TRANSFER LEARNING FOR CLOUD IMAGE CLASSIFICATION

Mayank Jain[1,2], *Navya Jain*[3], *Yee Hui Lee*[4], *Stefan Winkler*[5], *and Soumyabrata Dev*[1,2]

[1] The ADAPT SFI Research Centre, Ireland
[2] School of Computer Science, University College Dublin, Ireland
[3] Ram Lal Anand College, University of Delhi, India
[4] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[5] Department of Computer Science, National University of Singapore

ABSTRACT

Cloud image classification has been extensively studied in the literature, as it has several radio-meteorological and remote sensing applications. Recently, images from ground-based sky imagers (GSIs) are being widely used because of their high temporal and spatial resolution and low infrastructure cost as compared to satellites. To classify sky/cloud images obtained from such GSIs, this paper[1] examines the application of transfer learning using the standard VGG-16 architecture. The paper further analyzes the importance of adjusting the number of neurons in the top dense layers to improve the performance of the model. The reasons for the same are traced by conducting extensive experiments on multiple datasets exhibiting varied properties.

Index Terms— Cloud Image Classification, Transfer Learning, Deep Learning, CNNs, VGG-16

1. INTRODUCTION

Clouds are known to hinder the propagation of radio waves [1] and sunlight [2]. At the same time, different cloud types impact the rays in a different manner. Therefore, by accurately identifying and classifying cloud types, researchers and engineers can better understand their impact on radio wave transmission and develop strategies to mitigate any potential disruptions. Similarly, solar energy systems can adjust their operations according to cloud types that are present locally, optimizing energy production, and predicting periods of reduced solar irradiance. This information is valuable in ensuring reliable solar energy generation, especially in regions heavily dependent on renewable energy sources.

Ground-based sky imagers (GSIs) are now preferred over satellites to capture cloud images at high temporal and spatial resolution in a cost-effective manner [3]. However, their drawback is the presence of noise caused by factors such as sun glare, dust particles, and rain droplets [4]. This noise, along with the diverse shapes, sizes, and textures of the clouds, complicates accurate classification [5]. The complex nature of cloud images and the ever changing atmospheric conditions have led researchers to focus on developing precise and reliable classification algorithms [6–9].

1.1. Related Work

Conventional methods for classifying sky/cloud images rely on statistically/manually identified characteristics that describe the color and texture of the image [6]. Extending on this concept, Dev *et al.* [7] proposed a modified texton-based classification method to integrate color and texture information and reported an average accuracy of 95% on the SWIMCAT dataset of 5 classes.

Recently, deep learning techniques have been employed. In 2018, Zhang *et al.* [8] proposed CloudNet which reported accuracies of 98.33% on the SWIMCAT dataset and up to 88% on the newly released CCSN dataset with 11 cloud classes. Wang *et al.*'s [9] CloudA architecture raised the bar with 98.47% on SWIMCAT and up to 98.83% on another private dataset. In 2022, Liu *et al.* [5] achieved 84.3% accuracy on a huge 7 class GCD dataset using a context graph attention network, where CloudNet was claimed to achieve a mere 74.84% accuracy. Although deep learning models show potential, such varied accuracies on different datasets is primarily due to the small size of annotated datasets.

1.2. Contributions

Transfer learning (TL) has proven to be highly effective in training deep learning models on limited datasets [10], which has been the issue of main concern. Hence, the objective of this paper is to examine the application of TL in cloud-type

This research was conducted with the financial support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106_P2 at the ADAPT SFI Research Centre at University College Dublin. ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology, is funded by Science Foundation Ireland through the SFI Research Centres Programme.

Send correspondence to S. Dev: `soumyabrata.dev@ucd.ie`

[1]In the spirit of reproducible research, the code to reproduce the simulations in this paper is shared at `https://github.com/jain15mayank/Cloud-Classification-using-Deep-Nets`

recognition using GSI-obtained sky/cloud images. The key contributions are summarized as follows:

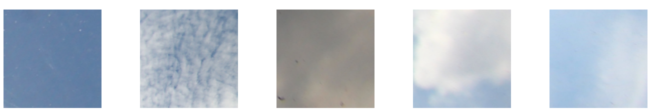
- Significantly reduced training time with high accuracy
- Effective usage of TL to compete with state-of-the-art
- In-depth analysis of the impact of number of neurons in the top dense layers while doing TL

2. DATASETS

This paper uses the following three publicly available datasets:

1. Singapore Whole-sky IMaging CATegories (SWIMCAT) dataset [7]
2. Cirrus Cumulus Stratus Nimbus (CCSN) dataset [8]
3. Ground-based Cloud Dataset (GCD) [5]

SWIMCAT has a total of 784 cloud images of size 125×125 pixels, classified into 5 classes. An image of each class is shown in Fig. 1 along with the number of images (img) in that class.



Clear Sky (224 imgs) | Patterned (89 imgs) | Thick-dark (251 imgs) | Thick-white (135 imgs) | Veil Clouds (85 imgs)

Fig. 1: Sample images from each class along with the number of images in each class of SWIMCAT dataset

CCSN is a dataset of 2543 cloud images of size 400×400, which are then divided into 11 classes. While the other two datasets were composed only of sky image patches obtained from the images captured by a GSI, CCSN dataset is composed of landscapes, sceneries, both day and night images, and sky image patches. Such a large variety of images and multiple classes with $139 - 340$ images per class makes this dataset more difficult to train on.

GCD is the largest ever annotated dataset of sky patch images captured by a GSI. It consists of $19,000$ cloud images, of size 512×512 pixels, which are divided into 7 classes. One of the classes contains 'mixed' clouds with significantly fewer images and is not considered in the experiments of this paper. Some sample images from both the CCSN and GCD datasets are shown in Fig. 2.

(a) CCSN | (b) GCD

Fig. 2: Some sample images from the CCSN, and GCD datasets.

In the pre-processing stage, images from all three datasets are resized to 125×125 pixels and then subjected to central area extraction of 100×100 pixels. Traditional augmentation layers for random flipping and rotation are added to make the models more robust.

3. METHODOLOGY

As noted before, most sky/cloud image classification datasets have small cardinality. Hence, transfer learning (TL) is used in this paper to train a largely successful deep learning model for image classification, namely, VGG-16 [11]. We use the pre-trained model weights on the IMAGENET dataset [12], which is one of the biggest and most complex dataset for image classfication. A model trained on a complex dataset like IMAGENET is expected to perform better on a relatively simpler task of cloud image classification. The study is divided into two parts. While the first part assesses the effectiveness and benefits of TL, the second part aims to understand the impact of the number of neurons in the top dense layers.

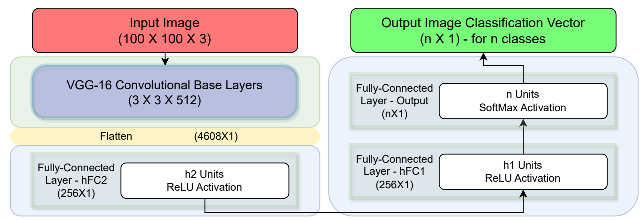


Fig. 3: Network architecture based on VGG-16 base convolutional layers used in this study. The number of units, i.e. h1 and h2 respectively, in hFC1 and hFC2 layers are variable. Number of units in the output layer (n) depend on the class count of the dataset.

For the first part, the SWIMCAT dataset was used for the experiments. According to the TL regime, the convolutional base layers of the VGG-16 architecture are used as is. In the original architecture of VGG-16, there are 2 fully connected (FC) hidden layers of 4096 neural units each and ReLU activation. We will refer to these layers as hFC1 and hFC2, as shown in Fig. 3. These layers are followed by another FC output layer with softmax activation. However, we have observed that completely removing the hFC2 layer and reducing the number of units in the hFC1 layer can significantly improve the performance of the network. The results are compared with the state-of-the-art CloudNet [8] and CloudA [9] architectures. Furthermore, the standard VGG-16 network [11] was trained for the task with and without TL. For all experiments, cosine decay with restarts was used after setting the initial learning rate $\eta_{init} = 10^{-6}$ and the minimum learning rate $\eta_{min} = 10^{-7}$. Eq. 2 shows the computation of the learning rate η in each epoch ep. Finally, early stopping was used to avoid overfitting by monitoring the categorical cross-entropy loss on the validation set. The held-out test set comprises randomly selected 25% images from the dataset.

$$t_{\max} = 100 \times (2.5^{\max((ep-100)\,,0)}) \tag{1}$$

$$\eta = \eta_{\min} + (\eta_{init} - \eta_{\min}) \times \frac{1 + \cos\left(\pi \frac{ep}{t_{\max}}\right)}{2} \qquad (2)$$

For the second part, experiments are performed with similar settings as before, on all three datasets with all possible combinations of $\{$hFC1, hFC2$\}$ with hFC1 $\in \{64, 128, 256, 512, 1024, 2048, 4096\}$ and hFC2 $\in \{0, 64, 128, 256, 512, 1024, 2048, 4096\}$. Here, 0 units in hFC2 would mean that we have completely omitted that layer. These experiments are done to fully understand the role of top dense layers in the case of cloud image classification using TL.

4. RESULTS

Table 1 shows that the transfer learning-based VGG-16 model gave results on par or better than the state-of-the-art models with significantly fewer epochs. Furthermore, reducing the number of neurons in the hFC layers also leads to a significant improvement in accuracy.

For the second part, Fig. 4 shows the heatmap of the accuracies obtained on the test set for the SWIMCAT dataset. Lighter hues (indicating better results) can be clearly seen in the top-left half of the heatmap. This means that for the SWIMCAT dataset, adding more neurons in the hFC layers of VGG-16 could lead to overfitting. Additionally, darker hues at the very top-left corner indicate that over-reducing the number of these neurons might lead to excessive information loss.

Similarly, the heatmaps that were obtained in the CCSN and GCD datasets are shown in Fig. 5. Remember that the number of classes is higher in the CCSN dataset and that it consists of more complex images. Therefore, in contrast to Fig. 4, lighter hues can now be seen in Fig. 5(a) when the number of neurons in the hFC layers was higher, that is, in the bottom-right half of the heatmap. Contrary to larger accuracy variations in SWIMCAT and CCSN, the variations for GCD dataset, upon changing the number of neurons in hFC1 and hFC2 layers, are very small. This is probably because of the large size of the GCD dataset.

Overall, it can be noted that the results for the optimized TL model are comparable to the state-of-the-art for GCD (88.3% on 6-classes by TL as compared to 84.3% on 7-classes by Liu *et al.* [5]); possibly worse on CCSN; but much better for SWIMCAT (99.5% by TL as compared to 98.47% by Wang *et al.*'s [9]). This shows that TL is more effective for datasets with low cardinality like SWIMCAT.

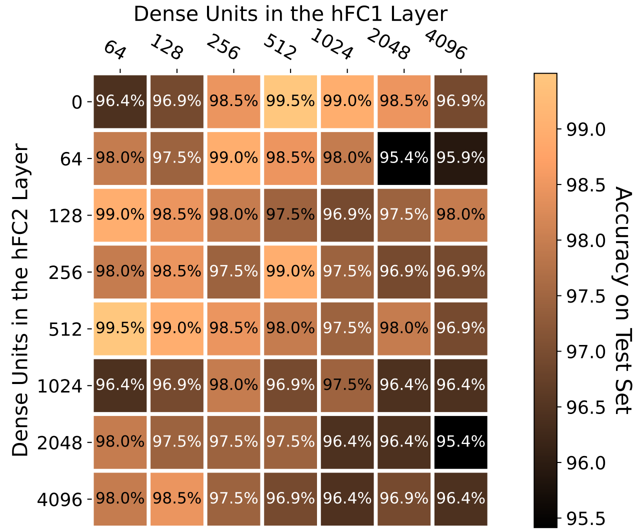


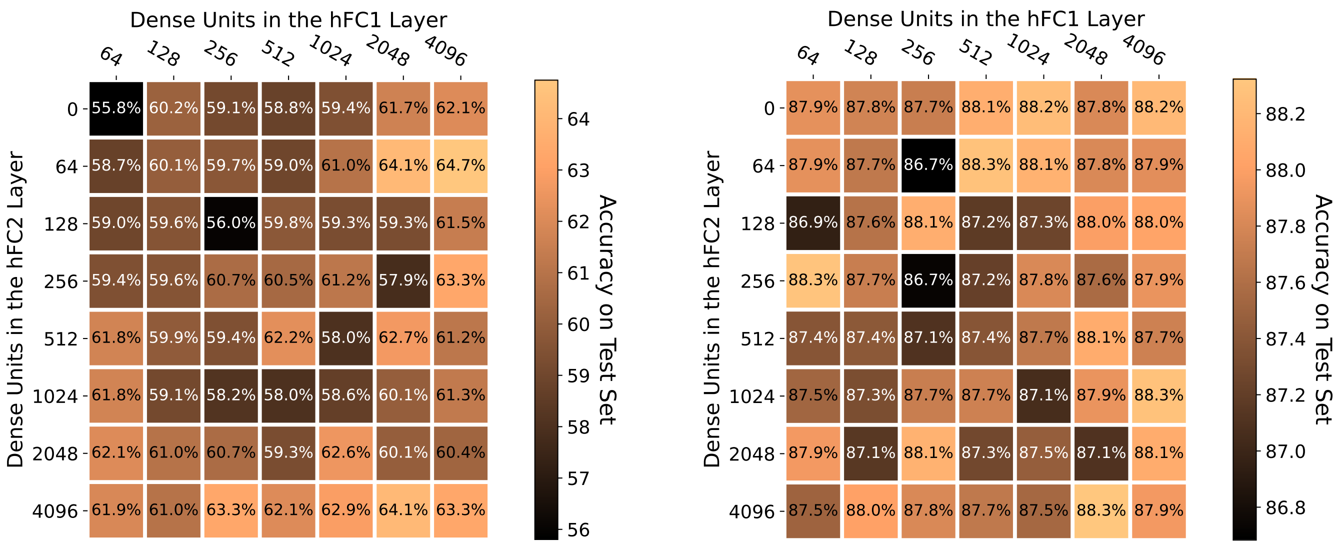
Fig. 4: Heatmap showing the grid-search results obtained on the SWIMCAT dataset when different models were trained with different combinations of neurons in the hFC1 and hFC2 layers.

5. CONCLUSION AND FUTURE WORK

This paper presents the effective use of transfer learning (TL) for the cloud image classification task. The paper notes that with TL, not only is the training time significantly reduced, but the results are also on par with or better than the state-of-the-art custom architectures for datasets with low cardinality. Additionally, the paper performs an extensive grid search study to understand the impact of the number of neurons in the top dense layers (hFC) on the performance of TL models. The paper notes that for simpler datasets and fewer classes, less number of neurons are preferred in the hFC layers. Whereas for more complex datasets and more number of classes, more number of hFC neurons produce better results. Consequently, this paper recommends tuning for hFC neurons in conjunction with other hyperparameters. Although this paper highlights the effectiveness of TL in cloud image classi-

Network Architecture	Avg. Training Time (/epoch)	Number of epochs	Accuracy on Test Set
CloudNet [8]	3.4 seconds	2480	96.97% (98.33%)
CloudA [9]	1.01 seconds	1000	97.47% (98.47%)
VGG$-$16 (hFC1 and hFC2 with 4096 units each)	1.15 seconds	1249	95.45%
TL-VGG$-$16 (hFC1 and hFC2 with 4096 units each)	1.15 seconds	843	96.46%
TL-VGG$-$16 (Only hFC1 with 512 units)	1.12 seconds	648	$\mathbf{99.49}\%$

Table 1: Average time per epoch and number of epochs that were required during training of the different deep CNN architectures on the SWIMCAT dataset. The accuracy of the test set is also reported in the last column. The prefix **TL-** means that transfer learning was used in that case. Values within parentheses '()' indicate the accuracy on the test set that is claimed by the original authors.



(a) Grid-search results over CCSN

(b) Grid-search results over GCD

Fig. 5: Heatmap showing the accuracy results obtained on the (a) CCSN and (b) GCD datasets when different models were trained with different combinations of neurons in the hFC1 and hFC2 layers.

fication problems with proper hyperparameter tuning, the authors would like to extend this study to other standard deep learning models apart from VGG-16. Also, it will be interesting to see if the importance of the number of hFC neurons persists in other architectures such as ResNet, Inception, and Xception.

6. REFERENCES

[1] F. Yuan, Y. H. Lee *et al.*, "Comparison of radio-sounding profiles for cloud attenuation analysis in the tropical region," in *Proc. IEEE International Symposium on Antennas and Propagation*, 2014, pp. 259–260.

[2] S. Dev, F. M. Savoy *et al.*, "Estimation of solar irradiance using ground-based whole sky imagers," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, pp. 7236–7239.

[3] M. Jain, I. Gollini *et al.*, "An extremely-low cost ground-based whole sky imager," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2021, pp. 8209–8212.

[4] M. Jain, N. Jain *et al.*, "Detecting blurred ground-based sky/cloud images," in *Proc. IEEE USNC-URSI Radio Science Meeting (Joint w/AP-S Symposium)*, 2021, pp. 62–63.

[5] S. Liu, L. Duan *et al.*, "Ground-based remote sensing cloud classification via context graph attention network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[6] T. Kliangsuwan and A. Heednacram, "Feature extraction techniques for ground-based cloud type classification," *Expert Systems with Applications*, vol. 42, no. 21, pp. 8294–8303, 2015.

[7] S. Dev, Y. H. Lee *et al.*, "Categorization of cloud image patches using an improved texton-based approach," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 422–426.

[8] J. Zhang, P. Liu *et al.*, "CloudNet: Ground-based cloud classification with deep convolutional neural network," *Geophysical Research Letters*, vol. 45, no. 16, pp. 8665–8672, 2018.

[9] M. Wang, S. Zhou *et al.*, "CloudA: A ground-based cloud classification method with a convolutional neural network," *Journal of Atmospheric and Oceanic Technology*, vol. 37, no. 9, pp. 1661 – 1668, 2020.

[10] M. Hussain, J. J. Bird *et al.*, "A study on cnn transfer learning for image classification," in *Advances in Computational Intelligence Systems*, A. Lotfi, H. Bouchachia *et al.*, Eds. Cham: Springer International Publishing, 2019, pp. 191–202.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. International Conference on Learning Representations (ICLR)*, 2015.

[12] J. Deng, W. Dong *et al.*, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.