Design Studies and Intelligence Engineering
L.C. Jain et al. (Eds.)

© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA220051

Deep Inside Feature Learning for Image Classification Using Transfer Learning Approach

Ranjini Surendran[1], J Anitha[1] and D. Jude Hemanth[1]
[1]*Dept. of ECE, Karunya Institute of Technology and Sciences, Coimbatore, India*

Abstract. A scene consists of manifold objects and some relations among them. Exploring the relation among objects to acquire a meaningful understanding find application in the field of remote sensing, robotics, self driving cars etc. Our study intend to explore one such scene activity that can assist in resciung during natural disasters for example, flood. We have used the concept of deep learning networks which is a sub class of machine learning to develop a model that can detect the vehicles which got drowned in flood. We have made our own small dataset consisting of four classes with 100 images relating vehicles in each class. In our work convolutional neural network based pretrained model resnet101 learn the features and Support Vector Machines(SVM) functions as the classifier. This approach has shown an overall success rate of 91% in classification.

Keywords: Convolutional neural networks, supervised learning, unsupervided learning, deep learning, transfer learning, pre-trained network.

1. Introduction

Every year we see reports regarding different natural calamities and disasters affecting different parts of the world which includes earth quakes, flood, forest fire, tsunami etc. These have in a large extend affected both animal and human genus. In places pretentious with calamities like flood many cases were reported, where the rescue people were not able to reach the affected areas by any normal means. Now a days, drones are assisting the rescue operations in a broader way. In our proposed approach we have developed a model employing convolutional neural networks (CNN) for feature extraction and a Support Vector Machine (SVM) for classification, that can lend a hand during rescue activites.

Scene understanding is one among the contemporary research topic contributing in the field of deeplearning [16]. Understanding a scene in deep representation have changed the revelation of computer vision tasks in classifying different objects in a complex scene by training a Convolutional Neural Network [1,2,3,4] which learns the features of the image . Understanding an image by classifying image level description as in past using hand-engineered features is meager as compared to machine learned features to classify the different images in the scene which are much complex. Classification of an image transpires accurately by the lower layers of the network. What features do they possess could be generated by training a CNN by varying their number

of hidden layers. The top level layers don't put in much contribution in feature learning in the classification progression. As the hidden layers are augmented various pertinent features may get condensed and the layers will get over fitted. Lower layers of the network learns more low level features which bear the minuscule details of the image and decides the feature learning effectiveness of the network. In visual recognition it is still the features from low – level to high- level which matters. Deep network can showcase its efficacy when large data set trains the network.

Developing a new model from scuff is hard as it requires large dataset, high processing period and powerful computing machines. Training a convolutional neural network for few images will not be good enough. As the number of image in our dataset is limited we are using a Transfer learning [15, 20] approach for training our network. Transfer learning is one of the methods of training machine learning algorithms unlike other methodologies like supervised learning, unsupervised training and semi-supervised learning. Transfer learning has a special feature that we train for one task and use that knowledge for other similar task.Existing pre-trained CNN models, AlexNet [11], GoogLeNet [12],VGGNet-19 [13], ResNet101[11] are already trained with millions of image datasets, ImageNet [17], PASCAL VOC[18],Microsoft COCO [19] . We can use a pre-trained network that can learn even with limited images. Since the models are already trained with many images thay have learned many features. So when these models are trained with more specific images, they can remember the features which they have already learned and can learn the features of new images with more precise learning.

2. Previous works

One of the key errands in the field of computer visual task is still image classification. This come up with the emerging developments in the feature learning tasks of CNN as compared to the traditional hand-engineered features using HOG [5], BoW [6], SIFT [6], Spatial Pyramid [7] to extract features to describe an image. These features were then given to any of the machine learning classifiers like Random Forest[14] or Support Vector Machine (SVM) [8]. Learning features automatically with large image set paved the way to increasing demand of CNN [10] with deep network. Deeper the layers more will be the features learned so that we can easily understand what pixels are on the background and what on the objects. Training a CNN [1] for extracting features of an image can improve the performance of image classification. Success of classification lies in the success of feature learning.

3. DeepLearning Network: ResNet101 Model

ResNet in short Residual Network is a particular category of Convolutional Neural Network. Resnet101 is a residual model having 101 weighted layers. Resnet eliminates the issues of deeper networks by introducing residual blocks. Here each layer is learning from a small factor known as residue, hence the name Residual Network. Resnet provides skip layers which are generally identity mappings that make available an alternate pathway for the gradients to flow and training is made easier. Weights of shallow network is used to create deeper network and wherever there is gap identity connection

is done. Resnet architecture has skip connection as shown in Fig.1 with two convolution layers with one taking bunch of input feature maps X, and other convolution layer with output denoted as H(X).

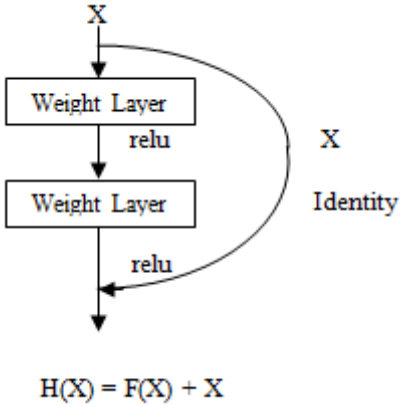

Fig.1. Residual Block of ResNet

In addition to the output, input is also copied here to the output by skipping two layers of convolution. Input of the first convolution layer is copied to the output of second convolution layer. So we can write the output equation as

$$H(X) = F(X) + X \tag{1}$$

Instead of learning H(X), we learn the residual. Each layer learns from the residual value, which ensures that even if the weights not get updated the layers will learns at least the input features. Gradient update step is additive and this additive gradient step leads to actual updates to the weights and thus eliminating vanishing gradient problem.3 x 3 convolution is used across all the layers, sub sampling carried out either using max pooling at beginning and end or by using striding convolution. Some of the deeper layers are modified with bottle neck modules, the 3 X 3 convolutions is replaced with 1 x 1 convolution to keep training sustainable.

4. Methodology

The proposed method in Fig.2 aims at developing a model using a pre-trained convolution neural network for learning the discriptive features. Convulutional Neural Networks learn the features of the images fed to its input image layer. Each layer in the network learns some features from the image. Features learned by the first hidden layer are transferred to the next hidden layer where it learns new features which are carried over to the next layer with the weights updated. In feature extraction, initial layers learn the low level features of the image like edges, corners etc and the top layers learn the mid – level and high level features of the image. Each convolutional layer is followed by nonlinear layer and pooling layer. Convolutional layer maps the input pixels to certain neurons in different regions. Non-linear layer turns negative values to zero and can fix vanishing gradients. The pooling layer down samples data to reduce the number of inputs to the next layer. SVM is fed with these feature discriptives to perform classification.

The steps for the projected system is as follows:

- Select the required image data set.
- Split the data set into testing and training categories.
- Preprocess the dataset before applying to the network.
- Feed preprocessed images to pre-trained model to learn features.
- These learned features are then classified by SVM.
- Test the model with new image and evaluate the performance.

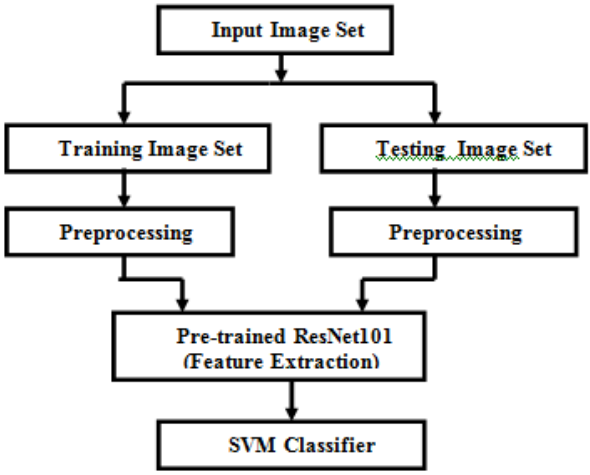

Fig.2. Proposed model

4.1. Image Dataset

Dataset which we have processed here is representing image of few vehicles drowned in flood. We have collected these photos from various sites and have made a small image dataset. Our dataset consists of four classes , with each class having 50 images each. All the imagess are seperated into four classes and these are car on road, rotated car on road, car in flood, rotated car in flood. All images are in JPEG and RGB format. İn two classe no rotation is given to the vehicles, but in other two rotation is carried out for images. Fig.3 shows a sample input image which belong to the third class i.e car in flood.

Fig.3. Sample input Image

4.2. Preprocessing

Image dataset is split into trainig and test data set. From each class 30% is taken as training data and remaining 70% is taken a testing data. Since the network implemented here accepts image of format having image size of 224 x 224 and RGB color, all the images are resized to 224 x 224 and grey to RGB conversion is also carried out. All the images fed to the input layer of network is of format 224 x 224 x3.

4.3. Feature Extraction

As the image dataset in our proposal is limited, use of conventional CNN models or developing a model from scratch will diminish the performance of the system. By using transfer learning approach we can enhance the performance of the model. More the network is deeper, larger will be the training and testing errors and the network accuracy is reduced. This happens because as we go deeper and deeper, vanishing gradient problem occurs. Among the esisting pre-trained deep learning models AlexNet [10], GoogLeNet[12], VGGNet-19 [13], performance of ResNet101[11] is better as it solves this problem by introducing a novel technique known as skip connection.

Deep CNN used is the pre-trained network, resnet101 that learns the feature parameters of the image dataset by using transfer learning [15,20] concept. Once the network is loaded we have to modify the layers of the network according to our requirement. The first layer of the deep learning layer is the image input layer. As the fully connected layer of pre-trained network is trained for 1000 classes we keep the middle layers of the pre-trained model unchanged and update the top three layers as shown in Fig.4 since our classification is for only four classes. We then train our modified model with input dataset

```
layers =

  4x1 Layer array with layers:

     1   ''   Image Input             224x224x3 images with 'zerocenter' normalization
     2   ''   Fully Connected         4 fully connected layer
     3   ''   Softmax                 softmax
     4   ''   Classification Output   crossentropyex
```

Fig.4. Modified layers of Pre-Trained Resnet101 model

4.4. Classification

CNN feature maps learned by the deep network model trains a multiclass Support Vector Machine (SVM) [8] by using fast sctochastic gradient descent algorithm. The training and testing features are passed to the classifier to evaluate its performance. New images are then categorised by this trained classifier.

5. Experimental Results

Deep CNN model is trained with four classes of vehicle image data: car on road, rotated car on road, car in flood, rotated car in flood. Our work aim at identifying the vechicle affected by flood and can discriminate among the vehicles on road and those affected in flood. Fig.5 shows plot of accuracy and loss function of training Progress for 20 epochs.Training graph shows that after 20 iterations the loss function comparively reduced. Fig.6 shows plot of accuracy and loss function of testing Progress for 20 epochs. Loss function is diminishing after 70 iterations. Training and testing accuracy is compared in Fig.7 which shows testing accuracy stabilizes after 120 iterations.

Performance of the system is evaluated using confusion matrix. The matrices are True Positive (Tp), True Negative (Tn), False Negative (Fn) and False Positive (Fp). If a class is correclty identified then its value belong to Tp and those belonging to different class if identified correctly belong to Tn. Incorrectly identified classes belong to Fn and Fp values. These matrices are used to calculate the following parameters for performance evaluation of the model.

$$Accuracy = Tp + Tn / (Tp + Tn + Fp + Fn) * 100 \quad (2)$$

$$Sensitivity = Tp / (Tp + Fn) * 100 \quad (3)$$

$$Specificity = Tn / (Tn + Fp) * 100 \quad (4)$$

$$Precision = Tp / (Tp + Fp) \quad (5)$$

$$Recall = Tp / (Tp + Fn) \quad (6)$$

$$F1\text{-}Score = (2 * Precision * Recall) / (Precision + Recall) * 100 \quad (7)$$

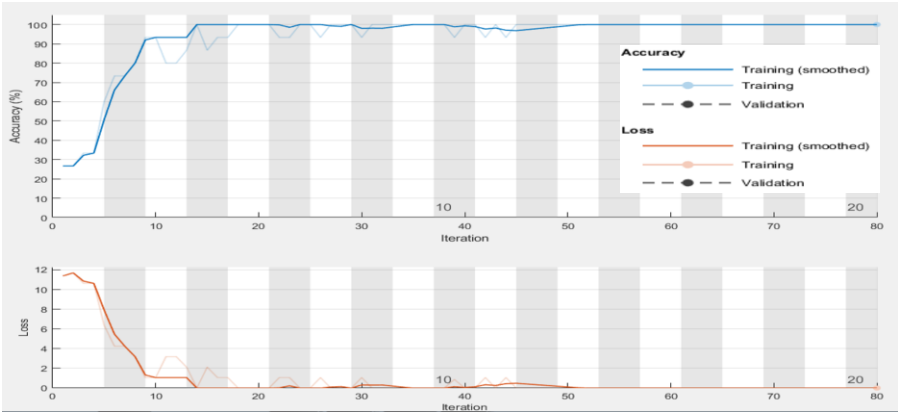

Fig.5. Training Progress of ResNet101

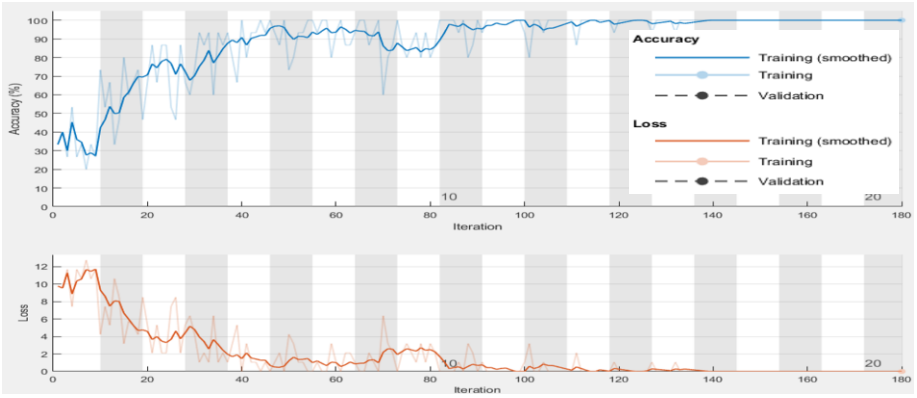

Fig.6. Testing Progress of Resnet101

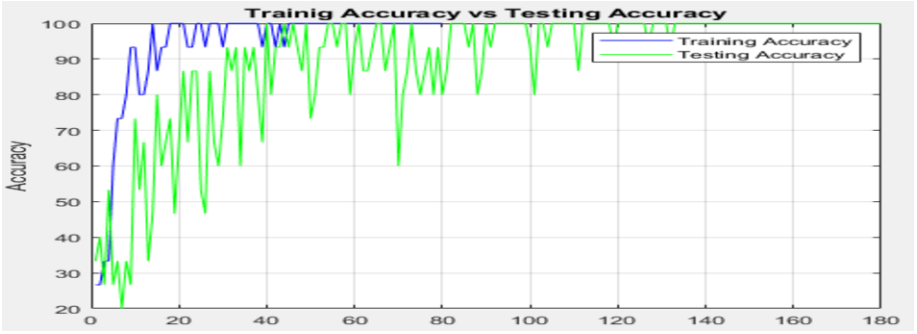

Fig.7. Comparison of training and testing accuracy of Resnet101

Confusion matrix in Fig.8 shows the values of Tp, Tn, Fp, Fn indices of our model obtained by experimentation with an accuracy of aggregate 90.7%. Performance parameters calculated using these indices with Sensitinity-87.14%, Specificity- 94.3%, Precision- 0.94%, Recall- 0.87 and F1-Score 90.3% are formulated in Table 1.

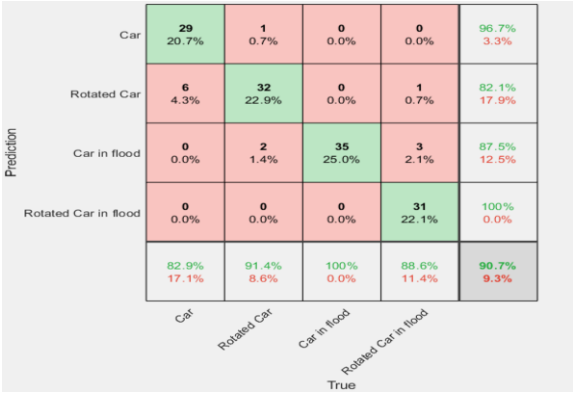

Fig.8. Confusion matrix showing 90.7 % accuracy for ResNet101 model

Table 1. Matrices calculated from Confusion matrix ResNet101 model

Accuracy	Sensitivity	Specificity	Precision	Recall	F1-Score
90.7143	87.1429	94.2857	0.9385	0.8714	90.3704

6. Conclusion

We have proposed a model to identify whether a vehicle in struck in water or flood. The model used pre-trained network ResNet101 as the feature extractor and SVM as classifier. The CNN network trained in this paper is carried out on limited number of images with resnet101 architecture. These hidden layers consume much less time of the machine in getting trained to the limited image set given. The model could succesfully predict the classes with the features learned by transfer learning [15,20] approach. The methodoly showed a success rate of 91 % even with the small image dataset. The proposal could be extended for detecting any vehicles, human and animals. Accuracy can be improved by increasing the image dataset classes to extract more features and train and test the network in less loss and high accuracy. In future the model has to be modified and improved for images with different scaling, blurring and rotation**.**

References

[1] Yim, J., Ju, J., Jung, H., & Kim, J. (2015). *Image Classification Using Convolutional Neural Networks With Multi-stage Feature. Robot Intelligence Technology and Applications 3, 587–594.* doi:10.1007/978-3-319-16841-8_52

[2] LeCun, Y., Huang, F.J., Bottou, L.: Learning methods for generic object recognition with invariance to pose and lighting. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 2, pp. II–97. IEEE (2004)

[3] Jarrett, K., Kavukcuoglu, K., Ranzato, M., LeCun, Y.: What is the best multi-stage architecture for object recognition? In: 2009 IEEE 12th International Conference on Computer Vision, pp. 2146–2153. IEEE (2009)

[4] National Y. LeCun, K. Kavukcuoglu, and C. Farabet. Convolutional networks and applications in vision. In Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, pages 253–256.IEEE, 2010.

[5] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005).

[6] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004).

[7] Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: Proceedings of the CVPR, 2006.

[8] Y.Lin, F.Lv, S.Zhu, etal., Large-scale image classification: fast feature extraction and svm training, in: Proceedings of the CVPR, 2011.

[9] D.Ciresan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, in: Proceedings of the CVPR, 2012.

[10] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the NIPS, 2012

[11] He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. CVPR (2016)

[12] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 1–9.
[13] Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[14] L. Breiman. Random forests. Machine learning, 45(1):5–32, 2001.
[15] D.C. Ciresan, U. Meier, J. Schmidhuber, Transfer learning for Latin and Chinese characters with deep neural networks, in: Proceedings of the IJCNN, 2012.
[16] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*,2015.
[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh,S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein,A. C. Berg, and L. Fei-Fei. Imagenet large scale visual Recognition challenge. *IJCV*, 115(3):211–252, 2015.
[18] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I.Williams, J. Winn, and A. Zisserman. The PASCAL visual object classes challenge: A retrospective. *IJCV*, 111(1):98– 136, 2015.
[19] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014.
[20] J.S.J. Ren, L. Xu, On vectorization of deep convolutional neural networks for vision tasks,in: Proceedings of the AAAI, 2015.