nidhin_b180948cs

by Nidhin_b180948cs Nitc

Medical Visual Question Answering

CS4099D Project
End Semester Report

Submitted by

Nidhin Mohan   (B180948CS)
Shaaheen A M   (B181134CS)
Abhinav B Naik   (B170297CS)

Under the Guidance of

Dr. Saidalavi Kalady
Dr. S. Sheerazuddin

Department of Computer Science and Engineering
National Institute of Technology Calicut
Calicut, Kerala, India - 673 601

May, 2022

NATIONAL INSTITUTE OF TECHNOLOGY CALICUT
KERALA, INDIA - 673 601

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

Certified that this is a bonafide report of the project work titled

MEDICAL VISUAL QUESTION ANSWERING

done by

Nidhin Mohan

Shaaheen A M

Abhinav B Naik

of Eighth Semester B. Tech, during the Winter Semester 2021-'22, in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering of the National Institute of Technology, Calicut.

(Dr. Saidalavi Kalady)

(Dr. S. Sheerazuddin)

05-05-2022

Date

Project Guide

DECLARATION

I hereby declare that the project titled, **Medical Visual Question Answering**, is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or any other institute of higher learning, except where due acknowledgement and reference has been made in the text.

Place : NIT Calicut
Date : 05-05-2022

Name : Nidhin Mohan
Roll. No. : B180948CS
Name : Shaaheen A M
Roll. No. : B181134CS
Name : Abhinav B Naik
Roll. No. : B170297CS

Abstract

The Visual Question Answering(VQA) domain explores the problem of answering questions regarding images. VQA in the medical domain(Med-VQA), has enormous potential and applications in the medical industry, such as assisting doctors in the field of radiology. Present Med-VQA frameworks employ deep learning approaches and image and question classification techniques. However, the quantity and diversity of questions and images in the medical domain, makes the task of Med-VQA challenging. Here we present our Med-VQA framework evaluated using the SLAKE dataset, which contains 642 images and 14,028 questions, which is larger compared to other datasets such as VQA-RAD which has 315 images and 3,515 questions. We use pre-trained VGG16 model for obtaining image features. GloVe word embeddings and LSTMs are used for obtaining question features. The questions are classified into two, the image and question features are then concatenated together, which is then passed to a classifier to obtain the answers. Our model has reached an accuracy of 62.91% and 61.08% for open and closed-ended questions respectively.

ACKNOWLEDGEMENT

We would like to express our sincere and heartfelt gratitude to our guides and mentor Dr. Saidalavi Kalady , Dr. S. Sheerazuddin and Lubna A, who have guided us throughout the course of the final year project. Without their active guidance, help, cooperation and encouragement, we would not have made headway in the project. We would like to thank our parents and the faculty members for motivating us and being supportive throughout our work. We also take this opportunity to thank our friends who have cooperated with us throughout the course of the project.

Contents

ii

List of Figures

List of Tables

Chapter 1

Introduction

Artificial Intelligence(AI) and Deep Learning has seen many advancements in the recent years, especially in the fields of computer vision and natural language processing. With advancements in these fields, more challenging problems are arising for which solutions are to be found. Many of those problems tend to combine image and language features, such as image captioning and visual question generation.

One such area of research is in the problem of Visual Question Answering(VQA). Visual Question Answering aims to correctly answer a question related to a given image. The task involves understanding both image and question features in order to give a correct answer. For example, given an image of a car, a person can ask "Which color is the car in the image?" or "What is shown in the image?". The VQA framework is expected to correctly answer the questions asked. VQA consists of both simple closed-ended questions like a question which answers either yes or no, or more complex open ended questions to which the answer can be anything. Certain questions also require better reasoning capabilities to be given a correct answer. Research is still being done to develop a highly efficient and accurate VQA models.

Visual Question Answering has many applications in the field of robotics,

assisting visually impaired people, answering questions regarding complex images and in the field of medicine. In medical visual question answering, a complex medical image is given as the input to which doctors can ask any clinical question regarding the image, in order to get a second opinion about the diagnosis. It can also be used by doctors to detect tumours or abnormalities in a given X-Ray, CT or MRI.
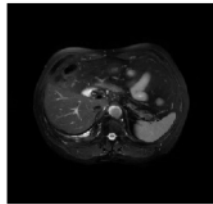
Q: What modality is used to take the image?
A: MRI
Q: Does the above picture contain kidney?
A: No

Tackling the problem of Medical VQA has its difficulties, mainly when procuring the medical datasets to train the VQA model. The literature available for Medical VQA is quite limited and the questions need to be answered with much greater accuracy. Med-VQA questions include closed ended questions such as asking for the presence of a particular organ in the image or detecting the presence of an abnormality, to more complex open ended questions such as identifying the abnormality or organs present in the given image.

In this paper, we are currently aiming to answer questions through classification means. The question features are obtained using GloVe word embeddings[1]. The questions are first classified using a Bidirectional Long Short Term Memory(Bi-LSTM) model classifier, into either open-ended questions with a single answer, or closed-ended questions with either yes or no as the answer. We then obtain the image features using Convolutional Neural Networks. Afterwards we concatenate the features together to obtain a single representation of the image-question pair, which is then given to either an

open-ended question classifier or a closed-ended question classifier, which predicts the answers from a given set of answers.

Chapter 2 mentions the problem statement for our project, Chapter 3 details the literature we've reviewed and the papers we have read to implement our project, Chapter 4 explains the details regarding our implementation and Chapter 5 contains the results obtained through our VQA framework.

Chapter 2

Problem Statement

To understand a given medical image, a question relevant to the image and to correctly predict the answer to the question based on the information present in the image, using deep learning techniques.

Chapter 3

Literature Survey

In this section, we describe the recent works related to VQA and Medical VQA.

[3] describes the AIML team's contribution to 2020 ImageCLEF Medical Domain Visual Question Answering (VQA-Med) challenge. The team approached VQA problem using a method known as Skeleton-Based Sentence Mapping(SSM). The method gave a common template to question whose structures were similar. Due to this technique's help, the VQA task was reduced to an image classification problem. They used a multi-architecture ensemble model for prediction of the answers. These methods helped them secure the first position in the challenge with a score of 0.496 in accuracy and 0.542 in BiLingual Evaluation Understudy(BLEU)[4] score.

[5] improves upon the reasoning part of a VQA framework. They treat the problem as a classification task. They introduce a Question Conditioned Reasoning(QCR) module and a Type Conditioned Reasoning(TCR) module, since the performance of VQA frameworks in open ended questions is relatively low compared to closed ended questions. The question conditioned reasoning module extract the abundant task information present in the questions. The type conditioned reasoning module classifies the questions into open-ended or closed-ended questions based on the emphasis given to specific

words that differentiate between an open-ended or closed-ended question. The introduction of QCR and TCR modules showed a significant increase in the accuracy of open ended questions, from about a percentage accuracy of 49.2 without either modules to a percentage accuracy of 60.0 with both modules.

[6] followed a question and image classification approach to tackle the VQA problem. The dataset used was generated from the MedPix database. It was observed that the category of question can be determined from the question words, and hence they classified the questions into four categories-plane, modality, organ system and abnormality type questions. Based on the output of question classification, the image was classified using the four image classification models used, and the output given as the answer. The modality classification model was more complex than the other three models as certain modality classes had various subclasses. The framework gave a percentage accuracy of 75.2 for organ system type questions, 77.6 for plane type questions, 72 for modality type questions and 18.4 for abnormality type questions.

[7] followed a similar approach to [6]. They used a dataset of 4000 training image-question-answer entries. The questions were of two types- closed-ended questions that asked whether the given image is normal or abnormal, and open-ended questions that asked which abnormality was present in the given image. The question was only used for determining the question type, closed or open-ended, and thus it was overall treated as an image classification task. Image features were extracted using pretrained VGG16 models using all but the final fully-connected classification layers. They used two models, one for closed-ended questions and another for open-ended questions. The model gave a best accuracy of 0.48 and a BLEU score of 0.511.

[8] follows an approach for MedVQA with a model that generates answers in a sequence of words for a medical image-question pair. Here a model was made to modify and combine both Machine Translation techniques and Image

captioning. Image features are obtained from CNNs and Gated Recurrent Units(GRUs) have been used for encoding and decoding. These helped them secure the first position in the challenge with a BLEU score of 0.188.

[9] proposed model follows a multi-modal approach of various levels, the first level being the separation of questions using SVMs and then using the results to predict the answers in the next level. Inception-Resnet is used for obtaining the image features and Bi LSTMs are used for obtaining question features. The questions are first segregated into yes/no or other types of questions. They have evaluated their model on the RAD and CLEF18 medical VQA datasets. For CLEF+RAD dataset, BLEU score of 0.0365 was obtained.

[10] proposes a model called CGMVQA (Classification and Generative Model for Medical VQA). The authors' motive was to use the CGMVQA model to answer questions containing various medical images by dividing the problem into multiple weak sub-problems. The dataset they used was the ImageCLEF 2019 VQA-Med dataset. Questions were divided into five different categories: yes-no, modality, plane, organ system and abnormality. Features were improved using augmentation of data and tokenization. Using ResNet152, they extracted the image features from different convolutional layers. Question features were extracted using three different types of word embeddings. They used a self-attention transformer with reduced parameters to learn sequence relations, to decrease costs. With this setup, the team was able to get a score of 0.64 in accuracy, 0.659 in BLEU and 0.678 in WBSS.

[11] proposes a solution using self-supervised pre-trained Transformers for NLP and Vision tasks. The team first pretrained their Multimodal Medical BERT(MMBERT) model on a VQA Dataset. They used the ROCO dataset for this task. Then they loaded the model with weights from pretraining and fine tuned it further on respective two medical VQA datasets, i.e, VQA RAD and VQA Med 2019 datasets. For Image feature extraction, they used Resnet152 and extracted features from different convolutional layers. They

experimented with 3 different settings for MMBERT, a general one, one with fine tuning for different question categories in the datasets, and a non-pretrained one. The second one achieved their best results with an accuracy of 0.672 and a BLEU score of 0.69.

We also looked into various meta learning models for training models to learn meta weights which can be used to learn weights of medical images using simple convolutional neural networks. [13] proposes a method of meta learning in which the parameters of a base model is updated using gradient descent from various adapted parameters and loss functions, in order to initialise the model parameters to learn according to a given task. However, it relies heavily on meta-annotation which is a limitation for the method.

[15] tries to overcome the problem of transfer learning weights for medical images and the limitations laid by MAML[13] by proposing a multiple meta-model quantifying process. In [15] multiple meta models are trained using MAML[13] and a list of candidate meta models is selected from the set of meta models based on their performance and difference of features from other meta models.

The overall framework of VQA used in these frameworks is similar- image feature extraction, question feature extraction, feature fusion and classification. Some of the papers have adopted simpler methods for modelling VQA frameworks. Other publications have added more modules for improving accuracy in areas like image and question feature extraction. On an average, the VQA models have an accuracy of around 0.6-0.7.
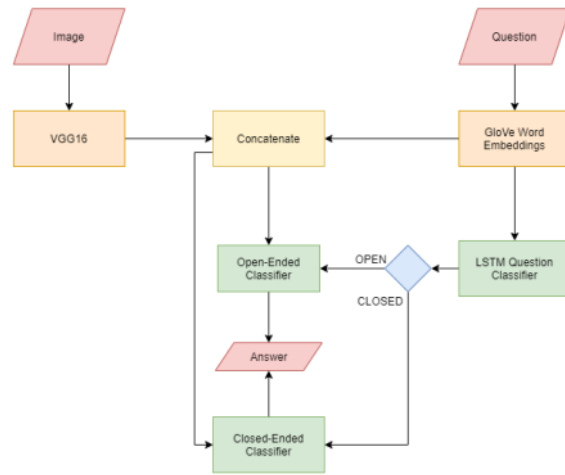
Chapter 4

Proposed Work

Figure 4.1: Our Proposed Med-VQA Framework

In our VQA framework shown in Figure 2, there are mainly four sections-
obtaining the necessary data, question feature extraction and classification,
image feature extraction, feature fusion and classification.

4.0.1   Obtaining the Necessary Data

We have used the SLAKE dataset[12] for our work, which actually consists of 14K image-question pairs and 642 images. Out of the 14K image-question pairs, the training data consists of 9,835 image-question pairs.

The questions in the dataset are divided into Chinese and English questions.

The questions are also divided into two: VQA and Knowledge-Aware VQA questions. Knowledge-Aware VQA questions required external knowledge along with the knowledge from the image to answer questions.

The questions are of four types based on their answers: open-ended single answer questions, open-ended multiple answer questions, multiple choice questions and closed-ended yes/no questions.

For our framework, we have used English VQA open-ended single answer questions and closed-ended yes/no questions. The training set consisted of a total of 3,949 questions, test set consisted of a total of 831 questions and validation set consisted of 814 questions.

4.0.2   Image Feature Extraction

Image features are extracted using VGG16[14]. A Convolutional Neural Network(CNN) is an algorithm which can be used to extract specific features regarding an image and learn the necessary weights which helps to distinguish images from one another. A CNN can be used to capture the relevant features in the image which could give a much better accuracy of answering to the VQA model.

VGG16 model takes images of fixed size input, 224x224x3, 3 being the number of channels. We used a VGG16 model that had been pretrained on the ImageNet dataset[16]. The full VGG16 model consists of three Dense layers and one Softmax layer at the end, which outputs a 1000-valued vector. For our model, we do not include the final Dense and Softmax layers, and

hence VGG16 outputs the extracted features of the image in a 7x7x512 array.

For our model, we load the images from file and pass through the VGG model to obtain the features of the images. Each image feature matrix is of size 7x7x512. This matrix is flattened into a 1x25088 array. Each image feature is combined into a single array with number of rows being the number of images, and each row being the image features obtained using VGG16.

The dimensions of the image, 1x25088, are in fact very large, but are not reduced so as to not avoid loss of information required for classification.

4.0.3 Question Feature Extraction and Classification

Question features are extracted using GloVe word embeddings. GloVe word embeddings converts each word into a 100-element vector. The Euclidean distance between each vector is a good measure for the semantic similarity between two words.

The questions are first passed through a tokenizer to obtain an encoded sequence for each question. Each word in the question is given an encoding by the tokenizer, which is an integer. The questions are then padded using zeros to obtain vectors of the same size. The size of the padded and encoded questions we have used is 16 since that is the maximum length of a question in the training set.

The padded questions are then passed through an embedding layer which assigns a 1x100 vector embedding to each encoded word in the sequence using GloVe word embeddings. The Embedding layer outputs a 16x100 vector for each question, which is then flattened to obtain a 1x1600 feature vector for each question.

In order to classify the question into either open-ended or closed-ended, an LSTM model is used.

Long Short Term Memory Networks have a feature of persistent memory, which enables them to learn the relevant features necessary and hence the features towards the end of the sequences can be learnt more accurately. This

enables them to make much more effective predictions when compared to a simple Recurrent Neural Network.

The question embeddings are passed through an LSTM layer and a Dropout layer to reduce overfitting which is then passed through a sigmoid classifier to obtain the final classification output.

4.0.4 Feature Fusion

In our model, we have performed the simplest feature fusion method, that is concatenating the two features together. For each question, the corresponding image feature vector is retrieved and concatenated to obtain a 1x26688 feature vector. Such a feature vector is obtained for each image-question pair and passed to the classification model.

4.0.5 Classification

We have used two separate classifiers for obtaining the final answers. An open-ended softmax classifier and a closed-ended sigmoid classifier. The image-question fused vector is passed to the corresponding classifier based on the output of the question classifier.

The softmax classifier classifies open-ended questions into 1 of 124 answer classes. The sigmoid classifier classifies closed-ended questions into either Yes or No.

Chapter 5

Experimental Results

Our model produced good results while training, testing and validation.

The question classification model did very well classifying the questions into closed and open ended questions. The classifier gave an accuracy of 100%.

```
What is the modality that this image is taken in?
Open

What is the main organ is this image?
Open

Is this a CT?
Closed

Is this an MRI?
Closed

Does this image contain lung?
Closed
```

Figure 5.1: Sample output when tested on individual question from the Question Classifier

The open ended image-question classifier model gave an accuracy of 62.91% when tested on 515 open-ended image-question pairs.

The closed ended image-question classifier model gave an accuracy of 61.08% when tested on 316 closed-ended image-question pairs.

We also tested the model manually on few sets of image question pairs for both closed and open-ended questions.

Table 5.1: Accuracies Obtained

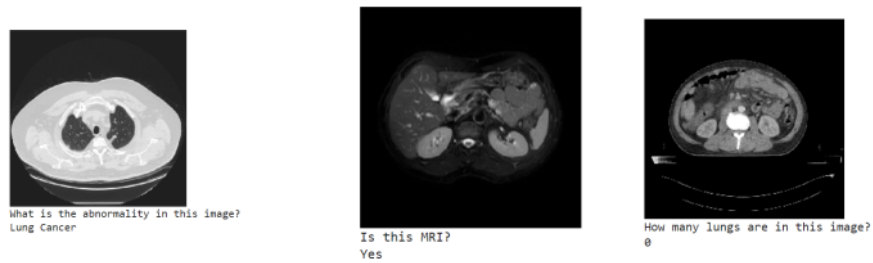| Model	Value
Question Classifier	100.00%
Open-Ended Classifier	62.91%
Closed-Ended Classifier	61.08%

Figure 5.2: Some examples of our manual testing of the model on image-question pairs, which gave correct answers
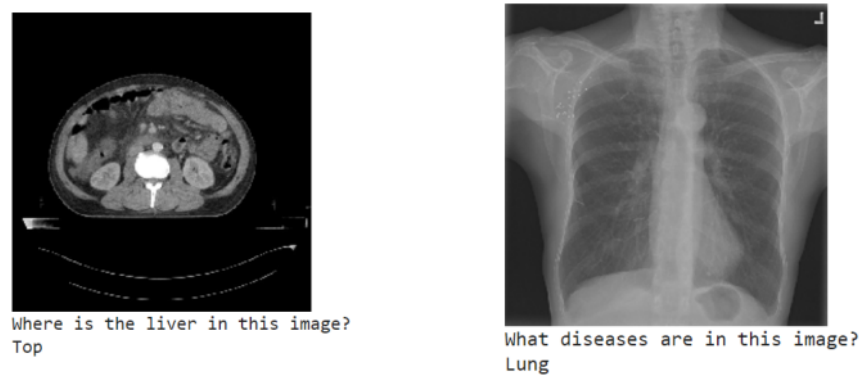

Figure 5.3: Some examples of our manual testing of the model on image-question pairs, which gave incorrect answers

Chapter 6

Conclusion

The use of Medical VQA frameworks can help ease the process of diagnosis by providing additional knowledge to medical practitioners when examining X-Ray, CT or MRI images. Another option can be given to help in the diagnosis and for more accurate detection of tumors and abnormalities.

Our model has given good accuracies on the training, testing and validation datasets. However the accuracies can be improved, especially during manual testing.

The questions could be segregated based on the content and appropriate information could be retrieved from the images to answer the questions.

The use of better feature fusion methods can improve the accuracy of the model.

In such ways, a Visual Question Answering model can be improved upon to give better accuracy when executing.

References

[1] Jeffrey Pennington, Richard Socher and Christopher D. Manning. "GloVe: Global Vectors for Word Representation". In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP).

[2] Jin-Hwa Kim, Jaehyun Jun and Byoung-Tak Zhang. "Bilinear Attention Networks". In arXiv:1805.07932.

[3] Zhibin Liao, Qi Wu, Chunhua Shen, Anton van den Hengel and Johan W. Verjans. "AIML at VQA-Med 2020: Knowledge Inference via a Skeleton-based Sentence Mapping Approach for Medical Domain Visual Question Answering." CLEF (2020).

[4] Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu. "BLEU: a Method for Automatic Evaluation of Machine Translation". In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002.

[5] Li-Ming Zhan, Bo Liu, Lu Fan, Jiaxin Chen, and Xiao-Ming Wu. 2020. Medical Visual Question Answering via Conditional Reasoning. In Proceedings of the 28th ACM International Conference on Multimedia (MM '20). Association for Computing Machinery, New York, NY, USA, 2345–2354. DOI:https://doi.org/10.1145/3394171.3413761

17

[6] Aisha Al-Sadi, Mahmoud Al-Ayyoub, Yaser Jararweh, and Fumie Costen, "Visual question answering in the medical domain based on deep learning approaches: A comprehensive study", Pattern Recognition Letters, vol. 150, pp. 57–75, 2021. doi:10.1016/j.patrec.2021.07.002.

[7] Aisha Al-Sadi, Hana AL-Theiabat, Mahmoud Al-Ayyoub. (2020). The Inception Team at VQA-Med 2020: Pretrained VGG with Data Augmentation for Medical VQA and VQG.

[8] R. Ambati and C. Reddy Dudyala, "A Sequence-to-Sequence Model Approach for ImageCLEF 2018 Medical Domain Visual Question Answering," 2018 15th IEEE India Council International Conference (INDICON), 2018, pp. 1-6, doi: 10.1109/INDICON45594.2018.8987108.

[9] Deepak Gupta, Swati Suman, Asif Ekbal. (2020). Hierarchical Deep Multi-modal Network for Medical Visual Question Answering.

[10] F. Ren and Y. Zhou, "CGMVQA: A New Classification and Generative Model for Medical Visual Question Answering," in IEEE Access, vol. 8, pp. 50626-50636, 2020, doi: 10.1109/ACCESS.2020.2980024.

[11] Y. Khare, V. Bagal, M. Mathew, A. Devi, U. D. Priyakumar and C. Jawahar, "MMBERT: Multimodal BERT Pretraining for Improved Medical VQA," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021, pp. 1033-1036, doi: 10.1109/ISBI48211.2021.9434063.

[12] Bo Liu, Li-Ming Zhan, Li Xu, Lin Ma, Yan Yang, Xiao-Ming Wu. "SLAKE: A Semantically-Labelled Knowledge-Enhanced Dataset For Medical Visual Question Answering". In:arXiv:2102.09542v1(18 Feb 2021).

[13] Chelsea Finn, Pieter Abbeel, Sergey Levine. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks". In:arXiv:2106.15367(19 Jun 2021).

[14] Karen Simonyan, Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In:arXiv:1409.1556(4 Sep 2014).

[15] Tuong Do, Binh X. Nguyen, Erman Tjiputra, Minh Tran, Quang D. Tran, Anh Nguyen. "Multiple Meta-model Quantifying for Medical-Visual Question Answering". In:arXiv:2105.08913(19 May 2021).

[16] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, doi: 10.1109/CVPR.2009.5206848.

nidhin_b180948cs

23% SIMILARITY INDEX

19% INTERNET SOURCES

18% PUBLICATIONS

14% STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Natonal Institute of Technology Calicut Student Paper	**4**%
2	arxiv.org Internet Source	**2**%
3	icaeta.aiplustech.org Internet Source	**2**%
4	Submitted to Rochester Institute of Technology Student Paper	**1**%
5	Usman Naseem, Matloob Khushi, Jinman Kim. "Vision-Language Transformer for Interpretable Pathology Visual Question Answering", IEEE Journal of Biomedical and Health Informatics, 2022 Publication	**1**%
6	Submitted to University of Bristol Student Paper	**1**%
7	Submitted to University of Wollongong Student Paper	**1**%

8 Submitted to The Hong Kong Polytechnic University
Student Paper
1%

9 Aisha Al-Sadi, Mahmoud Al-Ayyoub, Yaser Jararweh, Fumie Costen. "Visual Question Answering in the Medical Domain Based on Deep Learning Approaches: A Comprehensive Study", Pattern Recognition Letters, 2021
Publication
1%

10 repo.lib.tokushima-u.ac.jp
Internet Source
1%

11 dblp.uni-trier.de
Internet Source
1%

12 researchr.org
Internet Source
1%

13 Soumya Priyadarsini Panda, Navin Chandra. "A Visual Question Answering System using YOLO Model", 2021 19th OITS International Conference on Information Technology (OCIT), 2021
Publication
1%

14 Submitted to AUT University
Student Paper
1%

15 hekyll.services.adelaide.edu.au
Internet Source
1%

upcommons.upc.edu

16   Internet Source   <1%

17   Li-Ming Zhan, Bo Liu, Lu Fan, Jiaxin Chen, Xiao-Ming Wu. "Medical Visual Question Answering via Conditional Reasoning", Proceedings of the 28th ACM International Conference on Multimedia, 2020
Publication   <1%

18   deepai.org
Internet Source   <1%

19   aclanthology.org
Internet Source   <1%

20   medinform.jmir.org
Internet Source   <1%

21   dblp.org
Internet Source   <1%

22   Submitted to Imperial College of Science, Technology and Medicine
Student Paper   <1%

23   M Penza. "Identification and quantification of individual volatile organic compounds in a binary mixture by SAW multisensor array and pattern recognition analysis", Measurement Science and Technology, 06/01/2002
Publication   <1%

24   bi.snu.ac.kr

Internet Source

<1%

25   Yash Khare, Viraj Bagal, Minesh Mathew, Adithi Devi, U Deva Priyakumar, CV Jawahar. "MMBERT: Multimodal BERT Pretraining for Improved Medical VQA", 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), 2021
      Publication

<1%

26   scholarworks.sjsu.edu
      Internet Source

<1%

27   docplayer.net
      Internet Source

<1%

28   virajbagal.github.io
      Internet Source

<1%

29   www.frontiersin.org
      Internet Source

<1%

30   www.nature.com
      Internet Source

<1%

31   Karim Gasmi, Ibtihel Ben Ltaifa, Gaël Lejeune, Hamoud Alshammari, Lassaad Ben Ammar, Mahmood A. Mahmood. "Optimal Deep Neural Network-Based Model for Answering Visual Medical Question", Cybernetics and Systems, 2021
      Publication

<1%

32 Ni Wang, Zheng Wang, Xing Xu, Fumin Shen, Yang Yang, Heng Tao Shen. "Attention-Based Relation Reasoning Network for Video-Text Retrieval", 2021 IEEE International Conference on Multimedia and Expo (ICME), 2021
Publication

<1 %

33 Silvio Barra, Carmen Bisogni, Maria De Marsico, Stefano Ricciardi. "Visual Question Answering: which investigated applications?", Pattern Recognition Letters, 2021
Publication

<1 %

34 Zhou Zhao, Shuwen Xiao, Zehan Song, Chujie Lu, Jun Xiao, Yueting Zhuang. "Open-Ended Video Question Answering via Multi-Modal Conditional Adversarial Networks", IEEE Transactions on Image Processing, 2020
Publication

<1 %

35 dias.library.tuc.gr
Internet Source

<1 %

36 dokumen.pub
Internet Source

<1 %

37 escholarship.org
Internet Source

<1 %

38 vdoc.pub
Internet Source

<1 %

39 Dhruv Sharma, Sanjay Purushotham, Chandan K. Reddy.
"MedFuseNet: An attention-based multimodal deep learning model for visual question answering in the medical domain", Scientific Reports, 2021
Publication

40 Fuji Ren, Yangyang Zhou. "CGMVQA: A new Classification and Generative Model for Medical Visual Question Answering", IEEE Access, 2020
Publication

41 Sepp Hochreiter, Jürgen Schmidhuber. "Long Short-Term Memory", Neural Computation, 1997
Publication

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		