CS772 MultiModal Machine Translation

Final Project discussion

Manas Jhalani, 22M0806
Vishal Tapase, 22M0807
Palash Moon 22M0816
Atul Kumar Singh 22M0823

30th April 2023

Github Link:

https://github.com/Atulkmrsingh/NLP_PROJECT

Course Code	Course Name	Credits	Tag	Grade	Credit/Audit
CS 635	Information Retrieval & Mining for Hypertext & the Web	6.0	Department elective	AB	C
CS 747	Foundations of Intelligent and Learning Agents	6.0	Department elective	CC	C
CS 768	Learning with Graphs	6.0	Department elective	BB	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 753	Automatic Speech Recognition	6.0	Department elective	BB	C
CS 769	Optimization in Machine Learning	6.0	Department elective	BB	C
CS 772	Deep Learning for Natural Language Processing	6.0	Department elective	AB	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

Report Problem

Year/Semester: 2022-23/Autumn

Course Code	Course Name	Credits	Tag	Grade	Credit/Audit
CE 396	Works Visits	0.0	Core course	W	N
CL 702	Lecture Series	2.0	Core course	W	N
CS 626	Speech and Natural Language Processing and the Web	6.0	Department elective	BB	C
CS 699	Software Lab.	8.0	Core course	BC	C
CS 725	Foundations of Machine Learning	6.0	Department elective	BC	C
GC 101	Gender in the workplace	0.0	Core course	PP	N

Problem Statement

To Develop a multimodal machine translation model for English to Hindi translation using text and images as inputs, with the challenge of effectively integrating visual information from images into the translation process.

Input: A text sentence in English and an image.

Output: A translated text sentence in Hindi that conveys the same meaning as the input text sentence, while also taking into account the visual information provided by the input image.

Related Work

- Multimodal Transformer for Multimodal Machine Translation (Yao & Wan et al,. ACL 2020). https://aclanthology.org/2020.acl-main.400.pdf.
- Multi30K: Multilingual English-German Image Descriptions (D Elliott et al ,. ACL 2016). https://aclanthology.org/W16-3210.pdf
- Multimodal Neural Machine Translation for English to Hindi (sahinur rs et al ,. ACL 2020). https://aclanthology.org/2020.wat-1.11.pdf

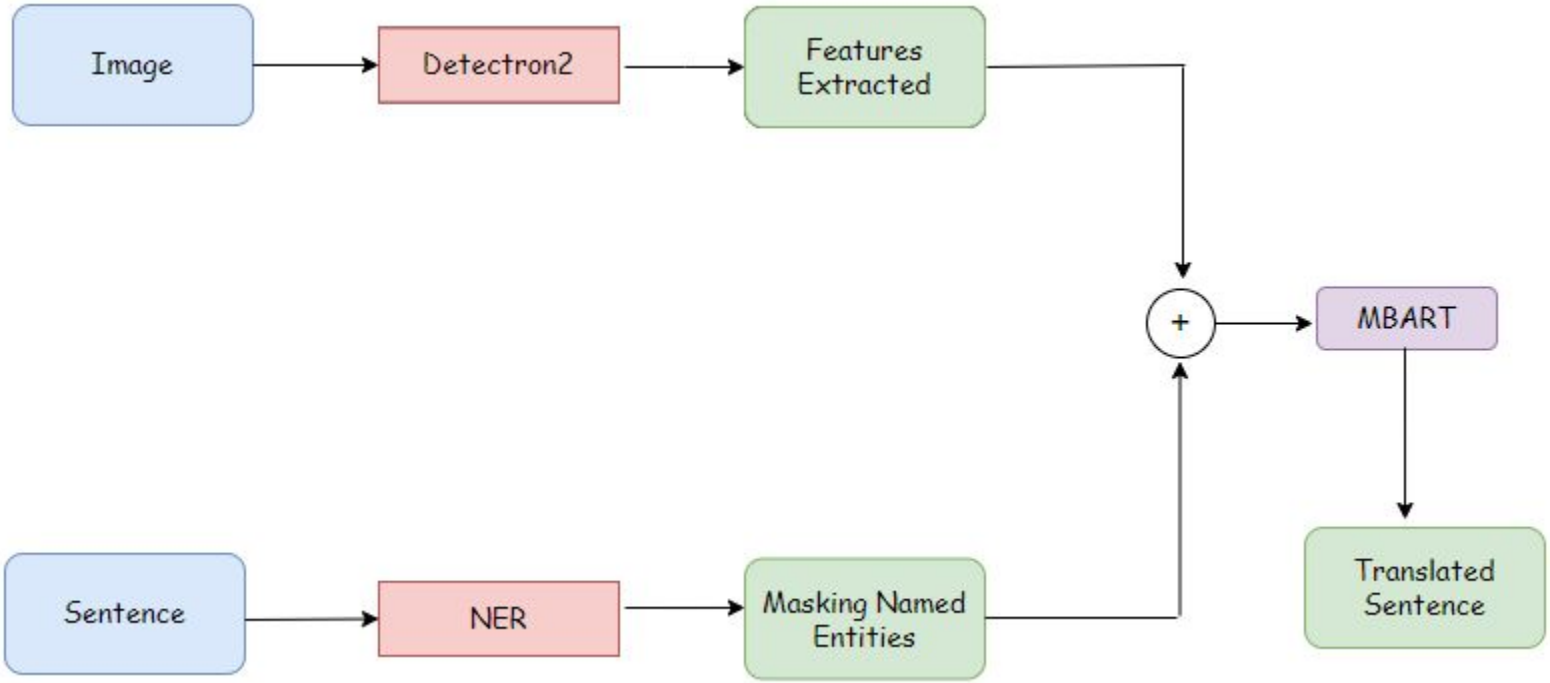
Dataset

- **HINDI VISUAL GENOME:** Hindi Visual Genome is a multimodal dataset that combines images and text in Hindi. It consists of over 29,000 images and corresponding text descriptions in Hindi, which were collected from various sources such as news articles, blogs, and social media. (http://hdl.handle.net/11234/1-2997).
-

Data Set	Items
Training Set	28,932
Development Test Set (D-Test)	998
Evaluation Test Set (E-Test)	1595
Challenge Test Set (C-Test)	1,400

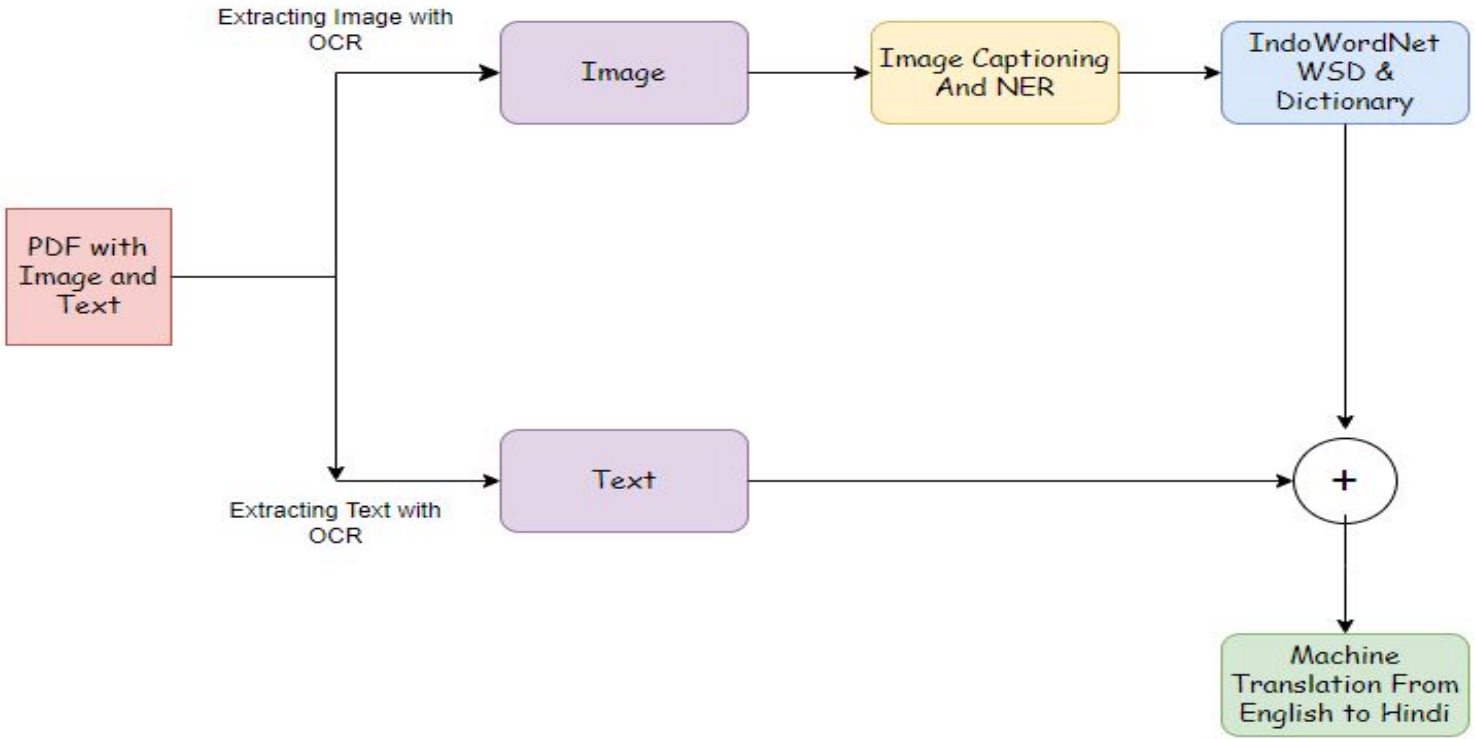
Table 1: Hindi Visual Genome corpus details.

Workflow, Architecture, Technique (1/2)



Architecture 1

Workflow, Architecture, Technique (2/2) (Future Work)



Results and Analysis

Our System	Blue Score
Text-only NMT	38.84
Multi-modal NMT	40.51
Proposed Model MMT (without Masking)	41.65
Proposed Model MMT (with Masking)	43.1

Results and Analysis (Quantitative)

Detectron:

- Framework for object detection and segmentation in images.
- Uses deep learning models like Faster R-CNN and Mask R-CNN for high accuracy.
- Can detect objects and segment them from the background.

Masking named entities:

- Improves accuracy of translations for named entities like people, places, and organizations.
- Involves masking named entities in input text.
- Helps the model learn to translate them more accurately and consistently.
- Named entities are often important for the overall meaning of the text.

Results and Analysis (Qualitative)

Example1:

Text : *“A person riding a motorcycle”*

Masking : *मोटरसाइकिल पर सवार व्यक्ति।*
mBart : *मोटरसाइकिल चलाने वाला व्यक्ति।*
Google Translate : *मोटरसाइकिल पर सवार एक व्यक्ति।*
Chat Gpt : *एक मोटरसाइकिल पर सवार व्यक्ति।*
GPT4: एक व्यक्ति मोटरसाइकिल सवार होता है।

Results and Analysis (Qualitative)

Example 2:

Text : “A large pipe extending from the wall of the court”

Masking : *कोर्ट की दीवार से निकलने वाला एक बड़ा पाइप।*
mBart : *आँगन की दीवार से निकलने वाला एक बड़ा पाइप।*
Google Translate : *कोर्ट की दीवार से निकला एक बड़ा पाइप।*
Chat Gpt : *अदालत की दीवार से बाहर फैला हुआ एक बड़ा पाइप।*
GPT4: *अदालत की दीवार से निकलता एक बड़ा पाइप ।*

Results and Analysis (Qualitative)

Example 3:

Text: *“March 7th is the date on the calender”*

Masking : *मार्च 7th कैलेंडर पर तारीख है।*

Google Translate : *7 मार्च कैलेंडर पर तारीख है।*

Chat Gpt: *७ मार्च कैलेंडर पर तिथि है।*

mbart : *मार्च 7 कैलेण्डर पर तिथि है।*

GPT4*: कैलेंडर पर 7 मार्च तारीख है*

Demo