

Thai culture image classification with transfer learning

Munlika Rattaphun, Kritaphat Songsri-in

Department of Computer Science, Faculty of Science and Technology, Nakhon Si Thammarat Rajabhat University,
Nakhon Si Thammarat, Thailand

Article Info

Article history:

Received Feb 2, 2023

Revised Apr 3, 2023

Accepted Apr 7, 2023

Keywords:

Convolutional neural network

Deep learning

Image classification

Thai culture

Transfer learning

ABSTRACT

Classifying images of Thai culture is important for a variety of applications, such as tourism, education, and cultural preservation. However, building a Machine learning model from scratch to classify Thai cultural images can be challenging due to the limited availability of annotated data. In this study, we investigate the use of transfer learning for the task of image classification on a dataset of Thai cultural images. We utilize three popular convolutional neural network models, namely MobileNet, EfficientNet, and residual network (ResNet) as baseline pre-trained models. Their performances were evaluated when they were trained from random initialization, used as a feature extractor, and fully fine-tuned. The results showed that all three models performed better in terms of accuracy and training time when they were used as a feature extractor, with EfficientNet achieving the highest accuracy of 95.87% while maintaining the training time of 24 ms/iteration. To better understand the reasoning behind the predictions made by the models, we deployed the gradient-weighted class activation mapping (Grad-CAM) visualization technique to generate heatmaps that the models attend to when making predictions. Both our quantitative and qualitative experiments demonstrated that transfer learning is an effective approach to image classification on Thai cultural images.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Kritaphat Songsri-in

Department of Computer Science, Faculty of Science and Technology, Nakhon Si Thammarat Rajabhat University

1 Moo 4, Tha Ngio, Mueang Nakhon Si Thammarat, Nakhon Si Thammarat Province, 80280, Thailand

Email: kritaphat_son@nstru.ac.th

1. INTRODUCTION

Deep learning is a subfield of machine learning that is inspired by the structure and function of neural networks in the brain. It involves training artificial neural networks on a large dataset, allowing the network to learn and make intelligent decisions on its own. One area where deep learning has been particularly successful is in the field of computer vision. This includes tasks such as image classification [1], [2], object detection [3], [4], and image segmentation [5], [6]. Deep learning has made significant progress in recent years with the development of convolutional neural networks (CNNs) and other novel deep learning architectures [7]–[9]. These networks are able to learn complex patterns and features in images, allowing them to achieve state-of-the-art results on a variety of tasks [10]–[14]. It has also been used to improve the efficiency of image and video processing, making it possible to run these tasks on mobile devices and other low-power platforms [15].

Although deep learning has been the method of choice for many tasks, it requires a large amount of data and computational power. Nevertheless, training a deep network on a new small dataset can be mitigated by using transfer learning. Transfer learning is a machine learning technique that allows a model trained on a large dataset to be fine-tuned for a specific task [16]–[18]. It has been widely used in a variety of applications,

including image classification [19], [20], natural language processing [21], [22], and speech recognition [23], [24]. In the field of image classification, transfer learning has been shown to be particularly effective when there is limited data available for the target task [25]. In this paper, we explore the use of transfer learning for classifying images of Thai culture.

Thailand is a country with a rich and diverse cultural heritage, and accurately classifying images of Thai culture is important for a variety of applications, such as tourism [26], education [27], and cultural preservation [28]. For example, a machine learning model that can classify images of Thai temples, traditional dress, and cultural festivals could be used to create a virtual tour of Thailand for tourists [7] or to help educators teach about Thai culture. However, building a machine-learning model from scratch to classify Thai cultural images can be challenging due to the limited availability of annotated data [29]. Many traditional machine learning models require large amounts of labeled data to achieve good performance [30], and it can be time-consuming and expensive to manually annotate a dataset for a specific task [31], [32]. In this paper, we proposed to utilize transfer learning for the Thai culture image classification problem. By exploiting models pre-trained on a large image dataset, we can leverage the knowledge learned by the models on our target task.

The main contributions of our paper include: i) Collecting a Thai culture image dataset which consists of 1,000 high-quality images from 10 well-known Thai cultures and traditions; ii) Investigating the performance of three famous CNNs, namely MobileNet, EfficientNet, and residual network (ResNet) as pre-trained models for Thai culture image classification; iii) Exploring how pre-trained models can be utilized for Thai culture image classification by comparing training the models with the random initialization, utilizing them as feature extractors, and fully fin-tuning them; and iv) Explaining the quantitative accuracy of the best-performing model, EfficientNet, with gradient-weighted class activation mapping (Grad-CAM), which confirms that the model focuses on relevant features in the input images. For the remainder of this paper, we presented our methodology including details about our Thai culture image dataset, preprocessing steps, and transfer learning approaches. We then report on the results of our experiments and discuss the limitations of our approach. Potential directions for future work are also considered, ensuring a comprehensive exploration of opportunities within this domain.

2. RESEARCH METHOD

In this section, we introduce our dataset for Thai cultural image classification. We then summarize the detail of our pre-trained models and how they could be fine-tuned for our dataset. Lastly, we explain the hyperparameter setting used in our research.

2.1. Dataset collection

We used a dataset of Thai cultural images that we collected and annotated ourselves. The dataset consists of a total of 1,000 images, split evenly into 10 classes. Figure 1 shows a sample of each Thai cultural tradition: Figure 1(a) *ประเพณีแห่ผ้าขึ้นธาตุ* (the merit-making ceremony of hoisting a cloth flag), Figure 1(b) *ประเพณีแห่เทียนพรรษา* (the merit-making ceremony of offering candles during the Buddhist Lent), Figure 1(c) *ประเพณีสารทเดือนสิบ* (the merit-making ceremony of offering food to monks on the tenth lunar month), Figure 1(d) *ประเพณีสงกรานต์* (Songkran festival), Figure 1(e) *ประเพณีวิ่งควาย* (running with a buffalo festival), Figure 1(f) *ลอยกระทง* (Krathong festival), Figure 1(g) *ประเพณีเซิ้ง* (the merit-making ceremony of offering candles to the spirits), Figure 1(h) *ประเพณีบุญบั้งไฟ* (the merit-making ceremony of lighting candles), Figure 1(i) *ประเพณีตักบาตรดอกไม้ม* (the merit-making ceremony with flowers), and Figure 1(j) *ประเพณีชักพระ* (the tradition of carrying the Buddha image in a procession). Each image was manually labeled with its corresponding class by four annotators, and any discrepancies were resolved through consensus. The images in the dataset were collected from various online sources, such as Google Image Search and Flickr. We made sure to include a diverse range of images within each class and to ensure that the images were representative of Thai culture. We also made sure to exclude any images that were unrelated to Thai culture or that were inappropriate for the target audience. To preprocess the dataset, we resized each image to 224×224 pixels and applied standard image augmentation techniques, such as horizontal flipping and random cropping. We also randomly split the dataset into training and test sets, with a ratio of 80:20.

2.2. Pretrained models

In this study, we investigated the use of transfer learning for image classification of Thai cultural images with three state-of-the-art convolutional neural networks: MobileNet, EfficientNet, and ResNet since they represent networks from three different scales. MobileNet is a lightweight architecture designed for mobile and embedded devices, making it efficient in terms of computational resources. EfficientNet is an architecture that aims to improve the accuracy and efficiency of CNNs. ResNet is a deep residual network that has achieved state-of-the-art performance on several image classification benchmarks. Table 1 describes the number of

parameters and the inference times on the central processing unit (CPU) and graphics processing unit (GPU) for all three networks. The CPU and GPU used to measure the time were an Intel(R) Xeon(R) with a clock speed of 2.00 GHz and an Nvidia Tesla P100 with a memory of 16 GB, respectively. From the table, we can notice that the number of parameters for ResNet was significantly larger than the previous two. Nevertheless, the inference time was less prominent, especially for the GPU. By comparing the performance of these architectures, we hope to gain a better understanding of how different CNN architectures affect the transfer learning process and the final classification performances.



Figure 1. Sample images from the Thai culture image dataset consist of 1,000 images of various aspects of Thai cultural traditions, (a) including the merit-making ceremony of hoisting a cloth flag, (b) the merit-making ceremony of offering candles during the Buddhist Lent, (c) the merit-making ceremony of offering food to monks on the tenth lunar month, (d) the Songkran festival, (e) running with a buffalo festival, (f) Krathong festival, (g) the merit-making ceremony of offering candles to the spirits, (h) the merit-making ceremony of lighting candles, (i) the merit-making ceremony with flowers, and (j) the tradition of carrying the Buddha image in a procession. The images were collected from various online sources and manually annotated by assigning each image to one of the 10 well-known traditions

Table 1. Comparison of MobileNet, EfficientNet, and ResNet in terms of parameters, CPU inference time, and GPU inference time

Models	Parameters (M)	Inference time (ms)	
		CPU	GPU
MobileNet	3.28	726	74
EfficientNet	4.11	1215	105
ResNet	23.66	2959	121

2.2.1. MobileNet

MobileNet [33] is a small, low-latency, and low-power convolutional neural network designed for mobile and embedded devices. It was developed by Google in 2017 and has been widely used in various applications, such as image classification, object detection, and semantic segmentation. MobileNet is based on the concept of depthwise separable convolutions, which decompose a standard convolution into two separate operations: a depthwise convolution and a pointwise convolution. This allows MobileNet to achieve a much smaller model size and faster inference times compared to traditional CNNs, while still maintaining a reasonable level of accuracy. MobileNet also uses a lightweight version of the Xception architecture, which consists of a series of depthwise separable convolutions and a few fully connected layers. This allows MobileNet to reduce the number of parameters and computations in the model, making it more suitable for mobile and embedded devices with limited resources.

2.2.2. EfficientNet

EfficientNet is a family of convolutional neural network models developed by Google in 2019 that aims to improve the efficiency of CNNs while maintaining a high level of accuracy [34]. The EfficientNet models are designed to be scalable in terms of model size, number of parameters, and computational cost, making them suitable for a wide range of applications and hardware platforms. EfficientNet models are built on top of the MobileNet architecture to achieve a high level of efficiency. The EfficientNet models also use a compound scaling method that scales the model size, depth, and width in a balanced manner to achieve optimal

performance. EfficientNet models have achieved state-of-the-art performance on various benchmarks, including image classification, object detection, and semantic segmentation. They have also been widely used in practical applications, such as mobile devices, internet of things (IoT) devices, and edge computing systems.

2.2.3. ResNet

Residual network (ResNet) is a convolutional neural network architecture developed by Microsoft researchers in 2015 [35]. It was designed to address the problem of vanishing gradients, which is a common issue in very deep CNNs where the gradients of the parameters tend to become very small during training, leading to slow convergence and poor performance. ResNet addresses the problem of vanishing gradients by introducing a residual block, which is a building block of the network that allows the input to be added to the output of the block. This allows the gradients to flow more easily through the network, enabling the model to learn deeper and more complex representations. ResNet has achieved state-of-the-art performance on various benchmarks, including image classification, object detection, and semantic segmentation. It has also been widely used in practical applications, such as image recognition and natural language processing.

2.3. Transfer learning

Transfer learning is a machine learning technique that involves using a pre-trained model on one task as the starting point for a model on a related task. The idea behind transfer learning is that the features learned by the pre-trained model on the original task can be useful for the new task, allowing the model to learn more efficiently and potentially achieve better performance. Transfer learning can be useful in situations where there is a limited amount of data or computational resources available for the new task. By using a pre-trained model as a starting point, the model can benefit from the knowledge and representation learned on the original task, allowing it to perform better on the new task with fewer data and resources. There are two main approaches to applying transfer learning when using a pre-trained model as a starting point for a new task: using the pre-trained model as a fixed feature extractor or fully fine-tuning the whole model on the new task [36].

2.3.1. Feature extractor

Using the pre-trained model as a fixed feature extractor involves using the output of the pre-trained model as input features for a new model trained on the new task. The weights of the pre-trained model are not updated during training, and the new model is trained only on the new task. For computer vision tasks, the baseline pre-trained model is usually trained on a large and diverse dataset such as the ImageNet dataset [29]. In this case, the features learned by the pre-trained model can be useful for the new task, allowing the new model to learn more efficiently and potentially achieve better performance. Figure 2 demonstrates the overview of using a pre-trained network as a fixed feature extractor. In the figure, the blue component represents the convolutional layers' parameters being frozen while the red components refer to fully connected layers being trained on a new small image dataset.

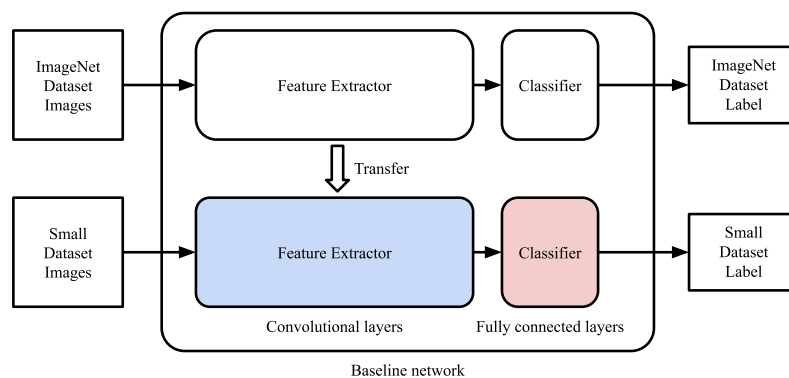


Figure 2. A pre-trained network can be used as a fixed feature extractor, where the weights of the feature extraction layers are not changed during training, but the weights of the classifier are updated during training on a new small dataset

2.3.2. Fully fine-tuning

Fully fine-tuning the pre-trained model on the new task involves adapting the model's parameters to the new data by training all layers of the model on the new task. This approach is typically useful when the new task

is significantly different from the original task and requires a more specialized model. Fine-tuning all layers of the model allows the model to learn task-specific features that may not be present in the pre-trained model. However, it also requires a larger amount of data and computational resources compared to using the pre-trained model as a fixed feature extractor. Figure 3 depicts the overview of fully fine-tuning the pre-trained model on the new task. In this case, both the feature extractor and the classifier were both trained on a new small dataset.

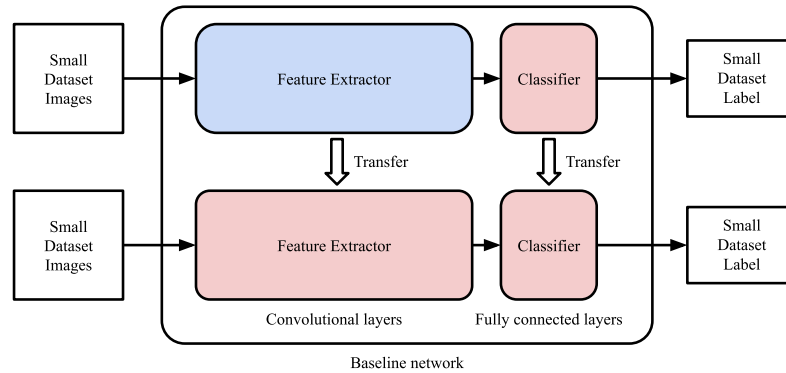


Figure 3. Fully fine-tuning the pre-trained model on the new task involves adapting the model's parameters to the new small data by training all layers of the model including the feature extractor and the classifier

2.4. Hyperparameters settings

To prevent an exhaustive search and reduce the time and resources required, we set the values of most hyperparameters to default values. Adam optimizer was used to adjust the networks' weights with a batch size of 32. For training the fixed feature extractors, the learning rate was set to 1×10^{-3} , and the β_1 and β_2 were set to 0.9 and 0.999, respectively. For fully fine-tuning the whole network, we lowered the learning rate to 1×10^{-5} in order to prevent significant changes to the models' parameters. All training was carried out for 100 epochs. All of our experiments were conducted using the framework Keras [37] in Python.

3. RESULTS AND DISCUSSION

To fully understand how transfer learning improves the performance of the models, we compared and discussed the proposed models in the following subsections. Specifically, the accuracy and inference time for the models were evaluated qualitatively. Lastly, the best model was then qualitatively examined with the Grad-CAM technique.

3.1. Quantitative results

To measure the performance of the models, we compared the quantitative results of three pre-trained models in terms of classification accuracy. Additionally, training time was used to evaluate the three different training strategies. We then discussed the best model when both metrics were taken into account.

3.1.1. Accuracy

We compared the performance of MobileNet, EfficientNet, and ResNet when trained with random initialization, used as fixed feature extractors, and fully fine-tuned for image classification on the proposed Thai culture image dataset. We used the same training and test sets for all models and training strategies. Their performances were evaluated using classification accuracy, which is defined as the fraction of correctly classified images over all tested images. The results of the experiments were shown in Table 2 which demonstrated that all three models performed better when they were used as a feature extractor followed by full fine-tuning and random initialization. Comparing three different networks, EfficientNet performed best at 95.87% followed by ResNet at 95.04%, and MobileNet at 92.56%. The results were similar to [36] which also pointed out that intermediate-size networks tend to perform better when they were repurposed for a new smaller dataset. Interestingly, fully finetuning the networks tends to have positive effects on the network, but in this case, it worsens the accuracies for all three networks. We argue that this is because of the size of the new dataset. The pre-trained models were trained on a large and diverse ImageNet dataset but the proposed dataset was significantly smaller. In this case, the features learned by the pre-trained models on the original task were more useful for the new task, when used as a feature extractor. When the models were fully fine-tuned, they were adapted to a new task by training all of their layers on the new dataset. This allows the models to learn

new-task-specific features that may not be present in the pre-trained models, but it also requires a larger amount of data and computational resources otherwise the models can overfit the training set. As a result, the models seem to overfit the training set when they were fully fine-tuned compared to when they were used as feature extractors. When the models were trained from random initialization, the best-performing model was ResNet with an accuracy of 60.33% followed by MobileNet and EfficientNet at 52.89% and 47.08%, respectively. The results highlighted the critical contribution of utilizing transfer learning on a small dataset, especially by the EfficientNet which increase its performance from 47.08% to 95.87%.

Table 2. Classification accuracy of MobileNet, EfficientNet, and ResNet when trained as random initialization, a fixed feature extractor, and fully fine-tuning on the Thai culture image dataset

Models	Accuracy (%)		
	Random initialization	Fixed feature extractor	Fully fine-tuned
MobileNet	52.89	92.56	91.74
EfficientNet	47.08	95.87	95.04
ResNet	60.33	95.04	94.21

3.1.2. Training time

We additionally compared the performance of the models based on the training time on the GPU. The results are shown in Table 3. From the table, we can notice that the training time of the fixed feature extractor was the smallest compared to random initialization and fully fine-tuning because the number of parameters used was much significantly lower. Besides as the random initialization and full fine-tuning updated the same number of parameters, their training time is relatively similar to each other. These results further favor using the pre-trained model as a fixed feature extractor on the proposed Thai culture image dataset as it was able to achieve the highest accuracy while requiring the least amount of training time.

Table 3. The training time of MobileNet, EfficientNet, and ResNet when trained as random initializations, a fixed feature extractor, and fully fine-tuning on the Thai culture image dataset

Models	Training time (ms/iteration)		
	Random initialization	Fixed feature extractor	Fully fine-tuned
MobileNet	31	19	32
EfficientNet	77	24	79
ResNet	72	31	69

3.2. Qualitative results

To better understand the reasoning behind the performance of our best model, an EfficientNet model trained as a fixed feature extractor, we used the Grad-CAM [38] visualization technique to generate heatmaps of the regions in the images that the model attended to when making predictions. Grad-CAM is a visualization technique for understanding the decisions made by a convolutional neural network. It produces a heatmap that indicates the regions of the input image that are most important for the prediction made by CNN. Grad-CAM can be used to explain the decisions made by CNN, which can be useful for debugging and improving the performance of the model by highlighting the most relevant parts of the input image for making predictions. Figure 4 shows ten randomly selected examples from each class of the Thai culture image dataset: Figure 4(a) ประเพณีแห่ผ้าขึ้นธาตุ (the merit-making ceremony of hoisting a cloth flag), Figure 4(b) ประเพณีแห่เทียนพรรษา (the merit-making ceremony of offering candles during the Buddhist Lent), Figure 4(c) ประเพณีสารทเดือนสิบ (the merit-making ceremony of offering food to monks on the tenth lunar month), Figure 4(d) ประเพณีสงกรานต์ (Songkran festival), Figure 4(e) ประเพณีวิ่งควาย (running with a buffalo festival), Figure 4(f) ลอยกระทง (Krathong festival), Figure 4(g) ประเพณียี่เป็ง (the merit-making ceremony of offering candles to the spirits), Figure 4(h) ประเพณีบุญบั้งไฟ (the merit-making ceremony of lighting candles), Figure 4(i) ประเพณีตักบาตรดอกไม้ (the merit-making ceremony with flowers), and Figure 4(j) ประเพณีชักพระ (the tradition of carrying the Buddha image in a procession). The results were heatmaps generated by Grad-CAM from our best model, the EfficientNet. The figures demonstrated that our model seems to focus on important features of the images. For example, the main building of the merit-making ceremony of hoisting a cloth flag (ประเพณีแห่ผ้าขึ้นธาตุ) was the main focus of Figure 4(a). The buffalo from the running with a buffalo festival (ประเพณีวิ่งควาย) were emphasized in Figure 4(f), and the lanterns from the merit-making ceremony of offering candles to the spirits (ประเพณียี่เป็ง) were highlighted in Figure 4(g). Overall, the heatmaps generated by Grad-CAM provide a qualitative understanding of the reasoning made by the EfficientNet model.

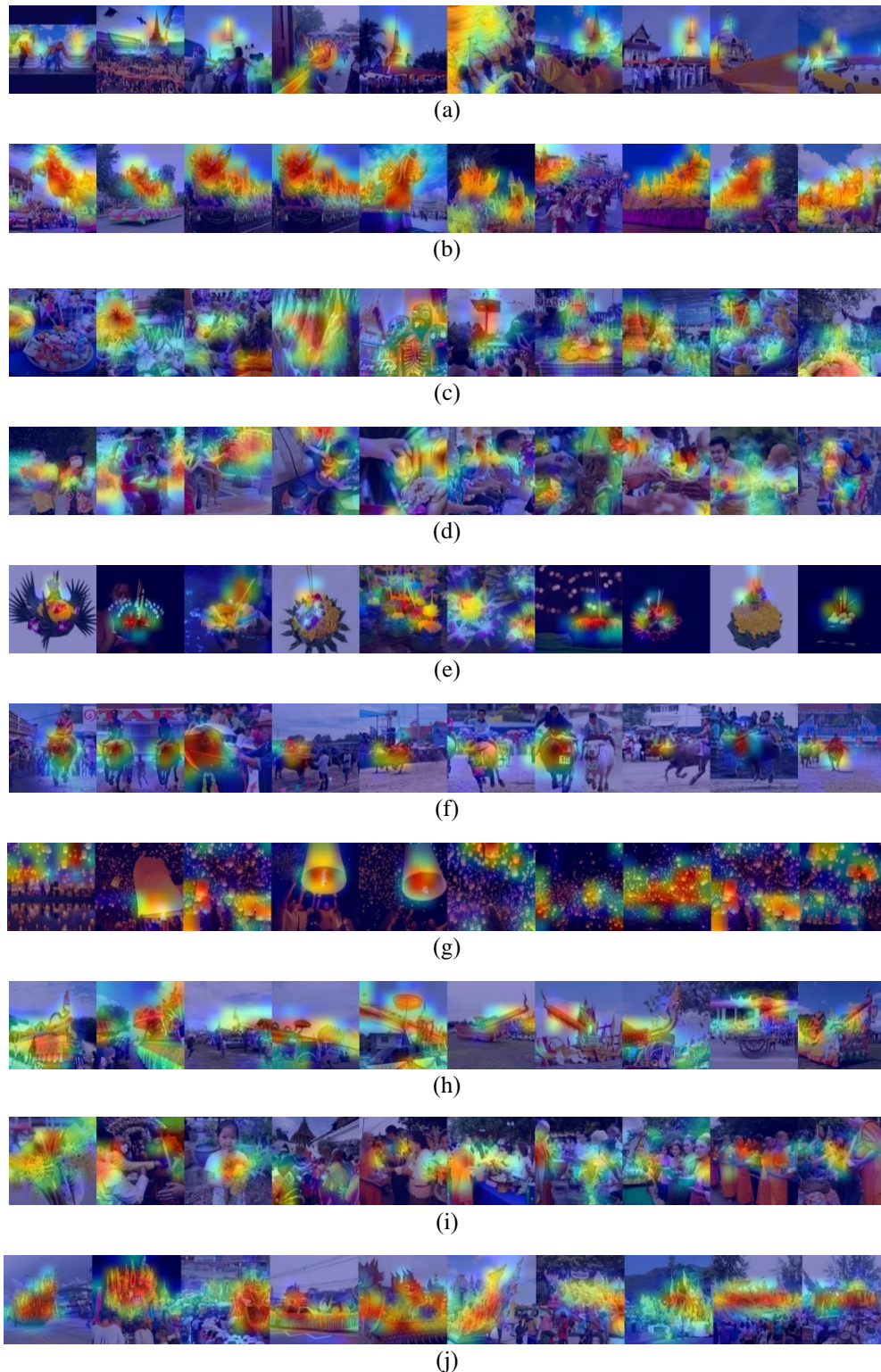


Figure 4. Visualizing the reasoning of our best-performing model, EfficientNet, using Grad-CAM on ten randomly selected Thai culture images from the test set for each class: (a) ประเพณีแห่ผ้าขึ้นธาตุ (the merit-making ceremony of hoisting a cloth flag), (b) ประเพณีแห่เทียนพรรษา (the merit-making ceremony of offering candles during the Buddhist Lent), (c) ประเพณีสารทเดือนสิบ (the merit-making ceremony of offering food to monks on the tenth lunar month), (d) ประเพณีสงกรานต์ (Songkran festival), (e) ประเพณีวิ่งควาย (running with a buffalo festival), (f) ลอยกระทง (Krathong festival), (g) ประเพณีฮีตสิบั้ง (the merit-making ceremony of offering candles to the spirits), (h) ประเพณีบุญบั้งไฟ (the merit-making ceremony of lighting candles), (i) ประเพณีตักบาตรดอกไม้ง (the merit-making ceremony with flowers), and (j) ประเพณีชักพระ (the tradition of carrying the Buddha image in a procession)

4. CONCLUSION

In this study, we investigated the use of transfer learning for the task of image classification on the Thai culture image dataset. The dataset was constructed from 1,000 images manually annotated into 10 well-known Thai cultures and traditions. We used MobileNet, EfficientNet, and ResNet as pre-trained models and evaluated their performance when they were trained with the random initialization, used as a feature extractor, and fully fine-tuned. The results showed that all three models performed better when used as feature extractors compared to random initialization and fully fine-tuning, with EfficientNet achieving the highest accuracy of 95.87%. Favorably, using pre-trained models as a feature extractor also takes the least amount of training time. In the case of EfficientNet, it only took 24 ms/iteration. Our qualitative findings from Grad-CAM additionally supported that the proposed method effectively localized the relevant information in the images for prediction.

Future work could focus on improving the performance of image classification on the Thai culture image dataset by exploring the use of different pre-trained models and fine-tuning strategies. This may include experimenting with different ratios of the networks that are allowed to be updated during fine-tuning. Additionally, novel neural network architectures such as vision transformers could be adapted to further enhance the performance.

ACKNOWLEDGEMENTS

The authors would like to thank Thanarit Chonrak, Thatpong Sangkhapan, Korapin Puttapong, and Natnaree Mombun for their contributions to dataset collection and annotation. The authors would like to also thank the Department of Computer Science, Faculty of Science and Technology, Nakhon Si Thammarat Rajabhat University for the hardware support for the project.




REFERENCES

- [1] J. Zhang, Y. Xie, Q. Wu, and Y. Xia, "Medical image classification using synergic deep learning," *Medical Image Analysis*, vol. 54, pp. 10–19, May 2019, doi: 10.1016/j.media.2019.02.010.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [3] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digital Signal Processing*, vol. 126, Jun. 2022, doi: 10.1016/j.dsp.2022.103514.
- [4] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Prepr. arXiv.1905.05055*, May 2019.
- [5] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 1280–1289, doi: 10.1109/CVPR52688.2022.00135.
- [6] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021, doi: 10.1109/TPAMI.2021.3059968.
- [7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
- [8] Y. Lecun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *Conference: The Handbook of Brain Theory and Neural Networks*, 1995, pp. 1–14.
- [9] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015, doi: 10.1016/j.neunet.2014.09.003.
- [10] Y. AUFAR and T. P. Kaloka, "Robusta coffee leaf diseases detection based on MobileNetV2 model," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 6, pp. 6675–6683, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6675-6683.
- [11] U. B. Patayon and R. V. Crisostomo, "Peanut leaf spot disease identification using pre-trained deep convolutional neural network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, pp. 3005–3012, Jun. 2022, doi: 10.11591/ijece.v12i3.pp3005-3012.
- [12] P. Patel and A. Thakkar, "The upsurge of deep learning for computer vision applications," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, pp. 538–548, Feb. 2020, doi: 10.11591/ijece.v10i1.pp538-548.
- [13] R. Ali, J. H. Chuah, M. S. A. Talip, N. Mokhtar, and M. A. Shoaib, "Structural crack detection using deep convolutional neural networks," *Automation in Construction*, vol. 133, Jan. 2022, doi: 10.1016/j.autcon.2021.103989.
- [14] C. Tian, Y. Yuan, S. Zhang, C.-W. Lin, W. Zuo, and D. Zhang, "Image super-resolution with an enhanced group convolutional neural network," *Neural Networks*, vol. 153, pp. 373–385, Sep. 2022, doi: 10.1016/j.neunet.2022.06.009.
- [15] M. Al-Shabi and A. Abuhamdah, "Using deep learning to detecting abnormal behavior in internet of things," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 2, pp. 2108–2120, Apr. 2022, doi: 10.11591/ijece.v12i2.pp2108-2120.
- [16] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.
- [17] M. Alswaitti, L. Zihao, W. Alomoush, A. Alrosan, and K. Alissa, "Effective classification of birds' species based on transfer learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 4, pp. 4172–4184, Aug. 2022, doi: 10.11591/ijece.v12i4.pp4172-4184.
- [18] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 2, pp. 151–166, Oct. 2020, doi: 10.1109/TAI.2021.3054609.
- [19] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: A literature review," *BMC Medical Imaging*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12880-022-00793-7.




- [20] M. Hussain, J. J. Bird, and D. R. Faria, "A study on CNN transfer learning for image classification," in *UKCI 2018: Advances in Computational Intelligence Systems*, 2019, pp. 191–202, doi: 10.1007/978-3-319-97982-3_16.
- [21] M. Mozafari, R. Farahbakhsh, and N. Crespi, "A BERT-based transfer learning approach for hate speech detection in online social media," in *COMPLEX NETWORKS 2019: Complex Networks and Their Applications VIII*, 2020, pp. 928–940, doi: 10.1007/978-3-030-36687-2_77.
- [22] S. Ruder, M. E. Peters, S. Swayamdipta, and T. Wolf, "Transfer learning in natural language processing," in *Proceedings of the 2019 Conference of the North*, 2019, pp. 15–18, doi: 10.18653/v1/N19-5004.
- [23] P. Gurunath Shivakumar and P. Georgiou, "Transfer learning from adult to children for speech recognition: Evaluation, analysis and recommendations," *Computer Speech & Language*, vol. 63, Sep. 2020, doi: 10.1016/j.csl.2020.101077.
- [24] D. Seo, H.-S. Oh, and Y. Jung, "Wav2KWS: transfer learning from speech representations for keyword spotting," *IEEE Access*, vol. 9, pp. 80682–80691, 2021, doi: 10.1109/ACCESS.2021.3078715.
- [25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.
- [26] R. B. Gozzoli, "Chapter 8: Sustainable tourism development and Thai cultural heritage," in *Contemporary Challenges of Climate Change, Sustainable Tourism Consumption, and Destination Competitiveness*, 2018, pp. 139–157, doi: 10.1108/S1871-317320180000015014.
- [27] C. Baker and P. Phongpaichit, *A history of Thailand*. Cambridge University Press, 2022.
- [28] S. Premrirat, "Redefining 'Thainess': Embracing diversity, preserving unity," in *Contemporary Socio-Cultural and Political Perspectives in Thailand*, Dordrecht: Springer Netherlands, 2014, pp. 3–22, doi: 10.1007/978-94-007-7244-1_1.
- [29] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.
- [30] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013, doi: 10.1109/TPAMI.2013.50.
- [31] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in *ECCV 2014: Computer Vision – ECCV 2014*, 2014, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.
- [32] J. Donahue *et al.*, "Decaf: A deep convolutional activation feature for generic visual recognition," in *ICML '14: Proceedings of the 31st International Conference on Machine Learning*, 2014, vol. 32, pp. 1–647–1–655.
- [33] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *Prepr. arXiv.1704.04861*, Apr. 2017.
- [34] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Prepr. arXiv.1905.11946*, May 2019.
- [35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [36] S. Ittisoponpisan, C. Kaipan, S. Ruang-on, R. Thaiphon, and K. Songsri-in, "Pushing the accuracy of Thai Food image classification with transfer learning," *Engineering Journal*, vol. 26, no. 10, pp. 57–71, Oct. 2022, doi: 10.4186/ej.2022.26.10.57.
- [37] TensorFlow Gardener, "GitHub repository," GitHub, 2023. <https://github.com/fchollet/keras> (accessed Feb. 7, 2023).
- [38] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

BIOGRAPHIES OF AUTHORS



Munlika Rattaphun    received a B.S. degree in computer science from Thaksin University, Songkhla, Thailand, in 2009, an M.S. degree in computer science from the Prince of Songkla University, Songkhla, in 2011, and the Ph.D. degree in computer science and information engineering from National Chiayi University, Chiayi, Taiwan in 2022. She is currently a lecturer at the Department of Computer Science, Faculty of Science and Technology, Nakhon Si Thammarat Rajabhat University, Nakhon Si Thammarat, Thailand. Her current research interests include machine learning, nearest-neighbor search, and recommender systems. She can be contacted at email: munlika_rat@nstru.ac.th.



Kritaphat Songsri-in    finished MEng and Ph.D. in computing from Imperial College London in 2011 and 2020, respectively. He has been a lecturer in the department of computer science at Nakhon Si Thammarat Rajabhat University since 2020. His research interests include machine learning, deep learning, and computer vision. He has published in and is a reviewer for multiple international conferences and journals such as *IEEE Transactions on Image Processing* and *IEEE Transactions on Information Forensics & Security*. Dr. Songsri-in was a recipient of the Royal Thai Government Scholarship covering his undergraduate and postgraduate degrees in 2010. He received the Best Student Paper Awards at the IEEE 13th International Conference for Automatic Face and Gesture Recognition (FG2018) and the 6th National Science and Technology Conference (NSCIC2021). In 2021, his Ph.D. thesis received an award from the National Research Council of Thailand (NRCT). He can be contacted at email: kritaphat_son@nstru.ac.th.