Title	Efficient Textual and Visual Question-Answering Systems for Scientific Documents
Author(s)	Le, Khang Nguyen
Citation	
Issue Date	2023-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18341
Rights	
Description	Supervisor: NGUYEN, Minh Le, , (

Efficient Textual and Visual Question-Answering Systems for Scientific Documents

2110205 LE, Khang Nguyen

Visual-Question-Answering (VQA) requires a VQA system to answer questions corresponding to visual information. VQA in mathematics abstract diagrams containing abstract objects instead of natural images requires diverse cognitive reasoning skills, posing many challenges to current VQA methods. Language-vision models whose vision features come from image patch tokens or object proposals may not efficiently capture and present the information about the type and number of objects, which is essential for answering the mathematical question. Object detection techniques are important for obtaining object proposals. However, existing object detection models are trained on natural images, and no dataset is available for fine-tuning object detection on abstract objects. This study proposes methods for detecting abstract objects and generating valuable descriptions that can be used to augment the training and inference process of abstract diagram VQA models. The experiments show that existing VQA models benefit greatly from the augmented descriptions. Moreover, Transformer models trained using only the descriptions without any visual information achieve state-of-the-art results in IconQA sub-tasks. Furthermore, the proposed abstract object detection method enables future research in abstract diagram VQA models that use features from object proposals.