|| विश्वशान्तिर्ध्रुवम ध्रुवा ||

DR. VISHWANATH KARAD'S MIT WORLD PEACE UNIVERSITY, PUNE

SCHOOL OF COMPUTER ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTER ENGINEERING AND TECHNOLOGY

BTECH CAPSTONE PROJECT 2024-25

Title: Anchor-Based Scene Graph Decomposition for Image Captioning

Authors:

1032211482	Joshwa Joy Philip
1032210890	Aditya Verma
1032211777	Arya Haldankar
1032210941	Jaydeep Lokhande

Group ID: SP4

Category of The Project : InHouse

Institution: MIT World Peace University, Pune, Maharashtra, India

Advisor: Dr. Sheetal Girase

Summary of the Work:

Anchor-Based Scene Graph Decomposition for Image Captioning

The project focuses on enhancing image captioning by integrating anchor-based methods with scene graph decomposition. It addresses the challenges of accurately describing complex visual scenes, emphasizing the relationships between objects and their contexts. By introducing "anchors" as key points within an image, the methodology constructs structured scene graphs. These graphs, refined through Graph Neural Networks (GNNs), enable detailed understanding and contextual representation of visual content.

The system involves several core components:

1. Scene Graph Generation: Using object detection models to identify objects and their relationships.
2. Subgraph Proposal Network (sGPN): Sampling meaningful subgraphs that focus on specific image components.
3. Caption Generation: Employing attention-based Long Short-Term Memory (LSTM) networks to produce accurate and diverse captions.
4. Evaluation: Assessing captions using metrics like BLEU, CIDEr, and SPICE to validate accuracy and diversity.

Results and Analysis:

Results demonstrate superior performance compared to existing methods, with notable improvements in contextual accuracy and diversity. Applications span assistive technologies, autonomous vehicles, e-commerce, surveillance, and creative content generation.The research highlights challenges such as computational demands and reliance on annotated datasets, presenting opportunities for optimization through transformer architectures and domain-specific training. This project provides a foundation for future advancements in automated visual understanding and captioning systems.

Applications:

- Assistive technology for the visually impaired.
- Enhanced perception in autonomous vehicles.
- Automated content creation for e-commerce platforms.
- Real-time surveillance for security systems.

Conclusion:

Anchor-based scene graph decomposition offers an innovative framework for bridging computer vision and natural language processing. By capturing relational and compositional structures, it significantly improves caption quality and diversity.