Figure Detection and Comparison

Bharath N[a], Devadath T M[a], Georgekutty Jose[a], Kevin Tom Sajeev[a], Anjaly Muraleedharan[b]

[a]UG Student, Artificial Intelligence and Data Science, Viswajyothi College of Engineering and Technology, Muvattupuzha, 686661
[b]Assistant Professor, Artificial Intelligence and Data Science, Viswajyothi College of Engineering and Technology, Muvattupuzha, 686661

ABSTRACT:

This system utilizes the complementary strengths of three deep learning models – YOLO v8, VGG19, and Inception v4 – to unlock the narrative potential within images. YOLO v8, known for its lightning-fast processing, takes the lead in real-time object detection. It efficiently identifies and pinpoints various objects in the image, providing a foundational understanding of the scene. VGG19, a pre-trained powerhouse, then steps in, leveraging its vast knowledge base to recognize objects with remarkable accuracy. It goes beyond simple detection, enriching the understanding by classifying these objects. Finally, Inception v4, the transfer learning specialist, refines the process. It analyzes the image through the lens of its pre-trained knowledge, identifying materials and textures that might be missed by the other models. This comprehensive approach yields a rich tapestry of keywords that capture the essence of the image.

These extracted keywords, along with any text or specific instructions provided by the user, are then fed into the Falcon 7b model. Falcon 7b, a large language model, acts as the system's creative engine. It devours this information – the visual details gleaned from the image and the user's specific input – and utilizes its knowledge and writing prowess to craft a compelling blog post. The generated content reflects the image's core elements, incorporating the user's desired direction or specific details they want to highlight. This system exemplifies the power of deep learning in the content creation domain. It demonstrates how the intelligent automation of tasks like object detection and material recognition can pave the way for innovative approaches in automated blog generation

Keywords: YOLO v8 , VGG19 , Inception v4

1. Introduction

The digital age thrives on captivating content, and visuals are key to grabbing attention. Imagine transforming those images into engaging blog posts!

This article explores an innovative image-to-blog converter system that bridges the gap between visual storytelling and automated content generation.

We'll delve into its workings, focusing on one that leverages cutting-edge deep learning models.

This system utilizes the combined strengths of YOLO v8, VGG19, and Inception v4 to unlock the narrative potential within images. YOLO v8, known for its speed, tackles real-time object identification, providing a foundational understanding of the scene. VGG19, pre-trained on colossal datasets like ImageNet, recognizes these objects with incredible accuracy. Finally, Inception v4 refines the process by analyzing the image for materials and textures. Importantly, the system focuses solely on objects, excluding people for privacy reasons.

The meticulous work of these models yields a rich set of keywords capturing the image's essence. Imagine a picture of snow-capped mountains bathed in a golden sunrise. Keywords like "majestic mountains," "golden sunrise," and "breathtaking landscape" might be generated. These keywords become the building blocks for the narrative.

The extracted keywords are then fed into Falcon 7b, a large language model acting as the system's creative engine. It devours this information - visual details and keywords - and crafts a compelling blog post. The generated content reflects the image's core elements, weaving a narrative around the identified objects.

Finally, a user-friendly Graphical User Interface (GUI) prioritizes the user experience. Users simply upload the image, and the intuitive interface transforms the visual content into an engaging written piece. This image-to-blog converter system exemplifies the transformative power of deep learning in content creation. It demonstrates how intelligent automation can elevate textual description generation from visual content, bridging the gap between captivating visuals and engaging written narratives

2. Literature review

2.1. Image-to-Text Generation

Several studies have explored the task of generating textual descriptions or captions for images. Early approaches employed handcrafted features and statistical language models to associate images with descriptive text.More recent works have leveraged deep learning techniques to learn implicit mappings between images and text. This includes methods based on CNN-RNN architectures, where CNNs extract image features, which are then fed into Recurrent Neural Networks (RNNs) for text generation. Attention mechanisms have been incorporated into image-to-text generation models to improve the alignment between visual and textual modalities, enabling more informative and contextually relevant descriptions.

2.2. Content Creation Automation

Beyond image-to-text generation, research efforts have explored broader applications of AI in automating content creation tasks. This includes systems for generating articles, product descriptions, social media posts, and marketing content.Content generation models based on large-scale language models (e.g., GPT) have shown promise in generating diverse and engaging text across various domains. Techniques for controlling the style, tone, and topic of generated content have been investigated to cater to specific user preferences and requirements. By building upon the insights and advancements from these diverse research areas, systems like IBAG aim to provide a holistic solution for automating the generation of textual content from images, thereby facilitating efficient and scalable content creation in the blogging domain.

3. Proposed work

WThe Image to Blog AI Generator is a comprehensive system designed to automate the process of generating textual content for blog posts based on input images. Leveraging state-of-the-art techniques in computer vision and natural language processing, The Image to Blog AI Generator offers a seamless and efficient solution for transforming visual stimuli into coherent and engaging narratives. The proposed system comprises several key components and functionalities:

A. Image Preprocessing Module:Upon receiving an input image, IBAG preprocesses it to enhance feature extraction and understanding. This may include resizing, normalization, and augmentation techniques to ensure consistency and improve model performance.

B. Image Feature Extraction using CNNs:

Employs Convolutional Neural Networks (CNNs) to extract meaningful visual features from input images. Pre-trained CNN models, such as ResNet, Inception, or EfficientNet, are utilized to capture hierarchical representations of image content.

C. Text Generation Module with GPT-based Models:

Extracted image features are then fed into a Generative Pre-trained Transformer (GPT) model for natural language generation. GPT models have been fine-tuned on large text corpora and possess the capability to generate coherent and contextually relevant textual content based on input prompts.

D. Semantic Understanding and Coherence Maintenance:

IBAG incorporates techniques for semantic understanding to ensure that the generated textual content accurately reflects the visual content of the input images. Additionally, coherence maintenance mechanisms are employed to maintain consistency and relevance throughout the generated blog post.

E. User Interaction and Customization:

The system provides interfaces for user interaction, allowing users to input preferences, prompts, or additional context to guide the content generation process. This may include specifying target audience demographics, tone of voice, or desired content structure.

F. Evaluation and Feedback Mechanisms:

IBAG includes evaluation and feedback mechanisms to assess the quality and relevance of generated blog content. This may involve metrics such as readability scores, relevance to input images, and user feedback loops to iteratively improve system performance.

G. Scalability and Integration:

The proposed system is designed to be modular and scalable, facilitating easy integration with existing blogging platforms and workflows. APIs and SDKs are provided to enable seamless integration with third-party applications and services.

H. Security and Privacy Considerations:

IBAG incorporates security and privacy measures to protect sensitive user data and ensure compliance with relevant regulations. This includes encryption techniques, access controls, and anonymization mechanisms to safeguard user privacy and data integrity.

Overall, the Image to Blog AI Generator (IBAG) represents a significant advancement in the automation of content creation tasks, offering a versatile and efficient solution for generating textual content from images in the blogging domain. By leveraging the synergies between computer vision and natural language processing, IBAG empowers content creators to streamline the content generation process while enhancing the quality and engagement of their blog posts.

4. Experimental results

Image to Blog Converter uses methodologies for automatically generating textual descriptions from images, enabling seamless integration of visual content into blog posts. This often uses a combination of computer vision and natural language processing techniques to achieve this goal. One common approach involves utilizing object detection algorithms like YOLO (You Only Look Once) to identify and localize objects within images. By detecting objects and their positions, the system can begin to understand the visual context of the image. Additionally, convolutional neural network (CNN) architectures such as Inception and VGG19 are employed for feature extraction. These models are pre-trained on large datasets for image classification tasks, allowing them to capture high-level visual features. The extracted keywords serve as rich representations of the image content and are subsequently fed into Falcon 7b model.

In the subsequent step, Falcon 7b model takes the extracted image features as input and generate textual descriptions. This model is pre-trained human-generated descriptions, learning to generate coherent and descriptive text based on the keywords from the image.

Ultimately, the aim is to develop robust and effective systems that can automatically generate blog-worthy textual descriptions from images. By seamlessly integrating visual content with textual narratives, these systems enhance the accessibility and engagement of blog posts, catering to diverse audiences with varying preferences for consuming content. While specific results may vary depending on the datasets used, the combination of object detection, feature extraction, and sequence generation techniques forms the foundation for advancing the field of image-to-blog conversion.

5. Conclusion

In this paper, we introduced the Image to Blog AI Generator (IBAG), a sophisticated system designed to automate the generation of textual content for blog posts based on input images. IBAG harnesses the power of cutting-edge technologies in computer vision and natural language processing to seamlessly transform visual stimuli into coherent and engaging narratives. Through a multi-stage pipeline involving image preprocessing, feature extraction using Convolutional Neural Networks (CNNs), and text generation using Generative Pre-trained Transformer (GPT) models, IBAG offers a versatile and efficient solution for content creation in the blogging domain.

Our proposed system addresses the growing need for automating content creation tasks, particularly in the context of visual-centric platforms where images play a crucial role in storytelling and communication. By leveraging pre-trained models and advanced techniques for semantic understanding and coherence maintenance, IBAG ensures the quality and relevance of generated blog content while accommodating user preferences and customization.

Moreover, IBAG's modular design and scalability make it suitable for integration with existing blogging platforms and workflows, facilitating seamless adoption and enhancing productivity for content creators. The inclusion of evaluation and feedback mechanisms enables continuous improvement and refinement of the system, ensuring that it remains adaptive to evolving user needs and preferences.

In conclusion, the Image to Blog AI Generator (IBAG) represents a significant advancement in content creation automation, offering promising prospects for enhancing productivity, creativity, and engagement in the blogging domain. By empowering content creators with the tools and technologies to efficiently generate compelling narratives from images, IBAG opens up new possibilities for storytelling and communication in the digital age

REFERENCES :

1. S. Lobry, J. Murray, D. Marcos and D. Tuia, "Visual Question Answering From Remote Sensing Images,"IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan,2019, pp. 4951-4954, doi: 10.1109/IGARSS.2019.8898891.

2. S. Saravanan and K. Sudha, "GPT-3 Powered System for Content Generation and Transformation," 2022 Fifth International Conference on
3. Computational Intelligence and Communication Technologies (CCICT),Sonepat, India, 2022, pp. 514-519, doi: 10.1109/CCiCT56684.2022.00096

4. M. -T. Nguyen, P. -T. Nguyen, V. -V. Nguyen and Q. -M. Nguyen, "Generating Product Description withGenerative Pre-trained Transformer 2," 2021 6th International Conference on Innovative Technology inIntelligent System and Industrial Applications (CITISIA), Sydney, Australia,
5. 2021, pp. 1-7, doi:10.1109/CITISIA53721.2021.9719940..

6. S. Värtinen, P. Hämäläinen and C. Guckelsberger, "Generating Role-Playing Game Quests With GPT Language Models," in IEEE Transactions on Games, doi: 10.1109/TG.2022.3228480..

7. Wu et al., "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development,"in IEEE/CAA Journal of Automatica Sinica, vol. 10, no. 5, pp. 1122-1136, May 2023, doi:10.1109/JAS.2023.123618.