

Smart Tour Guide in Makkah City based on Image Recognition

Amany Al Luhaybi¹, Salha Al Shaeri¹, Aisha Al Shaeria¹, Alyah Al Harbi¹, Bashayir Al Harbi¹, Khadija Mohammed¹ and Safaa Alraddadi²

¹ College of Computing in Al-Qunfudhah, Umm Al Qura University, Makkah 24382, Saudi Arabia

² College of Computer Science and Information Technology, Al-Baha University, Al-Baha 1988, Saudi Arabia

Abstract

Technology nowadays affects every aspect of our life and makes it easier, more convenient. This research aims to integrate Artificial intelligence (AI) in the development of a smart mobile application that works as a smart tour guide. An Android application was developed to do image recognition in almost real-time by using Convolutional Neural Network (CNN). After the image is recognized, reviews for the restaurant or the hotel are retrieved. The reviews were collected from various resources on the internet and saved in a database to be retrieved after the image is recognized. This research required collecting images to retrain a CNN model called Inception v3 from Google. A technique called transfer of learning was used, which is building CNN model that is trained on a small dataset. The data size is about 100 images for each of the 9 categories of the hotels and restaurants. The scope for this application is limited to a few numbers of hotels and restaurants in Makkah city to deliver to the visitors of this Holy city a pleasant experience while they are roaming around. They could take a photo or upload a photo of the hotel or restaurant and a list of reviews are shown to help them decide to go to the best places and save their time and effort rather than searching for this online. The application is available without an internet connection to make it convenient to use by users. The image recognition and the retrieval of the reviews from the database are successfully accomplished. The accuracy varies between 0.84 to 1 on different testing images.

Keywords

Artificial intelligence, Convolutional Neural Network, Android Application, Image Recognition.

1. Introduction

Deep learning is a subfield of machine learning, it tries to mimic the human brain when it receives a piece of information, and how this information is transferred from one neuron to the other one till they form a network. Deep learning depends on having what is known as a neural network and can be implemented in many areas. For instance, to do image recognition and sound recognition [1]. The main thing about these algorithms is how they are combined of multiple layers and these layers contain neurons, and that's how the term deep was generated [2,3]. We can have many architectures based on the decision of how many layers exist the programmer wants. This research is based on using one of the types of neural networks known as Convolutional Neural Network (CNN). There are many applications of using image recognition such as tagging images and searching for image content. One of the state-of-art techniques to do image recognition is CNN [4]. Image recognition is normal and effortless for humans, but it is a very difficult task to do with computers in which images are a matrix of numbers called pixels. Using CNN to do image recognition requires training. While training CNN requires to have a large dataset and a computer with high resources, the training could last for more than a week. Therefore, an approach called Transfer of Learning was used in research [5,6,7,8]. This approach depends on using CNN model that is already being used to solve some problem, and trained on a large amount of dataset, and retrain it to solve a different problem with a custom small dataset.

ACI'22: Workshop on Advances in Computation Intelligence, its Concepts & Applications at ISIC 2022, May 17-19, Savannah, United States
EMAIL: amluhaybi@uqu.edu.sa (A. 1)



© 2020 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)

This approach gave a high accuracy result in the results shown in [5,6,7,8]. For the methodology, this approach was followed as well. Inception v3 from Google is retrained to achieve the desired goal. A common problem any human can face is how can decide to go to some of the best places like hotels or restaurants. It is common to collect reviews of these places from those who already went to them, or search in Google and hopefully, it will show some reviews.

However, both ways are time-consuming. Technology nowadays affects every aspect of our life and makes it easier, more convenient, and efficient to do daily life's tasks. Images of data was collected along with the reviews of some places. Moreover, this task will be done in real-time and without an internet connection by applying image recognition using CNN and the technique of transfer of learning. After the training is over, this model can be deployed on an Android application. The user can simply take a picture or upload a photo to this application, and it will be able to provide the reviews. This research is highly beneficial to the visitors of Makkah city. It is difficult to find hotels and restaurants for visitors to Makkah, where they do not know the city and suffer from searching for a suitable place to stay in or to eat. As the visitors explore the city, they can take photos of the restaurant and decide to eat there or not based on the reviews that this application will provide, therefore, comfortable and enjoyable experience in Makkah without time and effort consuming was provided to them

1.2. Purpose of the Research

The objective of this research is to create an Android application that can do image recognition for the visitors to Makkah city, which makes it easier for users to find good places where the user uploads or takes a photo. For the authentication of the recognized image, it will provide the name it recognized just to show how accurate the application was to recognize that specific image and display the reviews review that people wrote about it. This research will contribute to the AI community by showing it is possible to build a smart system with limited dataset and computing resources.

2. Related Work

The application that was developed in [9] reads the square and barcodes, only a simple scan to already existing image, therefore, does not require collecting any personal information, it is one of the highly rated applications that has been used for this purpose. However, this application depends only on the photo that has a barcode only to scan.

The application in [10] allows you to reverse search images from the gallery or other media applications. It will display images with a general information about that image based on the search result. To the end, the app displays information about the images we supplied but does not show any reviews for example if the user searches for a particular product, this application performs a similar task like Google image search. To test this app, we search for some buildings and showed text with "Building" only. The same thing when we search for an unknown flag, and instead of providing us with information about it, only showed a "logo".

The application mentioned in [11] can search for a product through the image browser or camera. Using the image browser, the information and images appear correctly like the product that we searched about. The application in [12] aims to help Muslims easily to identify Halal products by allowing them to scan the barcode of various products. The application automatically highlights the food that is forbidden to Muslims based on the ingredients.

The application in [13] is mainly developed for people who are interested in fashion. The aim of this application is to identify clothing and accessories. The application recognizes the image by taking screenshot or video and display products that are in the online store like image that was taken. The application in [14] provides search by photo either upload it from the camera or the gallery. Then retrieve information about it from Google, Yandex.

The application in [15] is an image recognition based on mobile application that uses visual search technology to identify objects through a mobile device's camera. Users can take a photo of a physical object, then the information about the image retrieved from Google.

The application in [16] where the user can search for a product through the image browser or camera. The Search using image browser, take the user to the Google search engine where the search results appear there.

The application in [17] depend on some engines such as Google, TinEye, and Yandex. It allows searching through part of the image, searching by uploading pictures or taking a photo with a camera. Furthermore, it provides editing image choice; it is also used for searching for similar images and discovering the quality of them in terms of origin or falsification.

The application in [18] helps the traveler save effort and benefit from the trip. It allows publishing images by taking a picture or uploading it from the Internet. The application saves information, comments, and displays them on the map. Therefore, anyone can access the information to compare the prices and quality of hotels or restaurants.

The site [19] is a web tool used to search for images by uploading the image from folders, image libraries, or from the camera. After loading the image, the tool sends it to Google's image engine and starts searching for a match of this image.

The applications provided depend on searching images online using for example Google image search. All the represented applications require to be connected to the internet otherwise the application will be useless. Also, these applications' main feature is to search for similar images, no application according to our knowledge, was able to retrieve reviews about the image that the user search for by using image recognition technique.

3. Methodology

Image recognition is easy for humans and animals, but it is a very difficult task for computers. The task of image recognition can be done by using deep learning. CNN has a great ability to recognize the input images in this application we used already trained CNN model called Inception v3, then it is retrained to do the recognition on custom images. Python programming language was used to retrain the CNN model. The dataset was manually collected. It was divided into training dataset and testing dataset. Once the training is over, the model is smart enough to do image recognition. After that, the model was integrated into an Android application, in which users can take or upload photos. The application is using this retrained CNN model to recognize the input images from the users and retrieved the reviews from the database, the system overview is illustrated in Figure 1.

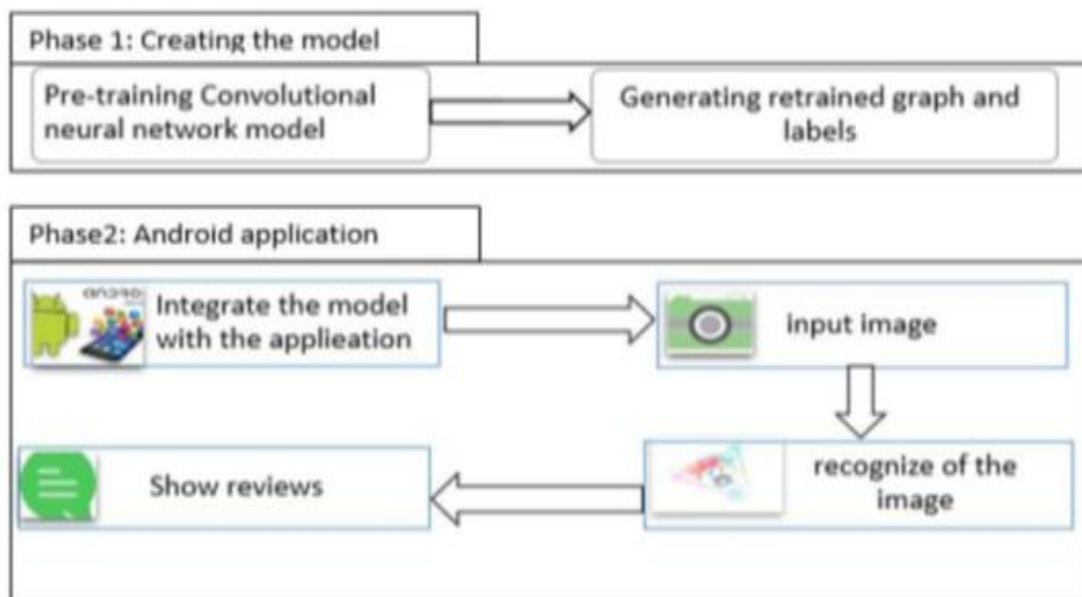


Figure 1: System Overview

3.1. Preparing the dataset

A dataset of four hotels and five restaurants was manually collected. In each category, there are about 100 images as illustrated in Table 1 and Table 2. To increase the data size and improve the performance of the model data augmentation was used, by feeding the model with images from different angles and at day or night as shown in Figure 2 a sample of the datasets. The images have been resized to 244 X 244. There are 100 images for training and 20 unique images for testing.

Table 1

The details of the dataset collection of hotels

The names of hotel	Number of images	Source of the images	Source of the comments
Shazamakkah	100	Google Maps; Twitter; Instagram	TripAdvisor
Daraltawhid	102	Google Maps	Google Maps
Elaf Almashaer	157	Google Maps	Google Maps
HolidayInn	100	Google Maps	Google Maps

Table 2

The details of the dataset collection of restaurants

The names of hotel	Number of images	Source of the images	Source of the comments
Afandim	102	Google Maps	Google Maps
Seeneez	100	Google Maps	Google Maps
Kabsa Hashi	117	Google Maps	Google Maps
Dartajindian	101	Google Maps	Google Maps; Twitter
Amo Hamza	100	Google Maps; Twitter; Instagram	Google Maps



Figure 2: (a) Dartajindian Restaurant; (b) Dartajindian Restaurant from different angle; (c) Elafalmsaer Hotel; (d) Elafalmsaer Hotel from different angle.

3.2. CNN Model Retraining

Building and training the CNN model from Scratch requires a huge amount of dataset to learn the important features. However, because of the small data size we have collected and limited computing resources, we used the transfer learning technique to do the image recognition task by using Inception v3[20], a model that was trained on ImageNet dataset with 1.2 million data size with more than 1000 labels. The knowledge is transferred from this model into the retrained model to recognize images for the new task which is in our case recognizing hotels or restaurants images. Recently, many CNN models have been developed. In this research, we used Inception v3 provided by Google as a pretrained model. This model has different sizes of filters in the same layer that significantly able to reduce computation complexity. The sizes of filters are 1×1 , 3×3 , and 5×5 one of these filters had to be chosen first then sent to the SoftMax layer to detect new features. This process will repeat until all features are extracted. The feature was extracted using both CNN fully connected layer and SoftMax layer to classify images. After that, new fully connected layers are added to extract the input features of the hotels and restaurants images. The top layer SoftMax that classified the original dataset was removed to train the next layer with our task. To enhance the performance of the model, 3 Dense layers were added. Two layers with Rectified Linear Unit (RELU) as an activation function and the output layer with Softmax, and categorical cross entropy loss function was used. This process lasted for almost two hours on a device with Windows 10 64-bit and Intel (R) Core (TM) i5 CPU.

3.3. Database Implementation

The database was implemented by using SQLite in Android Studio. This database is used to save the reviews. It has only one table in the database, the review table. The ID was set for each restaurant and hotel along with its name. After the recognition of images, the label was sent to the database to retrieve the corresponding reviews in the application.

4. Results and Discussion

When opening the application, it shows us an interface as shown in Figure 3. (a) through which the user can take or upload an image and when clicking on taking pictures, the image will appear in the main interface as shown in Figure 3. (b) after successfully taking the image and then clicking on reviews to do the image recognition and then show the reviews. When clicking on Upload Image, the gallery will open as shown in Figure 3. (c) after the image has been successfully uploaded, if the user clicks on Reviews, it will display the name of the recognized image and reviews as shown in Figure 3. (d). The results show that the application can recognize the image and retrieve the comments successfully with any angle the photos was taken.

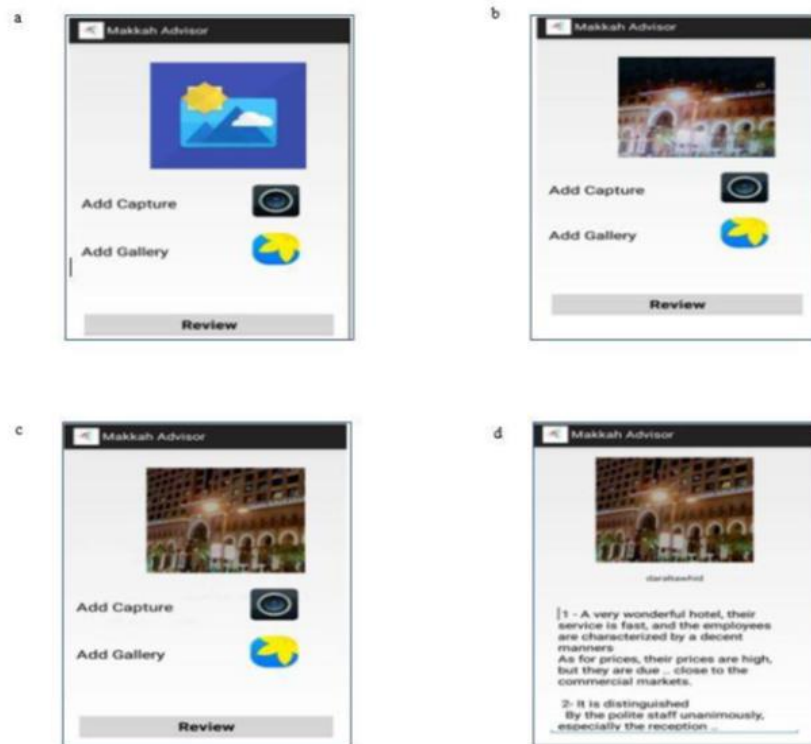


Figure 2: (a) Main User interface (b)Interface After clicking adding capture; (c) Interface After clicking adding Gallery; (d) Interface After clicking Review.

Many decisions regarding the number of layers were made to improve the performance. Moreover, a key factor to build an accurate model is the dataset. The decision of choosing the appropriate dataset was based on trial-and-error approach. Firstly, the model was trained on 50 images of each category, but overfitting problem was noticed as a result of the limited data. The problem is mainly happened because the model has learned the same patterns during the training, but it is not able to generalize the features already learned during the training phase on testing data. Therefore, the number of images were increased up to 100 for each one. As a result, the model recognized the data for testing successfully to our satisfaction.

Figure 4 demonstrate the integrating of the model with the android application. After the training is completed, two files are generated, one is the retrained model and the other one is the 9 classes or categories we have. The application used these files to do the image recognition in almost real-time.

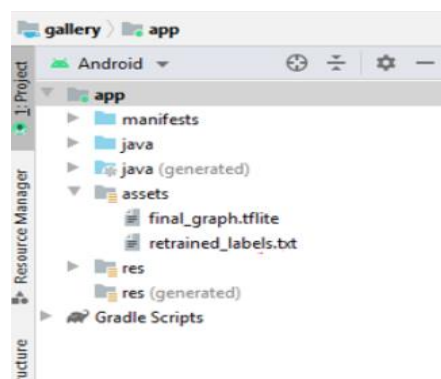


Figure 3: Integration model with the application

As it is illustrated in Figure 5 the loss or the error was about 24 at the beginning of the training and the error is reduced as the training continued until it reached to 0.03 by the end of epoch. Although the epoch was set to only 10, the accuracy got increased from 0.6 at the beginning of the training session to 0.99, the results are a proof of the effectiveness of using transfer of learning technique.

```
Epoch 1/10
10/10 [=====] - 128s 13s/step - loss: 24.1566 - accuracy: 0.6094
Epoch 2/10
10/10 [=====] - 131s 13s/step - loss: 1.6799 - accuracy: 0.8594
Epoch 3/10
10/10 [=====] - 138s 14s/step - loss: 0.4542 - accuracy: 0.9406
Epoch 4/10
10/10 [=====] - 151s 15s/step - loss: 0.1658 - accuracy: 0.9781
Epoch 5/10
10/10 [=====] - 144s 14s/step - loss: 0.1480 - accuracy: 0.9781
Epoch 6/10
10/10 [=====] - 137s 14s/step - loss: 0.1065 - accuracy: 0.9906
Epoch 7/10
10/10 [=====] - 144s 14s/step - loss: 2.9019e-05 - accuracy: 1.0000
Epoch 8/10
10/10 [=====] - 130s 13s/step - loss: 0.0658 - accuracy: 0.9937
Epoch 9/10
10/10 [=====] - 129s 13s/step - loss: 0.0326 - accuracy: 0.9844
Epoch 10/10
10/10 [=====] - 142s 14s/step - loss: 0.0383 - accuracy: 0.9906
```

Figure 4: Training results

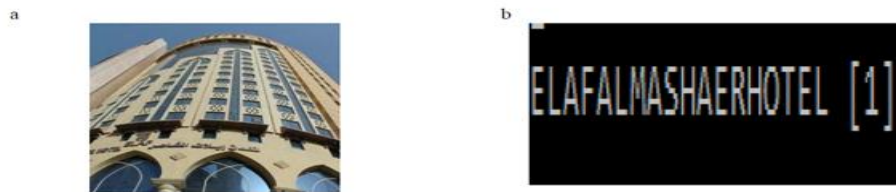


Figure 5: (a) The new image to be pass to the model; (b) Testing result

The figure above represented the result we got after passing to the retrained model a new test image it did not train on for a hotel called Ellaf Al Mashaer Hotel, the model perfectly identified it with 1 accuracy. In general, the accuracy varies between 0.84 to 1 on different testing images. While these results are satisfactory, by increasing the dataset size and adding more layers or different activation function the results could be enhanced. Moreover, the inception v3 model is an old model a better approach to increase the accuracy of our model is by using the-state-of-art CNN models.

5. Conclusion and Future Work

An Android application that works as a smart tour guide in Makkah city was developed by using a technique known as transfer of learning. This technique works well if we with limited datasets and computing resources. Inception v3 convolutional neural network model was retrained on our custom dataset. The trained model was converted from TensorFlow to TensorFlow Lite to be integrated within the Android application. SQLite database that contains all the reviews of some of the restaurants and hotels is used as well. All the functionalities were successfully implemented, and the image recognition was done in almost real-time.

As future work, an enhancement of the user interface of the application could be done, such as the way the reviews are displayed to the users. Another enhancement will be to add more hotels and restaurants that cover the important areas. Furthermore, sentiment analysis could be used in the reviews to show short comments if the hotels or the restaurants are good or bad. Finally, to improve the results a state-of-art CNN models are highly recommended to be used.

6. References

- [1] Wang Y, Sun Y, Liu Z, Sarma SE, Bronstein MM, Solomon JM. Dynamic Graph CNN for Learning on Point Clouds. *ACM Transactions on Graphics* 2019; 38:1-12, doi: 10.1145/3326362
- [2] Chollet F. *Deep Learning with Python*. 1st ed. United States: Manning; 2017.
- [3] SuperDataScienc Team. *The Ultimate Guide to Convolutional Neural Networks (CNN)*. SuperDataScience 2018.
Available online: <https://www.superdatascience.com/blogs/the-ultimate-guide-to-convolutional-neural-networkscnn>
- [4] Sorokina Ksenia. Image Classification with Convolutional Neural Networks. Medium 2017; Available online: <https://medium.com/@ksusorokina/image-classification-with-convolutional-neural-networks-496815db12a8>
- [5] Gatys Leon A, Ecker Alexander S, Bethge Matthias. Image Style Transfer Using Convolutional Neural Networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016; 2414-2423, doi: 10.1109/CVPR.2016.265.
- [6] Esteva A, Kuprel B, Novoa R, Ko J, Swetter S, Blau HM, Thurn S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; 542:115–118. <https://doi.org/10.1038/nature21056>
- [7] Al Luhaybi A, Alqurashi F, Tsaramirsis G, Buhari SM. Automatic Association of Scents Based on Visual Content. *Applied Sciences*. 2019; 9(8):1697. <https://doi.org/10.3390/app9081697>
- [8] Alraddadi S, Alqurashi F, Tsaramirsis G, Al Luhaybi A, M. Buhari S. Aroma Release of Olfactory Displays Based on Audio-Visual Content. *Applied Sciences*. 2019; 9(22):4866. <https://doi.org/10.3390/app9224866>
- [9] Find Apps & Games by Category - AppGrooves: Save Money on Android & iPhone Apps. AppGrooves. <https://appgrooves.com/rank/tools/qrcode-scanner-and-creator/best-apps-for-scanning-and-creating-qr-codes>. Published 2022.
- [10] Find Apps & Games by Category - AppGrooves: Save Money on Android & iPhone Apps. AppGrooves. <https://appgrooves.com/rank/tools/qrcode-scanner-and-creator/best-apps-for-scanning-and-creating-qr-codes>. Published 2022.
- [11] Qian Y. Reverse Image Search App. App Store. <https://apps.apple.com/sa/app/reverse-image-search-app/id1003144513>.
- [12] Scan Halal. App Store. <https://apps.apple.com/sa/app/scanhalal/id589534185>. Published 2022.
- [13] Mrukwa. (2018). Top image Recognition. Available online: <https://www.netguru.com/blog/11-top-image-recognition-apps-to-watch-in-2019>
- [14] App Grooves corporation. (2011). Available online: <https://appgrooves.com/search?q=Photo%20Sherlock>
- [15] Google Goggles - Wikipedia. https://en.m.wikipedia.org/wiki/Google_Goggles. Published 2022.
- [16] Hwang H. Search by Image! App Store. <https://apps.apple.com/sa/app/search-by-image/id1099395468>. Published 2022.
- [17] Dettmers T. The Best GPUs for Deep Learning in 2020 — An In-depth Analysis. Tim Dettmers. <https://timdettmers.com/2019/04/03/which-gpu-for-deep-learning>. Published 2020.
- [18] Yan T. Reverse Image Search Tool. App Store. <https://apps.apple.com/us/app/reverse-imagesearch-tool/id1375868438>. Published 2018.
- [19] AppGrooves Corporation. (2011). Retrieved 10 22, 2019, from App Store: <https://appgrooves.com/app/tripadvisor-hotels-restaurants-by-tripadvisor/negative> [20] Szegedy C, Vanhoucke V, Ioffe, S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2016; 2818–2826.