Libyan Landmark Recognition Using Deep Learning

Abdelhamid Elwaer [1*], Aisha Fayad [2] , Abdunnaser Diaf [3]

[1] elwaer@yahoo.com, [2] aishafayad.24@gmail.com, [3] aadiaf@*gmail.com*

[1] Department of Software Engineering, Faculty of Information Technology, Tripoli University, Libya
[2, 3] Department of Internet technology, Faculty of Information Technology, Tripoli University, Libya
*Corresponding author email:

ABSTRACT

With the increasing of computation power and amount of data, the recent years has seen tremendous advance in artificial intelligent techniques especially in the field of deep learning. Deep learning is revolutionizing several research fields like computer vision, autonomous driving, natural language processing, and speech recognition. In the area of computer vision such as image recognition a class of biologically inspired vision model called convolutional neural network outperformed human performance. To exploit this advance in deep learning, this paper introduce a trained model to identify Libyan landmarks such as Sabratha Theater, Alnaser arch, etc., in digital photos. The learning approach was based on convolutional neural network for image classification. A deep convolutional neural network was built using Keras and TensorFlow framework and trained on labeled dataset of thousand images from a variety of Libyan landmarks. This dataset used to train the network was collected from the internet and using different data augmentation techniques. The model obtained shows a classification accuracy of 99.0 %. This model can be used to develop various applications such as a tourist mobile app or web application, where users can take a photo of a Libyan landmark and upload it to the application, which will use the model to identify the landmark in the photo and shows some information to the user about it.

Keywords: Artificial Intelligence, Deep Learning, Convolutional neural Network, Landmark, Image classification

1. Introduction

Image recognition [1] is the process of identifying and detecting an object in an image or video. This concept is used in many healthcare systems, autonomous vehicles, and security surveillance. The spread of internet and smart phones produced new applications based on image detection and recognition such face recognition, road sign detection, landmark recognition. Many applications have been developed for landmark detection and recognition [2], [3] , these applications use different image recognition [4], [5] techniques.

With the increasing of computation power and amount of data, the recent years has seen tremendous advance in artificial intelligent techniques especially in the field of deep learning. Deep Learning uses algorithms inspired by the structure and function of the brain called artificial neural networks, Deep learning is revolutionizing several research fields like computer vision,

autonomous driving, natural language processing, and speech recognition. In the area of computer vision such as image recognition an important type of artificial neural network inspired biologically vision model called convolutional neural network (CNN) [6] outperformed human performance. CNN has received much attention for image recognition, object detection, and image description. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) has stimulated progress in the development of research on image recognition, new models have been developed, which are more effective than the previous image recognition models. CNN models are the most effective models. AlexNet [7], GoogLeNet [8], VGG[9], and ResNet [10] were the winners in the last few years. To exploit this advance in deep learning, this research introduce a model to identify Libyan landmarks such as Sabratha Theater, Alnaser arch, etc., in digital photos. The learning approach was based on convolutional neural network for image classification. In this paper, we adopt the idea for developing VGGNet, which consists of a number convolution with 3 × 3 filters to allows the network to learn more rich features. The proposed network can achieve high accuracy as well as reduced processing time.

2. Related Work

In [11] support vector machine (SVM) algorithm is used for recognizing different landmarks. It recognizes the picture using content matching. The information is extracted from the picture using the bag of words model. The bag of words model is used to reduce the computational cost. However, bag of words model has two major disadvantages first, it ignores the spatial layout information of the landmark image, and second it does not discriminate the different foreground in the image. In [12] to recognize different landmarks and to retrieve information from the picture, the picture is classified into different feature having global classification and local classification. Global feature differentiate picture according to its color and texture. SIFT algorithm [13] is used for retrieving information from the picture. SIFT algorithm improves the accuracy and improves the computational time. In [5] SVM is used for training data. The training dataset given to the SVM must contain vector that contain all the information of the images. The images were cropped to an aspect of 5:2 ratios to keep the same feature dimension for all images. To extract the feature from the image a Histogram of Oriented Gradients (HOG) [14] descriptor is used. The drawback of HOG is that it does not adjust to the invariant in orientation of the images.
In this paper, we create a Landmark classification system with deep learning that will help to recognize Libyan landmarks. Our approach is based on Convolutional neural network (CNN) to extract features[15] and to classify the images. Using convolutional neural network architectures like ResNet [10], it is possible to achieve a top-5 classification accuracy of 96.53% on ImageNet dataset [16]

.

3. Materials and Methods

The Libyan landmark dataset is a dataset of over 5000 images belong to five categories represent famous Libyan landmarks Fig(1). The images were collected from the internet using Google image search and Flickr API.

Fig 1. Famous Libyan Landmarks (Sabratha Theater, Alnaser arc, Benghazi Lighthouse, Acacus Mountains, and Lake gabroun)

2.1 Dataset Collection

To collect the dataset we used two methods the first one a python script was used to query the internet using google image search for the needed images. The second method used to query Flickr[17], flicker provide an API to the developers to query for images, we used this API to query for the Libyan landmarks that are available on flickr. The collected images were grouped by name to represent the landmarks.

2.2 Data Augmentation

The images collected from the internet were limited; we were able to collect around 100 images for each landmark. To train a deep learning network we need more images to avoid overfitting and get an accurate model. Data augmentation techniques [18] were used to increase the number of images in the dataset, a number of transformation are used for each image to produce new images. These transformations include randomly rotate, flip, shear, zoom, and rescale our training images as shown in figure 2.

Figure 2. New images using data augmentation

4. Network Architecture and Training

The convolutional neural network used in paper was inspired by VGGNet. In VGGNet, multiple CONV => RELU layers are stacked prior to applying a single POOL layer. Doing this allows the network to learn more rich features from the CONV layers prior to down sampling the spatial input size via the POOL operation. Our CNN consists of two sets of CONV => RELU => CONV => RELU => POOL layers, followed by a set of FC => RELU => FC => SOFTMAX layers. The first two CONV layers will learn 32 filters, each of size 3×3. The second two CONV layers will learn 64 filters, each of size 3×3. The POOL layers will perform max pooling over a 2×2 window with a 2×2 stride. A batch normalization layers after the activations along with dropout layers was inserted after each of the POOL layer and FC layers. The batch normalization and dropout layers are included in the network architecture will help reduce the effects of overfitting and increase our classification accuracy. The network architecture is detailed in Table 1, where the initial input image size is assumed to be 64 × 64 × 3 as we'll be training color images of size 64 × 64 pixel.

Table 1**:** *Summary CNN Architecture*

Layer Type	Output Size	Filter Size / Stride
INPUT IMAGE	64 × 64 × 3	
CONV	64 × 64 × 32	3 × 3, K = 32
ACT	64 × 64 × 32	
BN	64 × 64 × 32	128
CONV	64 × 64 × 32	3 × 3, K = 32
ACT	64 × 64 × 32	
BN	64 × 64 × 32	128
MAX POOL	32 × 32 × 32	2 × 2
DROPOUT	32 × 32 × 32	
CONV	32 × 32 × 64	3 × 3, K = 64
ACT	32 × 32 × 64	
CONV	32 × 32 × 64	3 × 3, K = 64
ACT	32 × 32 × 64	
MAX POOL	16 × 16 × 64	2 × 2
DROPOUT	16 × 16 × 64	
FLATTERN	16384	
FC	512	
ACT	512	
BN	512	
DROPOUT	512	
FC	5	
SOFTMAX	5	

5. Experiments and Results

In the experiment, we trained the networks with a high-performance computing (HPC) unit. It has the following specifications: Intel(R) Core(TM) i7-875HCPU @ 2.20GHz 8 Cores CPU, 16 GB

RAM, and nVIDIA-GeForce-GTX-1050. The operating system was Windows 10 64-bit .Our images were preprocessed before they were fed to the network. For every image in our dataset, we first resized it to 64×64 pixels. Once the image is resized, we then scaled them to the range [0,1]. When the data and labels are loaded figure 3, training and testing split are performed, 75% of data for training and 25% for testing.

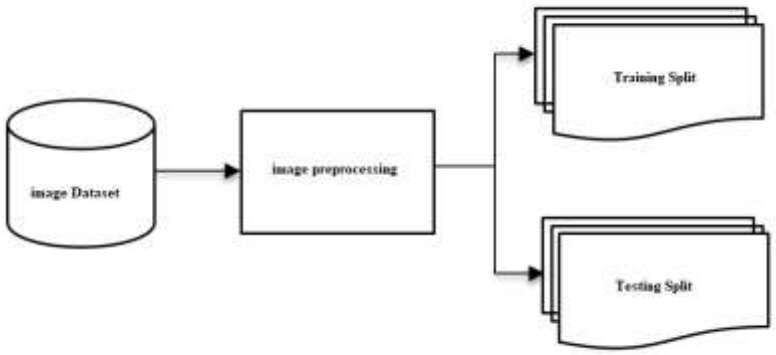

Figure 3. Dataset split for training and testing

The cross-entropy [19] is used as our loss function and stochastic gradient descent using a learning rate of 0.005 as our optimizer. The network was trained for 50 epochs using mini-batch sizes of 32. Table 2 shows a formatted classification report; we can see from the output that the network obtained 99% classification accuracy on our testing data

Table 2**:** *Network Evaluation*

[INFO] evaluating network...

	precision	recall	f1-score	support
acacus	1.00	0.97	0.99	263
alnaser	1.00	0.98	0.99	269
gabroun	1.00	0.99	1.00	269
lighthouse	0.99	1.00	0.99	252
sabratha	0.95	1.00	0.98	247
accuracy			0.99	1300
macro avg	0.99	0.99	0.99	1300
weighted avg	0.99	0.99	0.99	1300

The loss and accuracy plotted over time is displayed in Figure 4. On the x-axis we have the epoch number and on the y-axis we have the loss and accuracy. Examining this figure, we can see that our network is obtaining 99% classification accuracy. Furthermore, looking at our loss and accuracy plot over time in the figure demonstrates that our network is behaving quite well, after

only 10 epochs the network is already reaching ≈ 98% classification accuracy. Loss on both the training and validation data continues to fall with only a handful of minor "spikes" due to our learning rate staying constant and not decaying. At the end of the 25th epoch, we are reaching 99% accuracy on our testing set.

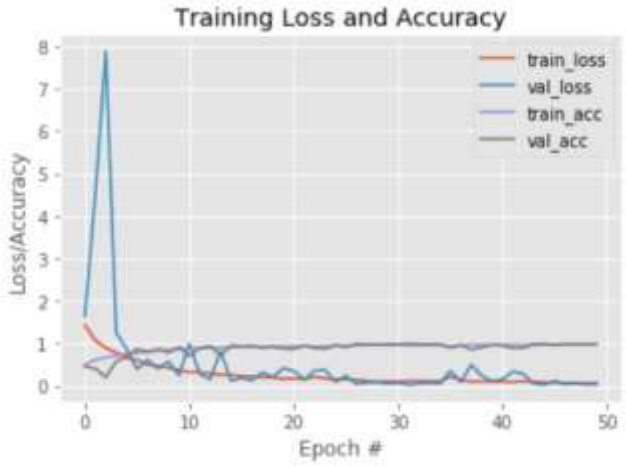

Figure.4: A plot of loss and accuracy over the course of 50 epochs for the Network Architecture trained on the Libyan Landmark dataset.

The above plot shows a quintessential graph as the loss decreases each time the accuracy increases, moreover the training and validation loss and accuracy mimic each other indicating that our network is learning the underlying patterns without overfitting.

6. Conclusions

Object recognition through CNN has interesting results. In fact, once the model is trained, making predictions on new images does not require intensive computational power, moreover it shows good performances and high accuracy. This paper presents CNN, which adopts the development idea of VGGNet to improve the network structure. The proposed network was trained on Libyan landmarks dataset to produce landmark classification model. The model shows a high accuracy in landmark recognition. This classification model could be incorporated into smartphone application to provide real-time feedback as images are taken, or utilized as a back end for landmark recognition systems. In future work, we will expand the model to include more Libyan landmarks and reduce model size to fit into more small devices to build offline landmark recognition applications.

References

[1] J. Magic and M. Magic, *Image Classification: Step-By-step Classifying Images with Python and Techniques of Computer Vision and Machine Learning (2; Python*. Amazon Digital Services LLC - Kdp Print Us, 2019.

[2] C. Termritthikun, S. Kanprachar, and P. Muneesawang, "NU-LiteNet : Mobile Landmark Recognition using Convolutional Neural Networks," no. 1, pp. 3–8, 2012.

[3] K. Yap, Z. Li, D. Zhang, and Z. Ng, "Efficient mobile landmark recognition based on saliency-aware scalable vocabulary tree," 2012, p. 1001.

[4] K. Bulkunde, A. Chakraborty, S. Kazi, and K. Dhumal, "Landmark Recognition using Image processing with MQTT protocol," pp. 2405–2407, 2017.

[5] T. Chen, K. Yap, and D. Zhang, "Discriminative Soft Bag-of-Visual Phrase for Mobile Landmark Recognition."

[6] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[7] A. Krizhevsky and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," pp. 1–9.

[8] C. Szegedy *et al.*, "Going Deeper with Convolutions," 2014.

[9] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," pp. 1–14, 2014.

[10] G. A. Hembury, V. V. Borovkov, J. M. Lintuluoto, and Y. Inoue, "Deep Residual Learning for Image Recognition Kaiming," *Cvpr*, vol. 32, no. 5, pp. 428–429, 2003.

[11] K. Banlupholsakul, J. Ieamsaard, and P. Muneesawang, "Re-ranking approach to mobile landmark recognition," in *2014 International Computer Science and Engineering Conference (ICSEC)*, 2014, pp. 251–254.

[12] K.-H. Yap, Z. Li, D.-J. Zhang, and Z.-K. Ng, "Efficient Mobile Landmark Recognition Based on Saliency-aware Scalable Vocabulary Tree," in *Proceedings of the 20th ACM International Conference on Multimedia*, 2012, pp. 1001–1004.

[13] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, vol. 2, pp. 1150–1157 vol.2.

[14] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, 2005, pp. 886–893.

[15] M. K. Benkaddour and A. Bounoua, "Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition," *Trait. du Signal*, vol. 34, no. 1–2, pp. 77–91, 2017.

[16] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-fei, "ImageNet : A Large-Scale Hierarchical Image Database," pp. 2–9.

[17] Flickr, "Flickr Services. The App Garden. API Documentation." [Online]. Available: https://www.flickr.com/services/api/. [Accessed: 01-Dec-2018].

[18] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, p. 60, 2019.

[19] K. Janocha and W. M. Czarnecki, "On Loss Functions for Deep Neural Networks in Classification," *CoRR*, vol. abs/1702.0, 2017.