

Chabbiimen at VQA-Med 2021: Visual Generation of Relevant Natural Language Questions from Radiology Images for Anomaly Detection

Imen Chebbi¹

¹*Higher Institute of Informatics and Communication Technologies (ISITCOM), Hammam Sousse, Tunisia*

Abstract

The new progress in the domain of artificial intelligence has created recent opportunity in support clinical decision. In exceptional, appropriate solutions for the automated analysis of medical pictures are increasing interest expected to their applications potential in picture research and in the diagnosis assisted. In addition, systems able to understand clinical pictures and answer the questions about its content can assist objective decision making, objective education. In my article, I will describe my technique for produce visual questions on radiology images. I have used augmentation techniques for increasing data and VGG19 for extraction of the feature from a picture and prediction. VQGR is implemented using Tensorflow. My model is quite similar to GRNN [1].

Keywords

Visual Question Generation, Natural Language, Anomaly Detection, Radiology Images, Medical Questions.

1. Introduction

Nowadays, the task of Visual Question Generation becomes the very important task in the medical domain, which seek to produce human-like questions from the picture and potentially other side information (e.g. answer type or answer itself). The last few years have seen a surge of interests in VQG because it is particularly useful for providing high-quality synthetic training data for Visual Dialog system and Visual Question Answering (VQA) [2,3]. It seems like a challenging task because the generated questions are not only required to be consistent with the image content but also meaningful and answerable to human.

Even with promising results have been achieved, previous works still encounter two major issues. First, all of existing methods significantly suffer from one image to many questions mapping problem rendering the failure of generating referential and meaningful questions from an image. Extant VQG methods can be generally categorized into three classes with respect to what hints are used for generating visual questions:

- 1) The whole image as the only context input [4].
- 2) All images and required answers [5].
- 3) The whole image with the desired answer types [6].

CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24, 2021, Bucharest, Romania

✉ chabbiimen@yahoo.fr (I. Chebbi)

>ID 0000-0003-4290-8415 (I. Chebbi)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

The picture is worth a thousand words, and it can be potentially mapped to many different questions, leading to the generation of diverse non-informative questions with poor quality. Even with the answer type or desired answer information, the similar one-to-many mapping issue remains, partially because the answer hints are often very short or too broad. As a consequence, these side informations are often not informative enough for guiding question generation process, rendering the failure of generating referential and meaningful questions from an image.

The second serious issue for the existing VQG methods is that they ignore the rich correlations between the visual objects in an image and potential interactions between the side information and image [6]. In the concept, the implicit relations among the visual objects (e.g., spatial, semantic) could be the key to generate meaningful and relevant questions. This is partially since when human annotators ask questions about a given image, they often focus on these kinds of interactions. Additionally, another important factor for producing informative and referential questions is about how to make entire use of side information to align with the targeted image. Such modeling potential interactions between the side information and an image becomes a critical component for generating referential and meaningful questions.

The mission is to generate relevant questions in natural language on the radiological images using their visual content. In my article, I will describe my method for produce visual questions on radiology images. I have used augmentation techniques for the augmentation of data and VGG19 convolutional neural network for image feature extraction and prediction. The structure of this article is arranged as follows: The part 1 and part 2 contain introduction and surveys work related, part 3 Methods, part 4 The Results Experimental, part 5 and part 6 contain discussion of my results, the Conclusion and my Future Work.

2. Related Work

Question generation, an increasingly important area, is the task of automatically creating natural language questions from a range of inputs, such as natural language text [7-9], structured data [10] and images [11]. Concerning VQG in the medical sector, Sarrouti et al. [12] suggested a technique for the task of VQG on radiological pictures named VQGR. His proposition was based on VAE. For augmenting the amount of the dataset, this author [12] used data augmentation for both pictures and questions [60]. In this work, I'm interested for the task of produce questions from medical images. VQG in the open-domain benefited from the available large annotated datasets [13-15]. There is a variety of work studying generative models for generating visual questions in the open domain [16-17]. Recent VQG approaches have used auto encoders architecture for the purposes of VQG [18-20]. The success of these systems is mainly the result of variational auto encoders (VAEs) [21].Conversely, VQG in the medical domain is still a challenging and under-explored task [22-24].

Despite the fact that a high-quality manually created medical VQA dataset exists, VQA-RAD [23], this dataset is too small for training and there is a need for VQG approaches to create training datasets of sufficient size. Generating new training data from the existing examples through data augmentation is an effective approach that has been widely used to manipulate the data poverty problem in the open domain [25,26]. In my paper, I will describe my method for

produce visual questions on radiology pictures named VQGR. I have used VGG19 for extraction feature from image and for prediction.

3. Methods

In this study, the goal is to generate natural language questions based on radiology image contents. The overview of VQGR is shown in Figure number three.

3.1. Technique Of Data Augmentation

Technique Of Data Augmentation is a method utilized for augmenting the size of data by adding on moderately adapted duplicate of already available data or recently generated artificial data from available data. It worked as a regularize and assist to lower over fitting when training a model of machine learning. In data analysis, data augmentation is nearly associated with oversampling. I have used this technique for augmenting the dataset of images and questions.

3.1.1. Questions.

I generated new training examples by using question augmentation technology. I create a dataset of new questions for medical question q. For my approach, during the augmenting process, I have use all the Dataset VQA-Med 2021/ VQA-Med 2020/ VQA-Med 2019/ VQA-Med 2018. For instance, for a given question “Are the kidneys normal?”, we produce the followings questions:

- “Were the kidneys normal?”
- “Are the pancreas normal?
- “Are the intestines normal?”
- “Are the isografted normal?”
- “Are the livers normal? ”
- “Are the lungs normal?”
- “Are the organs normal?”
- etc.

3.1.2. Images.

I have also created a new learning instance based on image enhancement technology. For this reason, in my approach I have applied all Datasets VQA-Med 2021 / VQA-Med 2020 / VQA-Med 2019/ VQA-Med 2018 educational image inversion, rotation, movement and blur techniques.

3.2. Visual Question Generation

To move work back founded on the representation exact of the figure, a technique for asking natural questions, from a picture, was created in 2016 [1]. This task of VQG is, therefore, a supplement of available VQA systems, to understand the method image questions might relate to inference and common sense reasoning. This system which has been developed presents models based on generation and retrieval to deal with the mission of generating questions

on visual pictures. A second application of the produce of Visual Question related with the identification of objects in a picture [27]. Because, is impossible to prepare picture models recognition with all feasible objects in the earth, the generation of visual questions founded on mysterious entity can be employed for class acquisition. Such class acquisition operation is accomplished by answering humans questions based on mysterious objects for every picture. I will present in this paper my approach to produce visual questions on radiology images quite similar to GRNN [1].

3.2.1. VGG-19 Model.

From Oxford University Simonyan and Zisserman developed a 19-layer CNN (3 fully connected, 16 conv.) that exclusively used 3×3 filters along pad and stride of 1, as well as 2×2 layers max pooling along stride 2, named model VGG19. If I compare VGG19 to AlexNet, the VGG19 (As seen in Figure 1) is a profound CNN along more layers. For lowing the parameters number in these deep networks, it utilize little 3×3 filters in every convolutional layer and is supreme used along its error rate of 7.3%. This model was not the champion in the ILSVRC 2014, nevertheless, the VGG Net is the most powerful articles since it strengthened the idea that CNNs continue to get a deep network of layers.

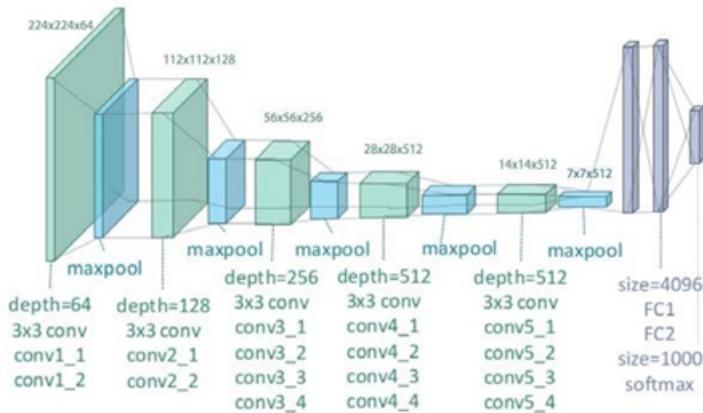


Figure 1: The architecture network of VGG-19 model [28].

3.2.2. Proposed Architecture.

My architecture contains two main parts (see Figure 2), namely the preprocessing unit and the question visual generation (VQG) part. The module of preprocessing needs picture processing techniques to train picture and transform it to an easier to use format. For the mining of the image features, I have utilize the convolutional neural network VGG19 [10]. In addition, preprocessing of annotations and questions is performed to produce vocabulary. The VQG unit will get the features visual that will study how to produce a question integration. Similar visual

characteristics are displayed in the model for all stride lengths, generating one question word at a time.

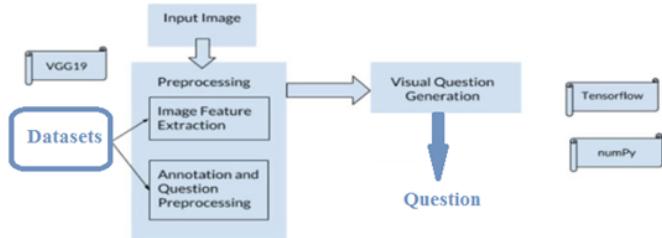


Figure 2: Complete Architecture.

3.2.3. Proposed Model.

Visual question generation is important task has currently been studied as a supplement of image caption. VQG techniques are utilized to produce questions that are dissimilar from image caption techniques that explain a picture in a some words (Eg, Are the organs normal?). The methods proposed in this article use a data augmentation and VGG19 (As seen in Figure 3) , so the module of question generation is the node of my approach, which employs a data augmentation and a VGG19 model which takes a picture as input and produce questions from its characteristics.

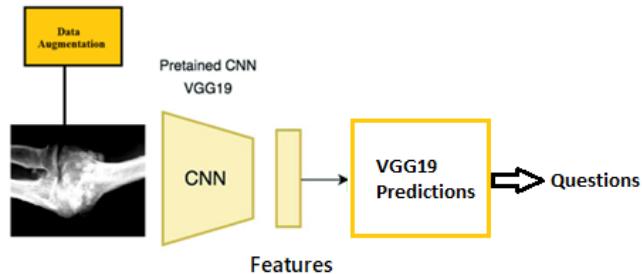


Figure 3: Proposed Model VQGR.

4. Experiments

In this study, I have used the dataset[31]:

4.1. Dataset

For training set I have explored the existing datasets used in VQG-Med 2020, VQG-Med 2019 and VQG-Med 2018 [31]. For validation set I have used the VQG 2021 validation [31] set contains 200 questions associated with 85 radiology images and for Test Set the VQG test[31] set includes 76 radiology images.

4.2. Evaluation Metrics

For evaluating efficiency of VQGR approach, I will utilize bilingual evaluation understudy BLEU.

4.2.1. BLEU score

It captures the resemblance allying the response generated by the model and ground truth response. BLEU scores are between zero and one. Scores closer to the value of one indicate more resemblance allying the answer and ground truth answer.

4.3. Results and Discussion

The table number 1 represents all my runs submitted in the Task VQG ImageCLEF2021[32]. My best score of Bleu metric from Aicrowd system for ImageCLEF2021 [32] is 0.383. My score is the best one in Task VQG for ImageCLEF2021 [32].

In first run number 133658 I get score 0.383 in this run I have used 2300 questions and I have generate seven questions for each image. In second run number 133896 I have used the same database of questions but I have generate three questions for each picture and I have get 0.212 score. In the third run number 134438 I have used 3,200 questions but I have generate seven questions. The fourth and fifth runs represents the same runs with run number 134438. I have noticed that in all my runs for every dataset of question used I get score 3.383 but when I change the number of predictions my score change.

Table 1

All my runs in ImageCLEF2021 Task VQG[32].

Run	Score
133658	0.383
133896	0.212
134438	0.383
134446	0.383
134449	0.383

Figure 4 represents examples of some images with natural questions and automatically generated questions by using my model VQGR.

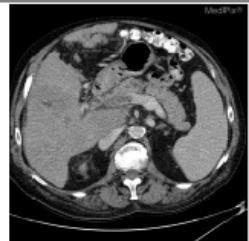
Images	Question
	<ul style="list-style-type: none"> -What does the image contain? -What do ultrasound images of the right kidney demonstrate duplicated collecting system with? -What demonstrates homogeneous density of thickened bowel wall with intestinal lumen surrounded by edematous mucosa? -What demonstrate large? -What is in the right parietal bone above the squamosal suture? -What infiltrates the neurforamen at C3-4 and displaces and compresses the spinal cord? -What does axial CT image through the mandible demonstrate?
	<ul style="list-style-type: none"> -What is layering in the aneurysm? -What is compressed by collection? -What causes mild compression and leftward shift of the spinal cord? -What compresses 4th ventricle? -What does axial non-contrast CT demonstrate? -What does t2 weighted axial MR image reveal a well-margined mass with smooth borders in the left lung at? -Who does Chest CT reveal basilar-predominant hyperlucent lungs with?
	<ul style="list-style-type: none"> -What is compressed by collection? -What can be seen in pincer-type femoral acetabular? -What is layering in the aneurysm? -What is eccentric to the right with superimposed uncovertebral spurring causing mass impression and effacement of the ventral thecal sac and cord stenosis? -What is produced by metallic dental braces? -What causes mild compression and leftward shift of the spinal cord? -What does cerebellar lesion extend into stretching 7th and 8th cranial nerves?

Figure 4: Examples of some images with natural questions and automatically generated questions.

5. Discussion

For every result of my model in Figure 3, I can view that for one image I can get one to seven questions in output. More interestingly, I have presented a model to generate visual questions in the medical domain. The results of automatic and manual assessment shows that VQGR surpass the baseline model for generating fluency and related questions. In this work, I have presented the initial trial to generate visual questions in the medical domain by using VGG19.

6. Conclusion and Future Work

I have presented in this article my method for generating visual questions in the medical domain by using VGG19. I showed a data expansion method for generating new training questions and images from the VQA-RAD dataset [31]. Next, I introduced the VQGR model, which generates questions from radiographs. Automated and manual evaluation results showed that the VQGR

fluency-related questions outperformed the baseline model. I will develop this model in the future and investigate the use of produced questions intended to process VQA in healthcare.

7. Acknowledgments

This research has received funding according to the funding agreement number LR11ES48 of the Tunisian Ministry of Higher Education and Scientific Research.

References

- [1] Mostafazadeh.N., Misra.I., Devlin.J., Mitchell.M., He.X.,and Vanderwende.L.. Generating natural questions about animage.arXiv preprint arXiv:1603.06059, (2016).
- [2] Li.Y, Duan.N., Zhou.B, Chu.X, Ouyang.W., Wang.X. Visual Question Generation as Dual Task of Visual Question Answering.In2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6116–6124(2018).
- [3] Jain.U., Lazebnik.S., Schwing.A.G.. Two can play this game: Visual dialog with discriminative question generation and answering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2018).
- [4] Mora. I. M., S. P. de la Puente, Giro-i Nieto.X. Towards automatic generation of question answer pairs from images.CVPRW, (2016).
- [5] Lau.J.J., Gayen.S., Ben Abacha.A., Demner-Fushman.D.: A dataset of clinically generated visual questions and answers about radiology images. Scientific data5(1), 1–10 (2018).
- [6] Krishna.R., Bernstein.M., Fei-Fei.Li. Information maximizing visual question generation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages2008–2018 (2019).
- [7] Kalady.S., Elikkottil.A., Das.R. Natural language question generation using syntax and keywords. In Proceedings of QG2010: The Third Workshop on Question Generation, volume 2, pages 5–14 (2010).
- [8] Kim.Y.,Lee.H., Shin.J.,Jung.K. Improving neural question generation using answer separation. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 6602–6609 (2019).
- [9] Li.J., Gao.Y., Bing.L., King.L., Michael.R. Lyu. Improving question generation with to the point context. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the9th International Joint Conference on Natural Language Processing, volume 33, page 3216–3226 (2019).
- [10] Serban.I., Alberto.G., CaglarGulcehre, Sarath.C., Aaron (2011).
- [11] Mostafazadeh.N., Misra.I., Devlin.J., Mitchell.M., He.X., Vanderwende.L. Generating natural questions about an image. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages1802–1813, Berlin, Germany. Association for Computational Linguistics (2016).
- [12] Sarrouti.M., Ben Abacha.A., Demner-Fushman.D.: Visual question generation from radiology images. In: Proceedings of the First Workshop on Advances in Language and Vision Research. pp. 12–18 (2020).

- [13] Agrawal.A., Lu.J., Antol.S., Mitchell.M., Zitnick.C.L., Batra.D., Parikh.D. VQA: Visual Question Answering. In2015 IEEE International Conference on Computer Vision (ICCV), pages2425–2433 (2015).
- [14] Goyal.Y., Khot.T., Agrawal.A., Summers-Stay.D., Batra.D., Parikh.D. Making the V in VQA Matter: Elevating the Role of Image Understanding in Visual Question Answering. Int. J. Comput. Vision, 127(4):398–414 (2019).
- [15] Johnson.J., Hariharan.B., DerMaaten.L.V., Fei-Fei.Li., Zitnick.C.L., Girshick.R. Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. In2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages1988–1997 (2017).
- [16] Masuda-Mora.I., Pascual-deLaPuente.S., Gir o i NietXX. Towards automatic generation of question answer pairs from images. In Visual Question Answering Challenge Workshop, CVPR , Las Vegas, NV, USA (2016).
- [17] Zhang.S., Qu.L., You.S., Yang.S., Zhang.J. Automatic generation of grounded visual questions. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), pages 4235–4243 (2016).
- [18] Jain.U., Zhang.Z., Schwing.A. Creativity: Generating diverse questions using variational auto encoders. In Proceedings -30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, pages 5415–5424, United States. Institute of Electrical and Electronics Engineers Inc (2017).
- [19] Li.Y., Duan.N., Zhou.B., Chu.X., Ouyang.W., Wang.X. Visual Question Generation as Dual Task of Visual Question Answering. In2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6116–6124 (2018).
- [20] Krishna.R., Bernstein.M., Fei-Fei.Li. Information maximizing visual question generation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages2008–2018 (2019).
- [21] Kingma.D.P., Welling.M. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013).
- [22] Hasan.S.A, Ling.Y., Farri.O., Liu.J., M ulle..Hr, Lungre.M.Pn .Overview of image clef 2018 medical domain visual question answering task. In Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, September 10-14, volume 2125 of CEUR Workshop Proceedings. CEUR-WS.org (2018).
- [23] Lau.J.J, Gayen.S., Ben Abacha.A., Demner-Fushman.D. A dataset of clinically generated visual questions and answers about radiology images. Scientific Data, 5(1) (2018).
- [24] Ben Abacha.A., Hasan.S.A., Datla.V.V., Liu.J., Demner-Fushman.D., M uller.H. Vqa-med: Overview of the medical visual question answering task at image clef 2019. InWorking Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019, volume 2380 of CEUR Workshop Proceedings. CEUR-WS.org (2019).
- [25] G ozde G l. Srk SteedmaM.n. Data augmentation via dependency tree morphing for low-resource languages. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics (2018).
- [26] Kobayashi.S. Contextual augmentation: Data augmentation by words with paradigmatic relations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2

- (Short Papers). Association for Computational Linguistics (2018).
- [27] Uehara.K., Tejero-de-Pablos.A., Ushiku.Y., Harada.T.: Visual question generation for class acquisition of unknown objects. In: Computer Vision ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, pp. 492–507 (2018).
 - [28] Zheng.Y., Yang.C., Merkulov.A.: Breast cancer screening using convolutional neural network and follow-up digital mammography. In: Computational Imaging III. DOI:10.1117/12.2304564 (2018).
 - [29] He.K., Zhang.X., Ren.S., and Sun.J., “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 9, pp. 1904–1916, (2015).
 - [30] Ahsean.M., Based.MD.A, Haider.J., Kowalski.M. COVID-19 Detection from Chest X-ray Images Using Feature Fusion and Deep Learning. DOI:10.3390/s21041480 (2021).
 - [31] Ben Abacha.A., Sarrouti.M., Demner-Fushman.D., A. Hasan.S., Müller.H., Overview of the VQA-Med Task at ImageCLEF 2021: Visual Question Answering and Generation in the Medical Domain, CLEF 2021 Working Notes, CEUR Workshop Proceedings, 2021,CEUR-WS.org September 21-24 Bucharest, Romania (2021).
 - [32] Ionescu.B., Müller.H., Peteri.R., Ben Abacha.A., Sarrouti.M., Demner-Fushman.D., A. Hasan.S., Kovalev.V., Kozlovski.S., Liauchuk.V., Dicente.Y., Pelka.O., Herrera.A.G.S., Jancutprakart.J., Friedrich.C.M., Berari.R., Tauteanu.A., Fichou.D., Brie.P., Dogariu.M., Stefan.L.D., Constantin.M.G., Chamberlain.J., Campello.A., Clark.A., Oliver.T.A., Moustahfid.H., Popescu.A., Deshayes-Chossart.J., Overview of the ImageCLEF 2021: Multimedia Retrieval in Medical, Nature, Internet and Social Media Applications, Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 12th International Conference of the CLEF Association (CLEF 2021), LNCS Lecture Notes in Computer Science, Springer, September 21-24 Bucharest, Romania (2021).