Convolutional Neural Networks for Different Locations Classification

Mahmoud Elwan
Master of Engineering
Ottawa university
Cairo, Egypt
melwa053@uottawa.ca

Salma Sultan
Master of Engineering
Ottawa university
Mansoura, Egypt
ssult103@uottawa.ca

Mahmoud Yahia
Master of Engineering
Ottawa university
Cairo, Egypt
maboe017@uottawa.ca

Yasmeen Moahmed
Master of Engineering
Ottawa university
Qena, Egypt
ymoha090@uottawa.ca

Aya Reda
Master of Engineering
Ottawa university
Alexandria, Egypt
ayaaredaa1@gmail.com

***Abstract*— Refinements to the training process are mainly responsible for the current advancements in image classification research [1].**

This paper suggests a straightforward convolutional neural network (CNN)-based model for classifying various areas in low-resolution photos using feature extraction methods. The presented models are tested on a new dataset containing tiny and poor-quality countries' images to evaluate its efficacy. The accuracy and complexity of the baseline model are then contrasted with well-known VGG16-based CNN models. The findings show that the well-known models had higher accuracy [2].

As a result, it is considered that this study would offer the best model among the ones offered for classifying the locations in images.

I. INTRODUCTION

In many domains, including image classification, Deep Learning (DL) techniques outperformed cutting-edge algorithms in terms of predicted performance. By introducing more basic intermediate representations that may be used to create complex concepts, DL addresses the issue of data representation. A classic example of a DL model is an artificial neural network (ANN). The learning algorithm used to train a network for image classification justifies the idea of employing a higher number of layers, or a multilayer network [3]. Deep convolutional neural networks (CNNs) have demonstrated excellent performance in a variety of applications, including image, speech, and sound identification. Network engineering has been one of the major topics of image recognition research [4]. There have been several CNN architectures put out, including VGG, NiN, Inception, ResNet, DenseNet, and NASNet [5].

Image classification, localization, image segmentation, and object detection are examples of major problems in computer vision. Among these, image classification is the most fundamental problem. It forms the basis for other computer vision problems. In this project, we assessed a model for recognizing a country name by providing an image of a famous place in it, attempting multiple models, comparing them, and seeking the champion model.

In building a model for country recognition based on images, low image quality and data augmentation are used. Various models (VGG16, ResNet) are then used to compare the accuracy and the best model is chosen.

II. RELATED WORK

Researchers have recently begun to investigate the application of CNN-based deep learning methods in order to address the aforementioned issues. With a CNN-based vehicle classification and detection system, it is suggested to use a low-quality real-time monitoring camera. To evaluate CNN's relevance progressively applications, location and arrangement execution times are looked at utilizing both the computer processor and GPU. In order to identify and classify faraway vehicles in real-time applications, a different strategy based on Faster R-CNN architecture was proposed in a study. The study's proposed design's performance in a variety of weather conditions is evaluated.

In addition, the problem of classifying small objects with low resolution is investigated. A strategy utilizing two CNNs and generative adversarial networks (GAN) is proposed to address this issue. The proposed approach creates high-resolution images from low-resolution images in order to provide the classifier with images that are more accurate.

There are few studies on the use of deep learning-based vehicle classification techniques for low-resolution images, as the brief discussion above indicates. It should be noted that tracking camera systems with a depression-specific angle and/or dashcam views set to ROI are used to collect the data for these studies. In addition, the ROI is not sufficiently separated from the cameras in the studies.

Coates and Ng designed an algorithm for automatically choosing a receptive field and selecting similar groups of features by using similarity metric. During training, they used unsupervised learning for adjusting weights and the connected layers. They used CIFAR and STL-10 datasets for testing, and their results showed 60.1% accuracy. Dundar et al. used modified K-means to minimize correlated parameters while training a DCNN. They used labelled data for training and supervised learning before moving on to unsupervised learning to extract useful features. The input layer of 3*96*96 pixels is passed through 5*5 filters to obtain the feature maps (96*96*92) and add the Connection matrix between the model parts to improve accuracy by 67.1%.

Dong and Tan used unsupervised learning to train single-layer K-mean networks with various layers of unlabeled data and learning features. They chose 20.000 unlabeled images at random, and the square pooling size was 4*4. The pooling size is determined by the number of clusters and shallow learning (SIFT-representation rather than deep learning). They increased classification accuracy by 68.23% while overcoming the dataset's cluttered object background. Alexey et al. used unlabeled data of some transformations on a set of surrogate classes to train raw data images that are expressed by data augmentation (translation, scaling, rotation, and contrast) (samples of image patches). They extracted features from large amounts of data using supervised feature learning and unsupervised learning. The classification accuracy of the STL-10 dataset is 72.8%.

III. DATA

Data from different countries were collected from a variety of websites on Google. The total amount of data was 650 images from different 10 classes.65 images per class. Each image was a different size than the others. We split our data to train, validation, and test. Each class had 30 images for training, 20 images for validation, and 15 images for testing. Some pre-processing was made. The size of all the images was standardized to be 3x64x64.

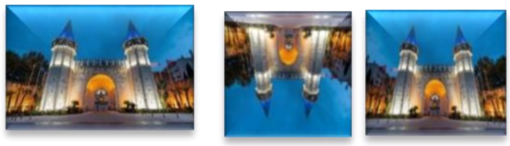
As our dataset includes many features which means it will not be easy for models to obtain good accuracy from classification. Besides, the dataset size is small. Data augmentation was applied with some spinning, flipping, and shifting in order to increase our data size and accuracy.

Examples of Data before and after augmentation

Class one - Egypt

Class two – USA

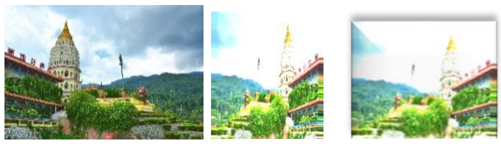
Class three – Turkey



Class four – UK

Class five – Spain

Class Six – Malaysia



Class Seven – Italy

Class Eight – Greece

Class Nine – India

Class Ten – France

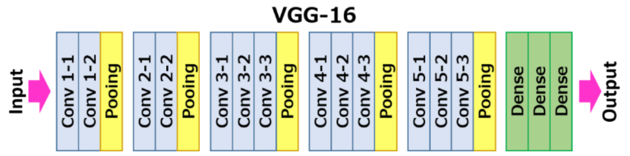
IV. METHODS

As our dataset was low-quality resolution images with many features, we were not sure how the used algorithms will perform. Four different models were applied to the dataset to see the accuracies in principle without using the augmented data.

A. Vgg-16

It is one of the top computer vision models to date. This model's developers evaluated the networks and increased the depth using an architecture with incredibly tiny (3 x 3) convolution filters, which showed a significant improvement on the prior-art configurations. The depth was increased to 16–19 weight layers, yielding approximately 138 trainable parameters.

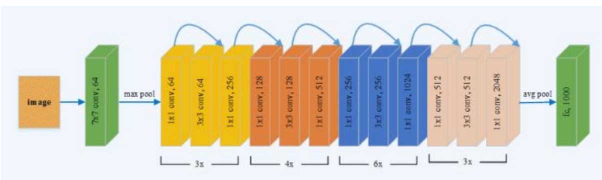
VGG16 is an object detection and classification algorithm that has a 92.7% accuracy rate when classifying 1000 images into 1000 different categories. It is a popular algorithm for classifying images and is simple to use with transfer learning [6].



VGG16

B. ResNet50

ResNet-50 is a convolutional neural network with 50 layers. The ImageNet database contains a pre-trained version of the network that has been trained on more than a million photos. The pretrained network can categorize images into 1000 different item categories [7]. It was proposed to solve the issue of diminishing gradient. The idea is to skip the connection and pass the residual to the next layer so that the model can continue to train. With Resnet models, CNN models can go deeper and deeper [8][9].

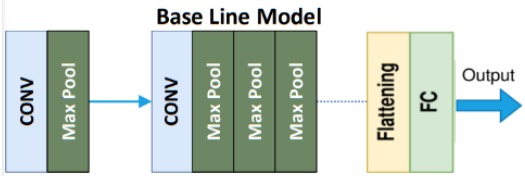


C. Baseline Model

It was a proposed model for solving the problem of low-quality images. [2]

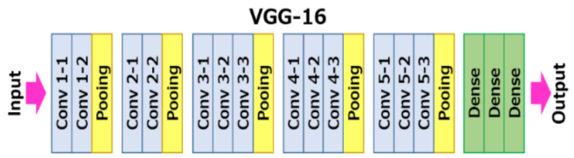
It is composed of two convolutional layers (Conv2D) and four max pooling layers (MaxPool2D). The Conv2d layers has 16 filters each 5 × 5 filter size, and both layers used activation function of Rectifier Linear Unit (ReLU). The MaxPool2D layers has 2 × 2 filter size and a stride value of 2.

Then the MaxPool2D layers are followed with flatten layer, fully connected layer of 16 units using L2 regularization with rate 0.008, then, dropout layer with value 0.3 and finally, fully connected layer that consists of 10 nodes (classes) with Softmax activation function.

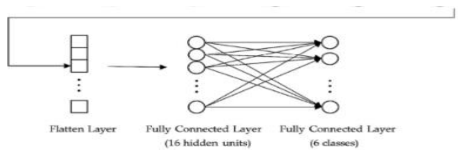


D. Vgg-16 + Baseline

This another proposed model for solving the problem of low-quality images of the same paper [2]. The vgg-16 model was fine-tuned by unfreezing some of the top layers. By doing this, the last layers of the base model + the first layers of the Vgg-16 model will be trainable layers which will present the features in a better way for the given task.



VGG16



V. EXPERIMENTS

I. Evaluate four CNNs: Baseline Model, ResNet-50 Model, VGG-16 Model and VGG combined with Baseline.
We resized the input images for all model to 64x64.
For Baseline:

Model	Acc		loss			Ephoc
	train	valid	test	train	valid	
BASELINE MODEL	74%	56%	39%	1.1	1.7	600

With 600 epoch and adam optimizer with learning rate = 0.0001 the model was able to learn some classes.

II. For ResNet-50:

Model	Acc		loss			Ephoc
	train	valid	test	train	valid	
ResNet-50	46%	35%	21%	1.85	2.05	600

With 600 epoch and same optimizer with same learning rate=0.0001 the model wasn't able to converge with that number of epochs the ResNet didn't fit our problem.

III. For VGG-16:

Model	Acc			loss		Ephoc
	train	valid	test	train	valid	
VGG-16	100%	63%	47%	0.006	1.6	600

With same Settings the model was able to learn most of classes with a little overfitting during the training process.

IV. For VGG-16 with Baseline:

Model	Acc			loss		Ephoc
	train	valid	test	train	valid	
VGG-16&Baseline	100%	70%	48%	0.03	1.4	600

With same Settings the results improved from the previous experiment.

V. For VGG-16 with Baseline with augmented data:

Model	Acc			loss		Ephoc
	train	valid	test	train	valid	
VGG-16&Baseline	100%	94.4%	49%	0.1	0.1	50

Training the model with Augmented data increase the model ability to converge more knowledge during training and was able to identify more classes the test accuracy wasn t high enough as the test set quality wasn't the same.

VI. Conclusion

To sum up, we used" Deep Learning-Based Vehicle Classification for Low-Quality Images" paper as a reference to build our models with an extra step by using data augmentation with the champion model which is "VGG-16 + BL" in our case. Furthermore, we aim to generalize our model by adding more different locations of different countries. Also, try to enhance our model test accuracy by adding more places of the same country to train our model with. Over and above we intend to build captioning model to describe the image which will help in recognizing the location.

References

1. He, Tong, et al. Bag of Tricks for Image Classification with Convolutional Neural Networks. arXiv, 5 Dec. 2018. arXiv.org, https://doi.org/10.48550/arXiv.1812.01187.
2. Tas, Sumeyra, et al. "Deep Learning-Based Vehicle Classification for Low Quality Images." Sensors (Basel, Switzerland), vol. 22, no. 13, June 2022, p. 4740. PubMed, https://doi.org/10.3390/s22134740.
3. Affonso, Carlos, et al. "Deep Learning for Biological Image Classification." Expert Systems with Applications, vol. 85, Nov. 2017, pp. 114–22. ScienceDirect, https://doi.org/10.1016/j.eswa.2017.05.039
4. [PDF] Between-Class Learning for Image Classification | Semantic Scholar. https://doi.org/10.1109/CVPR.2018.00575. Accessed 30 Nov. 2022.
5. He, Tong, et al. Bag of Tricks for Image Classification with Convolutional Neural Networks. arXiv, 5 Dec. 2018. arXiv.org, https://doi.org/10.48550/arXiv.1812.01187.
6. Learning, Great. "Everything You Need to Know about VGG16." Medium, 23 Sept. 2021, https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918.
7. Garg, Aryan. "Image Classification Using Resnet-50 Deep Learning Model." Analytics Vidhya, 20 Sept. 2022, https://www.analyticsvidhya.com/blog/2022/09/image-classification-in-stl-10-dataset-using-resnet-50-deep-learning-model/.
8. Zheng, Rachel Zhiqing. "Beginners' Guide to Image Classification: VGG-19, Resnet 50 and InceptionResnetV2 with TensorFlow." Medium, 29 Apr. 2020, https://towardsdatascience.com/beginners-guide-on-image-classification-vgg-19-resnet-50-and-inceptionresnetv2-with-tensorflow-4909c6478941.
9. "Convolutional Neural Networks for Multi-Class Histopathology Image Classification." DeepAI, 24 Mar. 2019, https://deepai.org/publication/convolutional-neural-networks-for-multi-class-histopathology-image-classification.
10. Bautista, C.M.; Dy, C.A.; Mañalac, M.I.; Orbe, R.A.; Cordel, M. Convolutional Neural Network for Vehicle Detection in Low Resolution Traffic Videos. In Proceedings of the 2016 IEEE Region 10 Symposium (TENSYMP), Bali, Indonesia, 9–11 May 2016; pp. 277–281.
11. Huilgol, P. *Top 4 Pre-Trained Models for Image Classification—With Python Code*; Analytics Vidhya: Gurgaon, India, 2020.
12. Krause, J.; Stark, M.; Deng, J.; Fei-Fei, L. 3D Object Representations for Fine-Grained Categorization. In Proceedings of the 2013
13. IEEE International Conference on Computer Vision Workshops, Washington, DC, USA, 2–8 December 2013; pp. 554–561.
14. El-Ashmony, E. & El-dosuky, Mohamed & Elmougy, Samir. (2016). CLASSIFICATION OF LOW QUALITY IMAGES USING CONVOLUTIONAL NEURAL NETWORK AND DEEP BELIEF NETWORK. International Journal of Intelligent Computing and Information Sciences. 16. 19-28. 10.21608/ijicis.2016.19822.