Project Proposal Reference No. : 46S_BE_1009

Title of the project :
Knowledge-based Scene Graph Generation in Medical Field

Name of the College and Department:
St. Joseph Engineering College, Vamanjoor
Department of Computer Science and Engineering

Student Names :
Ms. Jessica Naomi D Souza
Ms. Aleema PK
Ms. Clita Fernandes
Ms. Dhanyashree S

Guide Name:
Dr. Chandra Naik, Associate Professor, Dept of CSE, SJEC

Keywords:
Scene graph generation: Scene Graph Generation (SGG) aims to extract entities, predicates, and their semantic structure from images, enabling a deep understanding of visual content, with many applications such as visual reasoning and image retrieval.
Knowledge graphs: A knowledge graph is a graph database that represents real-world entities and their relationships, providing a visual representation of interconnected information. It serves as a semantic network, capturing the connections between objects, events, situations, or concepts.
Scene understanding: Scene understanding aims to enable machines to comprehensively analyze visual scenes, achieving human-like abilities to understand the context, objects, and relationships within complex scenes.

Introduction :

Scene understanding is a crucial research area, particularly in the medical field, where object detection intersects with knowledge graphs. Traditional object detection lacks the ability to leverage contextual information and understand the scene's overall context. To address this, knowledge-aware object detection, which integrates external knowledge graphs, is needed.

Our project focuses on combining object detection and understanding object relationships to achieve a correlated understanding of medical scenes. By utilizing knowledge graphs, we create relationships between detected objects, generate scene graphs, and derive inferences to predict the scene's overall understanding. Our work stands out as one of the pioneering efforts in incorporating this approach specifically in the medical domain.

Objectives :

The objective of the proposed work is to understand scenes in the medical domain based on visual context using object detection and represent them in the form of knowledge graphs in order to derive conclusions or understand contexts.

The major contributions of our proposed work are:

(i) A Knowledge base represented in the form of scene graphs in the medical field.
(ii) Accurate analysis or understanding of the objects in the scene presented to it as an image.
(iv) Publication of the accomplished work in domain-specific Conference/Journal.

By reviewing 8-10 recent research papers, we developed a solution that extracts visual features and bounding boxes from input images to create knowledge graphs. We utilized Faster RCNN through transfer learning with a custom dataset of 880 images and 11 classes for object detection. Neo4j Graphs were used to store object relationships as nodes and edges in the form of SVO triplets, populated with predicate classes like USES, HAS, IN, RECEIVES, etc. The directed graph representation allows querying the database and deriving inferences in the form of SVO triplets based on the results.

Methodology :

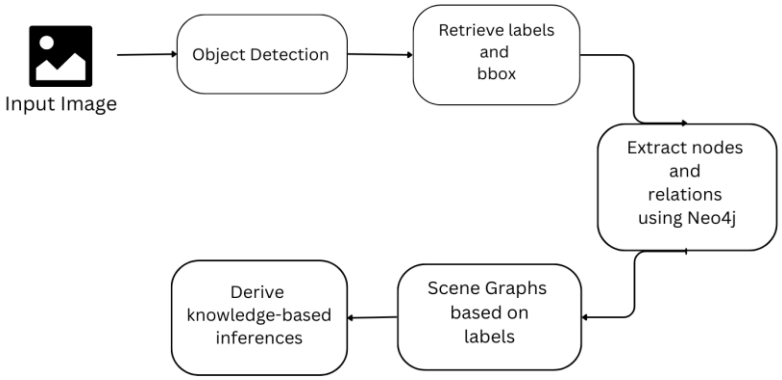

Figure 1 : Architecture Diagram using Neo4j

Neo4j is a powerful graph database ideal for storing and querying complex graph structures, including scene graphs. Using pre-trained object detection models like Faster R-CNN, we extract visual features from images to predict object relationships and generate scene graphs using tools like Scene Graph Benchmark.
These scene graphs are stored in Neo4j as nodes and relationships, with each object represented as a node containing class, coordinates, and features. Object relationships are represented as edges with labels and confidence scores.
By querying the stored scene graphs, tasks such as object retrieval, relationship detection, and scene understanding can be performed. Our project focuses on the medical field, utilizing a dataset with 11 classes encompassing objects found in hospitals.

To create nodes and relationships, we employ Subject Verb Object (SVO) triplets, including predicate and class markers along with bounding box values. Predicate classes such as USES, HAS, IN, LAY_ON, RECEIVES, CHECKS, READS, TREATS, INJECTED_TO, and WEARS are used. Currently, our knowledge base comprises 12 nodes and 21 relationships, representing simple medical scenarios.

Figure 2: Knowledge Base

For each of the nodes specified, we differentiate them into subjects and objects to form the correct relationships. This allows us to derive inferences in terms of the node being a subject of the SVO triplet or the object. We represent the relationships in terms of directed graphs which makes it possible to query the database. Based on the matches found, the surrounding nodes make the knowledge base for the results and hence inferences are derived.

Results and Conclusions :

In our medical-focused project, we aim to accurately analyze and understand the presented scene, providing a summary in the form of a sentence based on the image data. By incorporating auxiliary information from external knowledge graphs and using scene graphs, we can extract contextual information and predict a comprehensive understanding of the scene. This knowledge reasoning greatly enhances image retrieval and improves object recognition capabilities.

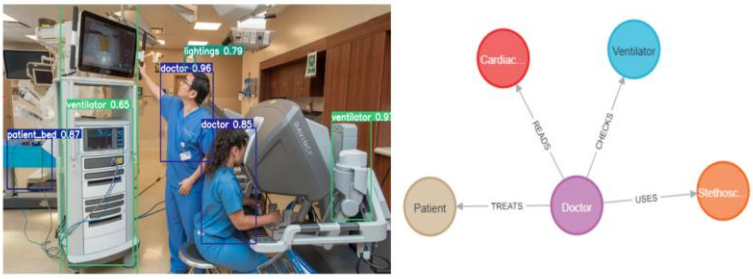

Figure 3: Example of an image and its scene graph

The steps followed are given below:

1. First, the image needs to be uploaded, this can be done by dragging or dropping the image as well as browsing the file from the computer. This image is the medical environment setup, for which we want to find the scene graph.

2. After this step, the object detection phase starts, and the result of this phase is shown below:

Figure 5: Example of object detection

3. Object proposals:
   Object proposals are candidate bounding boxes generated by an object detection algorithm, focusing on areas likely to contain objects of interest in an image. They help narrow down the search space for object detection and recognition tasks, reducing processing time. Object proposals can be created using algorithms like Selective Search, Edge Boxes, and Region Proposal Networks (RPNs). These proposals often serve as annotations for images, stored in the Pascal XML format.

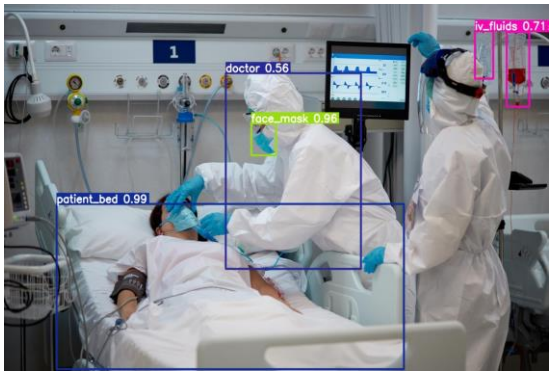

Figure 6: Example of an annotated and labeled image

4. Scene Graph Generation:

Generating scene graphs relies on the predicted object classes from object detection. Each object is treated as a node, and queries are executed to find matches labeled as the respective object. The surrounding nodes and relationships connected to the queried node are then interpreted to derive the final inference. This results in a set of SVO triplets that describe the scene, with varying degrees of accuracy, for each predicted class.

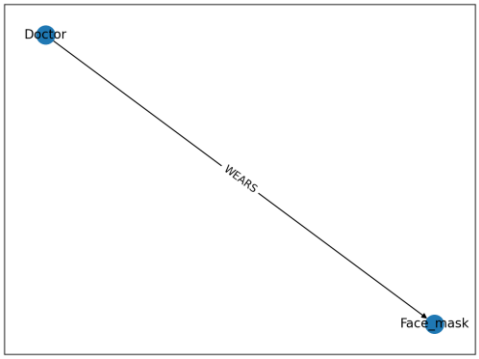

Figure 7: Nodes labelled "Face_mask" as object

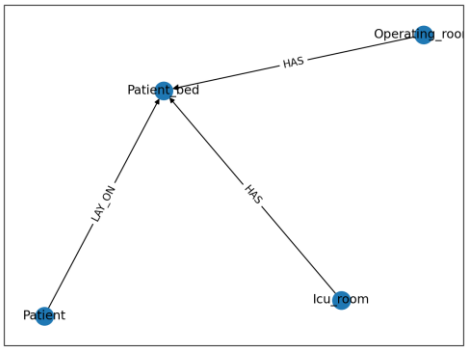

Figure 8: Nodes labelled "patient_bed" as object

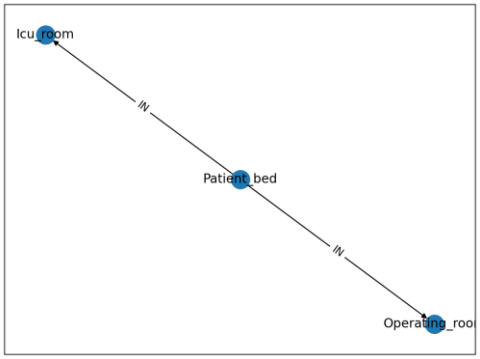

Figure 9: Nodes labelled "Patient_bed" as subject

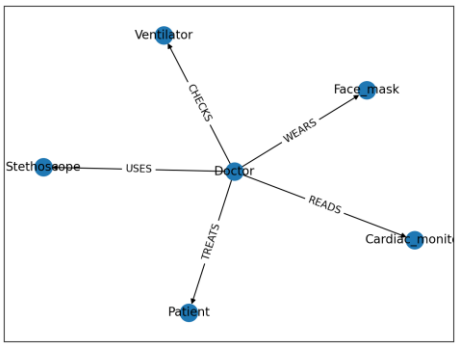

Figure 10: Nodes labelled "Doctor"

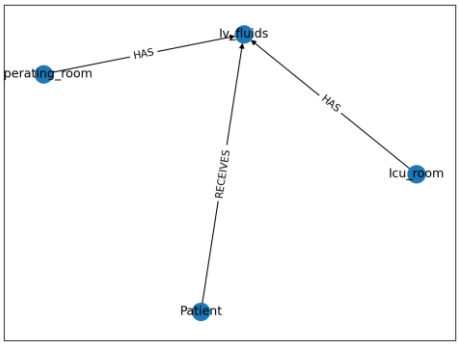

Figure 11: Nodes labelled "iv_fluids" as object

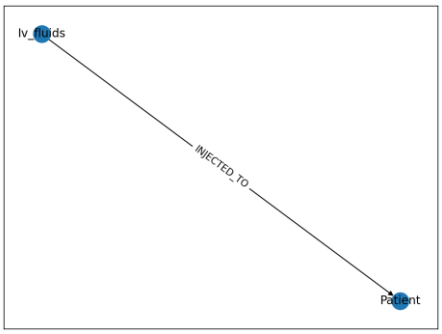

Figure 12: Nodes labelled "iv_fluids" as subject

6. Knowledge-Based Inferences:

SVO triplets in scene graph generation represent Subject-Verb-Object relationships in a scene. The Subject refers to the object being described, the Verb represents the relationship, and the Object is the object being described in relation to. For example, "Doctor - TREATS - patient" describes the relationship between a doctor and a patient.

These SVO triplets are valuable for tasks like image captioning, object detection, and visual question answering, as they capture the relationships between objects in an image.

Patient RECEIVES Iv_fluids

Icu_room HAS Patient_bed

Icu_room HAS Iv_fluids

Patient LAY_ON Patient_bed

Doctor WEARS Face_mask

Figure 13: Final Inference

The object detection model is evaluated using Precision, Recall, and mAP metrics with IoU. The Faster RCNN model trained for 18 epochs achieves an overall mAP of 0.55 and a training loss of 0.136, indicating moderate to good performance in object detection tasks. mAP is a widely used evaluation metric that considers precision and recall across various thresholds. However, the model's performance can vary depending on the use case and training data quality.

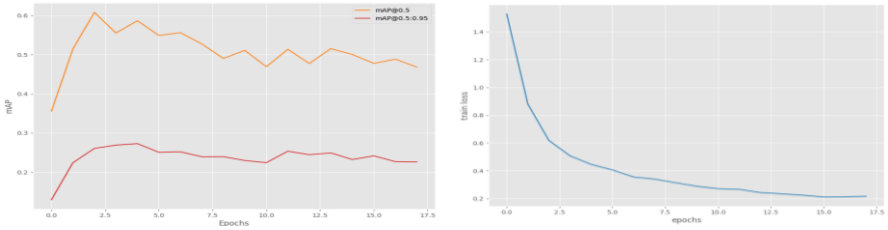

Figure 14: mAP

Future Scope :

Our project combines object detection and knowledge graphs into a single structure, aiming to improve understanding of the relationships between objects in a given scene. This approach integrates the state-of-the-art techniques in object detection and leverages deep learning, machine learning, and natural language processing. Our focus is on implementing this concept practically in hospital settings, which is a novel application compared to existing research papers. The target audience includes visually impaired individuals, the elderly, and future robot training, all of whom would benefit from a better understanding of their surroundings in the medical field.

By addressing the specific constraints and requirements of the mentioned areas, our project stands out as a unique and non-fungible contribution. Furthermore, the potential future application of this technology lies in training robots for healthcare environments, particularly in situations like disease outbreaks where human staff are at risk. Robots can perform health checkups on patients, unaffected by diseases like Covid-19. Additionally, visually impaired individuals can benefit from this project by receiving audio outputs in the form of sentences that describe the relationships between detected objects, enabling them to better comprehend their surroundings. This feature holds significant potential for future advancements in the field.