr^{-1} for rigid bar simulation and ≈ 7 M$_\odot$ yr^{-1} for self-consistent run. Even if a slow and weak decrease in the star formation rate is found also in the models where the converging flow criterion is not implemented, we conclude that the inclusion of random gas motions in star-formation prescriptions produces a rapid quenching phase rather than a slow star-formation rate decrease in models without such a prescription.

3.9. Quenching parameters

In this section we aim to quantify the star-formation quenching efficiency in our various models and to investigate how the efficiency depends on bar parameters, i.e., strength ε, timescale h. We introduce simple quenching parameters. First, we estimate the quenching timescale h_q as an exponential timescale for the star-formation rate decrease:

$$\mathrm{SFR}(t) \propto \exp(-t/h_q)\,. \quad (6)$$

To make the fit, we did not use the whole time span of the simulations, but only the period when star formation is suppressed. The second parameter is the quenching rate ζ which we take as the ratio between the maximum star formation rate before star formation suppression and its value after the quenching phase:

$$\zeta = \mathrm{SFR(before\ quenching)/SFR(after\ quenching)}\,. \quad (7)$$

Original Image

S. Khoperskov et al.: Bar quenching in gas-rich galaxies

Fig. 13. Maps of the gas surface density (left), surface star formation rate (center) and star formation efficiency (right) for barred galaxy model at a single time, 1.2 Gyr. Black circles correspond to the positions of inner Lindblad resonance (solid line) and corotation radius (dashed line) for the bar.

Fig. 14. Star-formation history in the different models. Unbarred galaxy SFR is shown by black line in all panels. *Left:* Comparison of models with two different bar strengths, ε_b = 0.1 and 0.2; *right:* Models with three different bar formation time-scales, h = 0.2, 0.5, 1 Gyr.

decreases only due to the conversion of gas into stars, with a slower decay than in our basic no-bar simulation.

In barred galaxy simulations, star formation decreases not as rapidly after the bar formation in comparison to our standard simulations. The SFR decreases because the gas is driven to central regions and converted to stars. In the standard model star formation is suppressed very rapidly right after the bar formation while in a model where star formation does not depend on the gas random motions star formation continues after the bar formation, but at lower level decreasing only due to gas consumption and depletion. Globally, in the case of no converging flow, bar formation reduces the star-formation rate by a factor of few which is much slower than in our fiducial runs. At the end of the simulation (after 2 Gyr), SFR is still high: ≈ 3 M$_\odot$ yr^{-1} for rigid bar simulation and ≈ 7 M$_\odot$ yr^{-1} for self-consistent run. Even if a slow and weak decrease in the star formation rate is found also in the models where the converging flow criterion is not implemented, we conclude that the inclusion of random gas motions in star-formation prescriptions produces a rapid quenching phase rather than a slow star-formation rate decrease in models without such a prescription.

3.9. Quenching parameters

In this section we aim to quantify the star-formation quenching efficiency in our various models and to investigate how the efficiency depends on bar parameters, i.e., strength ε, timescale h. We introduce simple quenching parameters. First, we estimate the quenching timescale h_q as an exponential timescale for the star-formation rate decrease:

To make the fit, we did not use the whole time span of the simulations, but only the period when star formation is suppressed. The second parameter is the quenching rate ζ which we take as the ratio between the maximum star formation rate before star formation suppression and its value after the quenching phase:

Article number, page 13 of 17

Processed Image

S. Khoperskov et al.: Bar quenching in gas-rich galaxies

Fig. 13. Maps of the gas surface density (left), surface star formation rate (center) and star formation efficiency (right) for barred galaxy model at a single time, 1.2 Gyr. Black circles correspond to the positions of inner Lindblad resonance (solid line) and corotation radius (dashed line) for the bar.

Fig. 14. Star-formation history in the different models. Unbarred galaxy SFR is shown by black line in all panels. *Left:* Comparison of models with two different bar strengths, ε_b = 0.1 and 0.2; *right:* Models with three different bar formation time-scales, h = 0.2, 0.5, 1 Gyr.

decreases only due to the conversion of gas into stars, with a slower decay than in our basic no-bar simulation.

In barred galaxy simulations, star formation decreases not as rapidly after the bar formation in comparison to our standard simulations. The SFR decreases because the gas is driven to central regions and converted to stars. In the standard model star formation is suppressed very rapidly right after the bar formation while in a model where star formation does not depend on the gas random motions star formation continues after the bar formation, but at lower level decreasing only due to gas consumption and depletion. Globally, in the case of no converging flow, bar formation reduces the star-formation rate by a factor of few which is much slower than in our fiducial runs. At the end of the simulation (after 2 Gyr), SFR is still high: ≈ 3 M$_\odot$ yr^{-1} for rigid bar simulation and ≈ 7 M$_\odot$ yr^{-1} for self-consistent run. Even if a slow and weak decrease in the star formation rate is found also in the models where the converging flow criterion is not implemented, we conclude that the inclusion of random gas motions in star-formation prescriptions produces a rapid quenching phase rather than a slow star-formation rate decrease in models without such a prescription.

3.9. Quenching parameters

In this section we aim to quantify the star-formation quenching efficiency in our various models and to investigate how the efficiency depends on bar parameters, i.e., strength ε, timescale h. We introduce simple quenching parameters. First, we estimate the quenching timescale h_q as an exponential timescale for the star-formation rate decrease:

To make the fit, we did not use the whole time span of the simulations, but only the period when star formation is suppressed. The second parameter is the quenching rate ζ which we take as the ratio between the maximum star formation rate before star formation suppression and its value after the quenching phase:

Article number, page 13 of 17

Ground Truths

Dataset for MT – English-Hindi Corpus

Preprocessing:-

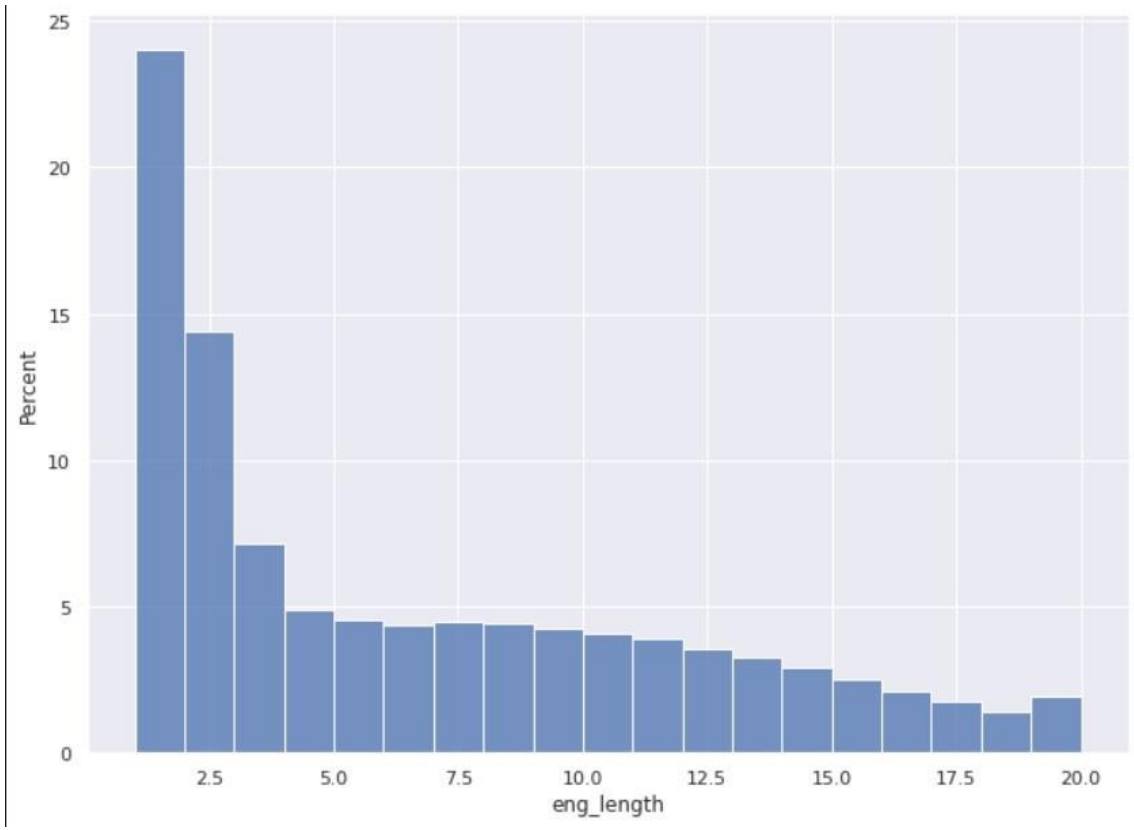
1. We had to removed punctuation, numerals, extra spaces and convert them to lowercase.
2. Hindi sentence also contain English character or word they need to be sanitized to remove using regex.
3. Since we use LSTM as encoder and decoder, we already decided the longest size(2) we take other we have removed.
4. Due to resources constraint, we also reduce the number of sentence that is we are just training them on 127605 sentences.

Preparing of Data:-

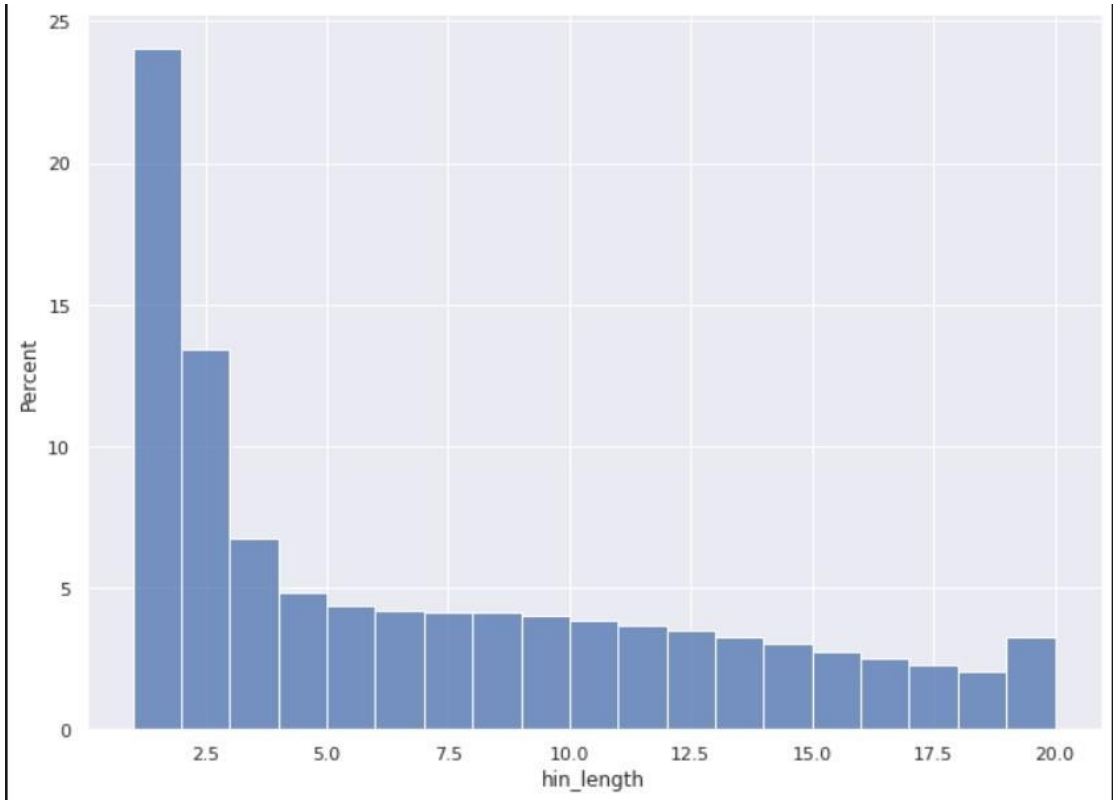
1. Prepend and append <START> and <END> tokens specifically to convey our model when to begin and end, added to target sentences.
2. We used TensorFlow inbuilt tokenizer and create a word-index representation for both the English and Hindi sentences.
3. Out-of-Vocabulary words [UNK] tag is there if it can't find a word in the translation process.

Dataset for MT – English-Hindi Corpus

Spread of English sentence length:-



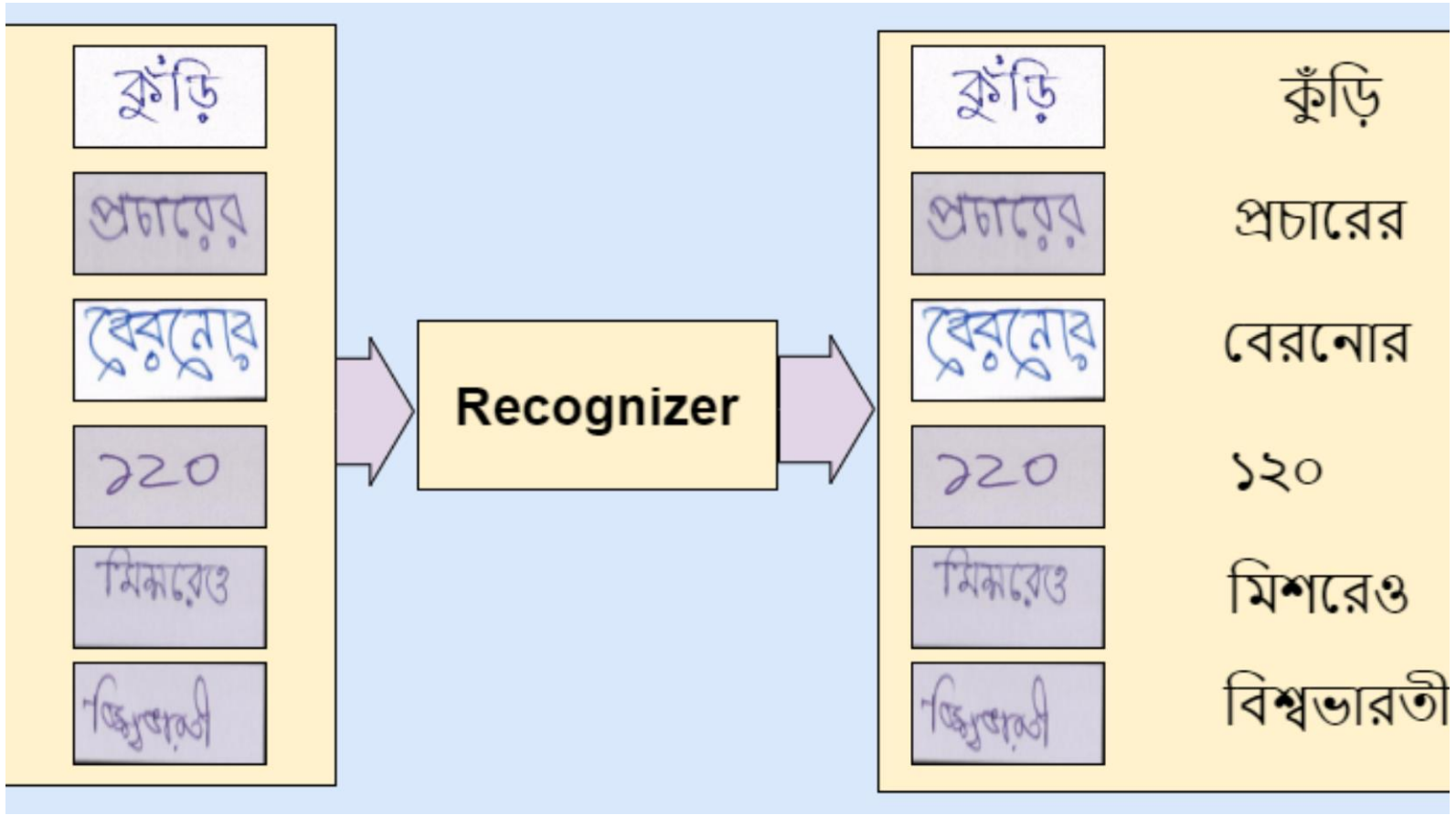
Spread of Hindi sentence length:-



Number of sentence/data point:- 1561841(**We trained On:-** 127605).

Data set Shape:- (1561841, 2)

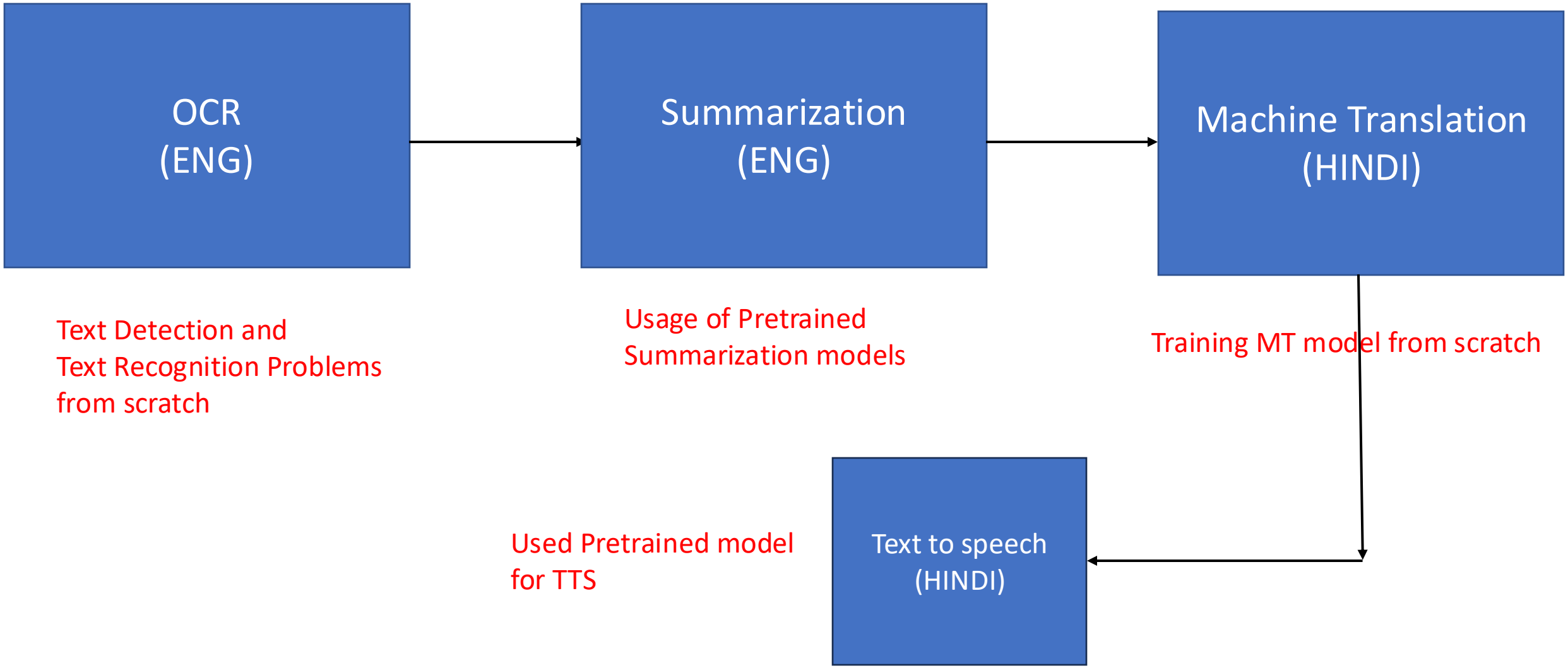
4



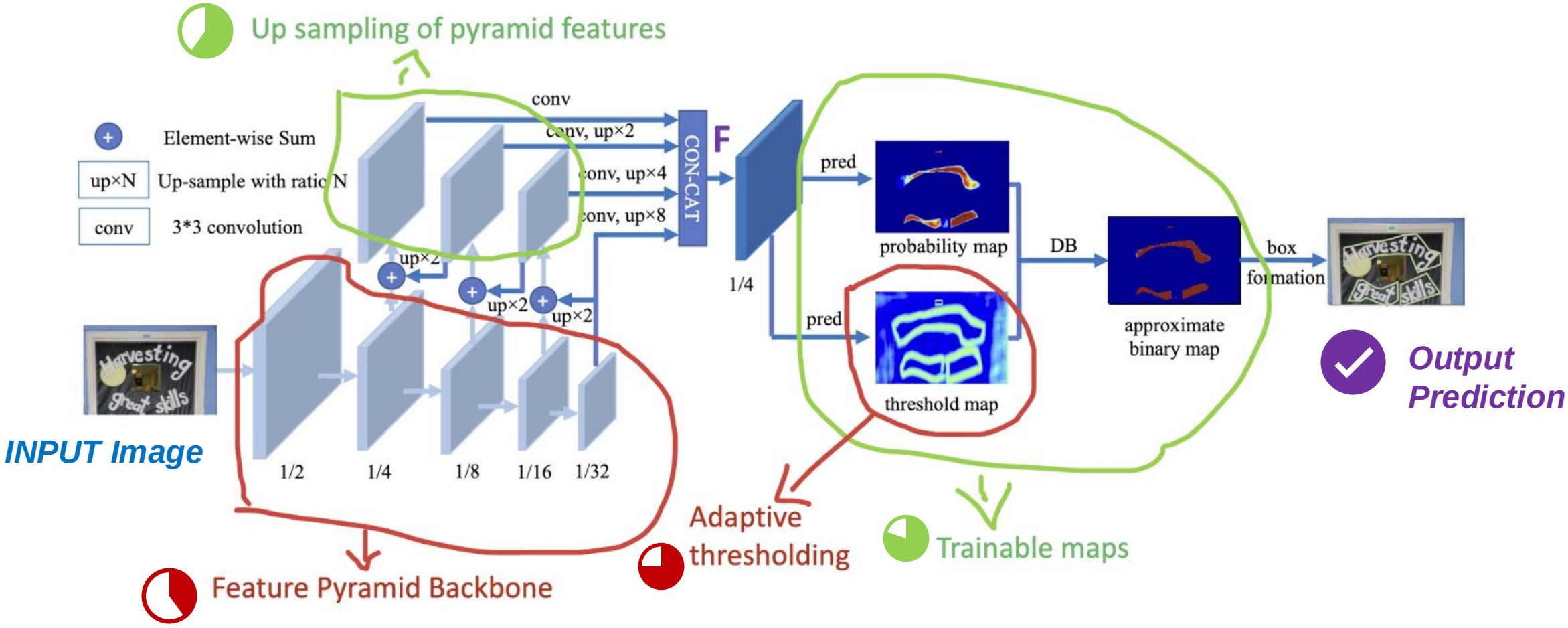
Workflow, Architecture and Technique

Datasets for OCR and MT tasks, Preprocessing and analysis

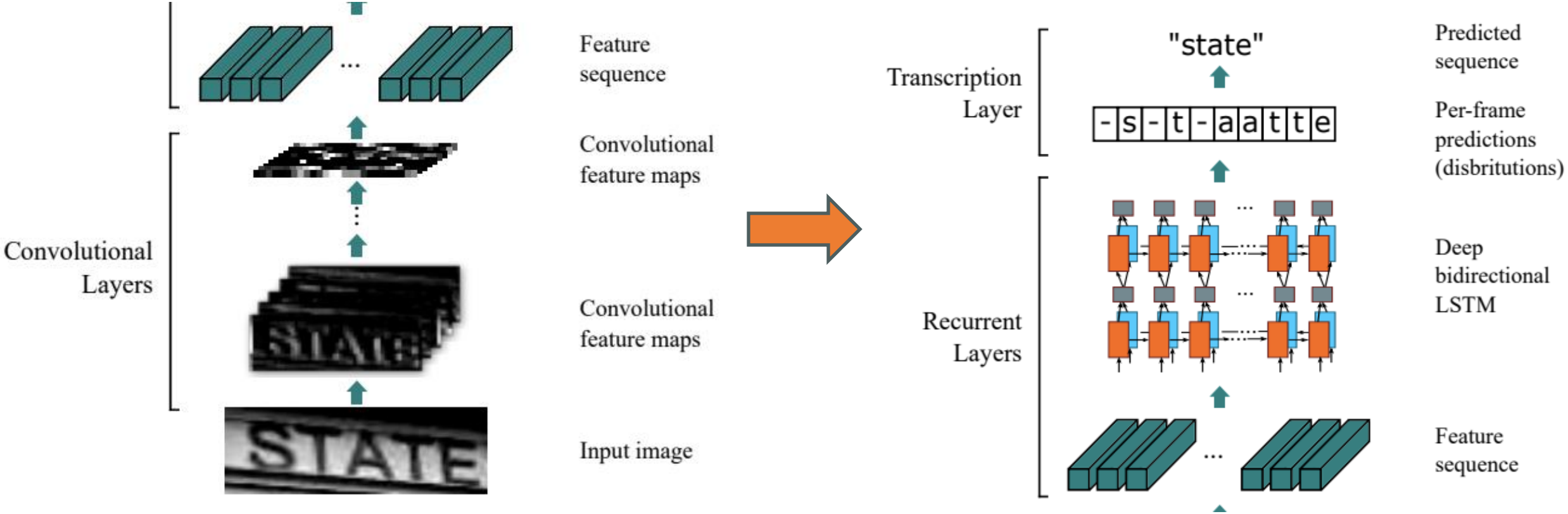
Workflow


OCR
(ENG)
Summarization
(ENG)
Machine Translation
(HINDI)
Text to speech
(HINDI)
Text Detection and
Text Recognition Problems
from scratch
Usage of Pretrained
Summarization models
Training MT model from scratch
Used Pretrained model
for TTS

Architecture – Text Detection



Architecture – Text Recognition



MT Architecture and Technique

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	(None, None)	0	
input_2 (InputLayer)	(None, None)	0	
embedding_1 (Embedding)	(None, None, 300)	4187400	input_1[0][0]
embedding_2 (Embedding)	(None, None, 300)	5261700	input_2[0][0]
lstm_1 (LSTM)	[(None, 300), (None,	721200	embedding_1[0][0]
lstm_2 (LSTM)	[(None, None, 300),	721200	embedding_2[0][0] lstm_1[0][1] lstm_1[0][2]
dense_1 (Dense)	(None, None, 17539)	5279239	lstm_2[0][0]

Total params: 16,170,739
Trainable params: 16,170,739
Non-trainable params: 0

Model- Encoder-Decoder(LSTM)
Avg. BLEU score-
0.13247268378107925

Model: "transformer"

Layer (type)	Output Shape	Param #	Connected to
encoder_inputs (InputLayer)	[(None, None)]	0	[]
positional_embedding (PositionalEmbedding)	(None, None, 128)	2562560	['encoder_inputs[0][0]']
decoder_inputs (InputLayer)	[(None, None)]	0	[]
encoder_1 (TransformerEncoder)	(None, None, 128)	1186304	['positional_embedding[0][0]']
model_1 (Functional)	(None, None, 20000)	6988448	['decoder_inputs[0][0]', 'encoder_1[0][0]']

Total params: 10,737,312
Trainable params: 10,737,312
Non-trainable params: 0

Model:- Transformer
Avg. Bleu Score:-
0.2729667965071246 3

Pretrained Models:-

- **Pegasus- For Summarization.**
  - **The model is transformer encoder-decoders with 16 layers in each component.**
  - **The implementation is completely inherited from BartForConditionalGeneration.**
  - **We Used that for Abstractive Summarization.**

- **Facebook/nllb-200-distilled-600M- For Machine Translation.**
  - **Facebook AI Research (FAIR) released their most recent model in the language generation field, especially language translation, called No Language Left Behind (NLLB). An open- source project capable of translating between 200 languages.**
  - **The model was trained with input lengths not exceeding 512 tokens.**
  - **NLLB-200 is a machine translation model primarily intended for research in machine translation, - especially for low-resource languages.**

Pretrained Models:-

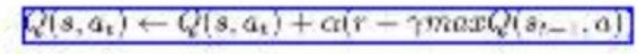
- **Vakyansh:- for text to speech in hindi.**
  - **This model was pre-trained using Nemo toolkit with 34,000 hours unlabeled audio in 39 Indian languages.**
  - **This includes 15,000 hours of news recordings available on the internet, 10,000 hours of YouTube audios and other audio data.**
  - **In addition, 9,000 hours of Indian English audio data was taken from NPTEL lectures open sourced by AI4Bharat.**
  - **This model was trained in collaboration with NVIDIA (NVIDIA Graphics Pvt Ltd).**

Current Work Details [OCR]

- We performed OCR on Research papers (DocBank data) so most of the text in papers could be classified as
  1. Normal Text
  2. Text with varying fonts (Title, Abstract, Section Headings, References)
  3. Tables and Text inside them
  4. Figures and possible text inside them
  5. Equations Text

- OCR on equations is not giving good results

Ground Truth



Predicted Text box

Work Details [OCR]

- Notebooks worked on (Pipeline setup and workflow)
  1. Pre-processing
  2. Opensource library (DocTR) usage to use OCR Models and get Predictions
  3. Evaluating the obtained results
  4. Analytical Study

```
block line confidence X1 Y1 X2 Y2 token
0 0 0.9995970129966736 247 206 337 238 While
0 0 0.9965320825576782 341 212 373 242 a
0 0 0.9999426603317261 376 204 413 242 is
0 0 0.8786391019821167 418 206 491 240 fixed
0 0 0.9036145210266113 496 206 537 242 at
0 0 0.9998965263366699 541 206 567 240 2
0 0 0.8580747246742249 572 201 645 244 A-1,
0 0 0.998884379863739 652 201 683 247 B
0 0 0.9994397759437561 687 204 722 244 is
0 0 0.9974278807640076 725 206 811 242 taken
0 0 0.9999362230300903 815 206 854 242 at
0 0 0.9999442100524902 859 206 914 240 the
```

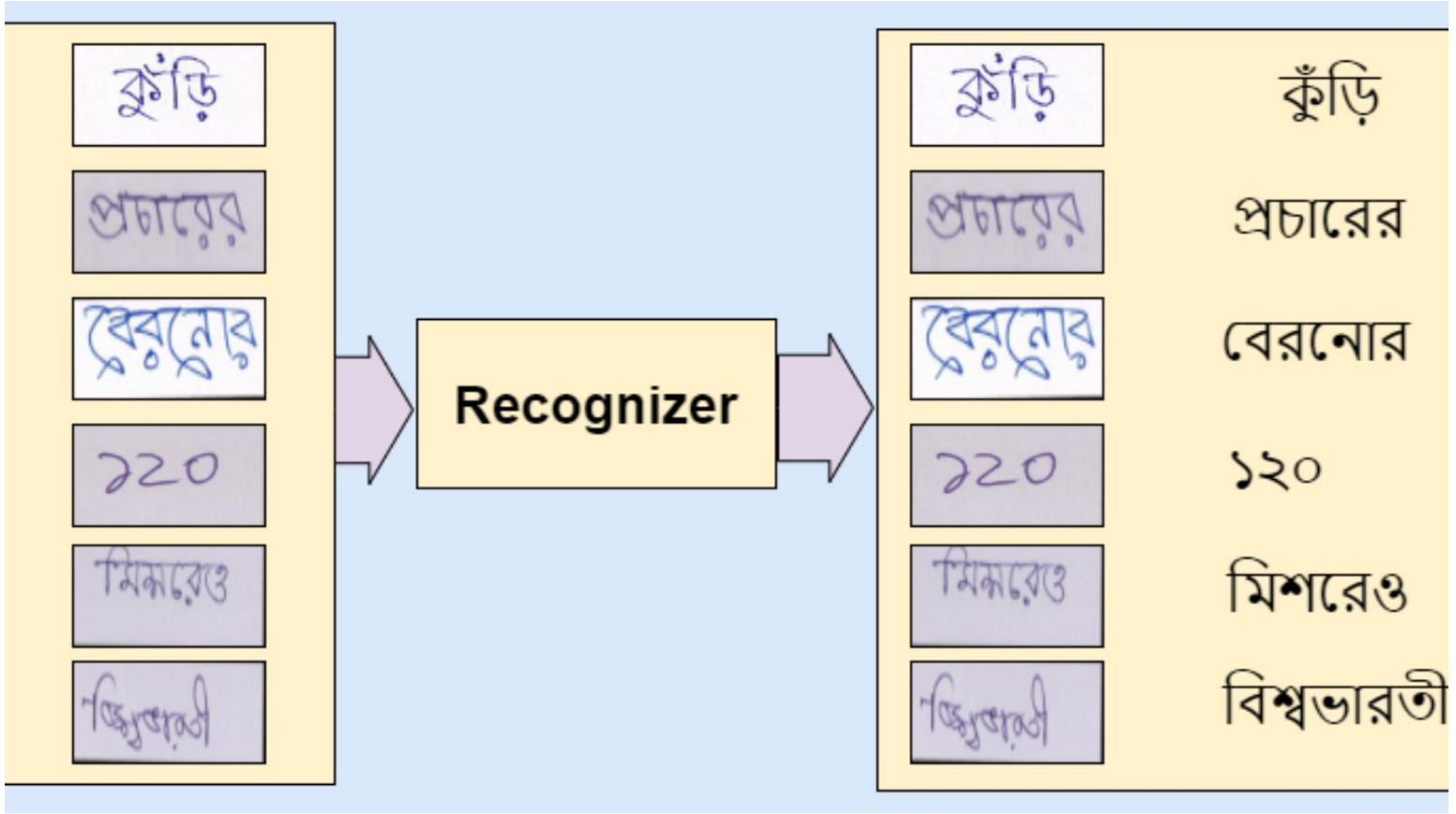
Work Details [OCR]

- Established the pipeline to OCR text using different Text Detection Models and Text Recognition models and making comparison between their performance on DocBank data
- Currently, used 100 labelled images of DocBank data for our OCR

- Text Detection Models Used
  1. DBNet Model
  2. MobileNet
  3. Tesseract

- Text Recognition Models Used
  1. CRNN_VGG Model
  2. MobileNet
  3. Tesseract

5



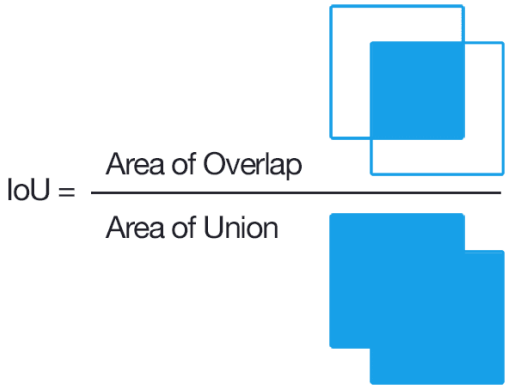
Results and Analysis

Analysis on various models' performance

Metrics to be used

OCR:

For Text Detection, we use IOU scores to evaluate quality of the predicted text boxes for OCR



For Text Recognition, we use WRR and CRR scores to evaluate the correctness of the recognized text

$$ER = \frac{S + D + I}{N} (1)$$

$$RR = 1 - ER(2)$$

$S = no.\ of\ Substitutions$

$D = no.\ of\ Deletions$

$I = no.\ of\ Insertions$

$N = no.\ of\ instances$

CRR = Character Recognition Rate
WRR = Word Recognition Rate

Machine Translation: BLEU score

Original

where $\boldsymbol{u}$, ρ and c are the background flow velocity, density and speed of sound, and t $\boldsymbol{\nabla}$ are the standard time and gradient operator of Newtonian mechanics. Conservation of m for the background flow tells us that,

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{u}) = 0. \tag{2}$$

Using this along with Eq. (1) we obtain a slightly altered expression for ϕ,

$$\left(\frac{\partial}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{u}\right) \frac{\rho}{c^2} \left(\frac{\partial}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla}\right) \phi - \boldsymbol{\nabla} \cdot (\rho \boldsymbol{\nabla} \phi) = 0, \tag{3}$$

where $\boldsymbol{\nabla}$ acts on *everything* to its right. This may seem like a needless complication , as pointed out by, for example, Visser [2], we can write this as the d'Alembertian of a curv 4-dimensional space-time of mixed signature,

$$\Delta \phi = \frac{1}{\sqrt{-g}} \partial_\mu \left(\sqrt{-g} g^{\mu\nu} \partial_\nu \phi\right) = 0. \tag{4}$$

$g^{\mu\nu}$ is the inverse of the metric $g_{\mu\nu}$ of the space, and g is the determinant of $g_{\mu\nu}$. Henc becomes clear that we can regard ϕ as a scalar field on a 4-dimensional space with the metric defined by the background flow. This space is the *acoustic space-time*. The authors who h pointed this out have largely used it as a tool to illuminate what parts of general relativity due to Einstein's field equations specifically, and what phenomena remain when the equati are changed but the form of the geometry remains (this is where black holes are of intere We, however, would like to turn this on its head, making use of the acoustic space-time illuminate how variable transformations can be used to understand sound propagation in presence of background flow.

The idea of these transformations is to use a change of variables to convert a challeng equation, such as Eq. (1), into, say, the classic wave equation. To the authors knowledge most general transformation to date of this kind was first presented by Taylor [5], and is v for irrotational, barotropic, low Mach number, steady background flows for acoustic fields wh the wavelength is of the same order of magnitude or smaller than the length scale of variati of the background flow [7]. The presentation of this transformation is fairly ad hoc, and would like to use the acoustic space-time to derive and generalise such transformations i more systematic way.

2 Calculus on Curved Manifolds

In order to proceed further we shall introduce some new tools to deal with calculus general manifolds. More precisely we shall introduce Geometric Algebra (GA).

GA deals with general manifolds through the concept of embedding. A flat vector sp of large dimension is first defined, within which the, possibly curved, manifold of interes placed. The embedded manifold then inherits a metric from the extrinsic space, and this all a study of Riemannian geometry. Some readers may wonder why we use GA instead of more widely used differential forms and differential geometry (see for example Nakahara [ultimately this a preference of the authors. We find that the approach gives more streamli and intuitive proofs of some key results, and also allows more to be expressed independentl coordinates. A debate of the relative benefits is beyond the scope of the current paper. As

DBNet

where $\boldsymbol{u}$, ρ and c are the background flow velocity, density and speed of sound, and t $\boldsymbol{\nabla}$ are the standard time and gradient operator of Newtonian mechanics. Conservation of for the background flow tells us that,

(2) $$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{u}) = 0.$$

Using this along with Eq. (1) we obtain a slightly altered expression for ϕ,

(3) $$\left(\frac{\partial}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{u}\right) \frac{\rho}{c^2} \left(\frac{\partial}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla}\right) \phi - \boldsymbol{\nabla} \cdot (\rho \boldsymbol{\nabla} \phi) = 0,$$

where $\boldsymbol{\nabla}$ acts on *everything* to its right. This may seem like a needless complication as pointed out by, for example, Visser [2], we can write this as the d'Alembertian of a cu 4-dimensional space-time of mixed signature,

(4) $$\Delta \phi = \frac{1}{\sqrt{-g}} \partial_\mu \left(\sqrt{-g} g^{\mu\nu} \partial_\nu \phi\right) = 0.$$

$g^{\mu\nu}$ is the inverse of the metric $g_{\mu\nu}$ of the space, and g is the determinant of $g_{\mu\nu}$. Hen becomes clear that we can regard ϕ as a scalar field on a 4-dimensional space with the metri defined by the background flow. This space is the *acoustic space-time*. The authors who pointed this out have largely used it as a tool to illuminate what parts of general relativit due to Einstein's field equations specifically, and what phenomena remain when the equa are changed but the form of the geometry remains (this is where black holes are of inte We, however, would like to turn this on its head, making use of the acoustic space-tim illuminate how variable transformations can be used to understand sound propagation i presence of background flow.

The idea of these transformations is to use a change of variables to convert a challen equation, such as Eq. (1), into, say, the classic wave equation. To the authors knowledg most general transformation to date of this kind was first presented by Taylor [5], and is for irrotational, barotropic, low Mach number, steady background flows for acoustic fields w the wavelength is of the same order of magnitude or smaller than the length scale of varia of the background flow [7]. The presentation of this transformation is fairly ad hoc, an would like to use the acoustic space-time to derive and generalise such transformations more systematic way.

2 Calculus on Curved Manifolds

In order to proceed further we shall introduce some new tools to deal with calculu general manifolds. More precisely we shall introduce Geometric Algebra (GA).

GA deals with general manifolds through the concept of embedding. A flat vector s of large dimension is first defined, within which the, possibly curved, manifold of intere placed. The embedded manifold then inherits a metric from the extrinsic space, and this a a study of Riemannian geometry. Some readers may wonder why we use GA instead o more widely used differential forms and differential geometry (see for example Nakahara ultimately this a preference of the authors. We find that the approach gives more stream and intuitive proofs of some key results, and also allows more to be expressed independent coordinates. A debate of the relative benefits is beyond the scope of the current paper. As

Tesseract

where $\boldsymbol{u}$, ρ and c are the background flow velocity, density and speed of sound, and t an $\boldsymbol{\nabla}$ are the standard time and gradient operator of Newtonian mechanics. Conservation of mas for the background flow tells us that,

(2) $$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{u}) = 0.$$

Using this along with Eq. (1) we obtain a slightly altered expression for ϕ,

(3) $$\left(\frac{\partial}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{u}\right) \frac{\rho}{c^2} \left(\frac{\partial}{\partial t} + \boldsymbol{u} \cdot \boldsymbol{\nabla}\right) \phi - \boldsymbol{\nabla} \cdot (\rho \boldsymbol{\nabla} \phi) = 0,$$

where $\boldsymbol{\nabla}$ acts on *everything* to its right. This may seem like a needless complication , bu as pointed out by, for example, Visser [2], we can write this as the d'Alembertian of a curved 4-dimensional space-time of mixed signature,

(4) $$\Delta \phi = \frac{1}{\sqrt{-g}} \partial_\mu \left(\sqrt{-g} g^{\mu\nu} \partial_\nu \phi\right) = 0,$$

$g^{\mu\nu}$ is the inverse of the metric $g_{\mu\nu}$ of the space, and g is the determinant of $g_{\mu\nu}$. Hence becomes clear that we can regard ϕ as a scalar field on a 4-dimensional space with the metric g_{μ} defined by the background flow. This space is the *acoustic space-time*. The authors who hav pointed this out have largely used it as a tool to illuminate what parts of general relativity ar due to Einstein's field equations specifically, and what phenomena remain when the equation are changed but the form of the geometry remains (this is where black holes are of interest We, however, would like to turn this on its head, making use of the acoustic space-time t illuminate how variable transformations can be used to understand sound propagation in th presence of background flow.

The idea of these transformations is to use a change of variables to convert a challengin equation, such as Eq. (1), into, say, the classic wave equation. To the authors knowledge th most general transformation to date of this kind was first presented by Taylor [5], and is vali for irrotational, barotropic, low Mach number, steady background flows for acoustic fields wher the wavelength is of the same order of magnitude or smaller than the length scale of variation of the background flow [7]. The presentation of this transformation is fairly ad hoc, and w would like to use the acoustic space-time to derive and generalise such transformations in more systematic way.

2 Calculus on Curved Manifolds

In order to proceed further we shall introduce some new tools to deal with calculus o general manifolds. More precisely we shall introduce Geometric Algebra (GA).

GA deals with general manifolds through the concept of embedding. A flat vector spac of large dimension is first defined, within which the, possibly curved, manifold of interest i placed. The embedded manifold then inherits a metric from the extrinsic space, and this allow a study of Riemannian geometry. Some readers may wonder why we use GA instead of th more widely used differential forms and differential geometry (see for example Nakahara [8] ultimately this a preference of the authors. We find that the approach gives more streamline and intuitive proofs of some key results, and also allows more to be expressed independently coordinates. A debate of the relative benefits is beyond the scope of the current paper. As wa

Work Details [OCR]

Text Detection preliminary Results Analysis on 100 images

IOU Threshold	Model	Precision	Recall
0.5	MobileNet	86.24	75.53
0.5	**DBNet**	**87.79**	**77.96**
0.5	Tesseract	76.51	69.54
0.7	MobileNet	46.17	40.94
0.7	**DBNet**	**50.84**	**45.56**
0.7	Tesseract	23.20	21.34

Text Recognition Results Analysis on 100 Images

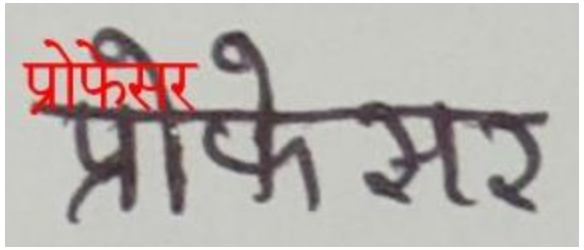
Model	CRR	WRR
MobileNet	64.85	55.80
CRNN_VGG16	65.19	56.02
Tesseract	64.18	52.07

Results and Analysis

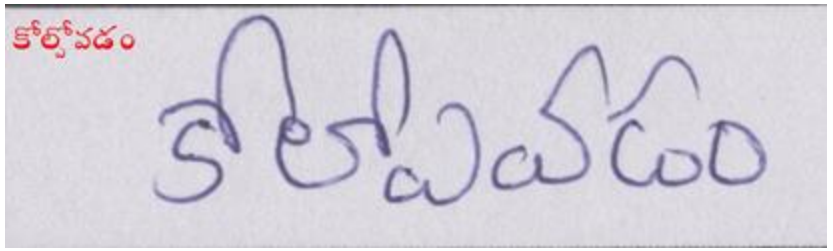
Language	CRR	WRR
Devanagari	92.53	58.34
Tamil	95.75	69.8
Telugu	93.08	56.43

Sample Predictions in different Languages →

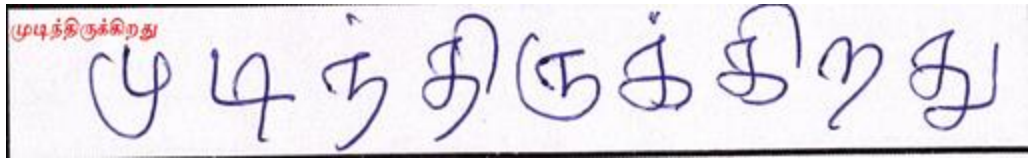
Devanagari



Telugu



Tamil

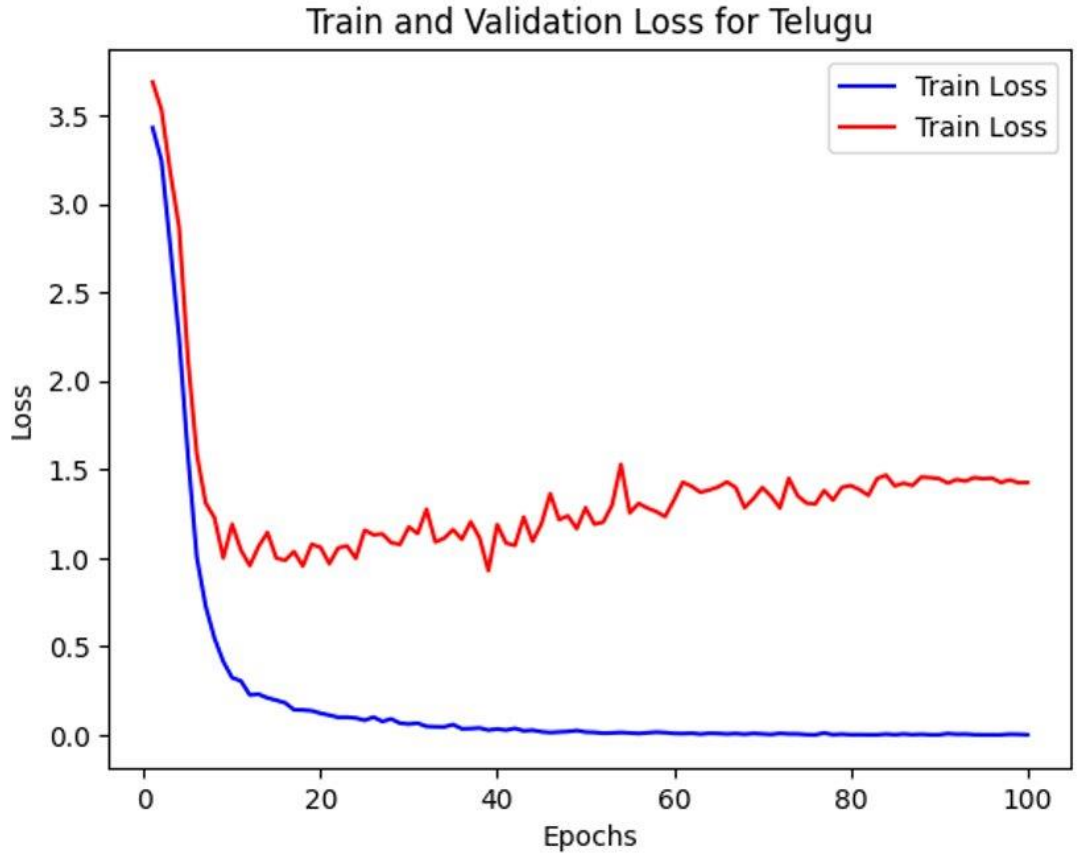


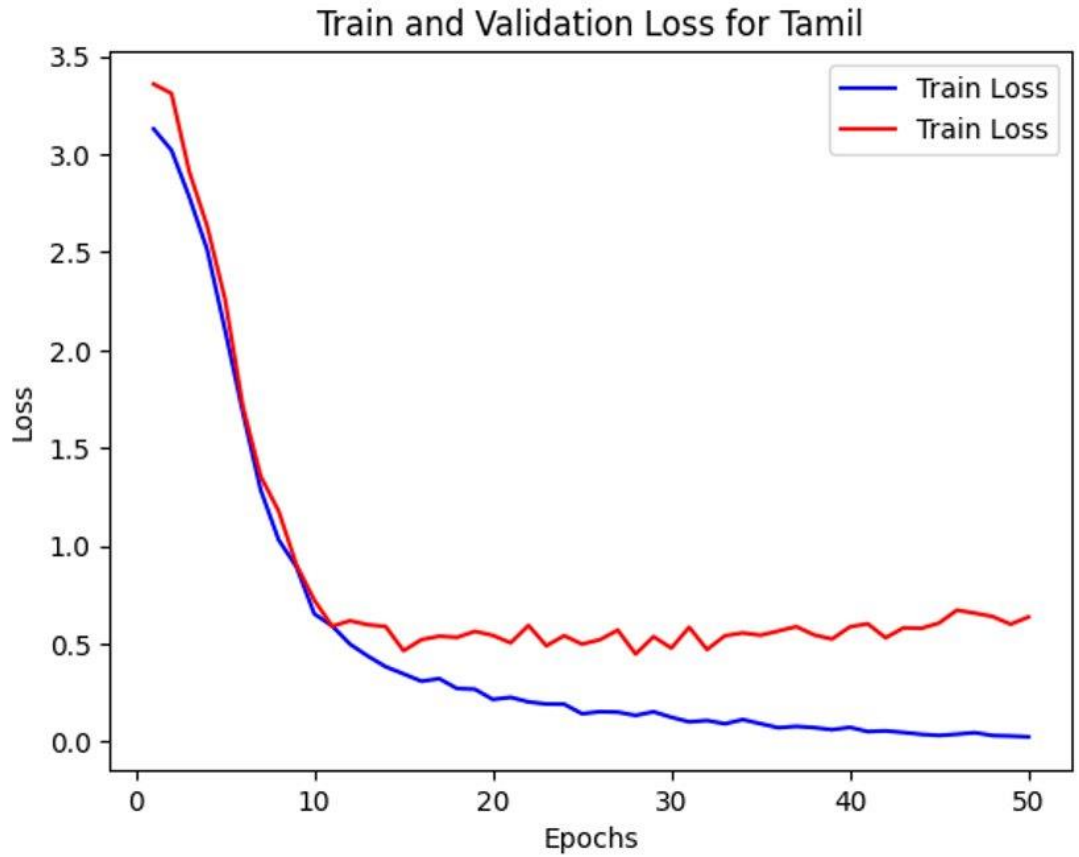
Results of Metrics score for Machine Translation of different models →

Model- Encoder-Decoder(LSTM)
Avg. BLEU score-
0.13247268378107925

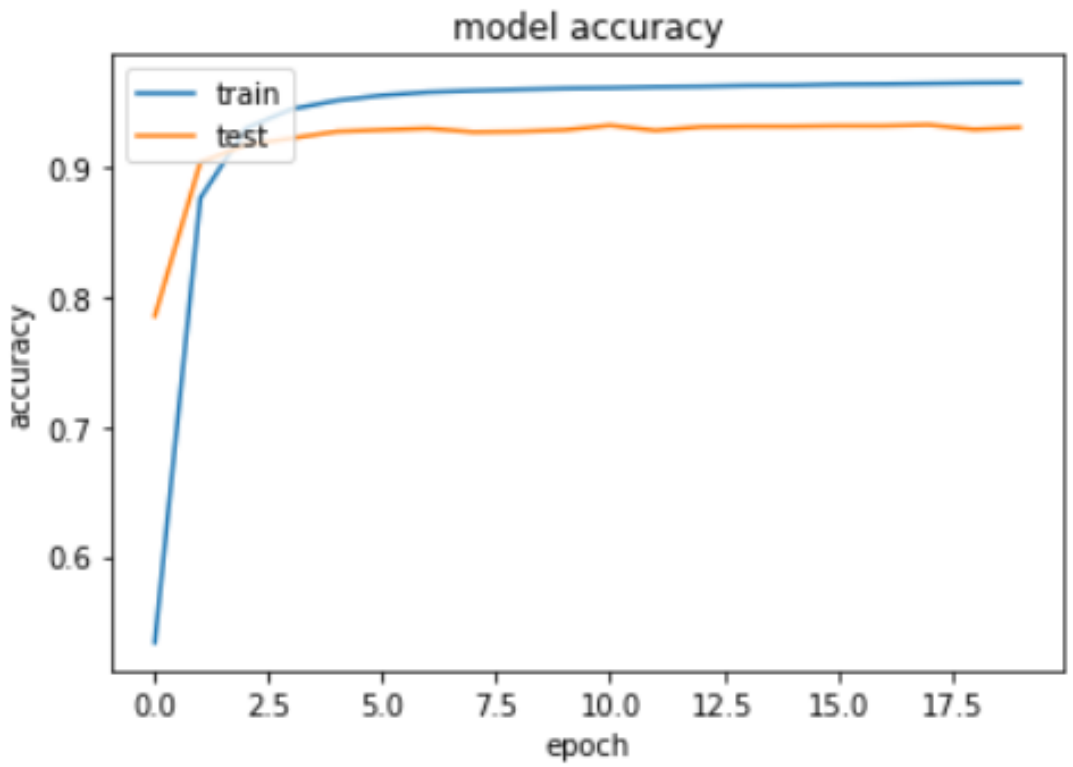
Model:- Transformer
Avg. Bleu Score:-
0.2729667965071246 3

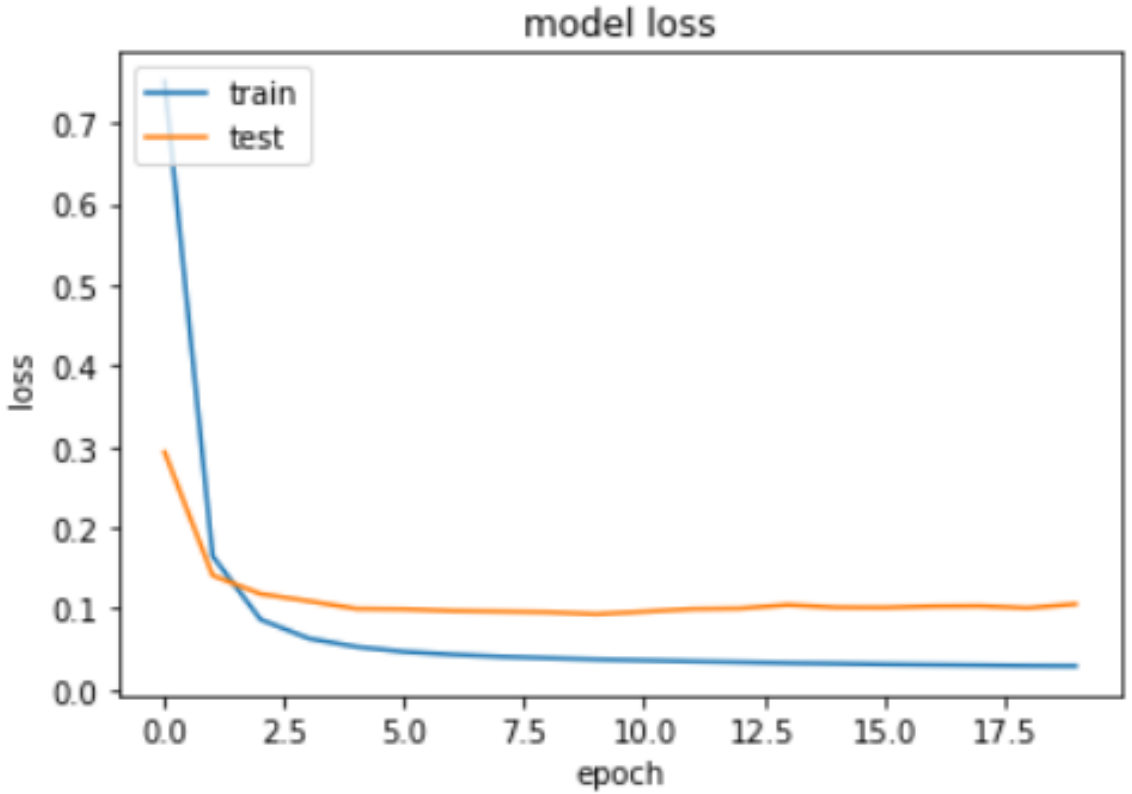
Results and Analysis (Text Recognition)





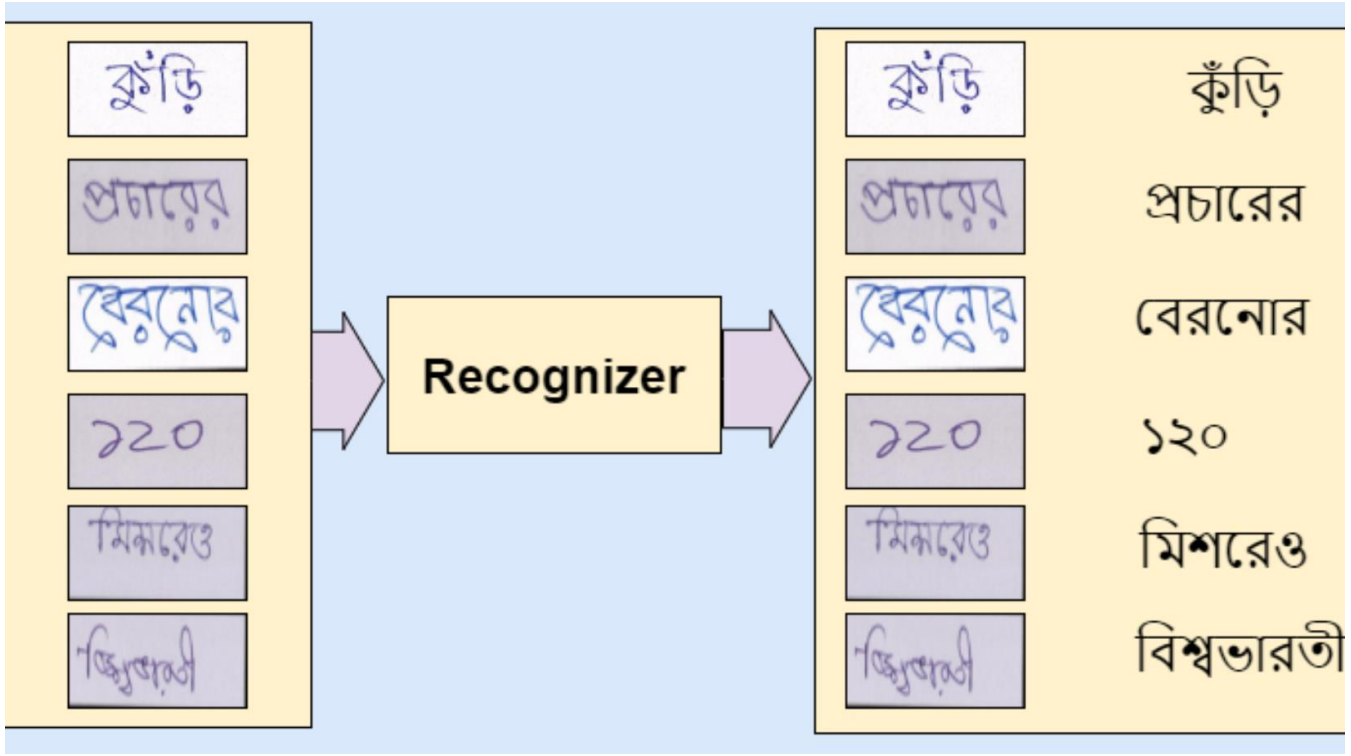
Results and Analysis (Machine Translation)





High Level Work Details and Conclusion

- Implemented of pretrained Summarization Model
- Created of pipeline for End-to-end OCR + Summarization + MT + TTS.
- Increased the scope of OCR data by taking larger data into consideration.
- Trained MT model from scratch and compared two models and also used pre-trained models for comparison for English-to-Hindi translation.
- Making analytical study on different models used and providing conclusive remarks and scope for improvements



6

Demo

A Running example demo for all cases

Evaluation Scheme

1. Attractive and complex problem: 5
2. Clarity in task and in input-output description: 5
3. Dataset effort- collection, annotation: 10
4. Workflow, Architecture, technique: 10
5. Results and analysis: 10
6. Demo working: 20