2024 International Conference on Computing and Data Science (ICCDS-2024)

Multilingual Transformer for Dynamic Cricket Commentary Generation

Gaurav Kamath
Department of Computer Scienceand Engineering (AI and ML)
Dayananda Sagar University
Bangalore, India
gauravkamath2002@gmail.com

Udayabhaskara N
Department of Computer Scienceand Engineering (AI and ML) Dayananda Sagar University Bangalore, India
udaya1918@gmail.com

2024 International Conference on Computing and Data Science (ICCDS) | 979-8-3503-6533-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCDS60734.2024.10560456

***Abstract*—The sport of Cricket is growing in popularity which necessitates the use of technology to adhere to popular game standards. Since this is a data rich and most elaborate game, understanding this for someone who's new to the gamehad been challenging. We are aiming to solve this unavailability of the game through native commentary to audience of all major languages. To outline our entire research work we ideate a Cricket Game AI through computer vision, use it to produce text commentary of game-play and perform Machine Translation(MT) to produce native commentary for a targeted language, As afirst part of the outlined process in this work we focus on the implementation of the machine translation component of this pipeline, evaluating various methods including Seq2seq and transformer based architectures. Our method of cross feeding Seq2Seq with Transformer model results in Bilingual Evaluation Understudy (BELU) score on Validation set of 2.08.**

***Index Terms*—Seq2Seq, Yolov9, Transformer architecture, Automatic Cricket Commentary, Machine Translation**

I. INTRODUCTION

Sports commentary plays a pivotal role in enhancing the viewer experience, offering insightful analysis, and a compelling narrative around the unfolding events. In cricket, commentary serves as a crucial medium through which fans connect with the game, gaining a deeper understanding of the strategies, nuances, and historical context that define each match. However, this immersive experience is predominantly limited by linguistic constraints, as the scarcity of native sports experts hampers the availability of diverse commentary options. Consequently, the current approach predominantly caters to specific linguistic demographics, leaving a significant portion of the global audience without access to comprehensive and engaging cricket commentary. Recognizing the significance of breaking down these language barriers, our re-search focuses on employing hybrid machine translation model by combining seq2seq with multi head attention layer from transformer architecture for the automated cricket commentary generation.

One of the common approaches for achieving this is to directlytranslate the commentary that's being broadcasted but such an approach suffers from translation delay issues considering the fact that there would be a delay from the commentary made bythe commentators to the translation of the commentary to the target language. Our work aims to mitigate this shortcoming by introducing computer vision powered Game AI (passive partof the work) that produces action parameters which is fed to the LLM trained with commentary dataset [1] that produces commentary for the desired language.

We propose a pipeline which not only focuses on the translation aspect of the commentary but also on dynamic commentary generation, The Game AI model takes a series of frames as input and an object detection model Yolov9 is used to describe the scene, which includes features such a line of ball, length of the ball, foot position of batsman, speed of ball, outcome of the ball, position the ball is placed by the batsman, etc. These serve as features for Machine Model which will be trained to understand features for the game of cricket, The machine model will generate prompts based on a knowledge base which has encoded game knowledge and fed into a large language model which generates the commentary. This commentary is then translated into other languages to allow a larger viewership. Fig 1 depicts the proposed pipeline.

By leveraging pre-trained language models and adapting them to real-time cricket scenarios, we aim to transcend the limitations imposed by linguistic constraints. This approach not only broadens the global appeal of cricket commentary but also democratizes access to the sport by creating a language-agnostic system. In this paper, we delve into the pressing need for such an innovation, its potential impact on the inclusivity of cricket commentary, and the methodologies employed. Through our exploration of multilingual transfer learning, we aim to pave the way for a more universally accessible and engaging cricket commentary experience.

II. RELATED WORKS

In this phase of the research, the primary emphasis lies on the machine translation aspect of generating multilingual cricket commentary. Throughout this stage, we examined research works specifically addressing machine translation and studies aimed at enhancing the efficacy of machine translation systems.

Recently the rise of Large Language Model 's(LLM) has led to LLM 's to become a popular choice for machine translation, One such research analyzes the difficulties of typical neural

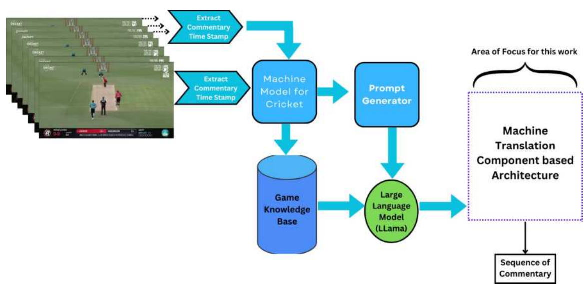

Fig. 1. Proposed pipeline for cricket commentary generation

machine translation systems in translating culturally unique information and presents a novel way to enhancing LLM-based machine translation with cultural awareness. The authors provide a data curation process for creating a culturally appropriate parallel corpus supplemented with annotations of cultural-specific items. They also develop suggesting tactics to aid with LLM-based translation. The experimental findings suggest that their techniques outperform typical NMT systems in translating culturally specific words. The studyalso discusses how to improve cultural awareness in LLM translation and introduces a new assessment tool for judging the translation quality of cultural notions [2].

Moving on to another study which proposes a method called DIPMT which incorporates bilingual dictionaries into prompting large language models for machine translation. This method improves translation quality for low-resource and out-of-domain scenarios. It involves adding word-level dictionary translations to the translation prompt, providing hints for rare words. Extensive experiments show significant improvements in translation quality using DIPMT compared to baseline methods [3].

LLM based MT systems are proven to be highly accurate but have a large number of parameters which requires a lot of training time and computational resources, In our study speed is of utmost importance in order to ensure the synchronization between video stream and commentary hence me move to a deep learning approach for machine translation.

A study by [4] highlights the central role of DNN, employing techniques like supervised learning with Recurrent Neural Networks (RNN), addressing word alignment challenges, and refining translation rule selection. Notable enhancements include the integration of word2vec and GPU support for improved performance.

Expanding on this shift, [5] explores the evolving landscape of machine translation dominated by deep learning. It distinguishes between component-wise methods, enhancing Statistical Machine Translation (SMT) components, and end- to-end methods directly mapping source and target languages. Deep learning addresses challenges in SMT, offering solutionsfor data sparsity and feature engineering. Improvements in word alignment, translation rule estimation, phrase reordering, language modeling, and feature combination highlight deep learning's transformative impact on traditional SMT.

Continuing this exploration into the integration of deep learning into machine translation, a study by [6] sheds light on its application in both Statistical Machine Translation (SMT) and Neural Machine Translation (NMT). With a focus on enhancing components like word alignment, transition rule estimation, phrase reordering, and language modeling within SMT, deep learning proves to be a valuable asset. The evolution towards Neural Machine Translation, employing theEncoder-Decoder architecture, represents a significant stridein the field. Attention Mechanism is introduced to tackle challenges posed by longer source sentences, marking a notable enhancement over traditional SMT methods.

Advancements in neural machine translation take center stage in recent studies, as explored by [7]. The paper delves into various models, including RNN-based, convolutional, and Transformer models, highlighting their individual strengths. Notably, the authors introduce the RNMT+ model, a hybrid approach combining techniques from these architectures. The RNMT+ model outperforms individual models on benchmark datasets, demonstrating the efficacy of hybrid architectures in pushing the boundaries of translation quality. The experiments, conducted on standard datasets with sub-word units, utilize BLEU scores on test data as the primary evaluation metric, providing a comprehensive assessment of the proposed model ‘s capabilities.

In a complementary effort to enhance neural language translation models, another study focuses on developing a robust sequence-to-sequence approach. The model leverages a transformer architecture with LSTM for both encoding and decoding, showcasing its adaptability and efficiency. The training, specifically on English Japanese translation datasets, yields high BLEU scores. Furthermore, the model's versatility is demonstrated on both GPU and CPU, achieving a notable BLEU score of 40.1. This work not only contributes to improved translation performance but also highlights the model's computational efficiency, a crucial aspect in real-world applications [8].

[9] Seq2Dep, a Sequence-to-Sequence model addressing challenges in Neural Machine Translation by incorporating syntactic dependency information. This approach transforms output into a linearized dependency tree, enhancing predictions with Depth-first pre-order traversal. Seq2Dep shows significant BLEU score improvements, highlighting the impact of syntactic dependency in refining translations.

In a comparative study, [10] evaluates Seq2Seq, ConvSeq2Seq, RNN, and MHA models for English-Indonesian translation. ConvSeq2Seq outperforms others based on BLEU and GLEU scores, emphasizing the significance of atten-tion mechanisms in NMT. The study explores fine-tuning strategies, contributing valuable insights to optimize NMTperformance.

finally [11] proposes a Sequence to Sequence Mixture Model to address translation diversity and quality. This innovative approach utelizes a committee of specialized models, optimizing training data selection to create clusters within the parallel corpus. Results demonstrate superior performance without additional computational costs, showcasing the potential of tailored model ensembles for enhanced translation quality.

Due to the lack of resources and papers available on sports commentary generation we focused on reviewing papers proposing real time commentary generation, Real time commentary systems has gained popularity over the years due tothe increase in hardware capability to handle translation in real time, one such study discusses the use of artificial intelligencein live stream chat translation. The author proposes a Seq2Seqmodel for two-way translation, Utilizing LSTM variant of RNN for feature extraction and a bidirectional language modelfor context identification. An attention network is used to aid in the prediction. The translation process involves tokenization, encoding, and decoding. The encoder captures meaningful features from the tokens, while the decoder processes the ground-truth translation. The combined hidden state utilized global attention for improving the prediction. [12]

Another study [13], describes a real-time neural machine translation framework that attempts to balance translation quality and delay. It introduces an agent that decides when to translate after interacting with a pre-trained NMT environment. When evaluated on two language pairs, the framework outperformed existing approaches. The proposed architecture consists of an encoder for input words and a decoder for translated words, with an agent determining whether to read or write based on observations. It utilizes reinforcement learning to train, with a reward function that considers translation quality and delay. Pre-training takes place within the NMT environment, and the incentive function includes BLEUscore evaluation.

finally [14], focused on making a real-time machine translation system aimed at translating English to local Indian languages, utilizing Neural Machine Translation (NMT) and OpenNMT. The system's principal goal is to overcome communication hurdles resulting from India's heterogeneous linguistic terrain. The suggested system uses a 2-layer LSTM model to achieve precise translation results and includes speech-to-text and text-to-speech modules for instant translation. The system stores its findings in a text file for future examination. It predicts providing faster and more accurate translations than traditional approaches such as RBMT and SMT.

III. METHODOLOGY

This work can mainly be divided into three phases 1) data acquisition and preprocessing phase, 2) Machine Translation phase and 3) post translation and feedback phase.

A. Data Acquisition and Preprocessing

Referring to the figure 1, the area of focus of this work has been highlighted where all the required dataset is acquired in two parts one is through Video commentary extraction.

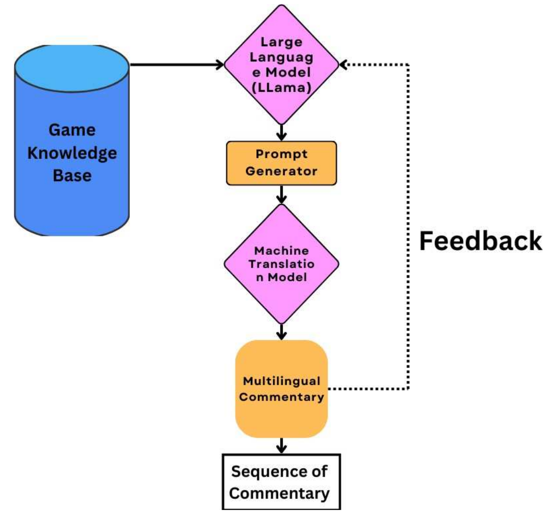

Fig. 2. Methodology Flowchart

and the other dataset utilized for this research comprises the commentary from the 2023 Indian Premier League (IPL) season, A total of 69 matches involving 10 different teams commentary was collected from the 2023 IPL season. This comprehensive collection encompasses both league matches and playoffs. The commentary for each match was recorded on a ball-by-ball basis and stored in text format, ensuring a detailed and organized dataset for analysis. Over 10,000 balls of commentary were collected for this study. The data was web scraped from cricket commentary websites. The dataset has been made available online for other researchers to utilize [1].

The dataset utilized for training the machine translation model 's was Multi30K, The English to German Data set was specifically used for this study but can be similarly adopted to other language translations [15].

Before feeding the collected ball-by-ball commentary data into a machine translation (MT) model, a series of preprocessing steps are employed to ensure optimal performance and coherence in language translation. Initially, the raw text undergoes tokenization, breaking it down into smaller units like words or subwords. This step facilitates a more granular analysis of the commentary, capturing nuanced linguistic structures. Following tokenization, common preprocessing techniques,lowercasing is applied to standardize the text, ensuring consistency in the translation process. Following this the removal of stop words, punctuation, and special characters further refined the input, eliminating irrelevant elements that mayhinder the model's comprehension. Additionally, we perform lemmatization to reduce words to their root forms, enhancing semantic understanding. These preprocessing steps collectivelycontribute to a cleaner and more linguistically coherent dataset, ultimately optimizing the machine translation model's ability to generate accurate and contextually appropriate commentary in the desired target languages.

TABLE I
ACTION PARAMETERS TABLE

Focus	Action	Vim score (100)	Player	Ball position	Shot
pitch	run up	30	A	hand	none
	audience	88	AZ	in air	hook
fielder	catch	98	B	player C	upper cut
	ball release	78	C,A,D	ground	square cut
shot	trajectory	50	C	before bat	defense
	umpire	60	D	player C	sweep

B. Machine Translation

The machine translation component for this project canbe divided into two components Passive component, which is mostly conceptual and not implemented for this stage of the study and Active component which consists of the modules implemented in this study.

The passive component consists of a game AI agent which is a conceptual part of this work assuming a Game AI agent learns on the game play of the event, e.g. sequence of action such as "release of ball", "shot played" and "position of the ball" could be determined through game AI Agent toproduces action parameters I. This enables the commentarygeneration system to have information before hand to allow machine translation to take place unlike waiting for the commentary to complete in broadcast language to translate it to target language, allowing a delay free experience to the viewer and a synchronization between the video stream and translated commentary. The Game AI agent contributes to knowledge base component, within this repository, cricket- specific terminologies are preserved, ensuring they are not mistranslated into other languages. For instance, "square leg" in cricket retains its meaning rather than being translated as separate entities "square" and "leg". By encoding cricket-related knowledge, biases in the machine translation process are reduced significantly.

Assuming that an AI agent which we describe as in the passive component, be able to generate action parameters as represented in Table I. Where 'Focus' signifies the action in focus in the game, 'Action' parameter identifies the action taking place in the focused frame. 'Vim' score ranges from 0 to 100 where it signifies the excitement level of the game through Game analysis. 'Player'- player/players in frame,
"Ball position" and 'shot' parameter provides more details to the commentary. Attributes given in the table could be further refined as the detail may increase.

In addition, the Knowledge Base integrates insights from other sports commentary such as football through transfer learning. Given the contrasting nature of these games, the commentary style differs significantly. Taking cricket as an example, the pace of the game tends to be more measured, with moments of intense action along with periods of calm.

The active component of our work is to use machine translation (MT) models to facilitate the multilingual generation of cricket commentary. Specifically, we utilize Seq2Seq, and transformer based Seq2Seq architectures, which have demonstrated exceptional performance in natural language processing applications.

1) Seq2Seq: Sequence-to-sequence (Seq2Seq) models have proven useful in natural language processing applications, notably machine translation. These models work by receiving a series of input data, such as a sentence in one language, and producing an output sequence, such as the translated sentence in another language. One of Seq2Seq model 's key features is their capacity to accommodate variable-length input and output sequences, which allows them to capture a wide range of linguistic patterns [16].

However, one of the primary drawbacks of Seq2Seq models is their susceptibility to issues related to context and information compression. These models tend to struggle with retaininglong-term dependencies within a sequence, which can lead to a loss of contextual information, especially in lengthy sentences. This limitation is often referred to as the "short- term memory" problem. As a result, Seq2Seq models may encounter challenges when dealing with complex languagestructures or tasks that require a deep understanding of context, such as capturing intricate nuances in sentiment or accurately translating idiomatic expressions. This proves to be crucial in the case of cricket commentary translation where sentiments of the commentator are essential for enhanced viewer experience.

2) Transformer Based Seq2Seq: Transformer-based Seq2Seq models have emerged as a transformative solution to address the contextual and information compressionissues inherent in traditional Seq2Seq architectures [17]. The introduction of transformers, as popularized by models like the Transformer architecture, has revolutionized the field of sequence-to-sequence learning. The key innovation lies in the attention mechanism, which enables the model to selectively focus on different parts of the input sequence when generating the output sequence.

In traditional Seq2Seq models, the entire input sequence is encoded into a fixed-size context vector, leading to challenges in handling long-term dependencies. Transformers, on the other hand, leverage self-attention mechanisms, allowing the model to assign varying degrees of importance to different parts of the input sequence. This dynamic attention mechanismsmenables the transformer to capture long-range dependencies more effectively, mitigating the "short-term memory" problem associated with traditional Seq2Seq models.

Moreover, transformers facilitate parallelization during training, enhancing computational efficiency and reducing training times compared to sequential processing. The ability to capture contextual information across the entire input sequence and parallelized computations makes transformer-based Seq2Seq models highly effective in a wide range of natural language processing tasks [18].

C. Post Translation & Feedback

In this phase of the work, we obtain the translations achieved by the Machine Translation phase and cross verify the Bilingual Evaluation Understudy (BELU) score on Validation set with translation achieved by that of Transformer model. The aim here is the use of seq2seq with multi-head attention layer from transformer model should convincingly minimize the BELU score on Validation set. through the feedback from this phase to the Machine translation phase and cross-validation of the same shows the decrease in the BELU score on Validation set from 4.27 to 2.08. This approach ensures that the generated commentary aligns with the natural linguistic expressions of native speakers, enhancing the overall viewership experience. Fig 2 depicts the flow of the machine translation pipeline.

IV. RESULTS & ANALYSIS

A. Results

Both Seq2Seq and Seq2Seq models with transformers underwent training on the Multi30K English-to-German dataset, utilizing the same GPU infrastructure for a fair comparison. Remarkably, the Seq2Seq model with transformers exhibited an average training time that was more than twice as longas that of the traditional Seq2Seq model. This discrepancy in training duration can be attributed to the increased complexity introduced by the transformer layers in the architecture. Table II depicts the Bilingual Evaluation Understudy (BELU) Score on validation set and training time per epoch for each model.

TABLE II
TABLE OF RESULTS

Model Name	BELU Score	Training Time Per Epoch (s)
Seq2Seq	4.272	23.43
Seq2Seq with Transformers	2.083	55.74

Despite the prolonged training time, the Seq2Seq model with transformers showcased a marked improvement in performance, achieving an overall BELU score that was significantlylower than its traditional Seq2Seq counterpart on the Multi30Kvalidation set.

B. Result Analysis

Based upon the results it is apparent that the Seq2Seq with Transformers model demands a considerably longer training time compared to its traditional Seq2Seq counterpart, a little over double the training duration. This extended training time can be attributed to the added complexity introduced by the transformer layers in the architecture. However, the promising aspect arises when considering the performance metric. However, the Seq2Seq with Transformers model showcases a significantly lower BELU score on the validation set compared to the Seq2Seq model, indicating superior effectiveness in capturing intricate linguistic structures and contextual dependencies within the dataset.

These preliminary findings suggest that, while the Seq2Seq with Transformers model incurs a computational cost in terms of training time, the trade-off results in a model that excels in accuracy and efficiency. A more comprehensive analysis, including hyperparameter tuning and testing on diverse datasets, is necessary to fully appreciate the potential advantages and drawbacks of each model in the context of language translationtasks.

V. CONCLUSION & FUTURE WORKS

A. Conclusion

The proposed pipeline and the various stages essential for the generation of cricket commentary. The implemented modules within this phase include the machine translation model, specifically designed to translate commentary from English into multiple languages, thereby expanding the viewership reach. The emphasis of this approach revolves around achieving efficiency and speed in the commentary generation process. To align with this objective, Seq2Seq models were chosen over larger language models. This selection aims to strike a balance between computational efficiency and the demand for real-time commentary generation, making Seq2Seq models well-suited for the task at hand.

B. Future Works

In the project's next phase, we aim to include application of transfer learning techniques, which would transmit style of commentary from one sport to another. This strategy aims to foster commenting styles that closely resemble those of native speakers. Using transfer learning allows the model to acquire insights and linguistic qualities from an expert commentator, resulting in a more genuine and contextualized output. By bridging the gap between diverse commentator styles, the study aims to increase the diversity and authenticity of produced cricket commentary, providing spectators with a more immersive and culturally relevant experience.

Following the integration of transfer learning, the project will extend its capabilities by incorporating YOLOv9 object detection which would be a fundamental component of the game AI agent. The game AI agent will be employed to create a detailed scene description, identifying key events within the cricket match. This includes capturing intricate details such as the batsman's foot position, the bowler's foot position,the speed of the ball, and the subsequent result of the ball.This would be an essential component for the game AI agent which would be used to generate the commentary, these fine- grained scene descriptions will serve as critical inputs for the commentary generation process.

ACKNOWLEDGMENT

We express our heartfelt gratitude to Dayananda Sagar University and the Department of Computer Science and Engineering AI & ML, for their invaluable support and guidance throughout the duration of our research. Our sincere thanks extend to the faculty of the AIML Department at Dayananda Sagar University for their support, valuable feedback, andunwavering guidance that have played a crucial role in shapingand enriching this research.

REFERENCES

[1] Gaurav Kamath, and Udayabhaskara N. (2024). IPL 2023 Match wise Ball by Ball Commentary [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/7918558

[2] Yao, Binwei, Ming Jiang, Diyi Yang, and Junjie Hu. "Empowering LLM-based machine translation with cultural awareness." arXiv preprint arXiv:2305.14328 (2023).

[3] Ghazvininejad, Marjan, Hila Gonen, and Luke Zettlemoyer. "Dictionary-based phrase-level prompting of large language models for machine translation." arXiv preprint arXiv:2302.07856 (2023).
[4] Singh, Shashi Pal, Ajai Kumar, Hemant Darbari, Lenali Singh, Anshika Rastogi, and Shikha Jain. "Machine translation using deep learning: An overview." In 2017 international conference on computer, communications and electronics (comptelix), pp. 162-167. IEEE, 2017.
[5] Liu, Yang, and Jiajun Zhang. "Deep learning in machine translation." Deep learning in natural language processing (2018): 147-183.
[6] Srivastava, Siddhant, Anupam Shukla, and Ritu Tiwari. "Machine translation: from statistical to modern deep-learning practices." arXiv preprint arXiv:1812.04238 (2018).
[7] Chen, Mia Xu, Orhan Firat, Ankur Bapna, Melvin Johnson, Wolf- gang Macherey, George Foster, Llion Jones et al. "The best of both worlds: Combining recent advances in neural machine translation." arXivpreprint arXiv:1804.09849 (2018).
[8] Gupta, Meenu, and Prince Kumar. "Robust neural language translation model formulation using Seq2seq approach." Fusion: Practice and Applications 5, no. 2 (2021): 61-67.
[9] Le, An Nguyen, Ander Martinez, Akifumi Yoshimoto, and Yuji Matsumoto. "Improving sequence to sequence neural machine translation by utilizing syntactic dependency information." In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 21-29. 2017.
[10] Puspitaningrum, Diyah. "A Study of English-Indonesian Neural Machine Translation with Attention (Seq2Seq, ConvSeq2Seq, RNN, and MHA) A Comparative Study of NMT on English-Indonesian." In Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology, pp. 271-280. 2021.
[11] He, Xuanli, Gholamreza Haffari, and Mohammad Norouzi. "Sequence to sequence mixture model for diverse machine translation." arXiv preprint arXiv:1810.07391 (2018).
[12] Sutandijo, Vincent Joyan, and Nunung Nurul Qomariyah. "Artificial Intelligence Based Automatic Live Stream Chat Machine Translator." Procedia Computer Science 227 (2023): 454-463.
[13] Gu, Jiatao, Graham Neubig, Kyunghyun Cho, and Victor OK Li. "Learning to translate in real-time with neural machine translation." arXiv preprint arXiv:1610.00388 (2016).
[14] Vyas, Raj, Kirti Joshi, Hitesh Sutar, and Tatwadarshi P. Nagarhalli. "Real time machine translation system for english to indian language." In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), pp. 838-842. IEEE, 2020.
[15] Elliott, Desmond, Stella Frank, Khalil Sima'an, and Lucia Specia. "Multi30k: Multilingual english-german image descriptions." arXiv preprint arXiv:1605.00459 (2016).
[16] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 (2014).
[17] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30 (2017).
[18] Egonmwan, Elozino, and Yllias Chali. "Transformer and seq2seq model for paraphrase generation." In Proceedings of the 3rd Workshop on Neural Generation and Translation, pp. 249-255. 2019.