International Journal of Computing and Artificial Intelligence

E-ISSN: 2707-658X
P-ISSN: 2707-6571
IJCAI 2022; 3(2): 60-65
Received: 06-05-2022
Accepted: 09-06-2022
www.computersciencejournals.com/ijcai

Madhura Prakash M
Research Scholar, Department of CSE, BNM Institute of Technology, Karnataka, Bangalore, India

Dr. Krishnamurthy GN
Principal, BNM Institute of Technology, Karnataka, Bangalore, India

Corresponding Author:
Madhura Prakash M
Research Scholar, Department of CSE, BNM Institute of Technology, Karnataka, Bangalore, India

A comparitive analysis of mobilenet and xception architecture for classification of endoscopy images

Madhura Prakash M and Dr. Krishnamurthy GN

DOI: https://doi.org/10.33545/27076571.2022.v3.i2a.56

Abstract
The deep architectures have been making their significant mark in solving the image classification problems. The usage of the standard deep learning architectures has been increasing at a greater pace because of the outcome of the state-of-the art accuracy and precision by incorporating these architectures in image classification problems. The classification of medical images from variety of modalities is also possible with better accuracy by using these deep architectures. The standard architectures that have achieved substantial results in general object classification can be used and their learning can be transferred to the required domain. Further, these architectures can be used in an ensemble structure thereby creating a customized architecture for solving the classification problem. Some of the challenges in including these architectures include, the enormous memory and processing power requirement. Current trends are focused towards generating optimized deep structures that has relatively lesser memory and processing capacity requirement. To this end, the two benchmark architectures that have obtained momentous results in object classification namely the MobileNetV2 and Xception are included in this study to perform binary classification of capsule endoscopy images to detect the presence of abnormality. Capsule endoscopy (CE) is a medical procedure where the patient swallows the pill-based camera that traverses the entire gastrointestinal (GI) path capturing numerous images. These images are to be assessed for detecting the presence of any abnormality. The architectures included in this work have been trained on the images from the publicly available CE datasets and an ablation study on the hyperparameter tuning of the architecture have been conducted. The results are compared, analyzed and presented here. The MobileNetV2 and Xception architecture have achieved a maximum accuracy of 85% and 82% in the abnormality detection in CE images.

Keywords: Medical image processing, deep architectures, MobileNet, Xception, capsule endoscopy

1. Introduction
The deep architectures are being used extensively in solving the computer vision problems like classification, object detection, segmentation, object tracking and further as these architectures have aided in achieving results with greater precision and accuracy. The standard deep architectures that are used as base structures in solving computer vision problems have emerged over a decade as a result of the Imagenet challenge [1]. Newer architectures have progressively improved their accuracy and reduced their complexities. The application of these architectures in a particular domain requires dealing with challenges like, enormous labelled data requirement, memory and processing units' necessity. The recent architecture design in deep learning arena have focused on generating structures that are relatively less dense and require nominal processing power. The idea is to be able to use the model generated from these structures to solve a computer vision task without the need of large units and also to be able to achieve these tasks on edge devices in certain cases.
The two noteworthy architectures in the list of standard deep architectures available for solving the classification problem are MobileNetV2 and Xception. They are relatively an efficient choice because of the lesser size and higher top-1 accuracy of more than 75% in solving the Imagenet challenge. The paradigm architectures are generally included as base structures in designing solutions for classification problem and they are customized by adding required layer designs. A key advantage of encompassing these reference structure in the solution design is to be able to have a better starting point for the model training and thereby providing an optimization in better training efficiency. In this work these two reference architectures have been used included and customized in designing the solution for abnormality classification on the capsule endoscopy images.

The capsule endoscopy [2] is a medical procedure that is gaining momentum in the analysis of the gastrointestinal tract. The patient swallows a pill-based camera and this traverses the entire GE tract capturing continuous frames. This procedure has several advantages like: no requirement of hospital stays for the patients, the patients being able to carry out the regular activity during the procedure, and no hassle during the process. The frames from areas of the GI tract that are difficult to reach in a traditional tube-based endoscopy can be easily captured by the capsule camera, for instance the frames from the small-intestine are captured easily in this process. The CE procedure generally lasts about 7 to 8 hours and it generates continuous frames about an average to 80k to 90k frames. The physician will have to examine these frames looking for the presence of any abnormality. The artificial intelligent solutions with the help of deep learning architectures backed by computer vision algorithms are being used extensively in the medical field to aid the doctors and to help them utilize their expertise in focused areas. These solutions try to avoid the manual time and labour-intensive tasks that have to be carried out by the doctors, instead they flag the areas of concern and help the experts focus on obligatory areas. The work in [3] provides a detailed review on the capsule endoscopy procedures and also about the various camera configurations and current imaging solutions available. Fig.1 shows a sample PillCam capsule by MedTronic used in wireless endoscopy.

Fig 1: Representative image of a PillCam Capsule used in Endoscopy [4]

The two base deep architectures used in this study are minimally customized and they have been trained on the publicly available CE data. The performance of these architectures has been analysed by tweaking several hyperparameters to assess the model's performance. The analysis and the results have been presented in the next sections.

2. Current trends in CE Classification

Several works are carried out in the literature for the classification of the capsule endoscopy images into normal and abnormal. Works have also been carried out in classifying the abnormal images into several categories of abnormality to detect lesion, tumours, bleeding, polyps or other areas of concern. Another important task in computer vision along with classifying the given images is to be able to segment the images to reflect the regions of abnormality in the CE frames. Many of the existing solutions perform the classification of the images by extracting relevant feature and key points from the images and then subjecting these features to a machine learning based classification model. However recent works have mostly concentrated on either using a hybrid approach based on the combination of the features extracted and off-the-shelf features extracted from the deep neural networks or exclusively based on the deep features.

The work in [5] had focused on designing an architecture for multiclass classification of CE images and further in segmenting the areas of abnormality using a segnet encoder-decoder based model. The authors in [6] have developed a deep saliency model to detect the anomaly points on the CE frames. This model is based on a Convolution Neural Network (CNN) that has been trained on a weekly annotated dataset. The authors in [7] have proposed a class labelling method that can be used in the deep architectures for CE classification purposes. The authors have provided a detailed analysis of the possible types of frames and the grouping of the frames in different clusters like normal, abnormal and indistinguishable has been provided in detail. The researchers in [8] have designed a customized neural network model that extracts features of the CE images on multi scale to extract salient feature points in the image. The authors in [9] have used an imbalanced dataset and analysed the performance of several CNN models in classification of CE frames. The researchers in [10] have focused on generating a customized and optimized deep architecture by using the standard CNN as base architecture. The work focuses on detecting ulcer images in the test set of CE images.

3. Methodology Followed

In this work the two standard architectures namely MobileNetV2 and Xception are used as base architecture for the purpose of identifying abnormal images in the CE image set. The performance of these two architectures on this classification task is analysed. A brief note on these two architectures is presented below:

3.1. Xception Architecture

An important feature of the Xception Architecture is the depth-wise convolutions. It is an extreme version of the inception module. The inception module convolves over the output from the previous layers with different dimensions of convolution filters to form spatial transformation of the output from the previous layers. These multiple output features are further assessed and the model decides which output feature is to be considered and by what proportion. Fig. 2. depicts this multi-dimensional convolution and filter concatenation.

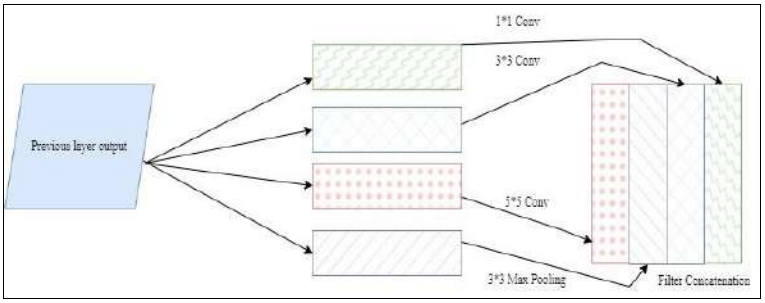

Fig 2: Multi-dimensional convolution and filter concatenation

The main disadvantage of this convolution is that it is computationally inefficient because of the convolutions the intermediary layers not happening spatially but across multiple-depths. To overcome this inefficiency 1*1 convolutions are applied across multiple channels of the input and later compressed to a lower-dimension. The Xception architecture first reduces the input by application of filter on each channel and then compresses the input space by using a 1*1 convolution. This architecture does not introduce any non-linearity in the execution of these two steps. Fig.3. depicts the architecture with layer details.

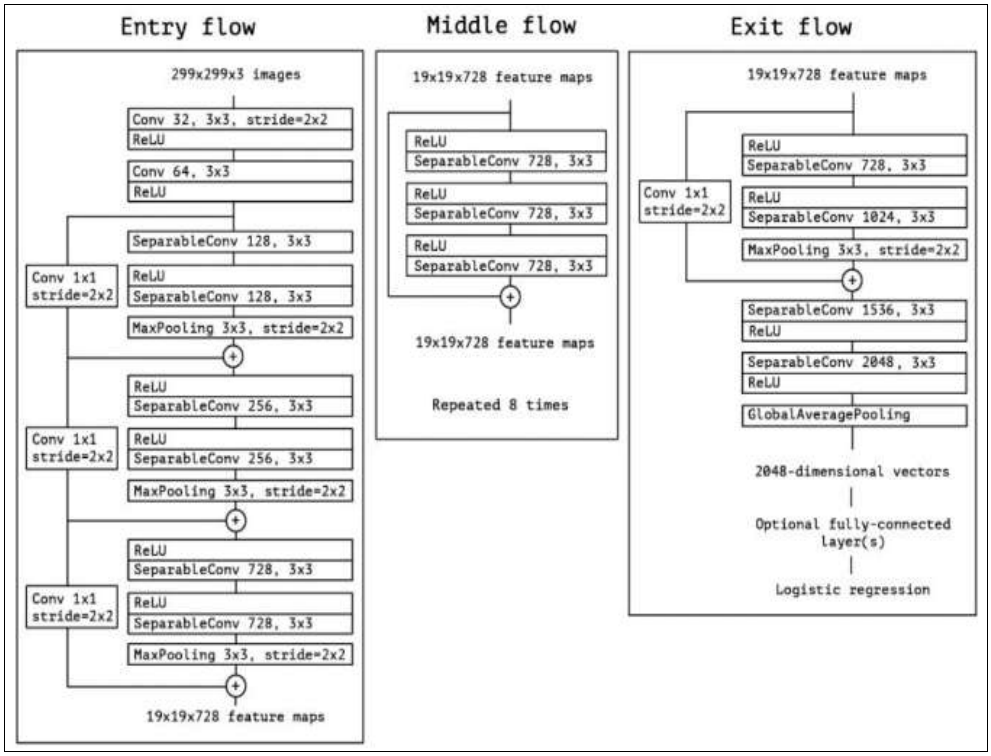

Fig 3: Xception Architecture [11]

3.2. MobileNetV2 Architecture

The major goal of MobileNet architecture is the reduction in the computation complexity and the memory size requirement by the model. These series of these architectures are developed with a goal of optimizing the performance of these structures on mobile devices. This is achieved by MobileNetV2 architecture by using inverted residual blocks that has fewer parameters, combined with skip connections. Fig. 4. depicts the architecture of Mobile Net V2.

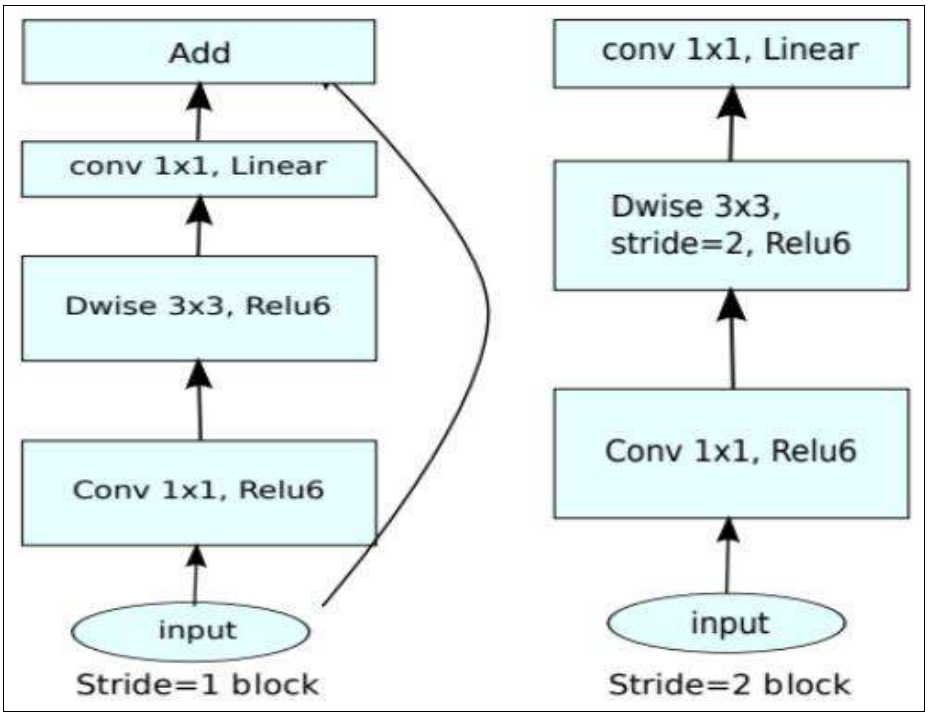

Fig 4: MobileNet V2 Architecture [12]

There are two types of blocks in this structure each with 3 layers. Both depth wise and 1*1 convolutions are incorporated with and without linearity in the layers. Linear bottlenecks are introduced in this architecture by squeezing the layers that are connected via skip connections.

3.3. Pipeline of the Proposed Work

This work on the binary classification of normal versus abnormal capsule endoscopy images progresses in multiple stages. The dataset consists of images collected from the widely available WCE image datasets. The abnormality images depict varied range of abnormality like bleeding, polyps, ulcer and cysts. Fig. 5. depicts sample images from the abnormality category.

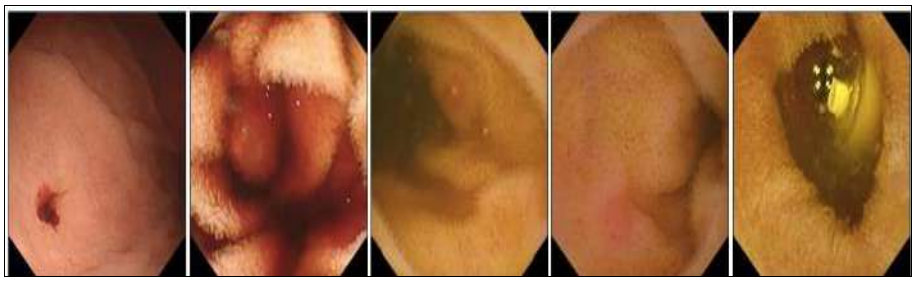

Fig 5: Sample Images under the abnormality category

The collected images from the publicly available libraries are divided into training and validation set in the ratio of 80:20. The images collected are augmented using various transformations along with color transformations like brightness, contrast, saturation and hue. This step is carried out to improve model generalizability along with creating a balanced dataset. The two base architectures are used with the concept of transfer learning and by using the model weights of these architectures that have succeeded in detecting generalized objects. These architectures are further customized by adding additional convolution and pooling layers. The complete pipeline followed is depicted in the Fig. 6.

Fig 6: Architecture Pipeline used in this work

The work includes experimentation with varied number of layers in the customization and comparing the accuracy of the model at different thresholds. The experiment included adding L1(Lasso Regression) and L2(Ridge Regression) regularization on both the activations and weights in the layers. This was done to avoid over fitting the model to the training data and to increase the validation accuracy. Ridge regression adds squared magnitude of coefficient as penalty term to the loss function namely binary cross entropy. Lasso Regression (Least Absolute Shrinkage and Selection Operator) adds absolute value of magnitude of coefficient as a penalty term. Lasso shrinks the less important feature's coefficient to zero thus, removing some feature altogether. So, helps in feature selection during the model training process that iterates through several epochs.

4. Model Performance and Discussions

The standard convolution neural network architectures that have been used in this work have been incorporated by transferring the weights of these models and using these weights as a standard reference point to begin the mode training. Both of these architectures have been customized by adding additional layers. Experiments have been conducted on the collected data by varying hyper-parameters. The L1 and L2 regularization techniques have been incorporated to add a penalty to both the weights and the activations occurring at each epoch of the model training to avoid over fitting and to increase model accuracy. The model accuracy and the area under the receiver operating characteristic curve (ROC) of the Xception model under one of the experiment trials is represented in the diagram below.

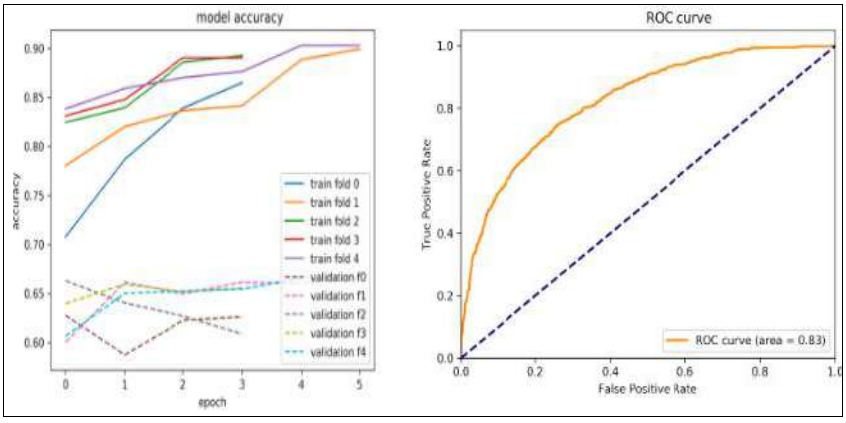

Fig 7: Model Accuracy and ROC curve for Xception

The model accuracy and the area under the receiver operating characteristic curve (ROC) of the MobileNetV2 model under one of the experiment trial is represented in the diagram below. It can be observed that the validation accuracy of the MobileNetV2 architecture improved considerably better as compared to the Xception architecture after applying regularization technique on the function activations and model weights.

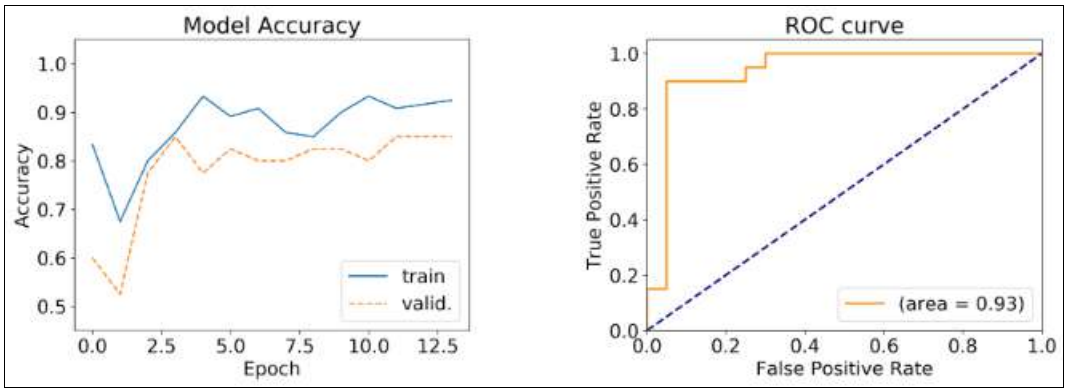

Fig 8: Model Accuracy and ROC curve for MobileNetV2 after Regularization

Fig. 9. indicates the performance of these architectures along with their accuracy, sensitivity and specificity under two different thresholds. The Time taken for inference by these two models are also presented. It can be observed that the MobileNetV2 has performed well with better accuracy.

	TRAIN						
MobileNet	Speed (imgs/sec)	Acc		Sens		Spec	
Xception	tot time	thresh1	thresh2	thresh1	thresh2	thresh1	thresh2
	0.9	93	80	100	100	85	60
cpu:	0.4	68	55	100	100	35	10
45 min	3 min						
2 hr	10 min						
	1.2	50	55	100	54	0.5	57
	0.2	50	45	100	57	0	34
	30 min						
	2 hr 15 min						
	1.4	36	46	100	97	8	24
	0.5	31	55	100	95	1	38
	2 hr						
	6 hr						
	INFER						
	13	95	70	100	100	90	40
	8	78	55	100	100	55	10
15 s	3 s						
45 s	5 s						
	40	50	50.5	100	52	0	49
	15	50	49	100	78	0	20
2 min 15 s	10 s						
7 min 30 s	30 s						
	59	30	45	100	96	0	22
	17	30	60	100	91	0	46
	50 s						
1 hr 50 min	3 min						

Fig 9: Comparative Analysis of the model performance

5. Concluding Remarks

In this work a transfer learning-based approach has been implemented to use the standard CNN models namely MobileNetV2 and Xception architectures for designing a customized architecture for the purpose of classifying the capsule endoscopy images into normal and abnormal. This work has focused on conducting experiments on the customized architecture to compare the classification performance on a dataset collected from the widely available libraries. Two customized architectures resulted by adding additional layers to MobilenetV2 and Xception architectures have been analysed. The accuracy of these models under various epochs after applying regularization techniques have been assessed. The highest accuracy achieved by MobileNetV2 and Xception architecture are 85% and 82% respectively after hyperparameter tuning and L1 and L2 regularization techniques.

Better accuracy, sensitivity and specificity can be achieved by applying customized pre-processing techniques on the image. Implementing post-processing techniques on the

model generated like model pruning can help in achieving optimized model that can be relatively light-weight and can be implemented on edge-based devices.

6. References

1. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, *et al*. Berg and Li Fei-Fei. (* = equal contribution) ImageNet Large Scale Visual Recognition Challenge. IJCV; c2015.
2. Iddan G, Meron G, Glukhovsky A, *et al*. Wireless capsule endoscopy. Nature; c2000. p. 405-417. https://doi.org/10.1038/35013140
3. Michael Vasilakakis, Anastasios Koulaouzidis, Diana E Yung, John N Plevris, Ervin Toth, Dimitris K Iakovidis. Follow-up on: optimizing lesion detection in small bowel capsule endoscopy and beyond: from present problems to future solutions, Expert Review of Gastroenterology & Hepatology; c2018. DOI: 10.1080/17474124.2019.1553616
4. Jain Samir, Seal Ayan, Ojha Aparajita, Yazidi Anis, Tacheci Ilja, Krejcar Ondrej. A deep CNN model for anomaly detection and localization in wireless capsule endoscopy images. Computers in Biology and Medicine. 2021;137:104789. 10.1016/j.compbiomed.2021.104789.
5. Medtronic-https://www.medtronic.com/covidien/en-us/products/capsule-endoscopy/pillcam-sb-3-system.html
6. Iakovidis DK, Georgakopoulos SV, Vasilakakis M, Koulaouzidis A, Plagianakos VP. Detecting and Locating Gastrointestinal Anomalies Using Deep Learning and Iterative Cluster Unification, in IEEE Transactions on Medical Imaging. 2018 Oct;37(10):2196-2210. doi: 10.1109/TMI.2018.2837002.
7. Park Ye-Seul, Lee Jung-Won. Class Labeling Method for Designing a Deep Neural Network of Capsule Endoscopic Images using a Lesion focused Knowledge Model. International Journal of Information Processing Systems. 2020;16:171-183. 10.3745/JIPS.02.0127.
8. Vasilakakis MD, Diamantis D, Spyrou E, *et al*. Weakly supervised multilabel classification for semantic interpretation of endoscopy video frames. Evolving Systems. 2020;11:409-421. https://doi.org/10.1007/s12530-018-9236-x
9. Filipe Fonseca, Beatriz Nunes, Marta Salgado, António Cunha. Abnormality classification in small datasets of capsule endoscopy images, Procedia Computer Science. 2022;196:469-476. ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.12.038
10. Wang S, Xing Y, Zhang L, Gao H, Zhang H. Deep Convolutional Neural Network for Ulcer Recognition in Wireless Capsule Endoscopy: Experimental Feasibility and Optimization. Comput Math Methods Med. 2019 Sep 18;2019:7546215. doi: 10.1155/2019/7546215. PMID: 31641370; PMCID: PMC6766681.
11. Chollet F, Xception: Deep Learning with Depthwise Separable Convolutions," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. p. 1800-1807. doi: 10.1109/CVPR.2017.195.
12. Sandler Mark, *et al*. MobileNetV2: Inverted Residuals and Linear Bottlenecks." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018. p. 4510-4520.