

A Comparative Study of Deep Learning based Image Classification Models for Material Classification

Atharva Ajay Gokhale
B.E. Student, Dept of Computer Engineering,
Pune Institute of Computer Technology
Pune, Maharashtra,
India

ABSTRACT

There are many instances in multiple fields where multiple objects are being used together and each of which can be composed of different materials. Hence the identification of these objects and their classification based on the material of which they are built is required. Due to recent developments in the field of Computer Vision and with the advent of multiple image classification technologies, it has enabled modern computer systems to play a role in this classification and identification process. A computer-based system making use of machine learning or deep learning models can be trained to differentiate between and hence classify objects into one of the multiple categories of materials under consideration. This can mainly be done in scenarios where objects of different materials are needed to be separated from one another in order to give further separate treatment to each one of them.

A number of image classification models and algorithms can be used for such types of use cases by predetermining the classes to be considered where each class can correspond to a unique kind of material. The models trained for these purposes can be used to predict the class pertaining to a material when an image of an object is given as input to it for identification. Further, in order to make the models or the learning algorithm more robust, the number of classes can be increased in order to cover a greater number of different types of materials. Following this, one can also compare among similar image classification models that can be used for the mentioned use case scenario in order to gain insights for determining the best model for the said purpose among a variety of models. Hence, this paper aims towards giving insights for the determination of a better model by considering a comparison or a comparative study of three different image classification models which are based on neural networks.

General Terms

Image classification, Deep Learning, Computer Vision, Python, Neural Networks

Keywords

Material classification, Transfer Learning, VGG16, ResNet50, Inception v3, Images' dataset, Keras

1. INTRODUCTION

Image classification techniques rendered by modern computer-based systems have opened new avenues where the tasks that are being carried out by a human can now be carried out by a computer. [10] These novel systems have enabled the achievement of improved classification rates, accuracy and precision with reduced additional time requirements and overheads. [9][8] As a result, it cannot be denied that these systems can also contribute towards the type of classification

required for distinction among different objects based on the material out of which they are made of. This requires training the models on labelled imagery data consisting of specific objects from different categories of materials. While classifying, these classes pertaining to different materials can denote materials like glass, paper, plastic, metals, wood, cloth, polymers and rubbers and fibers to name a few.

The comparative study being discussed in this paper considers the identification of the materials pertaining to the classes: glass, paper, metal and plastic. Such kinds of image classification systems for material classification can be embedded in existing systems and can be efficiently used in industry-based scenarios like assembly lines where the assembly of a particular product requires involvement of objects or components each of which comprises of a different material. Thus, for effective identification of these components based on the nature of material usage in them, such systems can play a vital role by introducing automation. Another major use case scenario where these systems can render service is that of waste management. The segregation of different waste materials based on their composition material from a mixture is important because each kind of material needs to be treated separately and no two types of materials can be treated in the same way. As a result, the automatic material classification systems can render a service in similar scenarios as well.

Due to the availability of a large number of image classification models available which can serve the purpose for the mentioned task, it is important to decide on the best models which render the required service by taking into consideration the accuracy and precision parameters. Hence this paper aims to present a comparative study between the performances of three distinct neural network-based image classification models by using transfer learning and considering their respective training and testing results on the same dataset. These neural network-based models are: the VGG16, ResNet50 and the Inception v3 models.

This paper aims towards discussing the details pertaining to the training and validation accuracies of each of these models along with certain performance parameters like recall, precision, f1 score and support values for each of the classes considered.

Thus, the overall goal of this comparative study is to provide insights regarding the models that can be used for image classification-based material classification and the selection of the best model amongst them.

2. DATASET

For the comparative study discussed in this paper, the imagery data pertaining to four distinct classes of materials were

considered. These classes are: glass, paper, metal and plastic. For preparing the dataset to train the image classification models, images containing multiple types of objects from each of the above-mentioned classes were collected from multiple sources which included around 400 to 500 images per class. All the images were used as RGB images and were resized to a size of 224x224. Out of the available data, 1563 images were used for training purposes, i.e., for the training of each of the models whereas 222 images were used for validation and testing purpose. Thus, the total count of the images used was 1785, considering a total of all the four mentioned classes. The input shape provided to each of the models based on the shape of these images was (224, 224, 3). The models and the subsequent classification can easily be made more robust by adding imagery data pertaining to additional classes to this dataset and then training the models on it. This can then be used to extend the services of the existing material classification system to an additional number of classes of materials.

3. METHODOLOGY AND MODELS

As mentioned earlier, the comparative study discussed in this paper considers three distinct neural network-based models which are: the VGG16 model, ResNet50 model and the Inception v3 model. Transfer learning has also been used as a part of implementation in order to make use of the pre-trained weights of each of the models and then to further train these models on the dataset under consideration for this particular study. This makes the training easier particularly in cases when the available data is less and hence the features that the model has already learnt can be reused by transfer learning. [7]

After training each of the mentioned models, the parameters which have been used for comparison purposes have been determined which include the training and validation accuracies along with recall, precision, f1 score and support for each of the classes used.

For each of the model, a certain number of layers of the neural network from the beginning were locked in order to make use of the weights pertaining to the locked layers to identify and extract certain features from the data. In addition to this, one or more of the lower layers for each of the models have been changed in order to suit the model for the dataset under consideration in this study. These changes and tweaking have been performed using the keras library available in Python language.

3.1 VGG16 Model

The VGG16 model is an image classification model which is based on the use of a deep convolutional neural network. In this model, stacks of convolution layer are used and the data images are made to pass through them.[1] These stacks of convolutional layers are followed by three fully connected layers. A special feature of this model is that it makes use of 3x3 filters that follow a stride of 1, in addition to which, the padding as well as the max pooling layers used everywhere follow the same uniformity consisting of a filter of 2x2 along with a stride of 2. These important features eliminate the need of tuning a large number of hyperparameters.[5]

As mentioned earlier, the starting layers have been locked in order to use them to extract certain features from the images in the dataset that the model has already learnt. The last Dense layer of the model has been tweaked in order to suit it to give the output probabilities for each of the 4 material classes under consideration. A softmax activation layer has also been

used along with this.

The model was trained on the dataset mentioned earlier. The optimizer used while training was Stochastic Gradient Descent (SGD) and the training was done using 100 epochs for the model.

Following chart represents the variations in the values of the training accuracy, validation accuracy, loss and validation loss as functions of the number of epochs for the training of the VGG16 model:

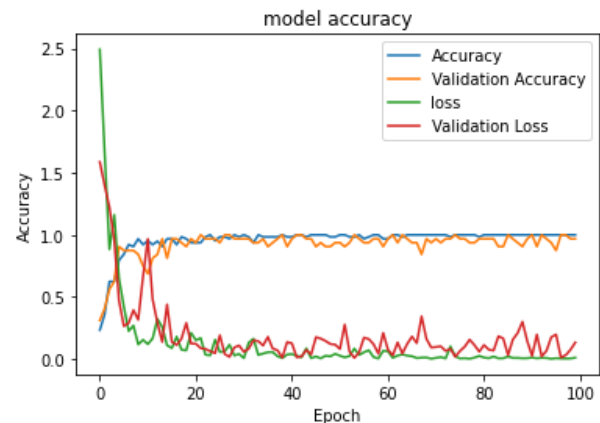


Fig 1: Variations in values of the parameters while training the VGG16 model

3.2 ResNet50 Model

ResNet stands for Residual Networks which work on the principle of Residual Learning. In Residual networks, the shortcut connections are added on an available plain network in order to convert the plain network into its corresponding residual version.[2]

In this study, the ResNet50 model has been used along with transfer learning. The reason for this is similar to that of the use of transfer learning in case of VGG16 model, which is to use the upper layers of the model by locking them in order to extract the features from the input dataset images for which the weights have already been trained and can be used for identification. It consists of 64 kernels with a stride of 2 having kernel size equal to 7x7. In addition to this, it also has a max pooling layer which also follows a stride of 2.[2]

As mentioned earlier, the upper layers of this model were obtained by transfer learning and the lower layers were tweaked according to the requirement of the output format and the nature of the data. For this, total 4 Dense layers were added and the last Dense layer comprised of 4 units in order to classify among the 4 classes under consideration.

Similar to the earlier model, a softmax activation layer was used in order to obtain the probability values for the 4 classes. The model was trained using Stochastic Gradient Descent (SGD) optimizer and the training comprised the use of 100 epochs for the model.

Following chart represents the variations in the values of the training accuracy, validation accuracy, loss and validation loss as functions of the number of epochs for the training of the ResNet50 model:

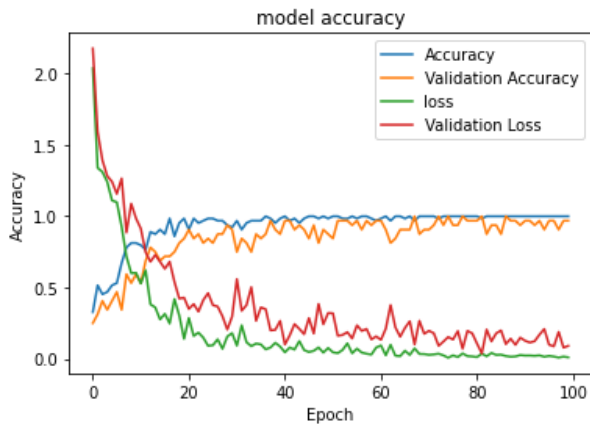


Fig 2: Variations in values of the parameters while training the ResNet50 model

3.3 Inception v3 Model

The Inception v3 is another famous neural network-based image classification model and the third one in the study discussed in this paper. This model makes use of 3 inception modules with 288 filters each with a size of 35x35.[4] A grid reduction technique is also used in this which basically steps from the earlier stage to 768 filters with a size of 17x17. As a next step, the earlier stage reduces to 1280 filters with a size of 8x8. This layer is followed by a pool layer which is then followed by a linear layer. The network ends with a softmax classifier in order to give probability values pertaining to the number of classes considered.[3][11]

Like the previous models, transfer learning has been applied to this model as well in order to lock the upper layers and use them for extracting the features.

Similar to the ResNet50 model, the lower layers of this model were tweaked in order to add a Dense layer consisting of 1024 units, followed by a Dropout layer. After this, the final Dense layer was added consisting of 4 units pertaining to 4 classes. A softmax activation function was also used to obtain probabilities for the 4 classes.

Similar to the earlier models, this model was also trained on the dataset using the Stochastic Gradient Descent (SGD) optimizer for 100 epochs.

Following chart represents the variations in the values of the training accuracy and validation accuracy as functions of the number of epochs for the training of the Inception v3 model:

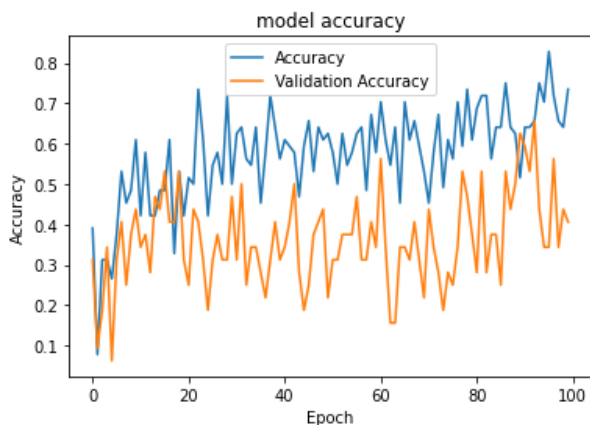


Fig 3: Variations in values of parameters (accuracy and validation accuracy) while training Inception v3 model

Following chart represents the variations in the values of the loss and validation loss as functions of the number of epochs for the training of the Inception v3 model:

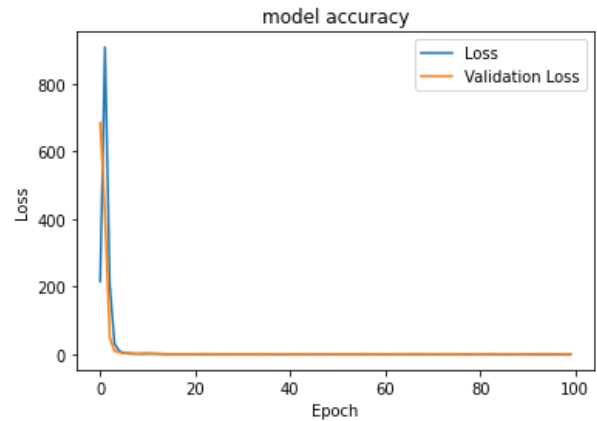


Fig 4: Variations in values of the parameters (loss and validation loss) while training the Inception v3 model

4. RESULTS

After each of the models was trained, the models were tested using the testing data and the parameters required for comparison between the models were determined.

The performance parameters considered for comparison between the three models are: training accuracy, validation accuracy, precision, recall, f1 score and support values for all four classes.

Following are the results that were obtained for each of the models:

Table 1. Table of performance parameters values for each of the models

Performance Parameter	VGG16 Model	ResNet50 Model	Inception v3 Model
Training Accuracy	96.33%	93.65%	57.37%
Validation Accuracy	91.34%	84.84%	36.03%
Precision (glass)	0.35	0.25	0.27
Precision (metal)	0.27	0.27	0.13
Precision (paper)	0.23	0.17	0.22
Precision (plastic)	0.29	0.20	0.19
Recall (glass)	0.37	0.23	0.10
Recall (metal)	0.25	0.29	0.07
Recall (paper)	0.22	0.15	0.20
Recall (plastic)	0.31	0.24	0.45
F1 score (glass)	0.36	0.24	0.15
F1 score (metal)	0.26	0.28	0.09

F1 score (paper)	0.22	0.16	0.21
F1 score (plastic)	0.30	0.22	0.27
Support (glass)	60	60	60
Support (metal)	56	56	56
Support (paper)	55	55	55
Support (plastic)	51	51	51

Following are some samples of the outputs for classification by the three models:

4.1 VGG16 Model

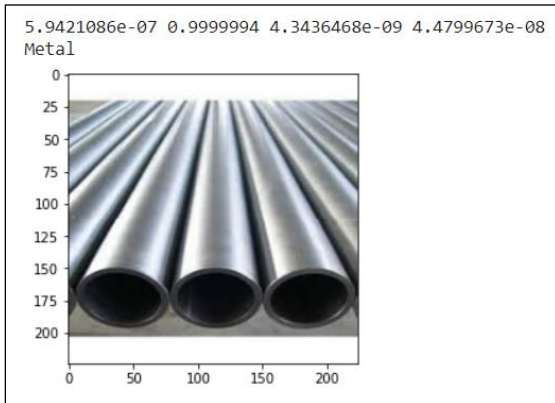


Fig 5: Image classified as 'Metal' by VGG16 model

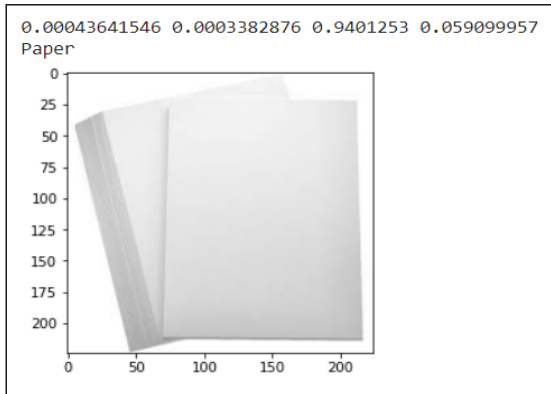


Fig 6: Image classified as 'Paper' by VGG16 model

4.2 ResNet50 Model

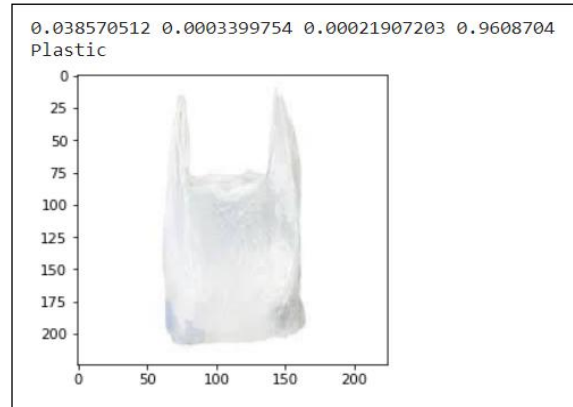


Fig 7: Image classified as 'Plastic' by ResNet50 model

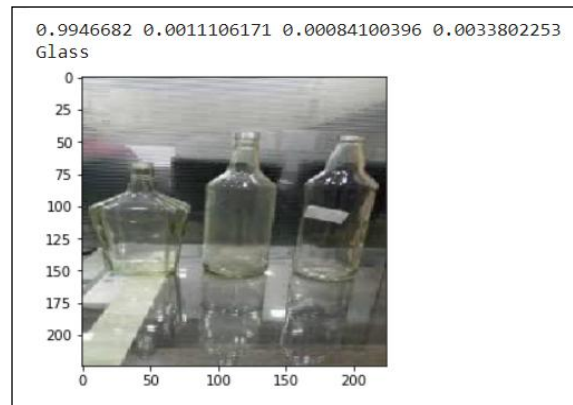


Fig 8: Image classified as 'Glass' by ResNet50 model

4.3 Inception v3 Model

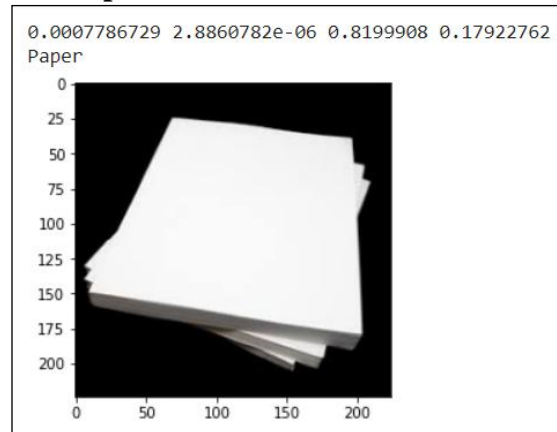


Fig 9: Image classified as 'Paper' by Inception v3 model

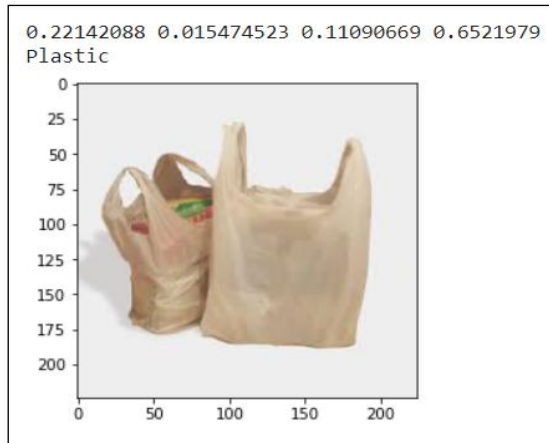


Fig 10: Image classified as 'Plastic' by Inception v3 model

5. CONCLUSION

Material classification can be an important requirement when a complex task requires use of components made up of diverse materials. This is because the components are required to be identified separately based on their material in order to put them to the required use. Deep Learning based image classification models can be efficiently used in such scenarios and can provide the required results according to the requirements. As discussed in this paper, three deep learning-based image classification models were successfully trained and tested considering the imagery data of four classes.

Considering the comparison between the three models discussed in this paper for the said purpose, we can infer from the obtained results that VGG16 model showed the best performance as the highest training and validation accuracies were obtained for this model. In addition to this, the other performance parameters including recall, precision and f1 score had comparatively better values in case of VGG16 model. After the VGG16, the model that showed the second-best performance among the three was ResNet50 model as its performance parameter values are better than that of the Inception v3 model but not as good as VGG16 model. Finally, the Inception v3 model showed the poorest performance among the three models. As a result, the maximum number of wrong classifications were obtained in case of the Inception v3 model.

Hence, the deep learning-based image classification models can be used for material classification for a predetermined set of material kinds. The system can easily be extended to include greater number of classes which can make the system more robust. The system can also be made custom to suit specific scenarios where a particular subset of classes of materials is to be considered and the other classes are not required. Thus, it can bring about overall efficiency in material classification and can obviate human efforts.

6. FUTURE SCOPE

The system can be enhanced by increasing the number of classes and adding more data pertaining to each of these classes. The classification efficiency can be enhanced by building ensemble models in order to make use of more than one models together. More than one models can also be used by constructing suitable pipeline structures and fitting multiple models in them in order to give the required treatment to the data in a sequential fashion and also to enhance the performance. The classification-based system can be embedded into existing hardware-based systems in the field where it is required. Systems like assembly lines can be

made efficient with robotic arms having facility to capture images of the objects and perform the classification. The classification system can also be integrated into software applications including web-based applications as well as android applications. The system can also be scaled by making use of cloud-based utilities for storage and analysis of data. IoT based devices and applications can also be used in order to improve hardware interactivity in situations where it is required. Thus, all these measures aim towards enhancing the overall system performance and to make it easier for humans to use by reducing human efforts.

7. REFERENCES

- [1] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- [2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
- [3] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).
- [4] Hussain, M., Bird, J. J., & Faria, D. R. (2018, September). A study on cnn transfer learning for image classification. In UK Workshop on computational Intelligence (pp. 191-202). Springer, Cham.
- [5] Tammina, S. (2019). Transfer learning using vgg-16 with deep convolutional neural network for classifying images. International Journal of Scientific and Research Publications (IJSRP), 9(10), 143-150.
- [6] Agarwal, T., & Mittal, H. (2019, August). Performance comparison of deep neural networks on image datasets. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 1-6). IEEE.
- [7] Krishna, S. T., & Kalluri, H. K. (2019). Deep learning and transfer learning approaches for image classification. International Journal of Recent Technology and Engineering (IJRTE), 7(5S4), 427-432.
- [8] Sharma, N., Jain, V., & Mishra, A. (2018). An analysis of convolutional neural networks for image classification. Procedia computer science, 132, 377-384.
- [9] Pak, M., & Kim, S. (2017, August). A review of deep learning in image recognition. In 2017 4th international conference on computer applications and information processing technology (CAIPT) (pp. 1-3). IEEE.
- [10] Lee, S. J., Chen, T., Yu, L., & Lai, C. H. (2018). Image classification based on the boost convolutional neural network. IEEE Access, 6, 12755-12768.
- [11] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.
- [12] Khan, A., Sohail, A., Zahoor, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review, 53(8), 5455-5516.