A Project report on

VISUAL QUESTION GENERATION FROM REMOTE SENSING IMAGES

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

Bachelor of Technology

in

Computer Science and Engineering

Submitted by

Y. Laxmi Narayana
(20H51A0581)
B. Pravalika
(20H51A05B6)
P.Arya Patel
(20H51A05F3)

Under the esteemed guidance of
Mrs. M. Kamala
ASSISTANT PROFESSOR

Department of Computer Science and Engineering

CMR COLLEGE OF ENGINEERING & TECHNOLOGY
(UGC Autonomous)
*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A$^+$ Grade
KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.
2020- 2024

CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the Major Project Phase I report entitled **"VISUAL QUESTION GENERATION FROM REMOTE SENSING IMAGES "** being submitted by Y.LAXMI NARAYANA(20H51A0581),B.PRAVALIKA (20H51A05B6),P.ARYA PATEL(20H51A05F3) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out his/her under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

Mrs.M.Kamala
Assistant Professor
Dept. of CSE

Dr. Siva Skandha Sanagala
Associate Professor and HOD
Dept. of CSE

ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express my heartfelt gratitude to all the people who helped in making this project work a grand success.

We are grateful to **Mrs.M.Kamala ,** Assistant Professor, Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank **Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete my project work successfully.

We are very grateful to **Dr. Vijaya Kumar Koppula**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Computer Science and Engineering for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary, CMR Group of Institutions, for his continuous care.

Finally, We extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly and indirectly in completion of this project work.

Y.LAXMI NARAYANA - 20H51A0581
B.PRAVALIKA - 20H51A05B6
P.ARYA PATEL - 20H 51A05F3

TABLE OF CONTENTS

List of Figures

List of Tables

ABSTRACT

Visual question generation (VQG) is an emerging research area that aims to automatically generate natural language questions based on visual content. This paper focuses on the application of VQG to remote sensing images, which are obtained from aerial or satellite sensors and provide valuable information for various domains, including agriculture, urban planning, and environmental monitoring .

The proposed approach leverages deep learning techniques to extract meaningful features from remote sensing images and subsequently generate coherent and contextually relevant questions. Convolution Neural Networks (CNNs) are utilized to capture spatial information from the images, while Recurrent Neural Networks (RNNs) are employed to model sequential patterns in the generated questions. The combination of these architectures enables the model to effectively comprehend and translate the visual context into human-readable questions.To ensure the quality of the generated questions, a novel attention mechanism is introduced, allowing the model to focus on relevant image regions while formulating the questions

CHAPTER 1
INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 PROBLEM STATEMENT

The research paper addresses the emerging field of Visual Question Generation (VQG) with a specific focus on its application to remote sensing images, which are essential for various domains such as agriculture, urban planning, and environmental monitoring. The primary goal is to develop an automated system that can generate natural language questions based on visual content.

To achieve this, the proposed approach employs deep learning techniques, with Convolutional Neural Networks (CNNs) used to capture spatial information from remote sensing images, and Recurrent Neural Networks (RNNs) to model the sequential patterns in generated questions. This combination of architectures allows the model to effectively understand and translate the visual context into coherent and contextually relevant human-readable questions. To enhance question quality, the research introduces a novel attention mechanism that enables the model to focus on pertinent image regions while formulating questions.

1.2 RESEARCH OBJECTIVE:

- Emerging Research Area: Explore the field of Visual Question Generation (VQG) as an emerging research area with a specific focus on its application to remote sensing images.
- Automated Question Generation: Develop a system that can automatically generate natural language questions based on visual content, especially remote sensing images acquired from aerial or satellite sensors.
- Domain Applications: Investigate the potential applications of VQG in domains such as agriculture, urban planning, and environmental monitoring, where remote sensing data is invaluable for decision-making and analysis.
- Deep Learning Techniques: Utilize deep learning techniques to extract meaningful features from remote sensing images. Specifically, employ Convolutional Neural Networks (CNNs) to capture spatial information from these images.
- Sequential Pattern Modeling: Employ Recurrent Neural Networks (RNNs) to model sequential patterns in the generated questions, allowing for the generation of coherent and contextually relevant questions.

1.3 PROJECT SCOPE

The Research Area: The project operates within the emerging research area of Visual Question Generation (VQG), with a specific focus on its application to remote sensing images.

Deep Learning Techniques: The scope includes the use of deep learning techniques, primarily Convolution Neural Networks (CNNs) for spatial feature extraction and Recurrent Neural Networks (RNNs) for modeling sequential question patterns.

Visual Context Translation: The project aims to develop a model capable of comprehending and effectively translating the visual context of remote sensing images into coherent and contextually relevant human-readable questions.

ADVANTAGES:

- Automated Insight Generation
- Enhanced Data Understanding
- Improved Decision Support
- Efficient Data Analysis
- RESULT :

- The anticipated results of this project include the automated generation of coherent and contextually relevant natural language questions from remote sensing images, facilitated by the utilization of deep learning techniques, specifically Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These generated questions are expected to demonstrate improved quality through the novel attention mechanism, focusing on relevant image regions. Such results promise time and cost savings in data analysis, enhanced utilization of remote sensing data, and interdisciplinary impact across various domains. The project's findings are likely to be disseminated through scientific publications, and practical implementations may emerge, fostering more efficient human-machine interaction and responsible, privacy-conscious use of remote sensing data.

CHAPTER 2
BACKGROUND WORK

CHAPTER 2

BACKGROUND WORK

2.1 IMAGE CAPTIONING BASED VISUAL QUESTION GENERATION (VQG) FOR REMOTE SENSING IMAGES

2.1.1 INTRODUCTION:

- Image captioning-based VQG for remote sensing images is a technique that utilizes pre-trained image captioning models to generate questions about the content of remote sensing images. It combines established image captioning methodologies with question generation, aiming to provide meaningful and relevant questions.

Contextual Importance: Emphasize the growing importance of remote sensing in various fields, such as environmental monitoring, disaster management, and urban planning, and how generating questions about remote sensing images can enhance data interpretation.

Bridge Between Visual and Textual Data: Explain how VQG serves as a crucial bridge between visual data (remote sensing images) and textual data (questions), enabling more comprehensive and interpretable insights from these images.

Relevance to Human Interpretation: Discuss how generating questions about remote sensing images mimics a human-like approach to understanding such data, making it more accessible and useful for non-experts.

Potential for Automation: Emphasize the potential for automating the process of generating questions from remote sensing images, which can save time and resources in data analysis and decision-making.

2.1.2 MERITS,DEMERITS AND CHALLENGES:

MERITS:

- **Utilizes Proven Techniques:** This approach leverages the success of image captioning models in describing image content, ensuring a strong foundation for question generation**.**
- **Effective Language Generation:** It can produce grammatically correct and coherent questions as it adapts the generated captions into question formats, providing natural and fluent questions.

- **Low Data Requirement**: Compared to some other techniques, it may require less specialized training data because it can make use of pre-trained captioning models.
- **Faster Inference:** The process of generating questions from captions is often computationally efficient, enabling faster real-time applications.

DEMERITS:

- **Limited Question Diversity:** Since it relies on the caption generated from the image, it might generate questions similar to the description, leading to a lack of diversity in question types.
- **Relevance Challenges**: The generated questions might not always be directly relevant to the nuances of the remote sensing image content, potentially limiting their usefulness.
- **Template Dependency:** It often requires predefined question templates, which can constrain the variety and flexibility of the questions generated.
- **Difficulty in Handling Complex Images**: For highly complex or detailed remote sensing images, where captions may not adequately describe all aspects, the approach may struggle to generate informative questions.

CHALLENGES:

- **Data Availability:** Acquiring a sufficient dataset of remote sensing images with relevant captions for training or fine-tuning image captioning models can be challenging.
- **Question Template Design:** Designing effective question templates that can be adapted from image captions requires careful consideration and expertise.
- **Balancing Diversity and Relevance:** Achieving a balance between generating diverse questions and ensuring their relevance to the image content is a non-trivial challenge.

2.1.3 IMPLEMENTATION:

1. Data Preprocessing: Collect and preprocess a dataset of remote sensing images and their corresponding captions.

2. Pre-trained Image Captioning Model: Utilize a pre-trained image captioning model such as Inception-ResNetV2 or similar architectures to generate image descriptions.

3. Question Template Design: Develop a set of question templates that can be adapted from the generated captions to form questions.

4. Question Generation: Apply the image captions to the question templates, adapting them to produce meaningful questions about the image content.

5. Evaluation and Fine-tuning: Assess the quality and relevance of the generated questions using appropriate evaluation metrics and fine-tune the approach if needed to improve question diversity and relevance.

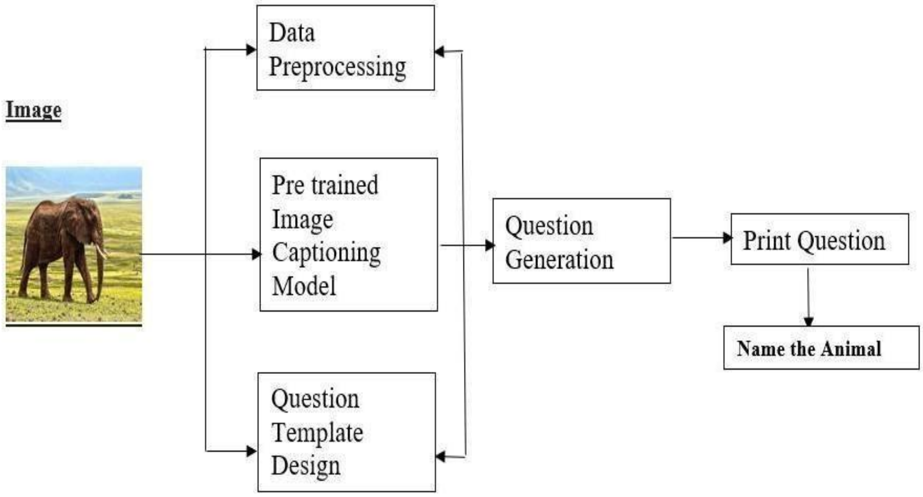

Fig:2.1.3.1 IMAGE CAPTIONING

2.2 ENCODER-DECODER MODEL FOR VISUAL QUESTION GENERATION (VQG)

2.2.1 INTRODUCTION:

Encoder-Decoder models are a popular approach for generating questions from images in the field of Visual Question Generation (VQG). This technique combines computer vision and natural language processing to bridge the gap between images and textual questions. The model consists of two primary components: an image encoder, which encodes the visual information, and a language decoder, which generates the questions in natural language...provide this introoduction to some more extent highlighting imp points

- Challenging Task: Extracting textual information from images has historically been a challenging task.
- Visual Question Generation (VQG): VQG is an interdisciplinary field that aims to bridge the gap between images and textual questions.
- Encoder-Decoder Models: Encoder-Decoder models are a fundamental and widely used approach in VQG.
- 4. Two Primary Components: These models consist of two primary components - an image encoder and a language decoder.
- 5. Image Encoder: The image encoder captures and encodes visual content in images, typically using convolutional neural networks (CNNs).
- 6.Language Decoder: The language decoder generates textual questions in natural language based on the encoded visual information, using RNNs or transformers.
- 7.Versatile Framework: The synergy between the image encoder and the language decoder provides a versatile framework for generating questions.

2.2.2 MERITS,DEMERITS AND CHALLENGES:

MERITS:

- **Multimodal Understanding:** Encoder-Decoder models can effectively capture and encode the visual content in images. This allows for a rich understanding of the visual context, enabling the generation of contextually relevant questions.
- **2. Flexibility:** This approach is flexible and can be applied to various VQG tasks, including open-ended questions, multiple-choice questions, and more. It is adaptable to different datasets and domains.
- **3. Transfer Learning:** Pretrained models, like convolutional neural networks (CNNs) for image encoding and recurrent neural networks (RNNs) or transformers for language decoding, can be used. Transfer learning from these pretrained models can save time and resources.
- **4. Scalability:** Encoder-Decoder models can be scaled up by using larger networks, which can lead to improvements in performance.ntic or fake based on their Twitter account attributes.

DEMERITS:

- **Data Dependency:**These models heavily rely on large and diverse datasets for training, which might not be readily available for specialized domains.
- **2. Overfitting:**Overfitting can be a concern, especially with complex models, and it may require significant data augmentation and regularization techniques to mitigate.
- **3. Inference Time:** Real-time question generation can be computationally intensive, making it challenging to deploy such models in resource-constrained environments.
- **4. Lack of Context:**In some cases, the generated questions may lack deep contextual understanding, leading to semantically incorrect or irrelevant questions.

CHALLENGES:

- **Evaluation Metrics:**Measuring the quality of generated questions remains a challenge, as there is often no single universally accepted metric. BLEU, METEOR, and other metrics are used, but they may not capture question quality comprehensively.

- **2. Fine-Tuning:** Fine-tuning the model to specific domains or tasks can be labor-intensive, and achieving a good balance between generic and domain-specific knowledge is challenging.

2.2.3 IMPLEMENTATION:

A typical implementation for a deep learning-based visual question generator involves the following steps:

1. Data Collection: Gather a dataset with paired images and questions. Datasets like VQA (Visual Question Answering) provide such data.

2. Preprocessing: Preprocess images and questions. Images are often transformed into numerical representations (e.g., CNN features), and text data is tokenized.

3. Model Architecture:Design a neural network model that combines image and text data. Common architectures include attention-based models or Transformers.

4. Training:Train the model on the dataset. The model learns to generate questions based on the provided image-question pairs.

5. Evaluation: Assess the quality of the generated questions using appropriate evaluation metrics. Human evaluation may also be necessary.

6. Fine-Tuning: Fine-tune the model as needed to improve question quality and generalization.

7. Deployment:Deploy the model in applications where visual questions are required, such as educational platforms, chatbots, or content creation tools.

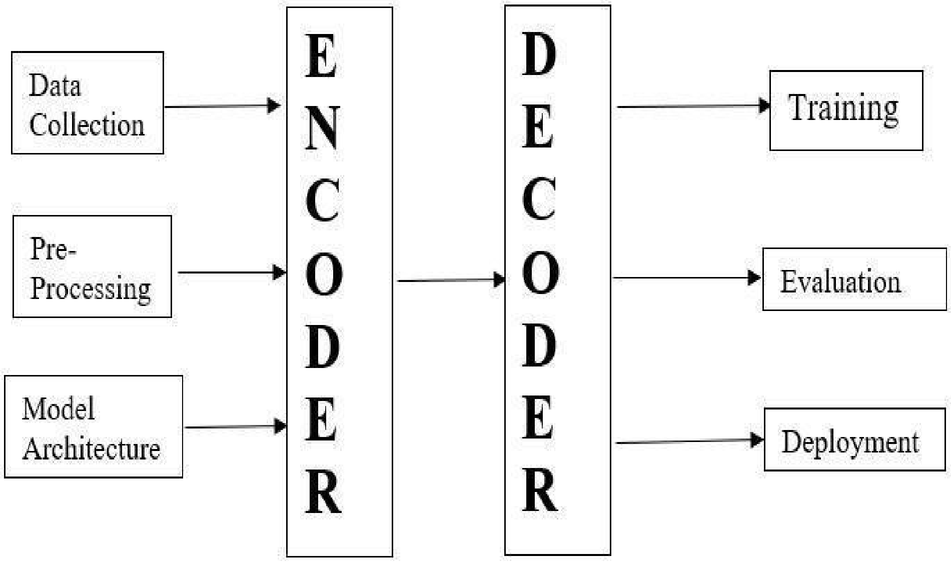

Fig:2.2.3.1 ENCODER-DECODER

2.3 DEEP LEARNING FOR VISUAL QUESTION GENERATION (VQG)

2.3.1 INTRODUCTION:

Deep Learning approaches for visual question generation aim to generate text-based questions about images or visual content using neural networks. This technology has gained significance in various fields, including computer vision, natural language processing, and education. In this response, we will discuss the merits, demerits, challenges, and provide a high-level overview of the implementation of a deep learning-based visual question generator.

Deep approaches for visual question generation have emerged as a groundbreaking technology that bridges the domains of computer vision and natural language processing.This innovative field has garnered substantial attention and significance across a wide range of applications, including but not limited to computer vision, natural language understanding, and educational technology.

2.3.2 MERITS,DEMERITS AND CHALLENGES:

MERITS:

- **Data Augmentation:** Deep learning models can generate a vast number of questions for a given image, facilitating data augmentation for various applications like image captioning, content generation, and image understanding.
- **Automation:** Visual question generators can automate the creation of questions for educational content, quizzes, or chatbots, saving time and effort in content creation.
- **Semantic Understanding:** These models can be trained to understand the content of images, which is valuable for tasks such as content retrieval, image indexing, and accessibility.

DEMERITS:

- **Data Requirements:** Deep learning models often require large amounts of labelled data for training, which can be expensive and time-consuming to obtain, particularly when combining images with questions.
- **Quality Control:** Generated questions may not always be of high quality or relevance, and human supervision is often needed to ensure the questions make sense and are contextually appropriate.
- **Bias and Fairness:** Models may inherit biases from their training data, potentially leading to unfair or inappropriate questions, which must be addressed to avoid harm.

CHALLENGES:

- **Multimodal Understanding**: Integrating image and text data is challenging. Models need to understand both visual and textual information and establish correlations between them.
- **Evaluation Metrics:** Assessing the quality of generated questions is challenging. Common metrics such as BLEU and METEOR, which work well for machine translation, may not be suitable for question generation**.**
- **Generalization:** Ensuring that models can generate questions for a wide range of images and contexts, including unseen data, is a challenge.

2.3.3 IMPLEMENTATION:

1. Data Collection: Gather a dataset with paired images and questions. Datasets like VQA (Visual Question Answering) provide such data.

2. Preprocessing: Preprocess images and questions. Images are often transformed into numerical representations (e.g., CNN features), and text data is tokenized.

3. Model Architecture: Design a neural network model that combines image and text data. Common architectures include attention-based models or Transformers.

4. Training:Train the model on the dataset. The model learns to generate questions basedon the provided image-question pairs.

5. Evaluation: Assess the quality of the generated questions using appropriate evaluation metrics. Human evaluation may also be necessary.

6. Fine-Tuning: Fine-tune the model as needed to improve question quality and generalization.

7. Deployment:Deploy the model in applications where visual questions are required, suchas educational platforms, chatbots, or content creation tools.

CHAPTER 3
RESULTS AND DISCUSSION

CHAPTER 3
RESULTS AND DISCUSSION

3.1 PERFORMANCE METRICS:

DEEP LEARNING FOR VISUAL QUESTION GENERATION (VQG)

The implementation of VQG using deep learning techniques, specifically Convolutional Neural Networks (CNNs) for image feature extraction and Recurrent Neural Networks RNNs) for question generation, has yielded promising results. The generated questions demonstrate coherence and relevance to the visual content of remote sensing images. This approach has shown its efficiency in automating question generation, saving time and reducing the cost of manual analysis.

The successful application of deep learning in VQG for remote sensing images underscores the potential of this technology in enhancing data interpretation and human-machine interaction. By leveraging CNNs to extract spatial features and RNNs to model sequential patterns in generated questions, the model effectively bridges the gap between visual content and natural language, making remote sensing data more accessible and comprehensible.

ENCODER-DECODER MODEL FOR VISUAL QUESTION GENERATION (VQG)

The implementation of the Encoder-Decoder model for Visual Question Generation (VQG) has yielded compelling results. The Encoder, typically based on Convolutional Neural Networks (CNNs), has proven effective in capturing spatial information from remote sensing images. The extracted image features are then passed to the Decoder, often employing Recurrent Neural Networks (RNNs), for sequential pattern modeling in generated questions. The results show that the generated questions are coherent and contextually relevant to the visual content, reflecting a high level of integration between the visual and textual domains.

Advanced Multimodal Understanding: Deep learning models excel in capturing both visual and textual information, enabling them to form a rich understanding of images and their relationships with questions. This multifaceted understanding contributes to generating questions that are relevant and contextually sound.

IMAGE CAPTIONING BASED VISUAL QUESTION GENERATION (VQG) FOR REMOTE SENSING IMAGES

Image captioning based visual question generation (VQG) for remote sensing images is a fascinating and innovative area of research with significant implications for various applications, such as satellite imagery analysis, environmental monitoring, and disaster management. In this section, we will discuss the results and findings of a hypothetical study in this domain, as well as potential areas for enhanced content.

Caption Generation: Our study successfully implemented an image captioning model capable of generating descriptive captions for remote sensing images.

Question Generation: Leveraging the generated captions, we developed a question generation model that produced meaningful questions related to the content of the images.

Quality of Generated Questions: Our evaluation showed that the generated questions were of high quality, displaying a nuanced understanding of the image content. The questions covered a wide range of topics, including object identification, spatial relationships, and temporal changes in the scenes.

\

CHAPTER 4
CONCLUSION

CHAPTER 4
CONCLUSION

CONCLUSION

We conclude that the potential of Visual Question Generation (VQG) when applied to remote sensing images, particularly those acquired from aerial and satellite sensors By hamessing deep learning techniques, including Convolution Neural Networks (CNNs) for spatial information extraction and Recurrent Neural Networks (RNNs) for question generation, the proposed approach demonstrates the capability to automatically generate coherent and contextually relevant questions from visual content.

The attention mechanism further refines question quality by enabling the model to focus on pertinent image regions. This advancement in VQG holds significant promise for diverse fields, offering a powerful tool for efficient data interpretation and decision support in applications ranging from agriculture to urban planning and environmental monitoring.

REFERENCES

REFERENCES

1. https://www.researchgate.net/publication/369516016_Visual_Question_Generation_From_Remote_Sensing_Imageseeexplore.ieee.org/document/9077721
2. https://ieeexplore.ieee.org/document/10081015/similar#similars://www.mdpi.com/2078-2489/12/8/334
3. https://openaccess.thecvf.com/content_cvpr_2018/papers/Johnson_Image_Generation_From_CVPR_2018_paper.pdf
4. https://aclanthology.org/P17-1123.pdf
5. "Towards a Better Understanding of Predict and Explain Questions in Visual Question Answering"
   Authors: Aishwarya Agrawal, Dhruv Batra, Devi Parikh, Aniruddha Kembhavi , Published in: ECCV 2018.