International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 3, June 2024, pp. 3434~3442
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i3.pp3434-3442

A multimodal machine learning approach to generate news articles from geo-tagged images

Abhay Gotmare[1], Gandharva Thite[2], Laxmi Bewoor[2]
[1]Department of Information Technology, Vishwakarma Institute of Information Technology, Pune, India
[2]Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, India

Article Info

Article history:

Received Nov 30, 2023
Revised Mar 7, 2024
Accepted Mar 9, 2024

Keywords:

BLOOM
Convolutional neural network
Journalism
Large language model
Long short-term memory
Multimodal machine learning
VGGNet16

ABSTRACT

Classical machine learning algorithms typically operate on unimodal data and hence it can analyze and make predictions based on data from a single source (modality). Whereas multimodal machine learning algorithm, learns from information across multiple modalities, such as text, images, audio, and sensor data. The paper leverages the functionalities of multimodal machine learning (ML) application for generating text from images. The proposed work presents an innovative multimodal algorithm that automates the creation of news articles from geo-tagged images by leveraging cutting-edge developments in machine learning, image captioning, and advanced text generation technologies. Employing a multimodal approach that integrates machine learning and transformer algorithms, such as visual geometry group network16 (VGGNet16), convolutional neural network (CNN) and a long short-term memory (LSTM) based system, the algorithm initiates by extracting the location from exchangeable image file format (Exif) data from the image. The features are extracted from the image and corresponding news headline is generated. The headlines are used for generating a comprehensive article with contemporary large language model (LLM). Further, the algorithm generates the news article big-science large open-science open-access multilingual language model (BLOOM). The algorithm was tested on real time photographs as well as images from the internet. In both the cases the news articles generated were validated with ROUGE and BULE score. The proposed work is found to be successful attempt in journalism field.

Corresponding Author:

Laxmi Bewoor
Department of Computer Engineering, Vishwakarma Institute of Information Technology
Pune, India
Email: laxmi.bewoor@viit.ac.in

1. INTRODUCTION

The manner that news is distributed and consumed has significantly changed with the advent of the digital age. Images have become a crucial component of this new environment because of their innate capacity to concisely express complex storylines. They frequently contain important information that supplements textual content and occasionally even replaces it. However, creating news stories manually from photos is a time-consuming task that requires a lot of labor and knowledge. Traditional machine learning (ML) algorithms have limitations with the uniform data format, but integrating more than one type of data is the need of the hour. Multimodal functionalities [1], [2] have recently garnered the attention of researchers to overcome this limitation. Multimodal applications have proven to be effective for hybridizing the models [3], [4] for text and image data types. The proposed work also represents an endeavor to develop a multimodal application and, in turn, step into the world of generative artificial intelligence (AI) [5], [6]. Therefore, the creation of automated systems capable of carrying out this activity could completely transform the journalism industry by increasing the productivity and scalability of news output. The development of artificial intelligence and machine learning technologies [7], [8] in recent years has opened significant opportunities for automating different parts of content generation. Particularly impressive improvements have been made in text generation models and image captioning methods. Transformer models have become a potent tool in the field of natural language processing, as introduced by Vaswani *et al.* [9]. These models have shown improved ability in comprehending the context and producing coherent text thanks to their self-attention mechanism. Later models have increased their use of transformers, demonstrating the value of pre-training on huge text corpora for a variety of downstream tasks with models like bidirectional encoder representations from transformers (BERT) [10]–[12] and big-science large open-science open-access multilingual language model (BLOOM) [13]. At the same time, there has been substantial development in the field of image captioning. The most common models [14]–[16] are those that combine convolutional neural networks (CNNs) [17]–[19] for visual feature extraction with recurrent neural networks (RNNs) or long short-term memory (LSTM) networks for caption production. These algorithms have proven to be excellent at producing pertinent and precise image captions. Despite these developments, creating news pieces from photographs still presents specific difficulties. News pieces are by nature complex; they frequently call for a thorough comprehension of the context, the aptitude to recognize important events, and the ability to write in a journalistic style that appeals to a wide audience. This duty entails more than just coming up with a caption for an image; it also entails turning that description into a lengthy news report that offers in-depth details about the portrayed event. In this research, we are aiming to bridge this gap by developing a novel model for generating news articles from images. Drawing upon the techniques and insights from the literature in both text generation and image captioning, we will explore the potential of integrating these technologies for our purpose. Our goal is to create a system that can automatically generate high-quality, contextually accurate news articles from images, thereby revolutionizing the process of news production and dissemination in the digital age. This research holds the potential to contribute significantly to the fields of automated journalism and AI-driven content generation [20].

2. RELATED WORK

The recent era has witnessed a significant shift from language understanding to language generation. Iqbal and Qureshi [20] presented a detailed survey on text generation models. Additionally, Amin-Nejad *et al.* [21] utilized transformers such as BERT and bidirectional encoder representations from transformers (BioBERT) on the medical information mart for intensive care-III (MIMIC-III) dataset. Their findings suggest that synthetic data can yield better results than poorer quality original data. Similarly, Abdelwahab and Elmaghraby [22] compared LSTM, RNN, gated recurrent unit (GRU), and Markov chain models, concluding that deep learning models handle incorrectly labeled reviews better than Markov chain-based text generators. A shift towards structuring data was observed with Marcheggiani and Perez-Beltrachini [23] using graph convolutional networks on the web natural language generator (WebNLG) and SR11Deep datasets, demonstrating the benefits of encoding structural information with s graph convolutional networks (GCN). Vaswani *et al.* [9] introduced the transformer architecture and self-attention mechanisms, eliminating the need for recurrent layers. The field of large language model (LLM) has also seen remarkable developments. Scao *et al.* [13] introduced BLOOM, a decoder-only transformer language model that showed competitive performance after multitask fine-tuning. Hu *et al.* [24] proposed low-rank adaptation (LoRA), a low-rank adaptation of large language models, which greatly reduced the number of trainable parameters for downstream tasks. Chai *et al.* [25] proposed a methodology that significantly reduced video random access memory (VRAM) requirements, enabling fine-tuning of a 7B LLM on a 6 GB consumer-grade graphical processing unit (GPU). The application of these models in image captioning has also been explored. Sharma *et al.* [14] suggested that multimodal image captioning can be effectively done using LSTM, as evaluated by bilingual evaluation understudy (BLEU) scores. Hossain *et al.* [15] illustrated an architecture for image captioning using CNN for feature extraction. Vinyals *et al.* [16] emphasized the use of LSTM networks with soft attention mechanisms. In conclusion, the field of text generation models and image captioning techniques has seen significant advancements and applications. However, there is still much to explore and understand in these areas.

3. PROPOSED METHOD

In this section, the use of multimodal algorithms is discussed. The multimodal ML integrates information from various modalities like text, image, and video. The proposed work inputs images and generates relevant text describing that image. The entire workflow of proposed multimodal ML algorithms is depicted in Figure 1 which includes following steps:

a. Data collection and preparation: the dataset primarily comprises images, encompassing both real-time images captured with mobile cameras and JPEG files obtained online. These images undergo preprocessing to a format suitable for subsequent analysis.
b. Choice of deep learning (DL) models: this step involves selecting the appropriate DL models for training. In the current work, the visual geometry group-16 (VGG-16), a CNN-based model, is employed for image processing, while LSTM and Transformer-based models handle textual data processing. Additionally, the large language model BLOOM-1b7 pre-trained model is utilized for text enhancement.
c. Choice of DL models: this step involves selecting the appropriate DL model for training. In the current work, VGG-16, a CNN-based model, is utilized for image processing, while LSTM and transformer-based models are employed for textual data processing.
d. Model hybridization: this step facilitates leveraging the advantages of multiple modalities. The model undergoes further training and fine-tuning by integrating features from both images and text.
e. Interpretation of the results: the results are assessed using well-established performance metrics: accuracy and loss rate. The accuracy assessment includes BLEU and recall-oriented understudy for gisting evaluation (ROUGE) scores [26].

Algorithm 1 provides detailed execution steps for converting input image to text.

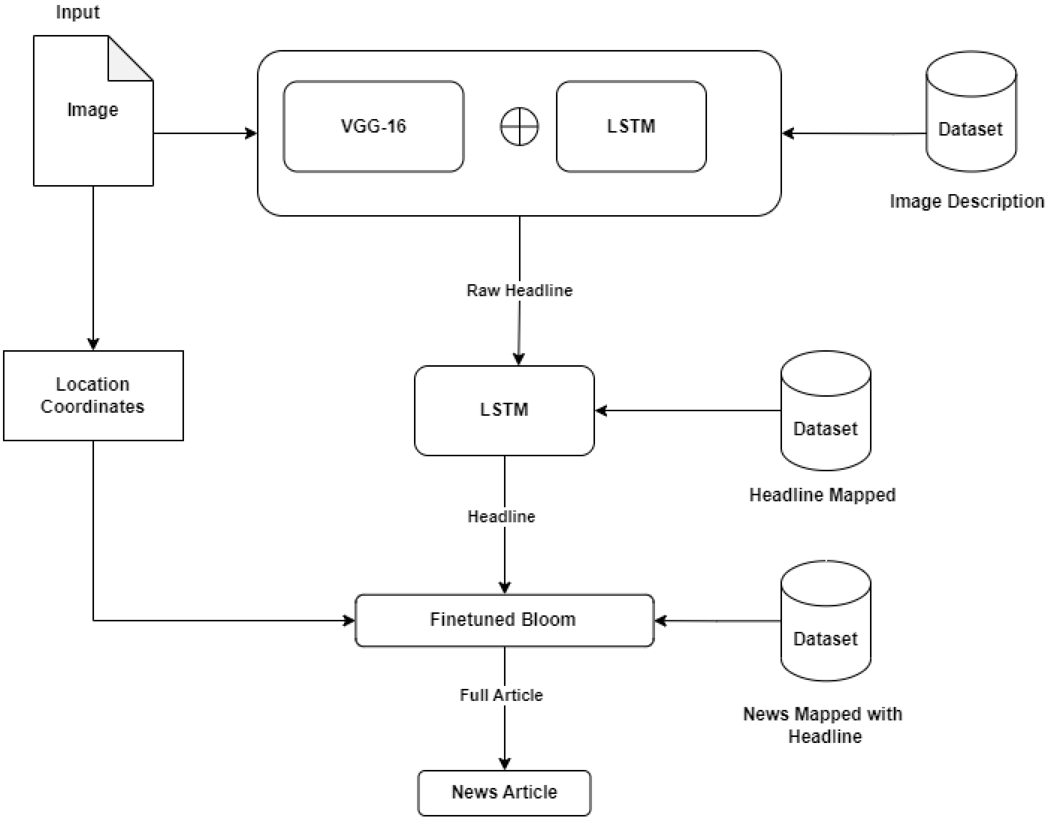

Figure 1. Proposed architecture diagram

Algorithm 1. Proposed attention-based hybrid CNN-LSTM model

```
Step 1. Upload Image: The user uploads a geo-tagged image.
Step 2. Extract Geo-tagged Locations: The exchangeable image file format (Exif) decryption algorithm is
      used to extract the geo-tagged locations from the uploaded image.
Step 3. Feature Extraction: The image is then processed for feature extraction using the
      VGGNet16 CNN model which is trained with a customized dataset.
Step 4. Tune Extracted Features: The features extracted from the image are then fine-tuned
      with the trained weights of the LSTM algorithm.
Step 5. Generate Raw Headline: The output of the LSTM is a raw headline that describes the
      image in a basic way.
```

```
Step 6. Convert Raw Headline: This raw headline is passed through another LSTM model to
      convert the sentence into a meaningful headline, which will serve as the actual
      headline for the image.
Step 7. Pass Headline Through finetuned LLM: The headline and the previously extracted
      location is then passed through a fine-tuned LLM Bloom-1b7 model. in this case, the
      Bloom-1b7 model. This model, fine-tuned using quantized low rank adaptation (QLoRA),
      is designed to generate news articles.
Step 8. Generate Description: The Bloom-1b7 model generates a meaningful description for
      the headline.
Step 9. Combine Elements to Form Article: Finally, this description, the headline, the
      location extracted from the Exif decryption algorithm, and the image are combined to
      form a full news article.
End Procedure.
```

4. LEARNING MODELS OF PROPOSED METHODOLOGY

4.1. Extraction of location co-ordinates

The geotagged image has an encrypted Exif file that is used to store the location and a variety of other image details. The Exif file is decrypted using the Python language functionalities to extract the location coordinates. The location coordinates are represented in the form of directions (NSEW) that are then converted into respective latitude and longitude with the following formulas:

$$latitude = Degrees + Minutes/60 + Seconds/3600$$
$$longitude = Degrees + Minutes/60 + Seconds/3600$$

The converted latitude and longitude are used to extract the exact location, and a real-time map is rendered with the help of folium, as shown in Figure 2.

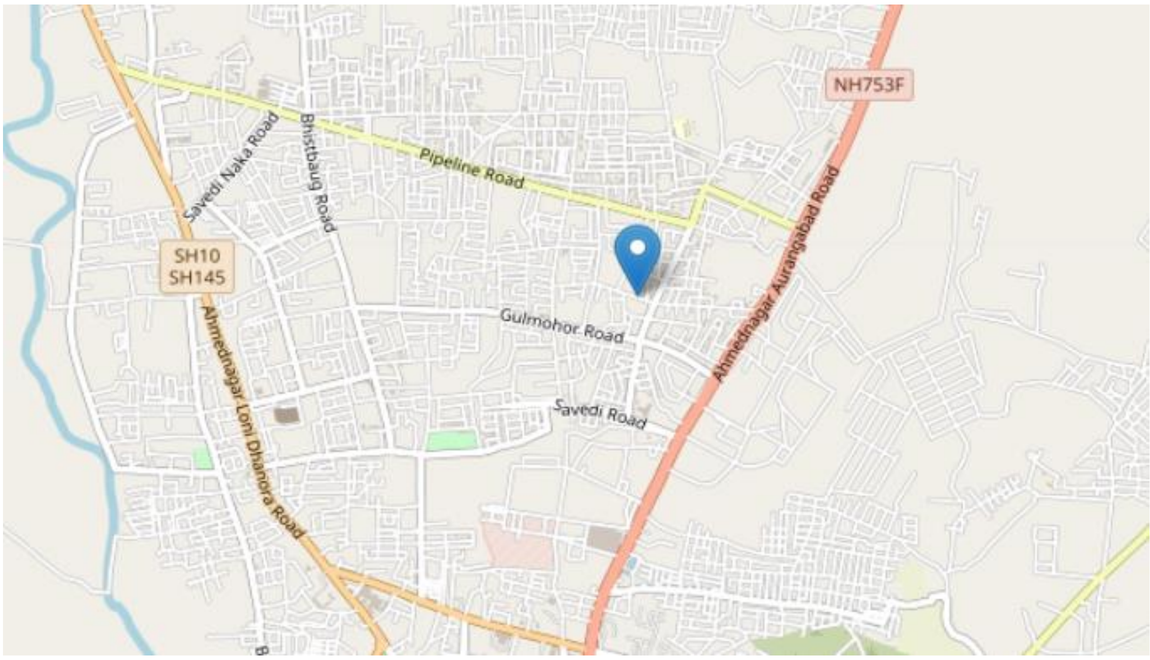

Figure 2. Exact extracted location on folium map with blue marker

4.2. VGG-16 and LSTM for generating caption from image

After extracting the location information from the Exif data of the image, the image is then passed to a VGGNet-16 model for feature extraction. The types and number of parameters are shown in Table 1. The extracted features are subsequently stored and passed into the long short-term memory (LSTM) model for training to extract captions from images. The VGGNet-16 inner model architecture, based on the required parameters and features, is illustrated in Figure 3. The components of the respective LSTM model are presented in Table 2.

Table 1. VGGNet-16 parameter distribution

Sr No	Parameter type	Number of parameters
1	Total Parameters	134260544
2	Trainable Parameters	134260544
3	Non-trainable Parameters	0

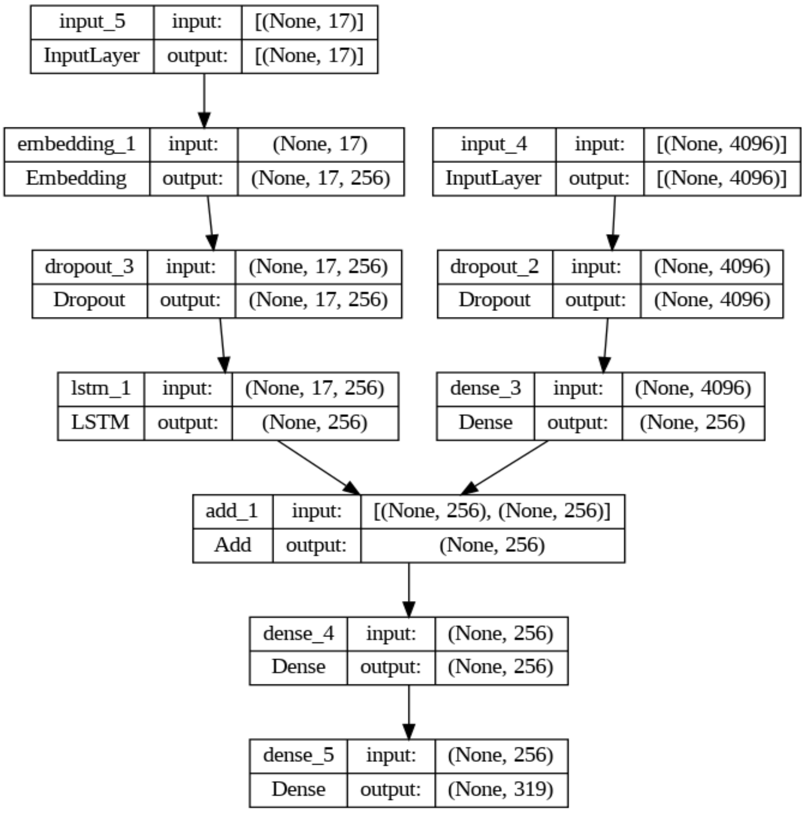
input_5 | input: | [(None, 17)]
InputLayer | output: | [(None, 17)]
embedding_1 | input: | (None, 17)
Embedding | output: | (None, 17, 256)
input_4 | input: | [(None, 4096)]
InputLayer | output: | [(None, 4096)]
dropout_3 | input: | (None, 17, 256)
Dropout | output: | (None, 17, 256)
dropout_2 | input: | (None, 4096)
Dropout | output: | (None, 4096)
lstm_1 | input: | (None, 17, 256)
LSTM | output: | (None, 256)
dense_3 | input: | (None, 4096)
Dense | output: | (None, 256)
add_1 | input: | [(None, 256), (None, 256)]
Add | output: | (None, 256)
dense_4 | input: | (None, 256)
Dense | output: | (None, 256)
dense_5 | input: | (None, 256)
Dense | output: | (None, 319)

Figure 3. VGGNet-16 inner architecture

Table 2. LSTM for caption generation attributes

Epochs	Batch Size	Loss	BLEU-1
60	3	0.2399	0.295

4.3. LSTM for headline generation

Upon the generation of an initial caption through the combined efforts of the VGGNet-16 and the primary LSTM model, it is then fed into a secondary LSTM network. This subsequent LSTM is specifically engineered to distill the preliminary caption into a contextually relevant and semantically coherent headline. The second LSTM model has the following attributes tabulated in Table 3. The compilation attributes of the LSTM model are tabulated in Table 4.

Table 3. LSTM for headline generation attributes

Layer (type)	Output shape	No of parameters
Embedding	(None,76,100)	54300
LSTM	(None,150)	150600
Dense	(None,543)	81993

Table 4. LSTM compilation attributes

Epochs	Batch Size	Loss	Accuracy
200	1	0.0130	0.9916

4.4. LLM for generating news from headline

After generating the headline, it is passed to the Bloom-1b7 pre-trained model, an open-source LLM. This model, with 1,722,408,960 parameters, has been trained on 45 natural languages and 1.5 terabytes of pre-processed text, converted into 350 billion unique tokens. The model comprises 513,802,240 embedding parameters, 24 layers, 16 attention heads, and 2048-dimensional hidden layers. It is designed to perform general-purpose tasks and can be fine-tuned for domain-specific tasks. The fine-tuning process was performed on a single Tesla T4 GPU with 16 gigabytes of GPU VRAM. Due to hardware constraints, we employed a QLoRA method. This method involved reducing the precision of model parameters to 4-bit integers, significantly reducing the amount of CPU and GPU memory required, making it feasible to perform the experiment on the available hardware. The dataset used for fine-tuning was self-generated and consisted of headlines and their respective news articles. Despite its small size, this dataset served as a prototype for the experiment. The fine-tuning process was guided by a prompt template, which was designed to generate a news article based on a given headline. The template was as follows: 'Based on the following headline and location, generate a news article: "Headline" "location" "news". The parameters for generation were set with a temperature of 0.1 and a maximum of 512 new tokens. The adapter type used was LoRA, and the quantization was set to 4 bits. These parameters were chosen to balance the trade-off between computational efficiency and the quality of the generated text.

5. RESULTS AND DISCUSSION

In the realm of waste management, the integration of multimodal approaches advocates a transformative shift in how news related to this critical field is generated and disseminated. The proposed work not only works on the images available online but also can collect real time images, identifies the location and relevant information from that picture is getting generated which will be supplied to higher authorities for further action. The detailed analysis of results is provided in subsequent subsections.

5.1. The results for the caption generated from the image

The methodology for this task involved the utilization of a combination of VGGNet-16 and LSTM. The combination underwent testing with various hyperparameters, and adjustments were made for optimal results. While the number of epochs remained constant at 60, we experimented with different batch sizes for the model, resulting in the findings tabulated in Table 5. Based on the BLEU-1 and BLEU-2 scores, the optimal batch size in terms of accuracy was three, where BLEU-1 score = 0.295 and BLEU-2 score = 0.

Table 5. Hyper parameter tuning results

Sr. no	Epoch	Batch size	Bleu1	Bleu2
1	60	10	0.29	0
2	60	20	0.35	0
3	60	11	0.29	0
4	60	5	0.30	0
5	60	15	0.32	0.07
6	60	7	0.31	0.07
7	**60**	**3**	**0.29**	**0**
8	60	4	0.29	0
9	60	6	0.32	0.07
10	60	8	0.31	0
11	60	9	0.30	0.07

5.2. The results for the headline generated from the caption

The methodology for this task involved using LSTM to train on the captions generated from the image, converting them into suitable headlines for news article generation. The model's training parameters, as well as its respective accuracy and loss factors, were obtained and tabulated. Table 6 depicts the number of epochs along with the corresponding loss factor and accuracy. Categorical cross-entropy is considered as the loss factor; hence, 200 epochs were chosen for model construction. The accuracy and loss factor per epoch is also plotted, as shown in Figures 4 and 5 respectively. Thus, the model's performance during training is evaluated.

5.3. The results for the article generated from the headline

The proposed algorithm also makes use of fine-tuning Bloom-1b7. As shown in Table 7, the findings encompass an array of accuracy metrics, including different versions of both BLEU and ROUGE scores. These results from all the three subdomains serve as the conclusive evaluation of the model’s performance.

Our research marks the first endeavor in the domain of generating text directly from images, pioneering a novel approach to bridging the gap between visual and textual modalities. Through extensive experimentation and evaluation, we introduce an innovative algorithm capable of autonomously generating descriptive text for images with remarkable accuracy and coherence. This groundbreaking achievement not only fills a significant gap in existing literature but also opens avenues for novel applications in image understanding, content generation, and accessibility enhancement. Our results showcase the transformative potential of multimodal learning techniques in unlocking new capabilities for understanding and interpreting visual content, laying the foundation for future advancements in this emerging field.

Table 6. Headline results

Sr No	Epochs	Categorical Cross-entropy	Accuracy
1	10	4.84	0.10
2	20	2.66	0.45
3	50	0.19	0.98
4	75	0.06	0.99
5	100	0.03	0.99
6	120	0.02	0.99
7	**200**	**0.01**	**0.99**

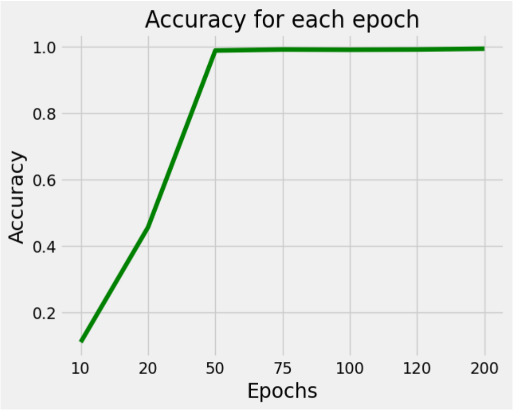

Figure 4. Accuracy per epoch

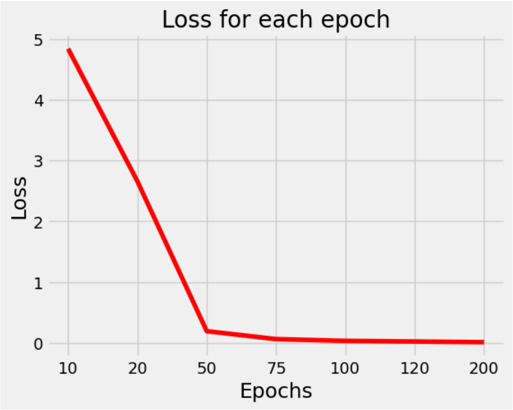

Figure 5. Loss per epoch

Table 7. Fine-tuned bloom model accuracy parameters

Sr No	Parameter	Score
1	Bleu1	0.26
2	char_error_rate	0.73
3	loss	2.24
4	next_token_perplexity	152971.76
5	perplexity	250849.78
6	rouge1_fmeasure	0.53
7	rouge1_precision	0.45
8	rouge1_recall	0.69
9	rouge2_fmeasure	0.35
10	rouge2_precision	0.28
11	rouge2_recall	0.46
12	rougeL_fmeasure	0.47
13	rougeL_precision	0.39
14	rougeL_recall	0.61
15	rougeLsum_fmeasure	0.51
16	rougeLsum_precision	0.42
17	rougeLsum_recall	0.66
18	combined_loss	2.24

6. CONCLUSION

This research aimed to develop a novel algorithm for generating news articles from geo-tagged images, leveraging advancements in machine learning, image captioning, and text generation models. The results of the study showed promising outcomes across the three subdomains of the algorithm. The use of

VGGNet-16 and LSTM for image captioning yielded optimal results with a batch size of three, as evidenced by the BLEU-1 and BLEU-2 scores. The subsequent conversion of the caption into a suitable headline using LSTM demonstrated high accuracy, with the model trained for 200 epochs. Finally, the generation of the article from the headline using a fine-tuned Bloom-1B7 LLM resulted in satisfactory accuracy metrics, including BLEU and Recall-Oriented Understudy for ROUGE scores. The study's findings highlight the potential of the proposed algorithm in revolutionizing news production and dissemination in the digital age. By automating the process of generating news articles from images, the algorithm can significantly reduce the manual effort and time required, thereby improving the efficiency and scalability of news production. This research represents a significant step towards the automation of journalism and AI-driven content generation. The proposed algorithm, with its multimodal combination of machine learning and transformer algorithms, offers a promising solution to the challenge of generating news articles from images. With further refinement and improvement, the algorithm has the potential to revolutionize the field of digital journalism.

The prospects of this research are broad and stimulating. The algorithm, while effective, presents opportunities for refinement. Enhancements in context understanding and journalistic style generation could be achieved by exploring advanced transformer models for text generation and more complex CNN models for image captioning. Further, the algorithm's real-world performance needs additional exploration. Future work could include testing with diverse images and in varied contexts. Moreover, the algorithm's potential extends beyond digital journalism to other fields requiring automated image-based content generation, such as social media management and public relations. In conclusion, this research opens vast possibilities. With further refinement, the algorithm could significantly contribute to automated journalism, AI-driven content generation, and other text and image-related automation scenarios.

REFERENCES

[1] C. A. K. A. Kounta, B. Kamsu-Foguem, F. Noureddine, and F. Tangara, “Multimodal deep learning for predicting the choice of cut parameters in the milling process,” *Intelligent Systems with Applications*, vol. 16, Nov. 2022, doi: 10.1016/j.iswa.2022.200112.

[2] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, “Multimodal machine learning: a survey and taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, Feb. 2019, doi: 10.1109/TPAMI.2018.2798607.

[3] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, “Challenges and applications in multimodal machine learning,” in *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 2*, Association for Computing Machinery, 2018, pp. 17–48.

[4] P. P. Liang, A. Zadeh, and L. P. Morency, “Foundations and trends in multimodal machine learning: principles, challenges, and open questions,” *arXiv preprint arXiv:2209.03430*, 2022.

[5] T. Brown *et al.*, “Language models are few-shot learners,” *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[6] R. Koncel-Kedziorski, D. Bekal, Y. Luan, M. Lapata, and H. Hajishirzi, “Text generation from knowledge graphs with graph transformers,” *arXiv preprint arXiv:1904.02342*, 2019.

[7] M. Z. Alom *et al.*, “A state-of-the-art survey on deep learning theory and architectures,” *Electronics*, vol. 8, no. 3, Mar. 2019, doi: 10.3390/electronics8030292.

[8] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, “Deep reinforcement learning: a brief survey,” *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, Nov. 2017, doi: 10.1109/MSP.2017.2743240.

[9] A. Vaswani *et al.*, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.

[10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “Bert: pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.

[11] P. Howlader, P. Paul, M. Madavi, L. Bewoor, and V. S. Deshpande, “Fine tuning transformer based BERT model for generating the automatic book summary,” *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 10, no. 1s, pp. 347–352, Dec. 2022, doi: 10.17762/ijritcc.v10i1s.5902.

[12] M. V Koroteev, “BERT: a review of applications in natural language processing and understanding,” *arXiv preprint arXiv:2103.11943*, 2021.

[13] T. Le Scao *et al.*, “Bloom: A 176b-parameter open-access multilingual language model,” *arXiv:2211.05100*, 2022.

[14] H. Sharma, M. Agrahari, S. K. Singh, M. Firoj, and R. K. Mishra, “Image captioning: a comprehensive survey,” in *2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC)*, Feb. 2020, pp. 325–328, doi: 10.1109/PARC49193.2020.236619.

[15] M. D. Z. Hossain, F. Sohel, M. F. Shiratuddin, and H. Laga, “A comprehensive survey of deep learning for image captioning,” *ACM Computing Surveys*, vol. 51, no. 6, pp. 1–36, Nov. 2019, doi: 10.1145/3295748.

[16] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, “Show and Tell: lessons learned from the 2015 MSCOCO image captioning challenge,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 652–663, Apr. 2017, doi: 10.1109/TPAMI.2016.2587640.

[17] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, “Inception-v4, Inception-ResNet and the impact of Residual connections on learning,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017, doi: 10.1609/aaai.v31i1.11231.

[18] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[19] P. Kale, M. Panchpor, S. Dingore, S. Gaikwad, and L. Bewoor, “Traffic sign classification using convolutional neural network,” *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 1–10, Nov. 2021, doi: 10.32628/CSEIT217545.

[20] T. Iqbal and S. Qureshi, “The survey: text generation models in deep learning,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 2515–2528, Jun. 2022, doi: 10.1016/j.jksuci.2020.04.001.

[21] A. Amin-Nejad, J. Ive, and S. Velupillai, "Exploring transformer text generation for medical dataset augmentation," *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pp. 4699–4708, 2020.
[22] O. Abdelwahab and A. Elmaghraby, "Deep learning based vs Markov Chain based text generation for cross domain adaptation for sentiment classification," in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, Jul. 2018, pp. 252–255, doi: 10.1109/IRI.2018.00046.
[23] D. Marcheggiani and L. Perez-Beltrachini, "Deep graph convolutional encoders for structured data to text generation," *arXiv preprint arXiv:1810.09995*, 2018.
[24] E. J. Hu *et al.*, "LoRA: low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2021.
[25] Y. Chai, J. Gkountouras, G. G. Ko, D. Brooks, and G. Y. Wei, "Int2. 1: towards fine-tunable quantized large language models with error correction through low-rank adaptation," *arXiv preprint arXiv:2306.08162*, 2023.
[26] S. Porwal, L. Bewoor, and V. Deshpande, "Transformer based implementation for automatic book summarization," *arXiv preprint arXiv:2301.07057*, 2023.

BIOGRAPHIES OF AUTHORS

Abhay Gotmare is a final year student at the Department of Information Technology, Vishwakarma Institute of Information Technology, Pune. His research interests lie in the fields of Natural Language Processing and Cyber Security. Abhay is currently engaged in developing innovative solutions in these areas. He can be contacted at email: abhay.22010481@viit.ac.in for further discussion or collaboration on his research interests.

Gandharva Thite is an undergraduate student at Vishwakarma Institute of Information Technology. His research and academic interests mainly include machine learning, natural language processing and mathematics. He can be contacted at email: gandharva.22010577@viit.ac.in.

Laxmi Bewoor holds PhD in artificial intelligence as major and optimization as minor. Currently she is an associate professor at Department of Computer Engineering, Vishwakarma Institute of Information Technology, Pune. Her research interests include natural language processing, metaheuristics techniques, deep learning, artificial intelligence, algorithms. She has successfully completed one funded research projects. 5 International Patents are granted in her name. She has published more than 35 research articles in reputable journals and conferences. She is a reviewer of leading journals of Springer, Elsevier and IEEE. She can be contacted at email: laxmi.bewoor@viit.ac.in.