OdiaGenAI's Participation at WAT2023

SK Shahid Silicon Institute of Technology, Bhubaneswar, India
Guneet Singh Kohli Thapar Institute of Engineering & Technology, Patiala, India
Sambit Sekhar Odia Generative AI, Bhubaneswar, India
Debasish Dhal NISER, Bhubaneswar, India
Adit Sharma Jaypee Institute of Information Technology, Noida, India
Shubhendra Khusawash ITER, Bhubaneswar, India
Shantipriya Parida Silo AI, Helsinki, Finland
Stig-Arne Grönroos Silo AI, Helsinki, Finland
Satya Ranjan Dash KIIT University, Bhubaneswar, India

Abstract

This paper offers an in-depth overview of the "ODIAGEN's" translation system submitted to the Workshop on Asian Translation (WAT2023). Our focus lies in the domain of Indic Multimodal tasks, specifically targeting English to Hindi, English to Malayalam, and English to Bengali translations.
The system uses a state-of-the-art Transformer-based architecture, specifically the NLLB-200 model, fine-tuned with language-specific Visual Genome Datasets. With this robust system, we were able to manage both text-to-text and multimodal translations, demonstrating versatility in handling different translation modes.

Our results showcase strong performance across the board, with particularly promising results in the Hindi and Bengali translation tasks. A noteworthy achievement of our system lies in its stellar performance across all text-to-text translation tasks. In the categories of English to Hindi, English to Bengali, and English to Malayalam translations, our system claimed the top positions for both the evaluation and challenge sets.
This system not only advances our understanding of the challenges and nuances of Indic language translation but also opens avenues for future research to enhance translation accuracy and performance.

1 Introduction

Machine translation (MT) is a well-established field within Natural Language Processing (NLP) that focuses on developing computer software to automatically translate text or speech between different languages. While significant progress has been made in achieving human-level translation for high-resource languages, challenges still remain, especially for low-resource languages. Additionally, recent research has explored the effective integration of other modalities, such as images, into the machine translation process.

The WAT is an open evaluation campaign focusing on Asian languages since 2013 (Nakazawa et al., 2020, 2022).The multimodal translation tasks in WAT2023 consist of image caption translation, in which the input is a descriptive source language caption together with the image it describes, while the output is a target language caption. The multimodal input enables the use of image context to disambiguate source words with multiple senses.

In this system description paper, we explain our approach for the tasks (including the sub-tasks) we participated in:

Task 1: English→Hindi (EN-HI) Multimodal Translation

- EN-HN text-only translation
- EN-HN multimodal translation

Task 2: English→Bengali (EN-BN) Multimodal Translation

- EN-BN text-only translation
- EN-BN multimodal translation

Task 3: English→Malayalam (EN-ML) Multimodal Translation

- EN-ML text-only translation

2 Data Sets

We used the data sets specified by the organizer for the related tasks along without any additional synthetic data.

Task 1: English→Hindi Multimodal Translation

For this task, the organizers provided HindiVisualGenome 1.1 (Parida et al., 2019) 3 dataset (HVG for short). The training part consists of 29k English and Hindi short captions of rectangular areas in photos of various scenes and it is complemented by three test sets: development (D-Test), evaluation (E-Test) and challenge test set (C-Test). Our WAT submissions were for E-Test (denoted "EV" in WAT official tables) and C-Test (denoted "CH" in WAT tables).
The statistics of the datasets are shown in Table 1.

Task 2: English→Malayalam Multimodal Translation

For this task, the organizers provided MalayalamVisualGenome 1.0 dataset4 (MVG for short). MVG is an extension of the HVG dataset for supporting Malayalam, which belongs to the Dravidian language family (Kumar et al., 2017). The dataset size and images are the same as HVG. While HVG contains bilingual English–Hindi segments, MVG contains bilingual English–Malayalam segments, with the English, shared across HVG and MVG, see Table 1.

Task 3: English→Bengali Multimodal Translation

For this task, the organizers provided BengaliVisualGenome 1.0 dataset5 (BVG for short). BVG is an extension of the HVG dataset for supporting Bengali. The dataset size and images are the same as HVG, and MVG, see Table 1.

3 Experimental Details

This section describes the experimental details of the tasks we participated in.

3.1 EN-HI, EN-ML, EN-BN text-only translation

For EN–HI, EN-BN and EN–ML text-only (E-Test and C-Test) translation, the study is rooted in the use of a pre-trained language model known as NLLB-200, which has been fine-tuned utilizing a HVG, BVG, MVG Datasets. The entire process, inclusive of the tokenization pipeline, has been managed internally by the model, all in a PyTorch environment.

Set	Sentences	Tokens			
		English	**Hindi**	**Malayalam**	**Bengali**
Train	28930	143164	145448	107126	113978
D-Test	998	4922	4978	3619	3936
E-Test	1595	7853	7852	5689	6408
C-Test	1400	8186	8639	6044	6657

Table 1: : Statistics of our data used in the English→Hindi, English→Malayalam, and English→Bengali task: the number of sentences and tokens.

The pipeline begins with the tokenization of both English input sentences and corresponding Hindi/Bengali/Malayalam labels. The tokenization, which forms an essential part of the preprocessing stage, converts the sentences into a format that is ingestible by the model. The English sentences, serving as the model's input-ids, and their corresponding HN/BN/ML translations, functioning as labels. The Transformer model learns to map the input English tokens to the HN/BN/ML tokens, learning the nuanced rules of translation in the process.

A crucial part of this system pipeline is the evaluation scheme used to assess the model's performance while training - the BLEU (Bilingual Evaluation Understudy) score. The SACREBLEU toolkit is used to standardize the computation of the BLEU scores, ensuring comparability with other translation models and research.

Post-training, the model's performance is validated on an evaluation set and a challenge set. The evaluation set provides a general performance measure, while the challenge set tests the model's ability to handle more complex and less frequent translation scenarios.

Overall, this pipeline encapsulates the entire process from preprocessing to evaluation, offering a streamlined method for training and validating an English to Hindi/Bengali/Malayalam machine translation model.

3.2 EN-HI, EN-BN Multimodal translation

This section discusses the multimodal translation pipeline for EN-HI and EN-BN. For EN-HI multimodal (E-Test and CTest) translation, we used the object tags extracted from the HVG dataset images for image features and concatenated them with the text. Similarly, For EN-BN (E-Test and C-Test) translation, we used object tags extracted from the BVG dataset.

We derive the extracted object tags using a pre-trained Faster RCNN with ResNet101-C4 backbone, which can recognize 80 object types that constitute the COCO Dataset ADD CITATION. In the next step, we select the top 10 tags based on the confidence scores, and in case the object tags are less than 10, we select all the detected tags. The original input English instance is concatenated with a '##' as a separator followed by a comma separated detected tags. This formatted input loaded with visual context from the object tags is fed into the mBART Encoder for processing.

4 Results

We report the official automatic evaluation results of our models for all the participating tasks in Table 2.

Following the fine-tuning process, these models were used to infer translations on two distinct sets for each language: the evaluation set and the challenge set. The translation quality was evaluated using the BLEU (Bilingual Evaluation Understudy) score, and RIBES (Ranking by Incremental Bilingual Evaluation System) score.

For the English to Hindi model, a BLEU score of 44.60 was achieved on the evaluation set, while a score of 53.60 was obtained for the challenge set. These results highlight the model's strong performance and its capacity to handle more complex or unusual translation tasks.

In the case of the English to Bengali model, a BLEU score of 49.20 was reached on the evaluation set, with a slightly lower score of 47.80 on the challenge set. This indicates a robust overall performance and a commendable capability to handle nuanced translations specific to the Bengali language.

Lastly, for the English to Malayalam model, the system achieved a BLEU score of 46.60 on the evaluation set and 39.70 on the challenge set. Despite a slightly lower score on the challenge set, the model still demonstrates a respectable performance in translating English to Malayalam.

Translation Model	**Translation Type**	**BLEU Score (Evaluation Set)**	**BLEU Score (Challenge Set)**
English to Hindi	Text-to-Text	44.60	53.60
	Multimodal	43.70	46.09
English to Bengali	Text-to-Text	49.20	47.80
	Multimodal	36.90	45.40
English to Malayalam	Text-to-Text	46.60	39.70

Table 2: BLEU scores of the text-to-text and multimodal translation models on the evaluation and challenge sets.

The lower BLEU score on the English to Malayalam translation task can be due to lot of possible factors , one of which is Linguistic Complexity, as Malayalam is a Dravidian language known for its complex grammatical structures and a rich set of linguistic phenomena, which may not be easily captured by the model. This complexity can make the mapping from English to Malayalam challenging.

5 Conclusion

In this system description paper, we presented our system for three tasks in WAT2022: (a) English→Hindi, (b) English→Malayalam, and (c) English→Bengali Multimodal Translation. We released the code through Github for research7.

These empirical results underscore the effectiveness of the methodology adopted for these machine translation models. Leveraging a fine-tuned NLLB-200 model with language-specific Visual Genome Datasets provides a robust solution to the machine translation task for the languages under study: Hindi, Bengali, and Malayalam. The results also pave the way for further enhancements and investigations in the realm of machine translation.

References

Grönroos, S.-A., Huet, B., Kurimo, M., Laaksonen, J. T., Mérialdo, B., Pham, P., Sjöberg, M., Sulubacak, U., Tiedemann, J., Troncy, R., and Vázquez, R. (2018). The memad submission to the wmt18 multimodal translation task. In *Conference on Machine Translation*.

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F. B., Wattenberg, M., Corrado, G. S., Hughes, M., and Dean, J. (2016). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, page 311–318, USA. Association for Computational Linguistics.

Parida, S., Bojar, O., and Dash, S. R. (2019). Hindi visual genome: A dataset for multi-modal english to hindi machine translation. *Computación y Sistemas*, 23(4):1499–1505.

Parida, S., Panda, S., Grönroos, S.-A., Granroth-Wilding, M., and Koistinen, M. (2022). Silo nlp's participation at wat2022. *arXiv preprint arXiv:2208.01296*.

Popel, M., Tomková, M., Tomek, J., Łukasz Kaiser, Uszkoreit, J., Bojar, O., and Žabokrtský, Z. (2020). Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, 11.

Rathi, A. (2020). Deep learning apporach for image captioning in hindi language. *2020 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, pages 1–8.

Sulubacak, U., Caglayan, O., Gronroos, S.-A., Rouhe, A., Elliott, D., Specia, L., and Tiedemann, J. (2019). Multimodal machine translation through visuals and speech. *Machine Translation*, 34:97 – 147.

Tang, Y., Tran, C., Li, X., Chen, P.-J., Goyal, N., Chaudhary, V., Gu, J., and Fan, A. (2020). Multilingual translation with extensible multilingual pretraining and finetuning. *ArXiv*, abs/2008.00401.

Team, N., Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., Kalbassi, E., Lam, J., Licht, D., Maillard, J., Sun, A., Wang, S., Wenzek, G., Youngblood, A., Akula, B., Barrault, L., Gonzalez, G. M., Hansanti, P., Hoffman, J., Jarrett, S., Sadagopan, K. R., Rowe, D., Spruit, S., Tran, C., Andrews, P., Ayan, N. F., Bhosale, S., Edunov, S., Fan, A., Gao, C., Goswami, V., Guzmán, F., Koehn, P., Mourachko, A., Ropers, C., Saleem, S., Schwenk, H., and Wang, J. (2022). No language left behind: Scaling human-centered machine translation.

Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In *NIPS*.

Yang, S., Wang, Y., and Chu, X. (2020). A survey of deep learning techniques for neural machine translation. *arXiv preprint arXiv:2002.07526*.

Yang et al. (2020) Team et al. (2022) Parida et al. (2019) Grönroos et al. (2018) Parida et al. (2022) Popel et al. (2020) Rathi (2020) Tang et al. (2020) Sulubacak et al. (2019) Vaswani et al. (2017) Johnson et al. (2016) Papineni et al. (2002)