Abstract

The Visual Question Answering (VQA) system represents an innovative fusion of computer vision and natural language processing, facilitating the capability of machines to respond to queries grounded in visual stimuli. In this scholarly task, we present a tailored VQA framework meticulously crafted for skeletal imagery, leveraging sophisticated techniques for extracting both visual and textual features. These techniques enable the system to comprehend the structural aspects of the skeletal system and gain insights from radiographic or medical imaging data. By effectively transforming the questions into textual features, the system gains a deeper understanding of the user's inquiries and can provide accurate answers. The fusion of both visual and textual features is achieved using sophisticated integration methods, ensuring a seamless correlation between the image content and the textual context. This integration empowers the system to reason effectively and formulate responses that are contextually relevant and adequate. To examine the effectiveness of our proposed VQA system, we conducted extensive experiments on a diverse dataset of skeletal images and corresponding textual queries. The results demonstrate the system's capability to provide accurate and insightful answers, showcasing its potential for applications in the healthcare domain, radiology of skeletal images, and beyond.

A novel skeletal image of the proposed approach is based on B12 FRCNN and Kai-Bi-LSTM approaches is introduced in this paper to address the different challenges. The proposed system aims to enhance communication between medical professionals and patients by providing accurate answers to visual questions related to medical images. The system uses advanced methods like B12 FRCNN for object localization and Kai-Bi-LSTM for sequential processing to try to understand and interpret medical image queries better. This should lead to better interactions between patients and doctors.