Thai Journal of Mathematics
Volume 19 Number 3 (2021)
Pages 1068–1081

http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

Deep Transfer Learning for Automated Artillery Crater Classification

Sani Aji[1,2], Poom Kumam[3,*], Punnarai Siricharoen[4], Ali Maina Bukar[5] and Mohammed Sani Adamu[6]

[1] *KMUTTFixed Point Research Laboratory, Room SCL 802 Fixed Point Laboratory, Science Laboratory Building, Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha-Uthit Road, Bang Mod, Thung Khru, Bangkok 10140, Thailand*
e-mail : ajysani@yahoo.com (S. Aji)

[2] *Department of Mathematics, Faculty of Science, Gombe State University, Gombe 760214, Nigeria.*

[3] *Center of Excellence in Theoretical and Computational Science (TaCS-CoE) & KMUTTFixed Point Research Laboratory, Room SCL 802 Fixed Point Laboratory, Science Laboratory Building, Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha-Uthit Road, Bang Mod, Thung Khru, Bangkok 10140, Thailand*
e-mail : poom.kum@kmutt.ac.th (P. Kumam)

[4] *Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Phayathai Road, Wang Mai, Pathumwan Bangkok 10330 Thailand*
e-mail :punnarai.s@chula.ac.th (P. Siricharoen)

[5] *Centre for Visual Computing, Faculty of Engineering and Informatics, University of Bradford, Bradford, West Yorkshire, United Kingdom*
e-mail :alimainabukar@gmail.com (A. B. Maina)

[6] *School Of Remedial and General studies, Department of General Studies, Federal Polytechnic Damaturu,. PMB 1006 Damaturu Yobe State, Nigeria*
e-mail :muhammadsani1019@gmail.com (M.S Adamu)

Abstract Analysis of artillery craters is an indispensable tool due to its importance in military investigation and other precautionary purposes. It is used to verify suspected locations of hostile fire, detect the presence of enemy ammunition as well as tracing the direction from which weapons are fired. Traditionally, crater analysis is done manually. Not only that the manual process consumes time, but also, they are prone to errors and sometimes, result in loss of crater information. Hence, the need for automation. Recently, deep neural networks have recorded great breakthroughs and achieved promising results in different applications. In this paper, we propose the use of deep convolution neural networks (DCNNs) for the automatic classification of artillery craters. Specifically, due to the small size of training data, transfer learning is employed. We used satellite images to fine tune five pretrained DCNN models, VGG16, VGG19, ResNet50, MobileNet and EfficientNetB0. Interestingly, all these pretrained models achieved impressive classification accuracy in classifying craters from other environmental potholes. Our proposed approach is automatic, and can assist military and other human right investigation in automatic crater analysis instead of the manual ways.

MSC: 68T07; 90C90
Keywords: Artillery Craters; Classification; Deep Learning; Transfer Learning

Submission date: 30.04.2021 / Acceptance date: 27.07.2021

1. Introduction

Artillery craters are marks left on the ground after explosion of heavy artillery projectiles. These craters play a vital role for investigating suspected locations of hostile fire, identifying the presence of other ammunition and tracing their trajectories to know the directions from which they are fired. Depending on the artilleries, craters are of different types including Low-Angle Fuze-Quick Craters (LAFCs), Low-Angle Fuze-Delay Craters, Mortar Craters, among others. LAFCs leave some distinctive arrow-head shape and are characterized by their movement in low angle and their quick explosion immediately they touch the ground [1].

Nowadays, a lot of attacks and bombardments occur in different places of the world, and sometimes, one challenging task is the ability to figure out the perpetrators behind the attacks. With crater analysis, it is possible to study and analyse the shapes left on the ground by the exploded projectiles and eventually, the direction from which the weapons are fired can be traced. To achieve this, there is need to firstly classify the craters accurately and discriminate them from other confusing environmental potholes.

One of the advantages of crater analysis is its applicability in military operation to investigate environments and safeguard themselves from enemy attacks in a battle field. Most importantly, with the distinctive arrow-head shape of LAFCs, investigation team can trace the direction of an unknown attack. Traditionally, there are two methods involved in identifying and classifying craters as well as tracing the weapons direction. The first method is the fuze furrow and center of crater method, and the second one is the side spray method [1]. Both of these methods involve using compass and stakes manually. Although the manual methods have been helpful, they are prone to errors and consume a lot of time. Moreover, manual crater identification lead to loss of crater information because craters exposed to weather and personnel erode overtime and blend into the earth surface [1].

Our motivation in this work is that, the distinctive features and shapes of the LAFCs make it possible to analyse them and figure out the direction from where artillery weapon emerged, however, the initial step to achieve this is to firstly classify the craters and discriminate them with other confusing environmental objects. The manual methods used are prone to errors, consume time and lead to loss of crater information. Therefore, there is need for better alternatives. With the recent advancements and the great breakthroughs in the field of computer vision, particularly, the success of DCNNs in these years, we propose the use of off-the-shelf pretrained models to automate the classification of LAFCs. The contributions of this work include the following:

- The use of machine learning training procedures to achieve excellent results in LAFCs analysis.
- Performance of five state of-the-art DCNNs is compared and evaluated on LAFCs dataset.

*Corresponding author.

Published by The Mathematical Association of Thailand.

2. Related Work

As mentioned earlier, craters are of different types and due to the importance of their analysis, a lot of works have been proposed to classify and detect different types of craters for different purposes. Wetzler et al [2] used planetary images to detect small impact craters using machine learning techniques. In their work, they employ different supervised learning algorithms such as support vector machine, continuously-scalable template matching (CSTM) algorithms and some ensemble methods (Bagging and Adaboost) to detect craters from groundtruth images. Their results were evaluated on a set of Viking orbiter images of Mars and 60% accuracy was achieved using SVM with image patches. Although this result is better compare to other methods, but it is just above chance.

Ding et al. [3] used boosting and transfer learning and presented an integrated autodetection approach of subkilometer craters. Their approach consist of three steps, first, identifying crater candidate using the concept of mathematical morphology. This is followed by extracting and selecting image features, then classifying the crater candidates with ensemble learning algorithms. The final step integrate transfer learning into boosting to enhance detection performance. They achieved more than 85% F1-score which is a significant improvement in comparison with other algorithms.

Currently, researchers have been employing the advanced techniques of DCNNs in classification and detection tasks to achieve better accuracy. Christoff et al. [4] proposed a novel approach for crater morphological classification based on 3D triangulated mesh of Mars sample. They used curvature analysis and local quantization method together with convolutional neural networks to classify impact craters into valid, secondary and degraded crater categories. With this approach, they achieved 88.3% accuracy in classifying the craters into the stated three catergories.

Emami et al. [5] proposed an extension of their work in [6] on crater classification and evaluated the prformance of deep neural networks in lunar crater detection. Specifically, they compare the performance of three pretrained models, VGG16, GoogleNet and Resnet models for the classification tasks. The classification accuaracy achieved by their approach demonstrate the strong capacity of DCNNs in classifying craters and non-craters. Although a classification accuracy of about 99% is achieved, the approach is specifically for lunar craters.

Recently, Aji et al. [7] used transfer learning to classify LAFCs where three pretrained models are used for feature extraction, and then, the extracted features are fed into a support vector machine for classification. It is clear from the literature that most of the studies focus specifically on the craters from celestial bodies. Our aim in this work is to use DCNNs, specifically, to employ the idea of transfer learning to fine tune diffrent pretrained models for LAFCs classification due to their interesting shape and the usefulness of their features in studying the direction of unknown attacks.

Deep neural networks have gained popularity in recent years due to their exceptional performance in image classification and object recognition, and they have contributed significantly in object detection, speech recognition [8–11] and medical imaging [12, 13]. In neural networks, an input image is passed from the input layer to some hidden layers before reaching the output layer for the classification results [14]. However, unlike the traditional neural networks, CNNs varies slightly as different inputs share weights instead of having single weight [15]. Therefore, in CNNs, hierarchical layers are used to identify low level to high level features which make them capable of learning complex abstractions used in making intelligent decisions. With all these, DCNNs depends largely on huge amount of data to be able to train and perform well. When there is no enough dataset to train DCNNs, data augmentation and transfer learning are used. Data augmentation is a method used to increase the diversity of the dataset by producing more training data from the existing dataset. Transfer learning involve using knowledge learned from a task for which there is a huge available labelled training data into a settings in which the labelled dataset is limited. To conduct transfer learning, two strategies can be employed.

- Use pretrained network for feature extraction and feed the extracted features to a machine leaning classifier
- Fine-tuning, that is retraining few layers of a pretrained network using the data under question.

Using pretrained networks have proved to be an effective approach in image classification and other deep learning applications. Researchers use models trained on some tasks having large amount of labelled data (usually the popular ImageNet dataset [16]) and transfer the learned features to other different tasks especially when there is limited amount of labelled data. Transfer learning solves the problem of data limitation and does not require large computation power because weights of pretrained models are used, so, only few last dense layers are trained. Moreover, despite training on small amount of data, transfer learned model seems to train at high speed. Usually, to perform transfer learning, models trained on the well-known ImageNet dataset (ILSVRC) [16] are used as base models. ImageNet is a large dataset of anotated images developed specifically for computer vision consisting of more than 14 million images with more than 21 thousand classes. Researchers used subsets of the Imagenet dataset for the annual ILSVRC competition using around 1 million images with 1000 categories. Some of the most popular models trained on Imagenet includes AlexNet [17], VGG16 and VGG19 [18], ResNet [9], MobileNet [19] and EfficientNets [20] among others.

In this paper, due to relative small size of the dataset and the fact that fine tuning involves retraining just few layers of a pretrained model, transfer learning trains at a high speed, thus, we propose using transfer learning-based approach to classify LAFCs using satellite images. This fully automated approach uses five pretrained models (VGG16, VGG19, ResNet50, MobileNet and EfficientNet-B0) as the base models, and after fine tuning these pretrained models, their performance on our crater dataset is also compared. Most importantly, this work is the first attempt to classify these special types of craters with end-to-end structure on satellite images without using any hand-crafted feature extraction.

3. Materials and Methods

In this section, the description of the pretrained models leveraged during training and details about the dataset used is given. Instead of training from scratch, these models were already trained to solve similar classification task. They are trained on the ImageNet dataset having more than 1 million labelled images with 1000 classes, and are also the state of-the-art networks used in computer vision to achieve different goals. Below, we give a brief description of the pretrained models fine tuned in this work.

3.1. Visual Geometry Group Network (VGGNet)

The VGGNet [18] was developed by VGG researchers at the university of Oxford. This model scored first in image localisation task and second in image classification task at the ILSVRC challenge in 2014. The architecture was designed by checking the effect of network depth on its performance. VGGNet-16 consist of 13 convolutional layers each equipped with filters of size 3 x 3, and 3 fully-connected layers. One of the advantage of VGGNet is that the use of smaller sized filters in the convolutional layer resulted in effective field increment and parameter reduction in the network. A more deeper version of the VGG16 is the VGG19 which has 16 convolutional layers and 3 fully-connected layers.

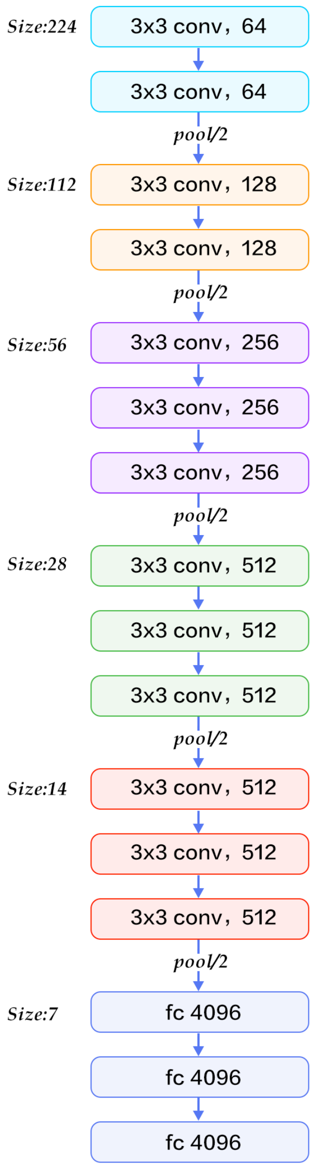

FIGURE 1. VGGNet Architecture

3.2. Residual Network (ResNet)

One major problem in training deeper networks is the vanishing gradient and degradation [21, 22]. As neural networks are trained through back propagation with gradient descent, when there are many layers, repeated multiplication make the gradient to become smaller and smaller until it disappears which results in performance saturation and degradation with each additional layer. ResNet architecture [23] developed by Microsoft research team in 2015 came with a solution to this problem. This architecture uses a residual networks that are easier to optimize, making a shortcut connection by skipping one or more convolution layers, taking the input of the previous layers and adding it to the result of the next skipped layers generating input for the remaining layers [9]. ResNet architectures gain more better accuracy by increasing the depth, thus producing better results compare to other shallower networks. ResNet50 is 50 layers deep and other ResNet variants have even more deeper layers such as 101 and 152.

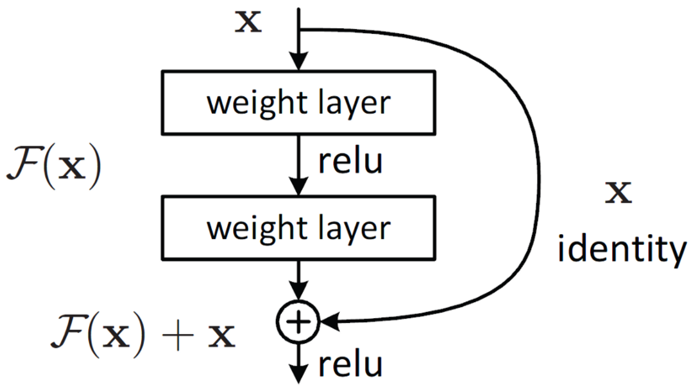

Figure 2. A Building Block of the ResNet Architecture

3.3. MobileNets

The MobileNets are developed by Google researchers [19] based on a streamlined architecture that uses the idea of depth wise separable convolutions in order to build a light weight DCNN. This light weight network provide an efficient models that can be used for mobile and other embedded vision applications. The depth wise separable convolution filters in MobileNets consist of depth wise convolution filters which perform a single convolution on each of the input channel, and a point wise convolution filters which combine the output of the depth wise convolution filters linearly with a 1×1 convolutions. Two parameters namely, a width multiplier and a resolution multiplier are introduced in the MobileNet to efficiently trade-off between latency and accuracy. The width and resolution parameters are responsible for controlling the input width of a layer and the resolution of the input image respectively. The depth wise separable convolution step and the architecture of the MobileNet are shown in Figures 3 and 4 respectively.

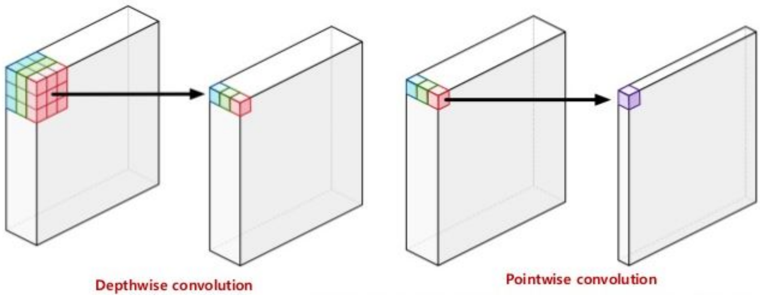

FIGURE 3. Depth wise separable convolution step

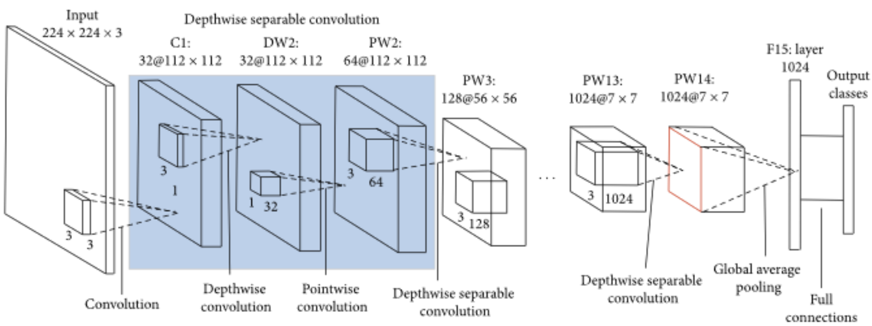

FIGURE 4. The MobileNet Architecture

3.4. EfficientNets

Since the introduction of ResNets, scaling up DCNNs proved to result in better accuracy. In ResNets [9], using the idea of residual blocks, the authors showed the possibility of going deeper and deeper by scaling the depth of the network. However, this process of scaling in ResNet was focusing only on the depth. Zagoruyko and Komodakis [24] proposed scaling convolution neural networks using the width. It is observed that network accuracy improves by increasing the depth, but every fraction of an increase lead to doubling the number of layers. This makes the training of deeper networks to have a diminishing feature reuse problem, which slow their training [24]. As a results, the authors in [24] proposed scaling up the ResNet model by increasing the width and reducing the depth. These types of scaling indicated that focus is given to only one factor. Tan and Le [20] observed that balancing the network depth, width and resolution can lead to better performance. Hence, they proposed a new scaling method which uniformly scales all the dimensions using a compound coefficient. Based on this, they design a new network using neural architecture search and scaled it up which give rise to the family of the EfficientNets (B0-B7). The baseline architecture of this family is the EfficientNetB0 which is a mobile size architecture that uses inverted residual blocks (MBConv) introduced in MobileNetV2 [25]. These blocks also use squeeze and excitation block along with swish activation function.

Stage i	Operator $\hat{\mathcal{F}}_i$	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels $\hat{C}_i$	#Layers $\hat{L}_i$
1	Conv3x3	224×224	32	1
2	MBConv1, k3x3	112×112	16	1
3	MBConv6, k3x3	112×112	24	2
4	MBConv6, k5x5	56×56	40	2
5	MBConv6, k3x3	28×28	80	3
6	MBConv6, k5x5	28×28	112	3
7	MBConv6, k5x5	14×14	192	4
8	MBConv6, k3x3	7×7	320	1
9	Conv1x1 & Pooling & FC	7×7	1280	1

Figure 5. EfficientNetB0 Network

3.5. Data Collection

The dataset consist of satellite images collected via google earth from the territory of Ukrain and its border with Russia between July and September, 2014. The positive (craters) and negative (non-craters) samples are 2,405 each, making the total amount of the dataset to be 4,810. The non-craters are also satellite images consisting of objects such as potholes, decayed grasses and stumps that can be confused as craters easily by human observer. Samples of craters and non-craters are given in Figure 6 and 7 respectively.

Figure 6. Crater samples from the dataset

Figure 7. Non-crater samples from the dataset

3.6. Data Augmentation

Due to the limited amount of crater dataset, data augmentation was utilised, thereby artificially increasing the dataset. Augmentation was leveraged in order to decrease over fitting and hence increase the models ability to generalise. Data augmentation is an essential deep learning technique that proved to reduce over fitting by increasing the diversity of the dataset. In our experiment, data augmentation techniques implemented include but not limited to random horizontal and vertical flipping, zooming, and rotations (through -10 to 10 degrees).

4. Experiments and Results

In this section, experiments and results obtained during evaluation are presented. It is worth mentioning that the experiments are implemented using Keras application programming interface on Google Colaboratory utilizing 12 GB NVIDIA Tesla K80 GPU. As mentioned earlier, Our dataset consists of 4,810 satellite images of craters and non-craters. This dataset is randomly split into 80% for training and the remaining 20% is used for testing.

Since our dataset is small in size, training a deep neural network from scratch will likely lead to overfitting. Thus, transfer learning was employed. Five sets of experiments are conducted via utilizing the weights of five pretrained models, the VGG16, VGG19, ResNet50, MobileNet and EfficientNetB0. As an initial step, the last fully connected layer of each of these models was discarded, then two dense layers separated by a ReLu activation function were appended to the networks. Afterwards, softmax activation function was used to output the classification predictions. During training, only these custom layers are trained. Hence parameters of subsequent layers are frozen, thus they are not updated using gradient descent. The optimization algorithm of choice was Adam, due to its proven efficiency. Table 1 shows the summary of the results obtained as well as the average time taken after fine tuning each of the pretrained models. Each of the models is trained for 20 epochs with batch size of 64.

Table 1. Accuracy and Average Time Taken by the Fine Tuned Models

Fine Tuned Models	**Accuracy(%)**	**Average Time(secs)**
VGG16	99.79	32.25
VGG19	99.48	40.25
MobileNet	98.13	7.00
ResNet50	100	54.9
EfficentNetB0	100	6.10

From Table 1, it can be observed that all the fine tuned models achieved impressive classification accuracy in discriminating craters from the non-crater samples in an average of some seconds. ResNet50 and EfficientNetB0 achieved excellent results, recording 100% classification accuracy as opposed to the other three models with lesser accuracy. Although VGG19 takes longer training time than its variant as it is deeper, its accuracy is not as that of the VGG16. This maybe as a result of the small size of our training data. MobileNet has advantage of training time, however, not as accurate as the rest of the models. This is not surprising as one of the advantages of the network is to trade off between latency and accuracy. Thus, despite being a very light model with fewer parameters, it trains at lesser time and the accuracy is just about 1% less compare to the VGG networks which have much more millions of parameters. In all, EfficientNetB0 happened to perform better with least training time and excellent classification accuracy than the rest of the models which may be due to the scaling method employed in the network which made it not only lighter but also efficient. In Figure 8 below, the graphical results of the models showing both accuracy and average time is presented.

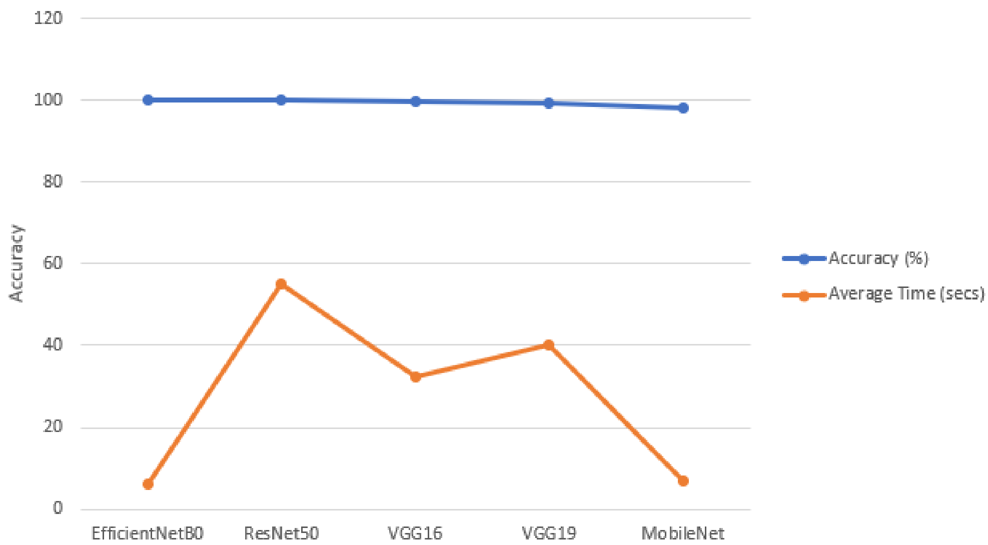

FIGURE 8. Models Accuracy and Average Time Taken

In order to visualize and see how all the fine tuned pretrained models did in making prediction of craters and non-craters on the validation data, we plot confusion matrices as shown in Figure 9, the confusion matrices show the respective accuracy of each of the five fine tuned models in predicting craters and non-craters using the 20% validation data.

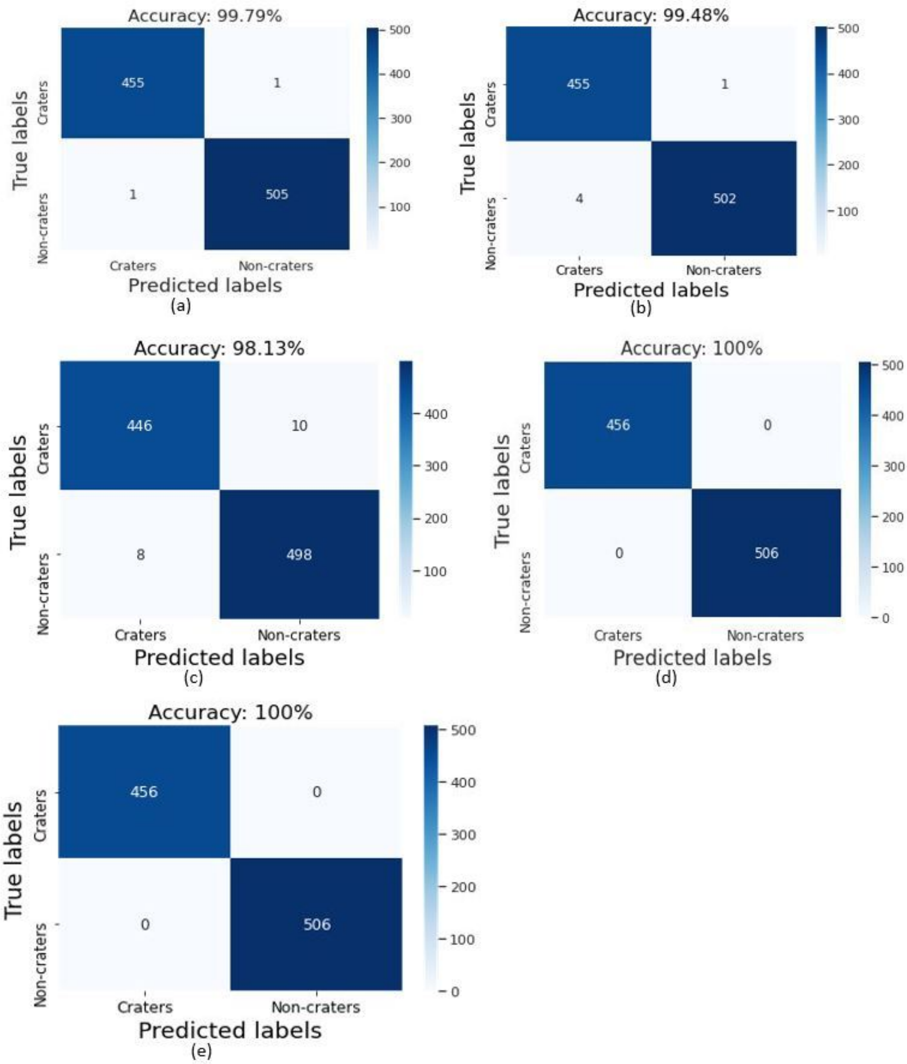

FIGURE 9. Confusion Matrices showing the Accuracy of each of the fine tuned models on Validation Data: a) VGG16 confusion matrix b) VGG19 confusion matrix c) MobileNet confusion matrix d) ResNet50 confusion matrix e) EfficientNetB0 confusion matrix

From the above plots, the blue cells in the diagonal show the number of accurately predicted validation data. While the other white cells represent the incorrectly classified samples. In total, our validation data consists 962 craters and non-craters, and from the confusion matrix in Figure 9 a), it is clearly observed that with VGG16, the model predicted 960 craters and non-craters accurately. Only two samples are incorrectly predicted. In b), VGG19 predicted 957 correctly and reported 5 misclassifications. Figure 4 c) shows the prediction by MobileNet which has slighly more misclassifications compare to the rest of the models. Finally, in Figure 4 d) and e), excellent prediction is obtained by ResNet50 and EfficientNetB0. Using these two as base models, all the 962 validation samples are correctly classified into their respective classes. Hence, for better accuracy, both of these networks can be utilized, however, the EfficientNetB0 has the advantage of fast training than the ResNet50.

In addition, precision, recall and F1-score are also calculated from the confusion matrices in order to further evaluate the performance of our approach. Precision simply refers to the proportion of true positives (craters predicted as craters) in all the samples that have been predicted to be craters. On the other hand, recall also known as sensitivity is the proportion of the identified positives (craters) from the samples that are positively craters. F1 Score calculates the weighted average of Precision and Recall. The computed metrics are shown in table 2 below.

TABLE 2. Precision, Recall and F1-Score

CNN Models	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
VGG16	99.78	99.78	99.78
VGG19	99.13	99.78	99.45
MobileNet	98.24	99.78	99.45
ResNet50	100	100	100
EfficientNetB0	100	100	100

In general, the proposed approach of using state of the art DCNNs achieved impressive accuracy in classifying craters from other environmental potholes despite it being a challenging problem to a human observer. All the fine tuned models performed very well on our crater dataset. However, the fine-tuned ResNet50 and EfficientNetB0 slightly outperformed the other three. Also, high precision, recall and F1-score are obtained from each of the fine tuned models. The results obtained in this work have shown that the manual crater classification can be automated using DCNNs and with the proposed transfer learning approach, the models are trained very fast and high accuracy is achieved. Thus, the approach can serve as an alternative to the manual crater analysis for faster and better accuracy.

5. Conclusions

In this paper, we employed the use of different pretrained deep learning models for the automation of artillery crater classification. Our proposed approach have achieved promising results in classifying artillery craters from other environmental potholes. Crater classification as the first step to computation of trajectories for tracing the direction of an unknown attack, we foresee this as an alternative to the error-prone and tedious manual crater classification methods. We are optimistic that this work will play a vital role in human right and other military investigations. For future research, our plan is on crater detection with eventual plan to compute the crater trajectory in order to trace the direction of unknown artillery projectiles attacks.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgements

The authors acknowledge the support provided by the Center of Excellence in Theoretical and Computational Science (TaCS-CoE), KMUTT and Center under Computational and Applied Science for Smart Innovation research Cluster (CLASSIC), Faculty of Science, KMUTT. The first author was supported by the Petchra Pra Jom Klao Ph.D. research Scholarship from King Mongkut's University of Technology Thonburi (KMUTT) Thailand (Grant No. 19/2562).

References

[1] Field Manual, Manuals Combined: Tactics, Techniques, and Procedures for Field Artillery Meteorology And Field Artillery Target Acquisition. Jeffrey Frank Jones, 2002.

[2] P.G. Wetzler, R. Honda, B. Enke, W.J. Merline, C.R. Chapman, M.C. Burl, Learning to detect small impact craters, IEEE, 1 (2005) 178–184.

[3] W. Ding, T.F. Stepinski, Y. Mu, L. Bandeira, Subkilometer crater discovery with boosting and transfer learning, ACM Transactions on Intelligent Systems and Technology (TIST) 2 (4) (2011) 1–22.

[4] N. Christoff, A. Manolova, L. Jorda, and J. Mari. Morphological crater classification via convolutional neural network with application on mola data, In ANNA18; Advances in Neural Networks and Applications (2018) 1–5.

[5] E. Emami, T. Ahmad, G. Bebis, A. Nefian, T. Fong, On crater classification using deep convolutional neural networks. In Lunar and Planetary Science Conference 49 (2018).

[6] E. Emami, T. Ahmad, G. Bebis, A. Nefian, T. Fong, On crater verification using mislocalized crater regions. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV) (2017) 1098–1104.

[7] S. Aji, P. Kumam, P. Siricharoen, A.B. Maina, Automatic Classification of Low-Angle Fuze-Quick Craters Using Deep Learning." In International Conference on Machine Learning, Optimization, and Data Science, Springer, Cham, (2020) 436–447.

[8] Y. LeCun, Y Bengio, and G, Hinton, Deep learning. nature, 521 (7553) (2015) 436–444.

[9] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (2016) 770–778.

[10] H.M. Fayek, M. Lech, L. Cavedon, Evaluating deep learning architectures for speech emotion recognition. Neural Networks 92 (2017) 60–68.

[11] S. Thongsuwan, S. Jaiyen, A. Padcharoen, P. Agarwal, ConvXGB: A new deep learning model for classification problems based on CNN and XGBoost, Nuclear Engineering and Technology 53 (2) (2021) 522–531.

[12] T.C. Chiang, Y.S. Huang, R.T. Chen, C.S. Huang, and R.F. Chang, Tumor detection in automated breast ultrasound using 3-d cnn and prioritized candidate aggregation, IEEE transactions on medical imaging, 38 (1) (2018) 240–249.

[13] K.H. Yu, F. Wang, G.J. Berry, C. Re, R.B. Altman, M. Snyder, I.S. Kohane. Classifying non-small cell lung cancer histopathology types and transcriptomic subtypes using convolutional neural networks, bioRxiv, (2019) 530360.

[14] K. Hornik, M. Stinchcombe, H. White, et al. Multilayer feedforward networks are universal approximators. Neural networks, 2 (5) (1089) 359–366.

[15] Y. Le Cun, O. Matan, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jacket, H.S. Baird. Handwritten zip code recognition with multilayer networks, In Proceedings. 10th International Conference on Pattern Recognition, 2 (1990) 35–40.

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge, International journal of computer vision 115(3) (2015), 211–252.

[17] A. Krizhevsky, I. Sutskever, G. E. Hinton. Imagenet classification with deep convolutional neural networks, In Advances in neural information processing systems (2012) 1097–1105.

[18] K. Simonyan, A. Zisserman. Very deep convolutional networks for largescale image recognition. arXiv preprint arXiv, (2014) 1409 1556.

[19] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv: (2017), 1704.04861.

[20] M. Tan Q.V. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, 2020.

[21] M. Habibzadeh, M. Jannesari, Z. Rezaei, H. Baharvand, M. Totonchi. Automatic white blood cell classification using pre-trained deep learning models: Resnet and inception. In Tenth International Conference on Machine Vision (ICMV 2017), 10696, 1069612. International Society for Optics and Photonics 2018.

[22] O. Wichrowska, N. Maheswaranathan, M.W. Hoffman, S.G. Colmenarejo, M. Denil, N. de Freitas, J. Sohl-Dickstein, Learned optimizers that scale and generalize, In Proceedings of the 34th International Conference on Machine Learning, 70 (2017) 3751–3760.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (2015) 1–9.

[24] S. Zagoruyko and N. Komodakis, Wide residual networks, In Edwin R. Hancock Richard C. Wilson and William A. P. Smith, editors, Proceedings of the British Machine Vision Conference (BMVC) (2016) 87.1–87.12.

[25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L. C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition, (2018) 4510–4520.