JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Transfer Learning-Based Feature Fusion Model for Classification of Maneuver Weapon Systems

Jinyong Hwang[1], You-Rak Choi[2], Tae-Jin Park[2], and Ji-Hoon Bae[1,*]

Abstract

Convolutional neural network-based deep learning technology is the most commonly used in image identification, but it requires large-scale data for training. Therefore, application in specific fields in which data acquisition is limited, such as in the military, may be challenging. In particular, the identification of ground weapon systems is a very important mission, and high identification accuracy is required. Accordingly, various studies have been conducted to achieve high performance using small-scale data. Among them, the ensemble method, which achieves excellent performance through the prediction average of the pre-trained models, is the most representative method; however, it requires considerable time and effort to find the optimal combination of ensemble models. In addition, there is a performance limitation in the prediction results obtained by using an ensemble method. Furthermore, it is difficult to obtain the ensemble effect using models with imbalanced classification accuracies. In this paper, we propose a transfer learning-based feature fusion technique for heterogeneous models that extracts and fuses features of pre-trained heterogeneous models and finally, fine-tunes hyperparameters of the fully connected layer to improve the classification accuracy. The experimental results of this study indicate that it is possible to overcome the limitations of the existing ensemble methods by improving the classification accuracy through feature fusion between heterogeneous models based on transfer learning.

Keywords

Artificial Intelligence, Deep Learning, Ensemble, Feature Fusion, Transfer Learning

1. Introduction

Artificial intelligence technology has been recently applied in various fields, including industry, education, medicine, and military, and related products are being researched and developed. In particular, deep learning-based technology is the most used in the field of detection, identification, and tracking using video images, and it has been extensively verified through various studies [1-3]. However, a large amount of high-quality training data and high-performance learning models are required in deep learning technology to obtain acceptable accuracy. In the case of civil and industrial fields, where large-scale server construction is possible, the requirements for model training using high-quality collected data can be sufficiently satisfied. Conversely, it is challenging to collect large amount of high-quality data in particular fields, such as military defense, because military operational environment involves restrictions on physical servers and Internet connection owing to security requirements, lack of information about

targets, and data collection limitations. Therefore, considerable limitations exist in the application of deep learning to existing weapon systems, and various studies and technologies are continuously being developed to overcome them [4].

Currently, various studies are being conducted in the military to apply artificial intelligence-related technologies to weapon systems. In particular, research and development for target detection, identification, and tracking is progressing to a certain level owing to the application of deep learning-based technology that uses sensor information, such as day/thermal cameras and radars, to surveillance/reconnaissance fields [5-7]. However, research is concentrated in specific fields in which data acquisition is easy, such as those using satellite images, and verification remains insufficient for application in actual military systems.

Maneuver weapon systems depend heavily on image information for foe identification and target detection/classification missions. In addition, as an important issue directly related to survival, high identification accuracy is required. Since all these processes solely rely on human eyes, there is a possibility of errors in the identification accuracy of the maneuver weapon system. In addition, efforts to introduce artificial intelligence-based technologies are currently underway to compensate for the decrease in operational personnel due to recent population decline and military force reduction. One of these is a technology for target detection and identification through deep learning. However, the challenges of the battlefield environment, such as urban areas, forests, and smoke, can make it difficult to acquire sufficient data for accurate identification. As a result, a technique that can improve identification accuracy with small-scale data is necessary.

Therefore, a transfer learning method using deep learning models pre-trained on large-scale data, such as ImageNet [8], has been applied, researched, developed, verified, and used in various fields [9]. In addition, research was conducted in the military field to effectively improve the performance of training models by applying representative ensemble techniques to improve the classification accuracy [10]. Existing ensemble methods require time and resources to obtain high-performance pre-trained models for their use. In general, additional performance tuning is not possible in the ensemble model, so there is a limit to performance improvement, and considerable resources are required to find a combination of models with optimal performance [11].

In this study, we propose a transfer learning-based feature fusion model that performs maneuver weapon system classification to improve the accuracy of the existing ensemble techniques. In the proposed method, the final high-level feature maps output by pre-trained heterogeneous models are concatenated, and a new discriminator composed of fully connected layers is added. The image-based maneuver weapon system classification requires high identification accuracy; therefore, the performance of the proposed technique was verified through various experiments, and it was confirmed that the classification performance was improved compared with traditional ensemble methods.

The main contributions of this study are summarized as follows:

- Image classification accuracy can be improved with small-scale data in situations where acquiring large amounts of high-quality training data is limited, such as in the military field.
- A new transfer learning-based feature fusion model is proposed, which leverages advanced models pre-trained on the ImageNet dataset to improve classification accuracy.
- Compared to the existing ensemble technique, our approach of fine-tuning through stepwise transfer learning for feature fusion, extracted from heterogeneous models, demonstrates superior classification accuracy.

The remainder of this paper is organized as follows: Section 2 present existing techniques for improving classification accuracy with small-scale data in the field of maneuver weapon systems in relation to this study, and explain the differences between these techniques and the proposed method. Section 3 describes the detailed structure of the transfer learning-based feature fusion model proposed in this paper. Section 4 describes an evaluation of the proposed model's performance through a comparison with existing ensemble methods using data collection and performance comparison analysis. The analysis confirms the superiority of the proposed model, which is presented in detail in this section. Finally, Section 5 presents conclusions and future research.

2. Related Works

Recently, there has been a growing interest in using deep learning technology for image classification in the military field, and as a result, research in this area is being actively conducted. In particular, maneuver weapon systems are expected to enhance the precision of surveillance and reconnaissance missions by leveraging deep learning-based technology. These systems can perform critical tasks such as recognizing the battlefield situational awareness, identifying friend and foe, and detecting and tracking targets using image-based analysis.

First of all, research has been conducted on object detection and classification of various weapon systems using deep learning technology [12]. The purpose of this research is to leverage advanced algorithms and neural networks to improve the accuracy and effectiveness of weapon system recognition and classification.

In addition, researchers conducted research to overcome the recognition and classification of weapons system through transfer learning in the situation where training data is limited due to the specific environment of the battlefield [13]. Studies have been also published aiming to improve the accuracy of classification for tanks, a representative ground weapon system, through data augmentation and transfer learning [14,15].

Further studies have been conducted to improve the accuracy of weapon system classification through the use of model ensembles based on transfer learning [10]. Additionally, there are studies in progress to improve the object detection and classification of various ground weapon systems by combining reinforcement learning with existing deep learning techniques [16].

Therefore, transfer learning is a powerful technique that enables the transfer of knowledge learned from one domain to another, which can help address the challenge of limited training data in the military field. In this paper, similar to previous studies, a transfer learning method that can effectively learn with a small amount of data is utilized to classify ground weapon systems in a battlefield environment where acquisition of training data is limited.

However, unlike other studies, we propose a transfer learning-based feature fusion model that can improve classification accuracy compared to existing techniques by using a method of concatenating and fusion the features of various models that have completed transfer learning as a means of improving classification accuracy.

To achieve this purpose, the main considerations of the proposed method are as follows:

- The proposed method leverages the diversity of high-level features automatically extracted through transfer learning from heterogeneous models that have completed various types of pre-training.

- By concatenating the features extracted from heterogeneous models and fine-tuning the fully connected layers, it is possible to maximize the improvement in classification accuracy.
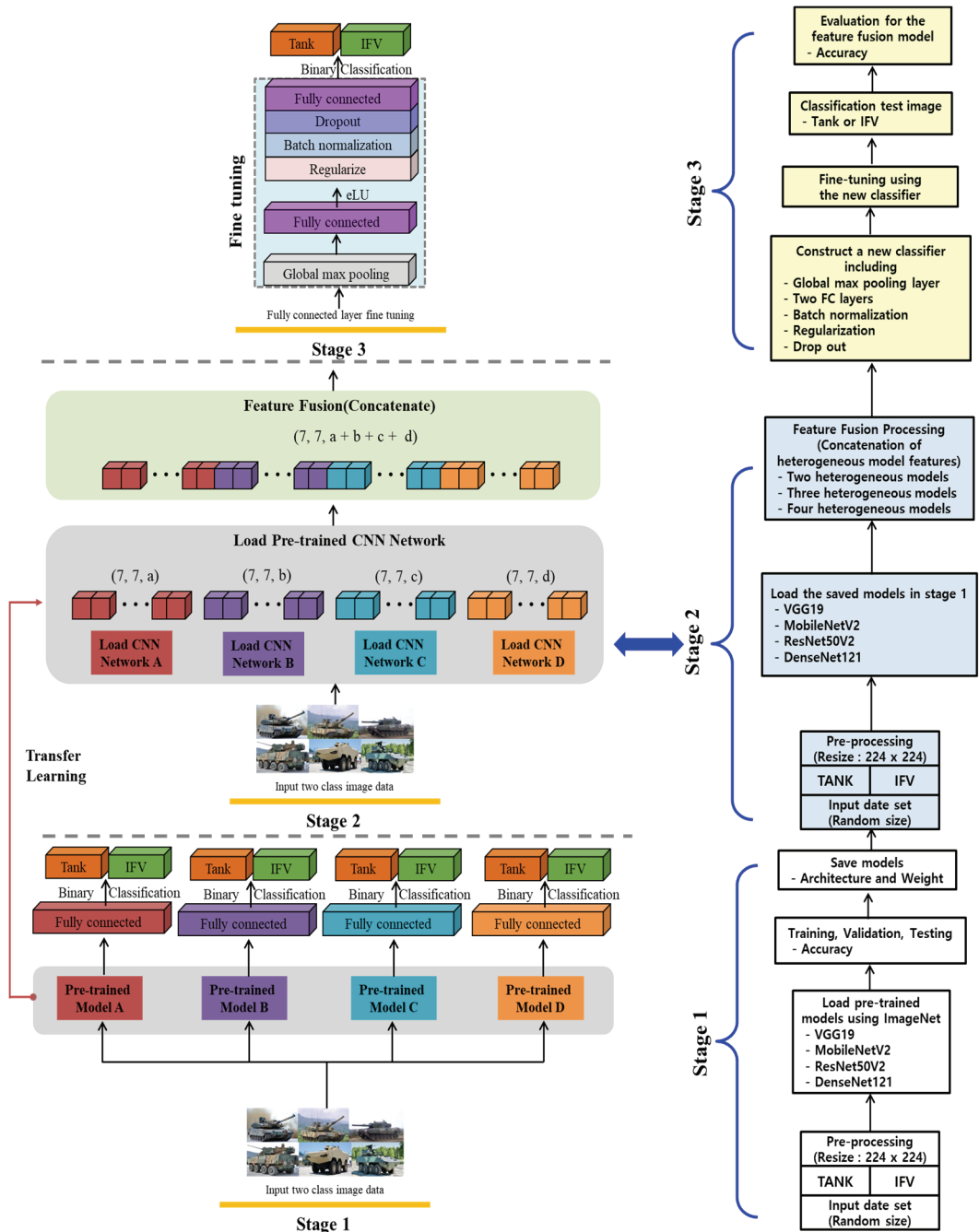
3. Proposed Transfer Learning-based Feature Fusion Model

To achieve better performance with small-scale image data, studies on training models to which pre-trained model-based transfer learning is applied are being actively conducted [17]. Among them, the most widely used method is to combine various trained models into one inference model, which has exhibited excellent performance in various fields. However, the effect can be confirmed in the case of a combination of excellent heterogeneous or homogeneous models, or an ensemble of multiple models. To do this, an adjustment process to find the models with the best performance and the time and effort required to find the optimal ensemble model combination are essential. In addition, there is a disadvantage in that accuracy cannot be improved with respect to the prediction result obtained after performing the ensemble. To overcome this problem, we propose a feature fusion method that concatenates the output feature maps of each model, for which transfer learning is performed through a pre-trained model to preserve the feature diversity of heterogeneous models, followed by fine-tuning of the fully connected layer to achieve improved accuracy. Fig. 1 shows the structure and detailed procedure of the proposed transfer learning-based feature fusion model.

The transfer learning-based feature fusion model is divided into three stages and is performed as follows:

Stage 1: After resizing the randomly collected data of tanks and armored vehicles, consisting of two classes, to a size of 224×224, each pre-trained model from ImageNet data is individually trained by inputting the resized data to each model. During this process, hyperparameters of each model are adjusted using validation data. Then, the classification accuracy is evaluated with test data, and each trained heterogeneous model is saved for the next stage.

Stage 2: Each trained heterogeneous model from Stage 1 is loaded and feature fusion is performed by concatenating the output feature maps of the heterogeneous models together, as shown in Fig. 2. The number of feature maps from the heterogeneous models after the implementation of the proposed feature fusion can be found in Table 1.

Stage 3: After applying global max pooling to the feature fusion layer generated in Stage 2 and adding a new fully connected classifier, fine-tuning training is performed.

To explain in more detail for Stage 1, first, binary classification training using the input data collected in Section 4.1 was performed with four representative deep learning models, VGG19, MobileNetV2, ResNet50V2, and DenseNet121 [18-21], which had been pre-trained with ImageNet. Next, we performed feature fusion by concatenating the final feature maps of all pre-trained models, excluding each previous binary classifier from the trained heterogeneous model, and then we added a classifier composed of new fully connected layers. Here, the exponential linear unite [22] was used as the activation function. In addition, a classifier was constructed to avoid overfitting by applying L1/L2 regularization, batch normalization, and dropout [23]. The classification accuracy was improved through the steps described above by fine-tuning the final connected layer parameters while maintaining the feature diversity of the transfer-learned heterogeneous models.

Fig. 1. Structure of the proposed transfer learning-based feature fusion model.

Next, Table 1 lists the output shape of the feature maps for each stage in the transfer learning-based feature fusion model, as shown in Fig. 1. The size of the input data was 224 × 224. In the case of Stage 1, 7×7×N feature maps were output for each model (Table 1). In Stage 2, by combining two, three, and four heterogeneous models, it can be confirmed that the feature maps of each model are connected during feature fusion.
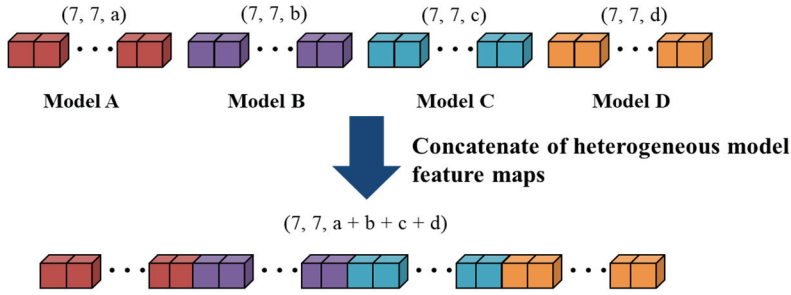
Fig. 2. Method of the proposed transfer learning-based feature fusion.

In the case of Stage 3, after the feature fusion step in which the feature maps of each transferred heterogeneous model are concatenated, global max pooling is performed on the feature fusion layer. This is done to extract the most meaningful and representative value from the fused feature map. The extracted representative values are then input into the fully connected layer of a new classifier for fine-tuning. Finally, the performance of the model is evaluated by checking the classification accuracy using test data.

As a result of observing the number of feature maps for each stage in the transfer learning-based feature fusion model, 512 feature maps for VGG19, 1,280 for MobileNetV2, 2,048 for ResNet50V2, and 1,024 feature maps for DenseNet121 were output for which transfer learning of Stage 1 was completed. Next, when the feature maps obtained from Stage 1 were concatenated with each other, the largest number of feature maps was obtained by combining the four heterogeneous models, as in Model #11 (Table 1). It can be seen that a large number of feature maps are fused in the order of Models #10, #7, and #9, all of which are combinations of three heterogeneous models. Among the three combinations of heterogeneous models, the case of Model #8 shows 3,000 or fewer feature map outputs, and the lowest feature map fusion combination is in Model #3, which fuses 1,500 feature maps. In Models #4 and #6, it can be seen that more than 3,000 feature maps are fused despite the combination of two heterogeneous models. Based on the above results, we will examine the correlation between the number of feature maps and the classification accuracy when performing transfer learning-based feature fusion in Section 4.

Table 1. Output shapes of feature maps for each stage in the transfer learning-based feature fusion model

Model No.	Feature fusion combination	Stage 1	Stage 2	Stage 3
1	VGG19 / MobileNetV2	(7, 7, 512) / (7, 7, 1280)	(7, 7, 1792)	1792
2	VGG19 / ResNet50V2	(7, 7, 512) / (7, 7, 2048)	(7, 7, 2560)	2560
3	VGG19 / DenseNet121	(7, 7, 512) / (7, 7, 1024)	(7, 7, 1536)	1536
4	MobileNetV2 / ResNet50V2	(7, 7, 1280) / (7, 7, 2048)	(7, 7, 3328)	3328
5	MobileNetV2 / DenseNet121	(7, 7, 1280) / (7, 7, 1024)	(7, 7, 2304)	2304
6	ResNet50V2 / DenseNet121	(7, 7, 2048) / (7, 7, 1024)	(7, 7, 3072)	3072
7	VGG19 / MobileNetV2 / ResNet50V2	(7, 7, 512) / (7, 7, 1280) / (7, 7, 2048)	(7, 7, 3840)	3840
8	VGG19 / MobileNetV2 / DenseNet121	(7, 7, 512) / (7, 7, 1280) / (7, 7, 1024)	(7, 7, 2816)	2816
9	VGG19 / ResNet50V2/ DenseNet121	(7, 7, 512) / (7, 7, 2048) / (7, 7, 1024)	(7, 7, 3584)	3584
10	MobileNetV2 / ResNet50V2/ DenseNet121	(7, 7, 1280) / (7, 7, 2048) / (7, 7, 1024)	(7, 7, 4352)	4352
11	VGG19 / MobileNetV2 / ResNet50V2 / DenseNet121	(7, 7, 512) / (7, 7, 1280) / (7, 7, 2048) / (7, 7, 1024)	(7, 7, 4864)	4864

Stage 1 indicates the output shapes of feature maps after transfer learning, Stage 2, the output shape of feature fusion, and Stage 3, the output shape of feature fusion after global max pooling.

4. Experimental Results

In this paper, a total of 3,600 images for maneuver weapon systems, consisting of 1,800 images each of two classes, tanks and armored vehicles, were collected through an Internet search. Next, for the deep learning model for transfer learning, among the models provided by the Keras application, an open-source software library for artificial neural networks, four models that have been pre-trained with the ImageNet dataset were selected. Finally, after the selected models were trained on the collected data using fine-tuning and data augmentation, the proposed transfer learning-based feature fusion model was confirmed for its classification accuracy. To this end, to evaluate the classification accuracy of the proposed model, Stages 2 and 3 of the step-by-step procedures mentioned in Section 3 are performed.

As hardware specifications for experiments, Intel Xeon 2.3 GHz central processing unit, 12.7 GB memory, and Tesla P100 graphic processing unit (GPU) were used, and CUDA Toolkit 11.2 was applied for model training using the GPU. In addition, as a software specification for model training and performance evaluation, the latest version of TensorFlow (2.11), a deep learning open-source framework library, was installed in the Ubuntu 18.04 LTS operating system environment.

4.1 Data Collection of Maneuver Weapon System

The maneuver weapon system operated by a military army is divided mainly into tanks and armored vehicles, and each performs different assigned missions. As shown in Fig. 3, the tank is equipped with a track-type hull, large-caliber cannon, and heavy armor to perform the mission of fighting enemy armored units at the forefront. The primary mission of the armored vehicle is to transport and protect the infantry. Although combat is not its main purpose, it is also equipped with various weapons to support firepower. In addition, although armored vehicles have a track type similar to a tank, the most distinctive feature is that they are mostly wheeled to maximize maneuverability.

(a)   (b)   (c)

Fig. 3. Examples of images collected for model training: (a) tank, (b) tracked armored vehicle, and (c) wheeled armored vehicle.

In this study, for the classification of maneuver weapon systems, related images were collected through an Internet search and then labeled as two classes: tanks and armored vehicles. First, tank images were obtained through search with the keyword "tank" from the Kaggle public image list, and data were additionally collected with the same keyword from web searches. All images of armored vehicles were collected through using web searches, using keywords such as "IFV," "AFV," or the specific model names of armored vehicles.

A total of 3,600 images were collected: 1,800 of tanks and 1,800 of armored vehicles. For the ex-

periments, 3,000 images consisting of 1,500 images for each class were used as training and verification data (Table 2). A total of 600 images, consisting of 300 images for each class, were used as test data for performance evaluation.

Table 2. List of the collected image data for the two classes

Class	Number of images		
	Training + validation	Test	Total
Tank	1,500	300	1,800
Armored vehicle	1,500	300	1,800
Total	3,000	600	3,600

4.2 Preparation of Deep Neural Network Models for Transfer Learning

In this section, as a preparatory step for the proposed method, transfer learning is performed using the most commonly used and representative convolutional neural network (CNN)-based pre-trained models, and their performance is evaluated. Because CNN-based deep learning technology generally requires a large amount of image data to derive an optimal model, considerable time and computational resources are required to train the model. In special cases, such as in the military field, there is a limitation to securing a large number of images; therefore, a method that can achieve excellent performance with only small-scale image data is necessary. To achieve this purpose, the utilization of transfer-learning technique that can effectively train models with only a small amount of data has been considered, and its performance has been proven and applied in various fields. In this study, we employed this transfer learning-based approach as a training method for the classification of maneuver weapon systems.

As the pre-trained model for transfer learning considered in this study, four deep neural network (DNN) models, VGG19, MobileNetV2, ResNet50V2, and DenseNet121 [18-21], were selected among the ImageNet-based pre-trained DNN models provided by the Keras application. In this case, the experimental environment for transfer learning for each DNN model was set as follows: a batch size of 32, Adam optimizer [24] with an initial learning rate of 10^{-5} as the learning algorithm, and epoch of 100. After applying the data augmentation and fine-tuning methods, the weights of the binary classifier were initialized each time, and training was performed. As a result of the transfer learning, the VGG19 and MobileNetV2 models achieved classification accuracy based on the test data of 95.3%, ResNet50V2 model of 96.5%, and DenseNet121 model of 97.1%. Thus, the DenseNet121 model exhibited the best performance in classifying tanks and armored vehicles.

4.3 Evaluation of the Proposed Method Using Deep Feature Diversity

In order to confirm the superiority of the proposed technique, a performance comparison is conducted with the ensemble method which is a widely used method for improving classification accuracy. To this end, ensemble is performed using the model for which transfer learning was performed in Section 4.2, and test accuracy evaluation is performed. Finally, the experiment on the transfer learning-based feature fusion model proposed in this paper is performed according to the procedure presented in Section 3, and then the ensemble method and test accuracy comparison analysis are performed.

4.3.1 Classification accuracy performance of ensemble technique

An ensemble technique is a method used to obtain better prediction performance by averaging the predictions of multiple inference models, with the purpose of improving performance based on the feature diversity of multiple independently trained models. In this regard, research and verification of ensemble techniques for CNN-based network structures are underway in various fields [25]. For a performance comparison with the proposed feature fusion model, the ensemble was performed using the transfer learning-based models selected in Section 4.2. The number of heterogeneous models for the ensemble was varied from two to four. As shown in Table 3, the ensemble accuracy for the test data ranges from 96.0% to 97.3%, and the average ensemble accuracy for the case where three or four models were combined was higher on average than when two models were combined.

Table 3. Ensemble accuracy results for the combinations of various heterogeneous models

Model No.	Model combinations for ensemble	Test accuracy (%)	
		Each model	Ensemble
1	VGG19 / MobileNetV2	95.3 / 95.3	96.0
2	VGG19 / ResNet50V2	95.3 / 96.5	96.6
3	VGG19 / DenseNet121	95.3 / 97.1	96.5
4	MobileNetV2 / ResNet50V2	95.3 / 96.5	97.3
5	MobileNetV2 / DenseNet121	95.3 / 97.1	97.3
6	ResNet50V2 / DenseNet121	96.5 / 97.1	97.5
7	VGG19 / MobileNetV2 / ResNet50V2	95.3 / 95.3 / 96.5	97.3
8	VGG19 / MobileNetV2 / DenseNet121	95.3 / 95.3 / 97.1	97.3
9	VGG19 / ResNet50V2/ DenseNet121	95.3 / 96.5 / 97.1	97.0
10	MobileNetV2 / ResNet50V2/ DenseNet121	95.3 / 96.5 / 97.1	97.3
11	VGG19 / MobileNetV2 / ResNet50V2 / DenseNet121	95.3 / 95.3 / 96.5 / 97.1	97.3

4.3.2 Classification accuracy performance of the proposed method

To verify the classification accuracy of the maneuver weapon system for the transfer learning-based feature fusion model proposed in this study, the performance of the traditional ensemble method was compared experimentally. First, for the proposed transfer learning-based feature fusion model, feature fusion was performed using the same two, three, and four heterogeneous models as in the existing ensemble method. Training was then performed by fine-tuning the hyper-parameters of a classifier composed of the newly added fully connected layers. Table 4 shows the test accuracy and accuracy ranking of the proposed transfer learning-based feature fusion model for the two-class maneuver weapon system dataset described in Section 4.1.

As shown in Table 4, as the combination of the feature fusion model with the best performance, the model combining three heterogeneous models, MobileNetV2, ResNet50V2, and DenseNet121 (Model #10 in Table 4), exhibited 98.1% classification accuracy. In contrast, the feature fusion model with the lowest performance was a combination of VGG19 and MobileNetV2 (Model #1 in Table 4), confirming a classification accuracy of 96.5%. We can observe that the combination of feature fusion models including ResNet50V2 or DenseNet121 with excellent individual performance as Models #3 and #7, respectively, or feature fusion models including ResNet50V2 and DenseNet121 simultaneously as Models #9 and #10, show high overall test accuracy. In addition, as in Models #3, #7, and #9, VGG19, whose individual performance is relatively lower than that of the other models, affects the performance improvement in the feature fusion model. Finally, the classification accuracy of the feature fusion model for the three-model combinations is higher than that for the heterogeneous two-model combinations.

However, in the case of the feature fusion combination of four heterogeneous models in Model #11, the combination of multiple models did not necessarily have a significant effect on the improvement of the feature fusion performance, as it showed a lower accuracy performance compared to the feature fusion combinations using three heterogeneous models.

Table 4. Test accuracy results for transfer learning-based feature fusion models according to 2, 3, and 4 heterogeneous model combinations

Model No.	Feature fusion combination	Test accuracy (%)		Rank
		Each model	Feature fusion model	
1	VGG19 / MobileNetV2	95.3 / 95.3	96.5	11
2	VGG19 / ResNet50V2	95.3 / 96.5	97.5	6
3	VGG19 / DenseNet121	95.3 / 97.1	97.8	3
4	MobileNetV2 / ResNet50V2	95.3 / 96.5	97.5	6
5	MobileNetV2 / DenseNet121	95.3 / 97.1	97.5	6
6	ResNet50V2 / DenseNet121	96.5 / 97.1	97.8	3
7	VGG19 / MobileNetV2 / ResNet50V2	95.3 / 95.3 / 96.5	98.0	2
8	VGG19 / MobileNetV2 / DenseNet121	95.3 / 95.3 / 97.1	97.5	6
9	VGG19 / ResNet50V2/ DenseNet121	95.3 / 96.5 / 97.1	97.8	3
10	MobileNetV2 / ResNet50V2/ DenseNet121	95.3 / 96.5 / 97.1	98.1	1
11	VGG19 / MobileNetV2 / ResNet50V2 / DenseNet121	95.3 / 95.3 / 96.5 / 97.1	97.3	10

In this regard, from the results of Tables 3 and 4, the correlation between the number of feature maps and ensemble performance for feature fusion can be analyzed. When sorting in the order of high classification accuracy in the combination of exactly three heterogeneous models (shown in Table 1), the number of fused feature maps are ranked at the top in the order of Models #10, #7, and #9. Next, Model 6, which ranks third in classification accuracy in Table 4, fused the second-largest number of feature maps in the case of a combination of two heterogeneous models. By contrast, in the case of Model #1, which has the lowest classification accuracy, the second-smallest feature maps were fused. In addition, it was confirmed that Model #8, which is a combination of the remaining three heterogeneous models, and Models #2 and #4, combinations of two heterogeneous models, were fused with a similar number of feature maps to obtain the same classification accuracy. However, the combination of four heterogeneous models (Model #11) showed relatively low classification accuracy despite the fusion of the largest number of feature maps. Similarly, among the two heterogeneous model combinations, Model #4 fused the most feature maps, but the classification accuracy was also low compared with the others. Exceptionally, it was observed that the combination of two heterogeneous models, as in Model #3, showed high classification accuracy, even though the smallest number of feature maps was fused.

Therefore, according to previous experimental results, it was confirmed that the classification accuracy improved as the number of fused feature maps increased. This is because the role of feature fusion is to reflect and complement the diversity of the features extracted from different models. However, it was confirmed that the classification accuracy improvement was insufficient when too many feature maps were fused.

4.3.3 Comparison of classification accuracy performance

Table 5 compares the classification accuracy of the traditional ensemble combinations for two, three, and four heterogeneous models performed in Section 4.3.1 and that of the transfer learning-based feature

fusion model combination presented in this section.

We can observe in Table 5 that the classification accuracy through the transfer learning-based feature fusion between the heterogeneous models proposed in this paper was improved by at least 0.2% to a maximum of 1.3% compared to the existing ensemble performance of two or three models. This is because, unlike the existing ensemble method, feature fusion specialized in the maneuver weapon system domain is reflected in transfer learning by concatenating the features between different heterogeneous models with each other and fine-tuning through the fully connected layer.

Table 5. Comparison of accuracy performance between the existing ensemble technique and the transfer learning-based feature fusion technique
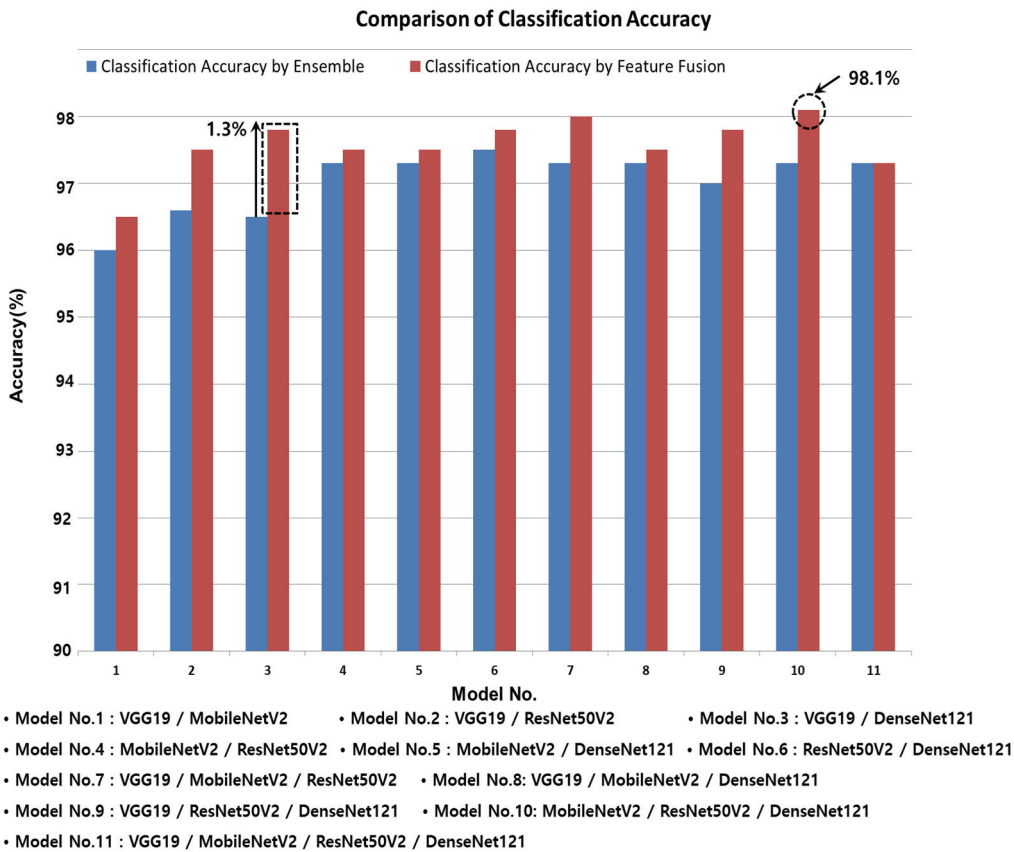
Model No.	Model combination	Test accuracy (%)			Variation (b − a) (%)
		Model	Ensemble[a]	Feature fusion[b]	
1	VGG19 / MobileNetV2	95.3 / 95.3	96.0	96.5	+0.5
2	VGG19 / ResNet50V2	95.3 / 96.5	96.6	97.5	+0.9
3	VGG19 / DenseNet121	95.3 / 97.1	96.5	97.8	+1.3
4	MobileNetV2 / ResNet50V2	95.3 / 96.5	97.3	97.5	+0.2
5	MobileNetV2 / DenseNet121	95.3 / 97.1	97.3	97.5	+0.2
6	ResNet50V2 / DenseNet121	96.5 / 97.1	97.5	97.8	+0.3
7	VGG19 / MobileNetV2 / ResNet50V2	95.3 / 95.3 / 96.5	97.3	98.0	+0.7
8	VGG19 / MobileNetV2 / DenseNet121	95.3 / 95.3 / 97.1	97.3	97.5	+0.2
9	VGG19 / ResNet50V2 / DenseNet121	95.3 / 96.5 / 97.1	97.0	97.8	+0.8
10	MobileNetV2 / ResNet50V2 / DenseNet121	95.3 / 96.5 / 97.1	97.3	98.1	+0.8
11	VGG19 / MobileNetV2 / ResNet50V2 / DenseNet121	95.3 / 95.3 / 96.5/ 97.1	97.3	97.3	0

By contrast, the existing ensemble method shows better performance on average with three heterogeneous models than with two heterogeneous models. However, in the case of the transfer learning-based feature fusion model, the accuracy performance is similarly leveled between the two heterogeneous model and the three heterogeneous model combinations.

Next, in the case of the existing ensemble method, it can be confirmed that there is almost no ensemble effect with imbalanced models, in which the classification accuracy between the combined models differs by 1% or more, such as Models #2, #3, #4, and #5. Conversely, the transfer learning-based feature fusion model demonstrates that classification accuracy improvement is achievable even for the combinations of the imbalanced models. In addition, the combination of VGG19, ResNet50V2, and DenseNet121 among the feature fusion model combinations mainly affects the improvement of classification accuracy performance. In general, it can be observed experimentally that the combination including MobileNetV2 has a relatively small performance improvement.

Finally, as shown in Fig. 4, the transfer learning-based feature fusion model combination that showed the highest classification accuracy improvement compared with the existing ensemble method was the combination of VGG19 and DenseNet121, which improved the classification accuracy by 1.3% while overcoming the imbalance in each model performance. The combination that showed the best classification accuracy among the transfer learning-based feature fusion models was that of three heterogeneous models, MobileNetV2, ResNet50V2, and DenseNet121, which recorded the highest classification accuracy of 98.1%.

Through the above results, it was confirmed for the transfer learning-based feature fusion model that the diversity of features fused from heterogeneous models rather than the combination between the number of heterogeneous models and models with excellent performance affects the performance improvement of the classification accuracy of the maneuver weapon system.

Comparison of Classification Accuracy

• Model No.1 : VGG19 / MobileNetV2 • Model No.2 : VGG19 / ResNet50V2 • Model No.3 : VGG19 / DenseNet121
• Model No.4 : MobileNetV2 / ResNet50V2 • Model No.5 : MobileNetV2 / DenseNet121 • Model No.6 : ResNet50V2 / DenseNet121
• Model No.7 : VGG19 / MobileNetV2 / ResNet50V2 • Model No.8 : VGG19 / MobileNetV2 / DenseNet121
• Model No.9 : VGG19 / ResNet50V2 / DenseNet121 • Model No.10: MobileNetV2 / ResNet50V2 / DenseNet121
• Model No.11 : VGG19 / MobileNetV2 / ResNet50V2 / DenseNet121

Fig. 4. Results of performance analysis between the ensemble method and the proposed transfer learning-based feature fusion model.

5. Conclusion

It can be experimentally observed that the transfer learning-based feature fusion model proposed in this paper can overcome the limitations of the existing ensemble methods, such as excessive resource input to find the optimal combination of performance models and the relatively small performance improvement effect when ensemble of two performance-imbalanced models is implemented. In addition, when the same model combination as the previous ensemble method was applied to the proposed method, the performance improvement effect was confirmed to be superior to that of the existing ensemble method. Thus, it was confirmed that the transfer learning-based feature fusion model was more effective in classifying the maneuver weapon systems than the existing ensemble methods.

In the near future, we plan to expand the performance analysis and validation of the proposed method

by applying the feature-fusion-based approach to a more diverse set of pre-trained models, in addition to the pre-trained models used in this paper. Next, Table 3 confirms that an imbalance in classification accuracy finally occurs due to the performance gap between heterogeneous models during transfer learning. As a result, it is difficult to achieve excellent performance through an ensemble that combines these models with imbalanced classification accuracy. However, for the proposed approach, it can be seen in Table 5 that better performance can be obtained despite the combination of heterogeneous models with imbalance in classification accuracy. This finding indicates that the proposed feature fusion model can effectively address the limitations of conventional ensemble methods, highlighting its potential to enhance performance in transfer learning applications. Therefore, we plan to conduct research on optimized and advanced ensemble approaches based on feature fusion to overcome performance degradation when combining heterogeneous models with classification accuracy imbalance.

Acknowledgement

References

[1] S. K. Singh, J. Cha, T. W. Kim, and J. H. Park, "Machine learning based distributed big data analysis framework for next generation web in IoT," *Computer Science and Information Systems*, vol. 18, no. 2, pp. 597-618, 2021. https://doi.org/10.2298/CSIS200330012S

[2] B. Li, X. Zheng, W. Guo, Y. Wang, R. Mao, X. Cheng, et al., "Radiation pneumonitis prediction using multi-omics fusion based on a novel machine learning pipeline," *Human-centric Computing and Information Sciences*, vol. 12, article no. 49, 2022. https://doi.org/10.22967/HCIS.2022.12.049

[3] E. Kristiani, Y. T. Tsan, P. Y. Liu, N. Y. Yen, and C. T. Yang, "Binary and multi-class assessment of face mask classification on edge AI using CNN and transfer learning," *Human-centric Computing and Information Sciences*, vol. 12, article no. 53, 2022. https://doi.org/10.22967/HCIS.2022.12.053

[4] H. Lee, J. Kim, J. Yu, Y. Jeong, and S. Kim, "Multi-class classification using transfer learning based convolutional neural network," *Journal of Korean Institute of Intelligent Systems*, vol. 28, no. 6, pp. 531-537, 2018. https://doi.org/10.5391/jkiis.2018.28.6.531

[5] Y. Kim, K. Yu, and J. Han, "Research for radar signal classification model using deep learning technique," *Journal of the Korea Institute of Military Science and Technology*, vol. 22, no. 2, pp. 170-178, 2019. https://doi.org/10.9766/KIMST.2019.22.2.170

[6] S. J. Hong, Y. G. Yi, J. W. Choi, J. Jo, and B. S. Seo, "Classification of Radar signals using machine learning techniques," *Journal of IKEEE*, vol. 22, no. 1, pp. 162-167, 2018. https://doi.org/10.7471/ikeee.2018.22.1.162

[7] H. Surrisyad and Wahyono, "A fast military object recognition using extreme learning approach on CNN," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 12, pp. 210-220, 2020. https://doi.org/10.14569/ijacsa.2020.0111227

[8] Stanford Vision Lab, "ImageNet overview," 2020 [Online]. Available: https://image-net.org/about.php.

[9] K. H. Kim and J. H. Bae, "Important parameter optimized flow-based transfer learning technique supporting heterogeneous teacher network based on deep learning," *Journal of the Korean Institute of Information Technology*, vol. 18, no. 3, pp. 21-29, 2019. http://doi.org/10.14801/jkiit.2020.18.3.21

[10] M. Jeong and J. Ma, "A study on the improvement of submarine detection based on mast images using an ensemble model of convolutional neural networks," *Journal of the Korea Institute of Military Science and Technology*, vol. 23, no. 2, pp. 115-124, 2020. https://doi.org/10.9766/KIMST.2020.23.2.115

[11] J. Y. Hwang, B. A. Choi, J. H. Lee, and J. H. Bae, "A study on transfer learning-based ensemble model for classification of maneuver weapon system," *Journal of the Korean Institute of Information Technology*, vol. 19, no. 12, pp. 1-10, 2021. http://doi.org/10.14801/jkiit.2021.19.12.1

[12] R. O. Lane, A. J. Wragge, W. J. Holmes, S. J. Bertram, and T. Lamont-Smith, "Object detection in EO/IR and SAR images using low-SWAP hardware," in *Proceedings of 2021 Sensor Signal Processing for Defence Conference (SSPD)*, Edinburgh, UK, 2021, pp. 1-5. https://doi.org/10.1109/SSPD51364.2021.9541497

[13] Z. Yang, W. Yu, P. Liang, H. Guo, L. Xia, F. Zhang, Y. Ma, and J. Ma, "Deep transfer learning for military object recognition under small training set condition," *Neural Computing and Applications*, vol. 31, pp. 6469-6478, 2019. https://doi.org/10.1007/s00521-018-3468-3

[14] D. Legendre and J. Vankka, "Military vehicle recognition with different image machine learning techniques," in *Information and Software Technologies*. Cham, Switzerland: Springer, 2020, pp. 220-242. https://doi.org/10.1007/978-3-030-59506-7_19

[15] T. Hiippala, "Recognizing military vehicles in social media images using deep learning," in *Proceedings of 2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*, Beijing, China, 2017, pp. 60-65. https://doi.org/10.1109/ISI.2017.8004875

[16] Y. Ouyang, X. Wang, R. Hu, H. Xu, and F. Shao, "Military vehicle object detection based on hierarchical feature representation and refined localization," *IEEE Access*, vol. 10, pp. 99897-99908, 2022. https://doi.org/10.1109/ACCESS.2022.3207153

[17] S. Chung and M. G. Chung, "Pedestrian classification using CNN's deep features and transfer learning," *Journal of Internet Computing and Services*, vol. 20, no. 4, pp. 91-102, 2019. https://doi.org/10.7472/jksii.2019.20.4.91

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, 2014, pp. 1-14.

[19] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetv2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, pp. 4510-4520. https://doi.org/10.1109/CVPR.2018.00474

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, 2017, pp. 2261-2269. https://doi.org/10.1109/CVPR.2017.243

[22] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (ELUs)," in *Proceedings of the 4th International Conference on Learning Representations*, San Juan, Puerto Rico, 2016.

[23] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929-1958, 2014.

[24] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, CA, 2015.

[25] S. W. Park, J. C. Kim, and D. Y. Kim, "A study on classification performance analysis of convolutional neural network using ensemble learning algorithm," *Journal of Korea Multimedia Society*, vol. 22, no. 6, pp. 665-675, 2019. https://doi.org/10.9717/kmms.2019.22.6.665

Jinyong Hwang https://orcid.org/0000-0002-3094-7436

He received the B.S. degree from the Department of Electronics and Telecommunications Engineering from Hannam University, Daejeon, Republic of Korea, in 2008, and the M.S. degree in electronics and telecommunications engineering from Chungnam National University, Daejeon, Republic of Korea, in 2011. Since March 2021, he has been with the Department of AI and Big Data Engineering from Daegu Catholic University as an M.S. candidate, and he joined the laboratory of the Artillery System Center, Hanwha Aerospace Corporation, Changwon, Republic of Korea, in 2019, where he is currently a senior researcher. His current research interests include deep learning, transfer learning, maneuver weapon systems, and digital hardware.

You-Rak Choi https://orcid.org/0000-0002-7868-257X

He earned his bachelor's and master's degrees in Computer Science from Chungnam National University in 1991 and 1993, respectively. In 1993, he started working as a researcher at the Korea Atomic Energy Research Institute (KAERI) as an exception to military service. In 2022, he earned his Ph.D. from the Department of Computer Science and Engineering, Chungnam National University under the guidance of Prof. Hoon Choi. Since 2011, he has been working as a principal researcher at KAERI. His research interests are smart detection and diagnosis for safety-related devices in nuclear power, cloud technology, and artificial intelligence technology.

Tae-Jin Park https://orcid.org/0000-0001-9057-9201

He earned a B.Sc. and M.Sc. in chemistry from Yonsei University in 1997 and 1999, respectively. After completing military service in 2001, he worked as a research scientist at the Korea Institute Science Technology (KIST). In 2007, he earned his Ph.D. in the Department of Chemistry at the State University of New York at Stony Brook under the supervision of Prof. Stanislaus S. Wong. After finishing a postdoctoral fellowship in NEAT ORU at University of California at Davis with Prof. Alexandra Navrotsky, he held the position, since 2011, of senior researcher at the Korea Atomic Energy Research Institute (KAERI). He is continuing his work at KAERI as principal researcher since 2017. His research interests include smart sensing and diagnosis for safety-related instrumentation in nuclear energy, safety in radwaste disposal, nano and materials sciences, thermochemistry of materials relevant to nuclear energy.

Ji-Hoon Bae https://orcid.org/0000-0002-0035-5261

He received the B.S. degree in electronic engineering from Kyungpook National University, Daegu, Republic of Korea, in 2000, and the M.S. and Ph.D. degrees in electrical engineering from Pohang University of Science and Technology, Pohang, Gyeongbuk, Republic of Korea, in 2002 and 2016, respectively. From 2002 to 2019, he was a Principal Researcher with the Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea. He joined the faculty of the Department of AI and Big Data Engineering, Daegu Catholic University, Gyeongsan, Republic of Korea in 2019, where he is currently an assistant professor. His research interests include deep learning, transfer learning, radar imaging, radar signal processing, and optimized techniques. He is a member of the Institute of Electrical and Electronics Engineers, the Korea Institute of Information Technology, and the Korea Institute of Electromagnetic Engineering and Science.