Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

May 2023, Volume 1, Issue 1

A New Model for Augmenting Retrieved Context-Based Information to Detected Objects in Augmented Reality

Maryam Moradi Shabestari [a] Code Orcid: 0000-0002-7118-3989,
Mojtaba Vahidi-Asl [b] ✉ Code Orcid: 0000-0003-4964-992X,
Monireh Abdoos [c] Code Orcid: 0000-0002-3106-503X
[a] Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran, Moradish@ce.sharif.edu
[b] Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran, mo_vahidi@sbu.ac.ir
[c] Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran, m_abdoos@sbu.ac.ir

ABSTRACT

Augmented reality technology integrates information with the environment in real-time. The capability to add information to objects and interact with users according to their requests can be an enjoyable and useful experience. Most Augmented Reality systems have a fixed static database containing registered objects. To enhance Augmented Reality applications, dynamic information sources can be used. An information source such as the web which has expanded and up-to-date data, could be an appropriate alternative. Using these information sources in real-time enables the use of Augmented Reality on a wider scale.

This paper introduces an Augmented Reality model based on online content is presented. In this research, case studies are human faces and context-based information is a virtual element, i.e., information about the target face. Using web information resources dynamizes the system in a variety of environments. Furthermore, the user interacts with the system through question-answering. Deep learning methods have been used for information retrieval and question-answering. Moreover, Answers to frequently asked questions that are confirmed by users are stored in the database for faster response time.

The experimental results reveal that users can interact with the Augmented Reality system through question-answering and augment the required information with a speed of 0.642 seconds and an F-score of 87.9% on the SQuAD dataset.

CC BY

KEYWORDS

Augmented Reality, Information Retrieval, Question-Answering, Object Recognition.

1. INTRODUCTION

Augmented Reality (AR) aims to improve the perception of reality. This technology allows for the real-time addition of valuable information to target objects. This technology has been noticed in many sciences and industries such as medicine, military, education, tourism, and entertainment. Initially, AR applications could only be implemented on systems with high processing power, but with the rise of smartphones, the spread of use among people, and the provision of complex communication infrastructure, AR applications have gained popularity. Increasing the processing power, extending the resources, and implementing deep neural networks in parallel on graphic processing units have improved the speed and accuracy of AR object recognition algorithms [1].

An essential step in an AR process is object recognition. The real environment and the virtual element must be coordinated with each other. Coordination of systems requires real-time execution. A typical AR database contains a limited number of static objects. After object recognition, the relevant information is retrieved from the database and augmented to the target object. As web resources continue to expand, including data can be used as a rich source of information. It is possible to dynamically add related information to the target object which improves the user experience with AR systems. Using questions to express the user's needs could be advantageous because the information is added to the object such that it could match the user's requirements.

In this research, using object detection, information retrieval, and question answering, a general model is presented to augment dynamic information in real-time to the environment. Based on our understanding of related works, this is the first time that an AR system with dynamic content generation based on user requirements at runtime, coupled with interaction through question-answering is suggested.

After recognizing the object, the user's requirement is entered into the system in the form of a question. The relevant information is retrieved by the information retrieval component. Then the answer to the question is extracted by the question-answering component. Finally, this information is augmented to the target object by AR technology.

This generalizable model has been tested on the SQuAD dataset and extracts the answer in 0.642 seconds with an 87.9% F-score. To further test the general applicability of this model, a set of questions about information on sports, media, and political

Submit Date: 2023-04-13
Revise Date: 2023-07-02
Accept Date: 2023-10-09
✉ Corresponding author

celebrities have also been collected. This test began with retrieving related documents from the web, and after ranking the documents, the answer to the question was extracted and user feedback was used to validate the result. As a result, the model can extract the answer in 2.4 seconds and the accuracy rate is 82.7%. The following presents several advantages of the proposed method versus customary AR:

- Providing an extensible model for various objects in the context of AR: The case study in this research is human face as a target object, but the presented model can be developed on various types of objects.
- Dynamic information in AR: The information, virtual elements, of the detected target object is dynamically retrieved from the web according to the user's needs. This information is extensive and up-to-date due to not being stored in a fixed database.
- User interaction with the AR through question answering: The user's need is entered into the system in the form of a question and the answer is extracted. This interaction between the system and the user leads to dynamic virtual elements.

The remainder of this paper is organized as follows. In section 2, basic concepts and related works are reviewed. In section 3, the proposed method and implementation of components are explained. In section 4, the components are evaluated. Finally, in section 5, conclusions and future works are presented.

2. BASIC CONCEPTS

In this section, a necessary background for a better understanding of the proposed model is provided.

2.1. literature survey

Following is the review of literature on augmented reality, face recognition, and question answering.

2.1.1. Augmented Reality

Adding valuable information to objects through augmented reality improves the user's perception of reality. Many industries and sciences have utilized this technology in their working environments. This process involves capturing an image with the camera and entering it into the system. Based on the settings of the virtual and real cameras, the camera position estimation module estimates the real camera's location and direction. In the end, the virtual element(s), e.g. text, audio, video, animated 3d or 2d models, which were registered in the database, is augmented at the estimated location or marker.

We have enriched the augmented reality database by automatically retrieving and processing data from web sources, aiming to enhance the user experience in a useful way. By incorporating online and real-time questioning capabilities, we enable users to interact with the augmented reality system and fulfill their diverse requirements for context-based information, including text, images, or videos.

There are two categories of augmented reality applications: image-based and location-based. There are two types of image-based applications, those that use markers and those without markers.

Many augmented reality programs were initially based on markers, but with the improvement of OS, software, and hardware power of smartphones, markerless methods have been extensively considered. Object detection is a crucial step in the augmented reality process. A fictitious feature or a natural feature is used to identify real objects in the environment. By detecting objects and tracking them continuously, the location of the camera is estimated. In classical object detection methods, fixed markers are added to the real environment by hand to make object detection easier. To extract local features and recognize natural features, markerless methods use features like boundaries, edges, and textures. Lighting variations, motion blur, and camera parameters can cause errors in natural point-based features. Typically, the classical methods of object detection do not properly work in AR. Instead, Artificial intelligence-based methods are suitable for AR. Raw images can be used and Various objects can be automatically learned using these methods [1].

2.1.2. Question-answering

Answering open-domain questions is an important task in natural language processing and answer a question in natural language format is based on large-scale unstructured documents. The answering system according to the type of information source can be divided into two parts: text question answering and knowledge base (KB-QA). Text question-answering extracts answers from unstructured textual documents.

KB-QA extracts answers from a pre-defined structured knowledge base that is often built manually. Textual question answering is generally more scaled than KB-QA since most of the structured text resources used to extract an answer, such as Wikipedia, news articles, and scientific books, are easily accessible. In this process, information is retrieved to get a list of relevant documents. Then, Questions are answered based on the top documents [2].

2.2. Related works

In the following, the related tasks in augmented reality, face recognition, and question-answering are reviewed.

2.2.1. Augmented Reality

Much research has been done in the field of augmented reality and its various applications. One of the AR applications is to improve the education process. In 2020, a group of researchers studied the application and impact of augmented reality in education. They examined the uptake of AR technology in educational settings, teachers' opinions about the need for continuous training, the process of creating 3D models, and the feasibility of implementing AR in schools and showed that AR can enhance teaching and learning motivation and effectiveness [3] [4]. Adopting and applying information at the right time and right place is important in education and training.

In 2020, Soltani et al. studied the benefits of AR in training and sports training by adding information to the players' reality [5]. Safety education is important in promoting a safe and healthy working environment in construction. In 2015,

a framework based on VR and AR was proposed to provide students with realistic and practical safety experiences and educate construction safety [6].

AR is one of the technological tools used for construction. In 1999, Reiners et al. described an Augmented Reality application for the task of door lock assembly into a car door that was developed trying to transport concepts to an observer [7]. Research in 2020, introduced architecture to augment information about the members of a university in the context of AR. A dataset containing 1200 facial images of 100 faculty members of the Shahrood University of Technology was used. The detection accuracy was 99.45%. This system contains fixed information that is limited to university faculty members. By recognizing people, information is retrieved from the database and augmented [8]. Another AR application proposed in 2016 is augmenting patient information such as vascular maps to facilitate surgery [9].

In 2021, Yang et al. presented a framework for designing websites that display more content and manage the main functions such as object detection, location detection, and interaction methods. Developers are responsible for placing information blocks in the AR world. With M2A, websites can display more content simultaneously, while making it easy for users to identify and isolate related content by exploiting the third dimension provided by the AR space. It simplifies user interaction and accelerates information access in this way. Users on this platform find information four times faster than they would on a normal smartphone and two times faster than they would on a standard web browser displayed in AR. Specific websites are used to retrieve this information [10]. In 2020, Gokhan introduced the ARgent Web-based Framework, which aims to facilitate the development of augmented reality systems with dynamic content. However, it is important to note that content selection occurs before system usage, as users select and pre-register the content from the web. This means that the content is not generated online in real-time based on the user's specific needs during system usage [11].

Face Recognition

Face recognition is one of the popular research fields in computer vision and pattern recognition. Some methods and databases have been proposed in recent years. Deep learning-based techniques are popular methods that use different deep neural networks. The DeepFace network, proposed in 2014, is a multi-stage approach. This network extracts face representation from a 9-layer deep neural network and has been trained on more than 4000 people. DeepFace was one of the first works that achieved very high accuracy in the LFW dataset using CNN [12]. In 2015, proposed two deep neural network architectures for face recognition called DeepID3 [13]. These two architectures consist of GoogLeNet and VGGNet. Joint surveillance signals of face identification verification during the training set are added to the final feature extraction and intermediate layers. The FaceNet network is a model from Google that was proposed in 2015. In this architecture, a 128-dimensional representation of deep complex networks is used. It is trained on 200 million face images using a triple loss function in the final layer [14].

2DPCANet is a deep-learning network for face recognition. It is employed to learn the filters of multistage layers. SVM and extreme learning machine (ELM) are used as the classifier (2018) [15].

In 2019, ArcFace function was proposed which added the margin to the cosine softmax loss for face recognition deep models based on the discriminative features [16]. Face recognition in low-quality face datasets is challenging. In 2022, AdaFace proposed a new loss function that emphasizes samples of different image quality. This method approximates the image quality with feature norms in the form of an adaptive margin function [17].

2.2.2. Question Answering

Many deep learning methods have been introduced to question answering. The Bidirectional Encoder Representation with Transformer (BERT) model was introduced as a language representation model in 2018. Bert is designed to pre-train deep bidirectional representations of unlabeled text by conditioning both left and right content across all layers. By adding an output layer to this model, it can be finetuned for a wide range of tasks, such as question-answering and language inference [18]. The BERT model ignores the dependence between the hidden positions and suffers from the challenges of pre-training and finetuning inconsistency.

In 2019, the XLNet model was proposed. A generalized auto-regression pre-training method that enables learning of two-way contexts by maximizing the expected likelihood over all permutations in sequence and overcomes the limitations of BERT [19].

In 2019, RoBERTa was introduced as a model to optimize and strengthen BERT with longer training, larger batches, and more and longer data. It also dynamically changes the masking pattern applied to the training data [20].

3. MAIN IDEA

Providing a general model for real-time augmentation of dynamic information in interactive AR is the main objective of this research. In this model, AR, object detection, information retrieval, and question-answering components are integrated, which is explained in detail below.

3.1. Implementation

AR can create an interactive experience between the user and the system, by retrieving dynamic virtual elements through questions and augmenting to real-world objects. Augmenting information about objects in the environment helps to better understand the user. This information varies based on the object type. People’s information is the case study that was tested in this research. This information can be required based on the person’s context in sports, politics, academics, media, music, etc. Information may contain data about time, place, job, honors, a field of expertise, and other items.

The architecture of this system according to Figure 1 can be generalized and extended to different objects.

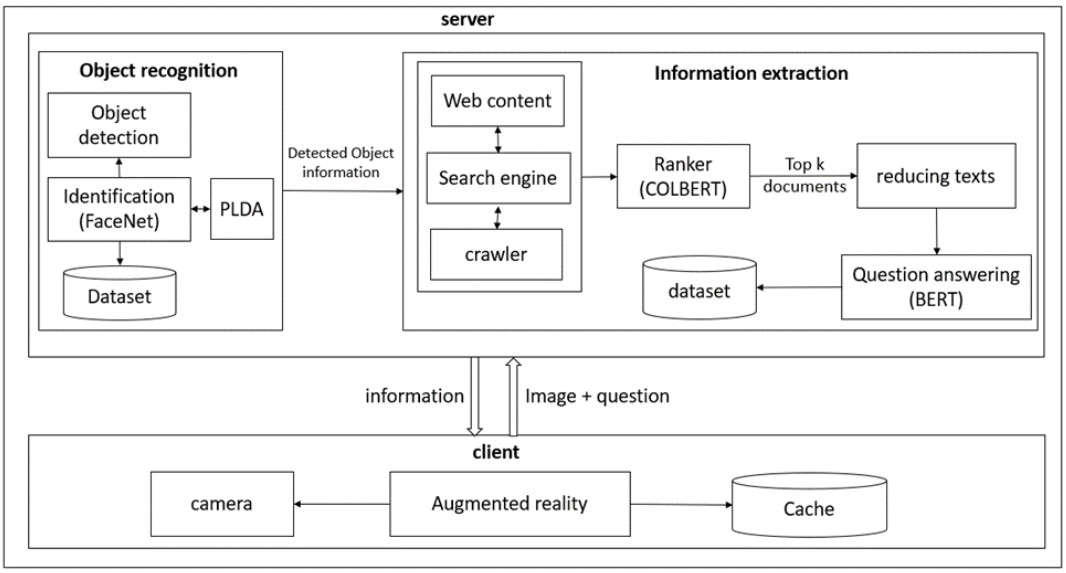

Figure 1. The proposed model.

Using the client-server architecture, this model can run on devices with limited processing power and memory, such as mobile phones.

The camera on the client side captures video and after analyzing the video frame on the server side, augments virtual elements to the target object. To accelerate the process, the information that the user has just accessed is stored in the internal temporary memory. Information extraction and retrieval are server-side processes.

The object recognition component includes the detector, the identification model, and a classifier. Objects are detected after capturing the image from the client's side. Next, the basic information about the object is sent to the information retrieval component along with the user's requirements (in the form of a question). This component extracts relevant resources from the vast information source of the web using search engines.

A crawler refines these sources, and the main parts of the document are separated and extracted in a suitable format. Based on the COLBERT model, these information sources are ranked. In the next step, unrelated sentences in the top *k* documents are removed to speed up the answering process. Answers are extracted from the top *k* documents and augmented by the client. Additionally, the server-side database is completed by receiving feedback and confirming answers from the client to increase the system's speed and accuracy. Data may change over time, so very old data is removed to replace updated information.

Figure 2 is an example of the proposed model result. First, the person's face is detected as the target object. His identity is confirmed as "Ali Dayi" in the face recognition step. The height of this person is requested as information required by the user. By knowing the name of the person, related documents are retrieved from the web and ranked. The answer to the question is extracted based on the top relevant documents. According to the answer, the height of Ali Daei is 1.92 meters. This data is augmented as a virtual element to the target object in the AR system.

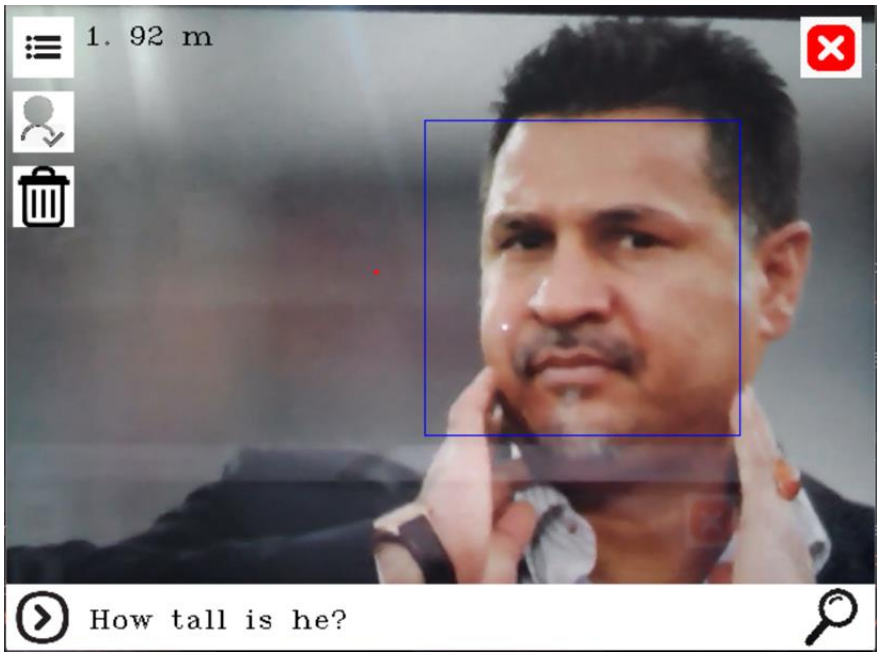

Figure 2. Sample output of the proposed model

3.2. Object Detection

In AR, augmenting virtual elements with the real environment is a major step. The detection of objects in real-time is important for system coherence. Human faces were studied in this research.

For facial recognition, a variety of methods and datasets have been proposed. In this research, FaceNet deep networks were used. FaceNet is a deep convolutional network that learns direct face embedding. The training of this network is end-to-end. The loss function for training the network is the Triplet loss function. The triplet function is given in Equation 1. $\propto$ is a margin between positive and negative pairs and f(x) is an embedding function. According to the model's purpose, the embedding vector of each person's face image should be close to the embedding vector of his other images and far from the embedding vector of other people's face images.

$$\sum_{i}^{N}\left[\left|\left|f(x_i^a)-f(x_i^p)\right|\right|_2^2-\left|\left|f(x_i^a)-f(x_i^n)\right|\right|_2^2+\propto\right] \quad \text{Equation 1}$$

Embedding vectors are calculated to describe the features of people's faces by FACENET [14]. Face recognition is equivalent to classifying these embedding vectors into classifies equal to the number of people in the collection. The classier is PLDA. PLDA is a generative probabilistic model that extracts features and uses its combination for identification [21]. So, the probability of assigning each face embedding to each person is calculated and the identity of the person is recognized.

3.3. Information Retrieval

Generally, open-domain question-answering systems require the retrieval of related documents. The COLBERT model is one method of retrieval. Scalable scoring is used in this model to determine the relationship between the question and the document. To calculate the similarity of questions and documents, the question *q* and the input document *d* are first coded separately into embedding vectors. The output dimensions are controlled by a linear layer after embedding vectors are generated. Then, each vector is converted into a unit float vector E_d and E_q. The final score $S_{q,d}$, is calculated based on the maximum similarity of the question embedding vector *Eqi* with all the document embedding vectors *Edj*, according to Equation 2.

$$S_{q,d}=\sum_{i=1}^{N}\max_{j=1}(E_{qi}.E_{dj}^{T}) \quad \text{Equation 2}$$

This model can be applied to billions of tokens and millions of documents. It can directly find the answer to the question from a large volume of documents. In this research, with the help of the COLBERT model, documents related to the question have been retrieved from a large volume of documents taken from the web [22].

3.4. Question-Answering

Accomplishing an automatic Question-answering process from unstructured databases is a challenging task in natural language processing. By using question-answering models, the required information about a question is extracted. The answer to the question is extracted after selecting *k*-related documents. The BERT is a language model. it is a multi-layer bidirectional encoder transformer. This model is designed to pre-train a deep two-dimensional representation of an unlabeled document.

The architecture of this model includes a transformer with only an encoder part. The encoder takes a sequence of *N* tokens as input and produces a representation vector for each token, and the output is a vector of the representation of each token. In this research, the implementation of the question-answering component is with the Bert model. This model is finetuned on the SQUAD dataset.

The model outputs the probability that the *i*-th word in the document is the beginning of the answer and the probability that it is the end word of the answer. The probability that each sequence of words in the document is the answer to the question is calculated based on the multiplication of the probability of the beginning words x_i and the ending words x_j according to equation 3. Based on these probabilities, *k*'s best answers are considered as output.

$$\text{prob\_answer}(x_i,x_j)=\text{prob\_start}(x_i)*\text{prob\_end}(x_j) \quad \text{Equation 3}$$

3.5. Document Reduction

In AR systems, it is important to have real-time augmenting objects. The coherence of the system is influenced by the speed of information retrieval and the question-answering model. To improve the speed of answering, it is recommended to reduce the volume of the documents by removing unrelated sentences.

Detecting the relevance of each sentence to the question is important in the correct final answer. The degree of relevance between the question and the answer is calculated with the output of the middle layer of the COLBERT model.

The question embedding vector *q*, E_q, and the representation vector of each document sentence *d*, E_d, are calculated. Then, the sum of the maximum score of the words for each word of the question for each sentence is obtained according to equation 4.

$$S_{q,d_i}=\sum_{j=1}^{N}\max_{k=1}(E_{q_j}.E_{di\_k}^{T}) \quad \text{Equation 4}$$

In this way, the importance of each sentence in the document is obtained. According to the type of questions in the dataset, by removing the sentences whose score is lower than the median of the total scores, the size of the resulting document becomes shorter. In this way, by reducing the volume of the document, the question-answering model predicts the answer faster.

4. EXPERIMENTAL RESULTS

The AR platform augments retrieved information with face recognition, information retrieval, and question-answering. Virtual objects must be augmented in real time by AR systems. Therefore, in the first step, it is necessary to evaluate the response component's speed. After that, the question-answering component is evaluated.

4.1. Experimental setup

The experiments are designed to examine the proposed model. the proposed model is designed based on the client-server architecture which can be used on mobile devices. The test of this model is implemented with the help of Python language in a Windows operating system with 16 GB RAM and Core i7 processor.

4.2. Face Recognition

Face recognition is an important component in the system due to retrieving the basic information about the object. It is evaluated based on the accuracy criterion.

4.2.1. Dataset

The face recognition component is evaluated on the LFW dataset. This dataset contains 13,233 images of 5,749 celebrities. 1680 people with at least two images in the dataset used for identification. The images of this dataset are varied in position, lightness, focus, facial expressions, age, gender, ethnicity, makeup, background, and quality [23].

4.2.2. Result of face recognition component

The parts of the LWF dataset that include people with at least 2 images have been selected. The data is augmented with the help of changing contrast, light, and flip methods and is divided into two parts, training, and testing, with a ratio of 80 to 20. The facial recognition component has achieved 92% accuracy. The calculated accuracy and speed are acceptable for the proposed AR system.

4.3. Question-Answering

4.3.1. Dataset

- SQuAD : Stanford Question Answering Large Dataset, is a reading comprehension dataset with over 100,000 questions asked from 500 Wikipedia articles. The answer to questions is part of the document. This dataset is divided into 80% for training, 10% for evaluation, and 10% for testing [24]. The answers are in the types of date, number, person, place, other entities, noun phrases, and present and partial sentences. The answers are more complex than other datasets and require more reasoning. Hence, it is suitable for evaluating the understanding of the model and its ability.
- Collected dataset: This study proposes a practical system that has to be evaluated in the real world according to the proposal of a practical system. The model has been evaluated using a new dataset, which includes questions and answers about famous people from sports, politics, music, academia, and cinema. This dataset contains 500 questions. The documents retrieving relevant information are retrieved from the web. It has tried to retrieve reliable sources in each field and not just rely on the Google search engine. Based on each person's context, various questions are randomly assigned to the documents. The answers are various types of location, time, number, noun phrases, and present phrases. Some of these answers are complex and require further reasoning.

  The collected dataset has been evaluated based on the user's feedback and his satisfaction with the answers and calculated according to Equation 5.

$$\text{accuracy} = \frac{\text{correct answers}}{\text{correct answers} + \text{incorrect answers}} \qquad \text{Equation 5}$$

4.3.2. Result of Question-Answering component

The results of the question-answering model evaluation on the SQuAD dataset are given in Table 1.

Table 1. The question-answering result on SQuAD dataset.

model	Time (s)	EM (%)	F1 (%)
BERT	0.956	83.6	91.7
BERT + Document reduction	0.642	79.8	87.9

The BERT model has had significant results in the task of question-answering; But according to the need for AR systems for real-time performance, this research tries to increase the speed of information retrieval. By reducing the document based on the importance of the sentence, the answering speed has improved by 1.48 times. This method has used the results of the intermediate layers of the COLBERT model in the document ranking step, and no additional pre-processing overhead has been added to the system.

The evaluation results on the collected dataset are given in Table 2. In addition to the top answer, the result of the test on the top 5 answers is also reported. According to the results, 10% of the wrong answers in the next top answers are correctly extracted.

Table 2. The question-answering result on the collected dataset.

model	Time (s)	Accuracy (%)	Accuracy (%) Top k answers
BERT	4.29	78.8	92.7
BERT + Document reduction	2.4	82.7	94.2

According to the results, the speed in the collected dataset has improved 1.78 times and the accuracy criterion has remained almost the same.

4.4. Discussion

The proposed model can be used in different frameworks for arbitrary objects with the proposed model and have operational applications in the real environment and be effective in improving the understanding of reality.

Validation of the collected data is dependent on the Internet. The related documents are retrieved from the web. The speed of the internet is effective. Also, the correctness of the answer is considered based on the expert's feedback and human error may be involved. In general, the conducted experiments show the proper performance of this model in the operational environment.

The advantages of the proposed model include the following:

- Providing a model for various objects in the context of AR.
- Dynamic information in AR system.
- User interaction with the AR system through question-answering.

The proposed model has many advantages and applications; But it also has disadvantages, which include the following:

- The inability to answer complex questions.
- The effect of the Internet on the speed of retrieving documents from the web
- Impossibility of face recognition in some conditions such as low light or face position

5. CONCLUSION

In this paper, AR real-time information augmentation is presented using a general model. Unlike previous AR systems that used static databases, this research uses dynamic information sources and creates user interaction through question-answering. In the AR system, three components were presented: object detection, context-dependent information retrieval, and question-answering. Due to the importance of being real-time, the appropriate speed of extracting information is one of the other goals of this research. Thus, document reduction based on the importance of sentences has been recommended. It has been suggested to reduce documents based on the importance of sentences.

In the future, by adding the speech-to-text conversion component, the user's interaction with the system can be made easier. The user's needs can be entered into the system through voice. Also, by providing the possibility of automatic image search on the web, this system could be more dynamic and may be closer to the operational environment.

6. REFERENCES

[1] Craig, Alan B. "Understanding augmented reality: Concepts and applications." Newnes (2013).

[2] Allam, Ali Mohamed Nabil, and Mohamed Hassan Haggag. "The question answering systems: A survey." International Journal of Research and Reviews in Information Sciences (IJRRIS) 2, no. 3 (2012).

[3] Iatsyshyn, Anna V., Valeriia O. Kovach, Yevhen O. Romanenko, Iryna I. Deinega, Andrii V. Iatsyshyn, Oleksandr O. Popov, Yulii G. Kutsan, Volodymyr O. Artemchuk, Oleksandr Yu Burov, and Svitlana H. Lytvynova. "Application of augmented reality technologies for preparation of specialists of new technological era." CEUR-WS (2020).

[4] Tzima, Stavroula, Georgios Styliaras, and Athanasios Bassounas. "Augmented reality applications in education: Teachers point of view." Education Sciences 9, no. 2 (2019): 99.

[5] Soltani, Pooya, and Antoine HP Morice. "Augmented reality tools for sports education and training." Computers & Education 155 (2020): 103923.

[6] Le, Quang Tuan, A. K. E. E. M. Pedro, Chung Rok Lim, Hee Taek Park, Chan Sik Park, and Hong Ki Kim. "A framework for using mobile based virtual reality and augmented reality for experiential construction safety education." International Journal of Engineering Education 31, no. 3 (2015): 713-725.

[7] Reiners, Dirk, Didier Stricker, Gudrun Klinker, and Stefan Müller. "Augmented reality for construction tasks: Doorlock assembly." In International workshop on Augmented Reality: Placing Artificial Objects in Real Scenes, pp. 31-46. 1999.

[8] Golnari, Amin, Hossein Khosravi, and Saeid Sanei. "Deepfacear: deep face recognition and displaying personal information via augmented reality." In 2020 International Conference on Machine Vision and Image Processing (MVIP), pp. 1-7. IEEE, 2020.

[9] Khor, Wee Sim, Benjamin Baker, Kavit Amin, Adrian Chan, Ketan Patel, and Jason Wong. "Augmented and virtual reality in surgery—the digital surgical environment: applications, limitations and legal pitfalls." Annals of translational medicine 4, no. 23 (2016).

[10] Lam, Kit Yung, Lik Hang Lee, Tristan Braud, and Pan Hui. "M2a: A framework for visualizing information from mobile web to mobile augmented reality." In 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom, pp. 1-10. IEEE, 2019.
[11] Gökhan, K. U. R. T., and İ. N. C. E. Gökhan. "ARgent: A web based augmented reality framework for dynamic content generation." Avrupa Bilim ve Teknoloji Dergisi (2020): 244-257.
[12] Taigman, Yaniv, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. "Deepface: Closing the gap to human-level performance in face verification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1701-1708. 2014.
[13] Sun, Yi, Ding Liang, Xiaogang Wang, and Xiaoou Tang. "Deepid3: Face recognition with very deep neural networks." arXiv preprint arXiv:1502.00873 (2015).
[14] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823. 2015.
[15] Yu, Dan, and Xiao-Jun Wu. "2DPCANet: a deep leaning network for face recognition." Multimedia Tools and Applications 77 (2018): 12919-12934.
[16] Deng, Jiankang, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. "Arcface: Additive angular margin loss for deep face recognition." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4690-4699. 2019.
[17] Kim, Minchul, Anil K. Jain, and Xiaoming Liu. "Adaface: Quality adaptive margin for face recognition." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18750-18759. 2022.
[18] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
[19] Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R. Salakhutdinov, and Quoc V. Le. "Xlnet: Generalized autoregressive pretraining for language understanding." Advances in neural information processing systems 32 (2019).
[20Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. "Roberta: A robustly optimized bert pretraining approach." arXiv preprint arXiv:1907.11692 (2019).
[21] Ioffe, Sergey. "Probabilistic linear discriminant analysis." In Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part IV 9, pp. 531-542. Springer Berlin Heidelberg, 2006.
[22] Khattab, Omar, and Matei Zaharia. "Colbert: Efficient and effective passage search via contextualized late interaction over bert." In Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, pp. 39-48. 2020.
[23] http://vis-www.cs.umass.edu/lfw. Accessed March 2022.
[24] Rajpurkar, Pranav, Jian Zhang, Konstantin Lopyrev, and Percy Liang. "Squad: 100,000+ questions for machine comprehension of text." arXiv preprint arXiv:1606.05250 (2016).

Maryam Moradi Shabestari
Faculty of Computer Science and Engineering, Shahid Beheshti University,
Tehran, Iran
Moradish@ce.sharif.edu

Mojtaba Vahidi-Asl
Faculty of Computer Science and Engineering, Shahid Beheshti University,
Tehran, Iran mo_vahidi@sbu.ac.ir

Monireh Abdoos
Faculty of Computer Science and Engineering, Shahid Beheshti University,
Tehran, Iran m_abdoos@sbu.ac.ir