Infrared Variation Optimized Deep Convolutional Neural Network for Robust Automatic Ground Target Recognition

Sungho Kim
Yeungnam University
280 Daehak-Ro, Gyeongsan, Gyeongbuk, 38541, Korea
sunghokim@ynu.ac.kr

Woo-Jin Song
Pohang University of Science and Technology
Pohang, Gyeongbuk, 37673, Korea
wjsong@postech.ac.kr

So-Hyun Kim
Agency for Defense Development
111 Sunam-dong, Daejeon, 34186, Korea
153074@add.re.kr

Abstract

Automatic infrared target recognition (ATR) is a traditionally unsolved problem in military applications because of the wide range of infrared (IR) image variations and limited number of training images, which is caused by various 3D target poses, non-cooperative weather conditions, and difficult target acquisition environments. Recently, deep convolutional neural network-based approaches in RGB images (RGB-CNN) showed breakthrough performance in computer vision problems, such as object detection and classification. The direct use of the RGB-CNN to IR ATR problem fails to work because of the IR database problems. This paper presents a novel infrared variation-optimized deep convolutional neural network (IVO-CNN) by considering database management, such as increasing the database by a thermal simulator, controlling the image contrast automatically and suppressing the thermal noise to reduce the effects of infrared image variations in deep convolutional neural network-based automatic ground target recognition. The experimental results on the synthesized infrared images generated by the thermal simulator (OKTAL-SE) validated the feasibility of IVO-CNN for military ATR applications.

1. Introduction

Automatic target recognition (ATR) in infrared (IR) images has been an active research topic historically because of its military applications with 24 hour operation capability [2, 17]. Although the original ATR covers target detection (region of interest extraction), classification, tracking, and threat assessment [2, 17], this paper focuses only the IR target classification problem to solve IR variations. IR images can visualize hot targets regardless of the time, but IR target images show wide range of intensities depending on the 3D target poses and weather conditions.

Since the 1980s, model-based approaches were popular and targets were recognized by alignment methods such as geometric hashing [11]. Since then, various image learning-based target recognition methods have been proposed by considering both the feature extractors and classifiers to cope with IR variations. The Markov tree feature [3], IR wavelet feature [16], scale invariant feature transform (SIFT) [6], histogram of oriented gradients (HOG) [24], and moment features [19, 25] are recently proposed infrared features and showed promising recognition results on their own applications. Simple machine learning-based classification methods, such as nearest neighbor classifier [6], Bayesian classifier, conventional neural network, Adaboost [13], and support vector machine (SVM) [24] are frequently used to discriminate the target features for classification.

Recently, deep learning-based algorithms have been proposed and showed excellent performance in RGB-based object classification on ImageNet [10] and CIFAR image database [21]. On the other hand, AlexNet requires 1.2 million high-resolution images to learn 60 million parameters and 650,000 neurons consisting of five convolutional layers [10]. ATR researchers have attempted to apply deep convolutional neural network (CNN) algorithms to IR target classification problems. Rodger et al. trained a CNN consisting of 5 convolutional layers with 10,000 long-wave IR instances sampled over 6 object classes [18]. Although the deep learning results shows promising results for the trained database, it lacks the analysis of an IR variation effect on the CNN. Akula et al. presented a CNN-based 4 category (auto, human,ambassador, and background) clas-

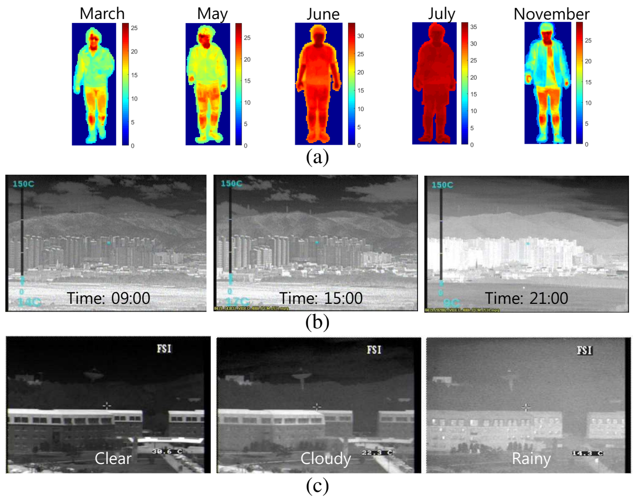


Figure 1. Examples of IR variations depending on (a) season, (b) day/night time, and (c) weather conditions.

sification and provided a classification accuracy of 88.15% using only 378 training images, which is too small to learn the complex CNN architecture [1].

The success of deep learning for object classification using RGB color images provides the starting point of IR ATR and recent deep learning applications to IR ATR highlight the feasibility. The key success of IR-based deep learning depends strongly on how to handle the infrared variations [9]. As shown in Fig. 1, the IR target images show quite different intensity variations caused by the target states (pose, heat), recording time, weather, and climate, which alters the image contrast, thermal noise, and blurring level. In general, if there are large image variations, deep learning systems need many training images to handle the variations. Furthermore, the infrared ground target recognition problem can be more difficult if the surveillance area cannot be accessed directly due to the restricted image acquisition.

This paper focuses on the IR variation problem in deep convolutional neural network-based target recognition and proposes a novel IR variation optimized convolutional neural network (IVO-CNN) by considering an IR variation reduction strategy. A synthetic IR image generator, OKTAL-SE, was adopted to make various IR images by varying the target pose and atmospheric parameters [12, 20, 5]. In particular, a 14-layered CNN (four convolutions) was constructed and the IR variation optimization block was inserted in front of the CNN architecture.

Section 2 presents the proposed IVO-CNN architecture, including variable DB generation. Section 3 validates the proposed method by applying this to the IR DB and Section 4 concludes the paper.

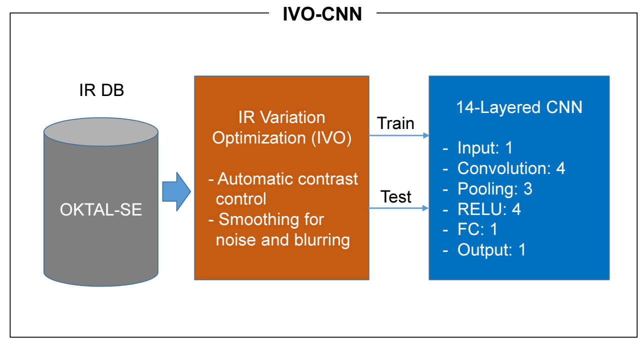


Figure 2. Proposed IR Variation Optimized deep Convolutional Neural Network (IVO-CNN) for automatic ground target recognition with synthetic image generator.

2. Proposed IR variation optimized deep convolutional neural network (IVO-CNN)

IR-based ATR for military applications has two critical issues: limited number of IR target images and large IR image variations. The small sized IR DB problem in deep learning can be mitigated using either transfer learning [14] or a synthetic IR image generator. The former can be implemented easily to IR ATR by extracting the features from AlexNet [10] and training an SVM classifier. This can be a feasible solution but it shows degraded performance on various IR images because the Alexnet has not been optimized to the IR images. The key idea of the second issue is to use the IR variation optimization (IVO) block during CNN training and testing, as shown in Fig. 2. Various IR target DB were constructed using OKTAL-SE by changing the geometric and environmental parameters. The core block, IR variation Optimization, is located in front of the 14-layered CNN to reduce the IR variations. Details of the proposed method will be explained in the following subsections.

2.1. OKTAL-SE: synthetic IR image generation

A synthetic IR image generator is useful for acquiring the target images in inaccessible areas and evaluating the effect of IVO block to the CNN-based target recognition quantitatively. Among the various IR image generators such as RadTherm-IR, IR-Workbench (OKTAL-SE), and F-TOM [20], the OKTAL-SE [12] was selected because it is the only simulator that can synthesize both passive (EO/IR) and active (synthetic aperture radar) data. Fig. 3 summarizes the overall flow of IR synthesis. Given the simulation parameters, such as weather and time, the atmospheric transmittance was calculated. The scenario program can select the background and target trajectory, and SE-RAY-IR then synthesizes the IR sequences using the ray tracing method. Fig. 4 shows the synthesized IR target images according to the depression (imaging) angle, weather condi-

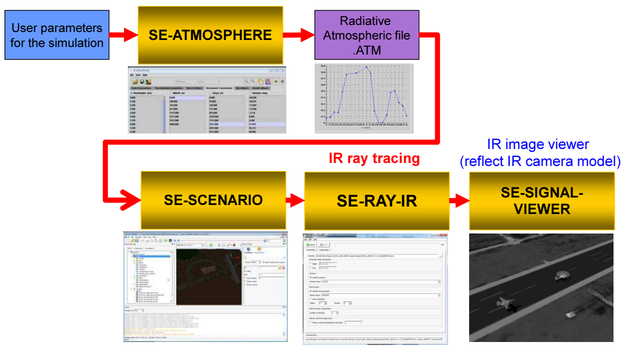


Figure 3. OKTAL-SE-based synthetic IR image generation flow: SE-ATMOSPHERE generates atmospheric data and SE-SCENARIO controls targets and background. SE-RAY-IR synthesize IR images by ray tracing and SE-SIGNAL-VIEWER visualizes the results.

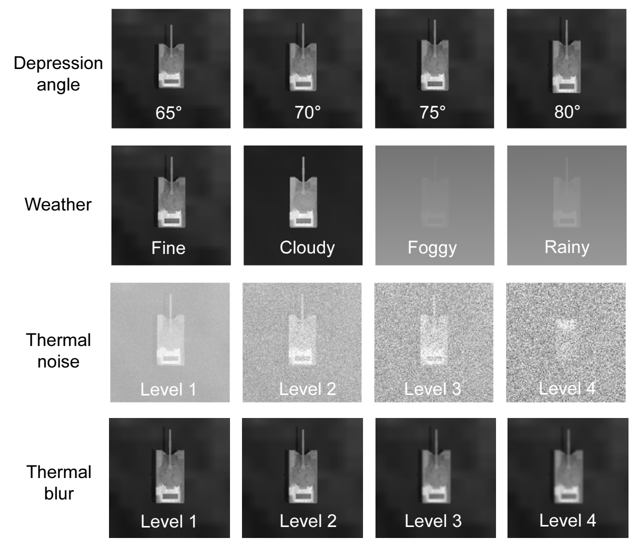


Figure 4. Synthesized IR target examples using the OKTAL-SE for a T72 target by varying the target attributes, imaging geometry, and atmospheric parameters.

tion, thermal noise, and thermal blur. Note that the same target (T72) shows different levels of image contrast, noise level, and image blur. IR image variations can be affected by the target attributes (pose, heat status, and thermal stealth coating), camera geometry (imaging angle (depression angle), distance), and atmospheric conditions (weather, noise, and blur).

2.2. Proposed IR variation optimization

- **IR variation reduction strategy I**: Training all the images

The simplest way to handle object variations in a deep convolutional network is to train all possible examples, such as AlexNet that used 1.2 million RGB images to train 1,000

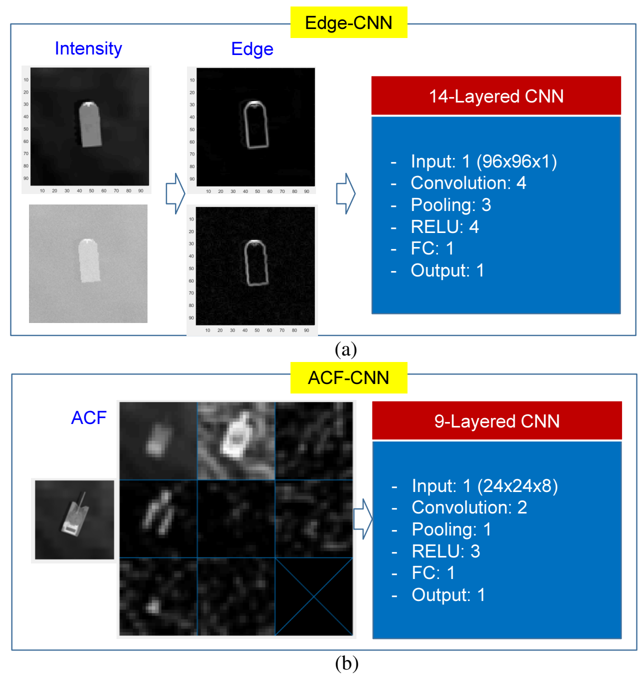


Figure 5. Basic flow of strategy II (IR variation reduction in feature space) using (a) edge information and (b) ACF information.

object categories [10]. If the same strategy is applied to the ground IR target recognition problem, the same CNN model cannot be trained (large residual error) because of the limited number of IR images.

- **IR variation reduction strategy II**: Preprocessing in the feature space

The next possible approach is to preprocess the IR images in feature space, which is used frequently in traditional ATR problems. For example, a gradient magnitude map (GMM) and gradient orientation map (GOM) are known to be robust to thermal contrast and used frequently in edge-based target matching [7]. Fig. 5(a) gives an example implementation flow using the edge information (Edge-CNN). The Aggregated Channel Feature (ACF) gave powerful pedestrian detection performance using color channels, a gradient magnitude channel, and gradient orientation channels [4]. Fig. 5(b) shows the implementation using ACF and 9-layered CNN (ACF-CNN). If the extracted feature map (GMM, GOM, ACF) is used to train the conventional CNN it works quite well with low contrast IR targets. On the other hand, the feature-based CNN shows a poor target recognition result to thermal noise variations due to gradient operation.

- **IR variation reduction strategy III (proposed method)**: Preprocessing in image space

This paper presents a novel IR variation optimization (IVO) in image space by introducing thermal contrast control (TCC, f_{TCC}) and a thermal smoothing filter (TSF, f_{TSF}), as defined in eq. (1), where $I(x,y)$ and $I_{IVO}(x,y)$ present an input image and a processed image at the pixel position (x,y), respectively.

$$I_{IVO}(x,y) = f_{TCC}(I(x,y)) * f_{TSF}(x,y,\sigma) \quad (1)$$

As defined in eq. (2), $f_{TCC}(I(x,y))$ is designed as an intensity transformation function and $t_{TSF}(x,y,\sigma)$ is a Gaussian smoothing kernel with a control parameter, σ (normally 1).

$$I_{IVO}(x,y) = \frac{255 \cdot (I(x,y) - I_{min})}{I_{max} - I_{min}} * G(x,y,\sigma) \quad (2)$$

The TCC can control the thermal contrast using the maximum intensity (I_{max}) and minimum intensity (I_{min}) estimated in a target chip, which produces consistent thermal contrast images. The TSF using the Gaussian kernel can remove the thermal noise effectively because the noise of infrared images follows a Gaussian distribution [8] and a Gaussian mean filter can provide an unbiased optimal signal estimation [15].

Fig. 6 demonstrates the effect of IVO for various IR images for the same target (T72). IVO processing can minimize IR variations by controlling the thermal contrast and removing thermal noise. The column sectional IR data indicated in the top-left (red line) of Fig. 6 is used to analyze the IVO effect quantitatively. Fig. 7(a) shows the original IR intensity curves for the normal, noisy, cloudy, and blurry target images. Fig. 7(b) shows the IR intensity curves after applying the proposed IVO processing for the same intensity curves. Note that the intensity curves show similar patterns after the IVO function. The performance of the IR variation optimization can be analyzed quantitatively, as shown in Fig. 7(c), by calculating the standard deviation per row. The proposed IVO can reduce the level of IR variation for each data and the average variation of the original curves is 51.7 and that of the IVO curves is 12.1, which is 4.3 times lower than the original one.

2.3. Architecture of 14-layered CNN

As a basic deep learning-based classifier, LeNet-based deep convolutional neural network architecture [23] is used by changing the input size and the number of layers for 16 IR target recognition, as shown in Fig. 8. The MatConvNet toolbox [22] was used for training and learning because this study focused on the IR variation effect on CNN.

The architecture consists of 14 layers (IR-CNN14): 1 input layer, 4 convolutional layers, 3 pooling layers, 4 REctified Linear Unit (RELU) layers, 1 fully connected (FC) layer, and 1 output layer. The IR-CNN14 architecture can be explained in terms of the data flow and operational flow.

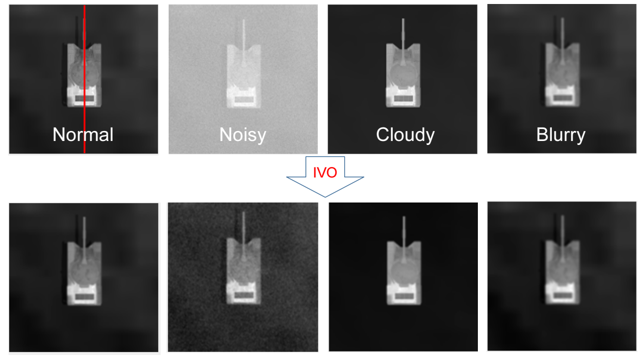


Figure 6. Examples of IR target images (top) before IVO and (bottom) after IVO processing.

An input layer receives an IR image with a size of 96 × 96. The first convolution operation using 32 kernels with a 5 × 5 support region, stride 1, and padding 2 produces 32 feature data with a 96 × 96 resolution, as shown in Fig. 8. Through the max pooling with the 3 × 3 support region, stride 2, padding ([top bottom left right]=[0 1 0 1]) and RELU, sub-sampled feature images are obtained. Two additional convolutions, RELU, and average pooling operations, produce 64 feature data with a 12 × 12 resolution. The fourth convolution with 64 kernels of size $12 \times 12 \times 64$, stride 1, padding 0 generates a feature vector of 1×64 that is fully connected to 16 output nodes, where the softmax function is used to calculate the probability distribution over 16 targets.

3. Experimental results

3.1. Preparation of IR database

OKTAL-SE can generate various IR target images by varying the camera setting (spectral band, detector size, field of view, depression angle, and height), atmospheric setting, and target pose (aspect angle). The simulation scenario is assumed to be the ground surveillance on an unmanned aerial vehicle. The spectral band is basically midwave IR (MWIR) and other camera parameters are set to produce a $5cm \times 5cm$ resolution per pixel.

Table 1 lists the composition of the target DB for training and testing for IVO-CNN. The total number of targets is 16; among them 10 targets are military targets (BMP3, T72, AMX10, AMX10RC, Leclerc, Jeep, TMM, Rada Camo, SA9 Inch Camo, and VAB OBS) and 6 targets are non-military targets (Audi, Bus, Clio, Firetruck, Oil tanker, and Ford transit). Three kinds of training DB were prepared, such as normal, noise, and long-wave IR (LWIR), for the IR variation analysis in CNN. In the case of a normal DB, each target template has a 96×96 image resolution with a depression angle of $65, 70, 75, and 80°$ and aspect angle of $5°$. The total number of normal training DB was 4,608 (16 targets ×

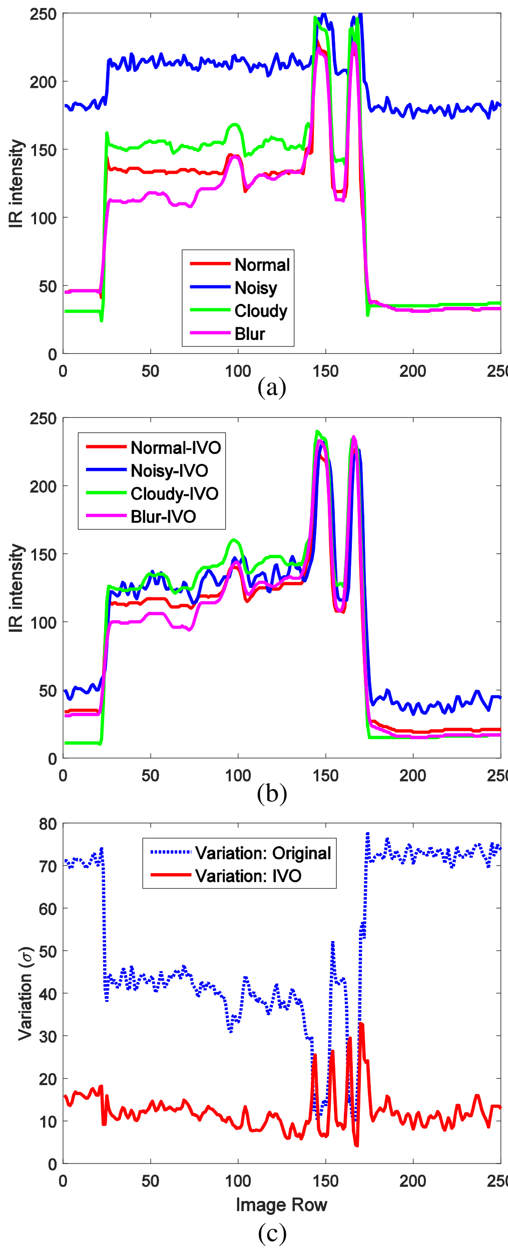


Figure 7. Quantitative analysis of the IR variations for the column sectional data: (a) original IR data, (b) IR data after IVO, (c) comparison of the IR variations in terms of the standard deviation.

72 aspect angles $\times$ 4 depression angles). Fig. 9 shows the 16 targets (top) and 72 aspect views for the T72 at a depression angle of 75° (bottom). The noise and LWIR DBs were prepared by adding thermal camera noise and changing the spectral band in the OKTAL-SE for IR variation analysis. In the testing phase, thermal noise was inserted and image blurring was applied artificially to the training DB. The total

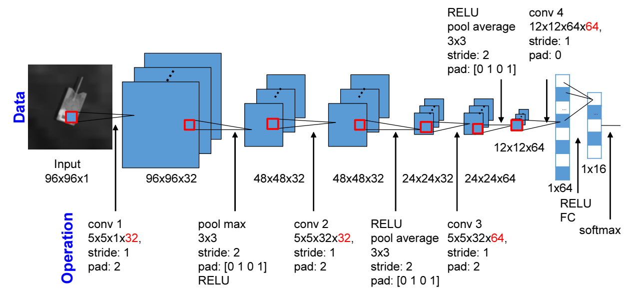


Figure 8. Details of 14-layered deep convolutional network for 16 IR target recognition.

Table 1. Composition of the training and test DB

Mode		No. DB	Size	Dep. angle	Asp. angle
Train	normal	4,608	96×96	65,70,75,80°	5°
	noise	1,152	96×96	75°	5°
	LWIR	1,152	96×96	75°	5°
Test		artificial noise adding and blurring			

number of noise DB was 1,152, where a depression angle of 75° was considered. Similarly, the LWIR DB has the same number of IR images as the noise DB.

3.2. Evaluation of the proposed IVO-CNN

In the first experiment, strategy I (simply increasing training DB in the CNN) was conducted to handle the IR variation problem and the learning result was checked. The basic training set is just a normal DB (aspect views with different depressions) in Table 1. Additional IR variation sets are noise and LWIR images, which should be learned in strategy I. Fig. 10(a) shows the limitation of direct IR variation learning in the IR-CNN14. The objective function cannot be reduced after epoch 10, which leads to a large top 1 residual error (15%). If the proposed IVO (strategy III) is adopted in the CNN learning, both the objective function and top 1 error curves show successful learning, as shown in Fig. 10(b) after epoch 25.

In the second experiment, the Edge-CNN, ACF (total, orientation only)-CNN, and the proposed IVO-CNN method were compared in terms of the IR image variations, such as blur level (controlled by σ_S) and thermal noise level (controlled by σ_N). These methods were trained using the train DB, as listed in Table 1. The test images were prepared by artificially adding Gaussian blur and thermal noise to the training DB. Fig. 11(a) presents the performance comparison results for the IR images blurred with different levels. The proposed IVO-CNN showed the best robustness to IR image blurring followed by ACF(tot)-CNN, ACF(ori)-CNN, and Edge-CNN. The Edge-CNN showed the worst target classification rate because image blur suppresses the details of edge information. Fig. 11(b) shows the performance comparison results for the noisy IR im-

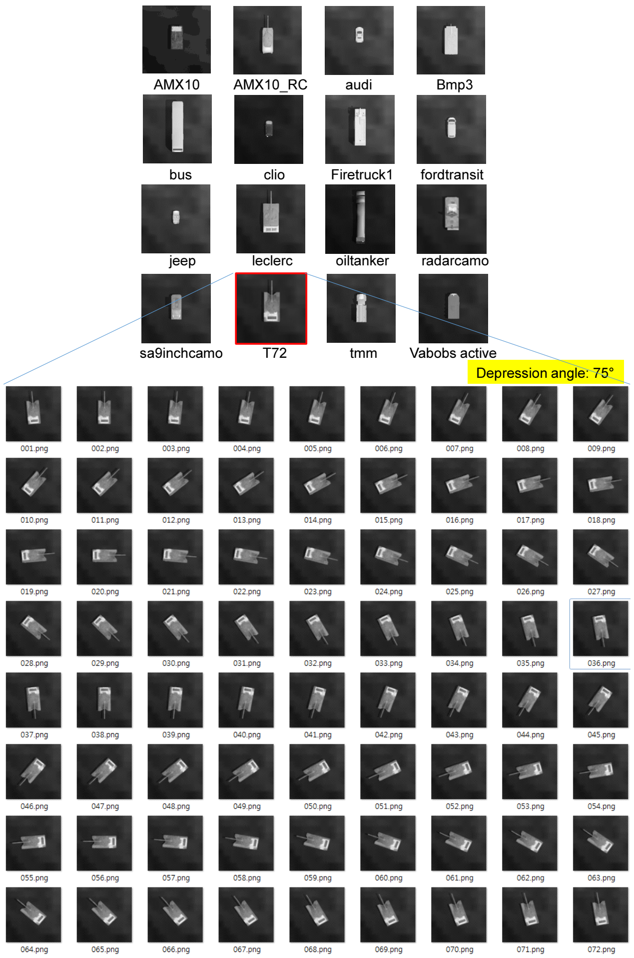


Figure 9. Composition of the IR target database: (top) 16 ground targets, (bottom) 72 aspect views of T72 at depression angle 75°.

ages. The proposed IVO-CNN showed the best robustness to noisy IR images followed by Edge-CNN, ACF(tot)-CNN, and ACF(ori)-CNN. The ACF(ori)-CNN showed the poorest target classification rate because the orientation information was obtained by the noise-sensitive gradient and arctangent operations.

4. Conclusions

This paper presented a novel Infrared (IR) Variation Optimized deep Convolutional Neural Network (IVO-CNN) for ground infrared target recognition. The IR variations are still an unsolved problem due to the target variations themselves and atmospheric condition variations. Recently, deep convolutional neural networks showed breakthrough performance in computer vision problems by learning a huge number of training images, which is difficult in IR-based target recognition due to the military specialty. This paper proposed a novel IR image variation optimization method by thermal contrast control (TCC) and a thermal smoothing filter (TSF) in CNN-based target recognition. The 16 IR target DB was prepared by changing the depression angle, aspect angle, and atmospheric parameters in the OKTAL-SE simulator. The proposed IVO-CNN can provide stable IR target recognition to a wide range of IR variations. In addition, it can remove the residual training error effectively compared to the other baseline methods (Edge-CNN and ACF-CNN). In the future, the IVO-CNN method will be fused with synthetic aperture radar (SAR-CNN) for more confident target recognition regardless of the weather conditions.

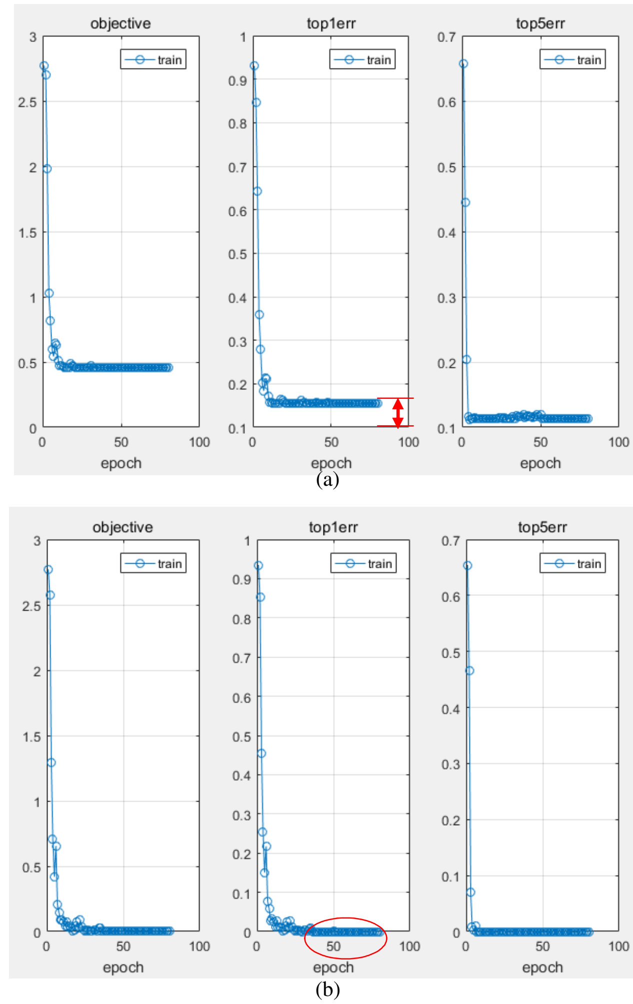
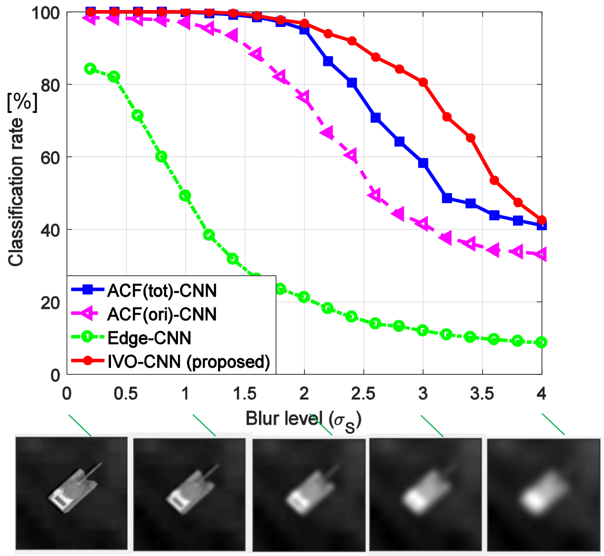
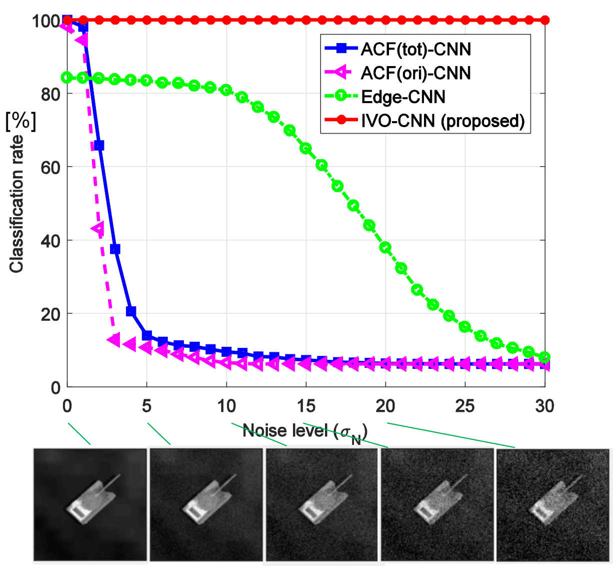


Figure 10. (a) Problem of strategy I (training IR variation by increasing DB) in CNN, (b) the effect of IVO in CNN training.

Acknowledgements: This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2016R1D1A1B03930076) and by the STRL (Sensor Target Recognition Laboratory) program of



(a)



(b)

Figure 11. (a) Problem of strategy I (training IR variation by increasing DB) in CNN, (b) effect of IVO in CNN training.

Defense Acquisition Program Administration and Agency for Defense Development. This research was also supported by the ETRI R&D Program [16Z3300, Advanced Development of Fusion-Platform and Local Parts Maker Support Project for Context-ware Smart Vehicle] funded by the government of South Korea.

References

[1] A. Akula, A. Singh, R. Ghosh, S. Kumar, and H. Sardana. Target recognition in infrared imagery using convolutional neural network. In *Proceedings of International Conference on Computer Vision and Image Processing*, volume 2, pages 25–34, 2017.

[2] B. Bhanu. Automatic target recognition: State of the art survey. *IEEE Trans. on Aerospace and Electronic Systems*, 22(4):364–379, 1986.

[3] P. Bharadwaj and L. Carin. Infrared-image classification using hidden markov trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1394–1398, 2002.

[4] P. Dollar, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1532–1545, 2014.

[5] A. L. Goff, T. Cathala, and J. Latger. New impressive capabilities of se-workbench for EO/IR real-time rendering of animated scenarios including flares. In *Proc. of SPIE*, volume 9653, page 965307, 2015.

[6] G. J. Gray, N. Aouf, M. A. Richardson, B. Butters, R. Walmsley, and E.Nicholls. Feature-based target recognition in infrared images for future unmanned aerial vehicles. *J. of Battle field Technology*, 14(2):27–36, 2011.

[7] S. Kim and I.-S. Kweon. 3D target recognition using cooperative feature map binding under Markov Chain Monte Carlo. *Pattern Recognition Letters*, 27:811–821, 2006.

[8] S. Kim, W.-J. Song, and S.-H. Kim. Robust ground target detection by sar and ir sensor fusion using adaboost-based feature selection. *Sensors*, 2016(16):1–27, 2016.

[9] S. Kim, S.-G. Sun, and K.-T. Kim. Analysis of infrared signature variation and robust filter-based supersonic target detection. *The Scientific World Journal*, 2014:1–17, 2014.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networkse. In *Advances in Neural Information Processing Systems*, pages 1–9, 2012.

[11] Y. Lamdan and H. J. Wolfson. Geometric hashing: a general and efficient model-based recognition scheme. In *2nd International Conference on Computer Vision*, pages 238–249, 1988.

[12] J. Latger, T. Cathala, N. Douchin, and L. Goff. Simulation of active and passive infrared images using the SE-WORKBENCH. In *Proc. of SPIE*, volume 6543, page 654302, 2007.

[13] L. Li and Y. Ren. Infrared target recognition based on combined feature and improved adaboost algorithm. *Advances in Intelligent and Soft Computing*, 105.

[14] M. Oquab and L. B. I. L. J. Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1717–1724, 2014.

[15] A. Papoulis and S. U. Pillai. *Probability, random variables, and stochastic processes*. McGraw-Hill, 4th edition, 2009.

[16] L. Quan and Z. Jianjun. Wavelet-based feature extraction and recognition of infrared target. In *Proceedings of the First International Conference on Innovative Computing, Information and Control*, pages 1–4, 2006.

[17] J. A. Ratches. Review of current aided/automatic target acquisition technology for military target acquisition tasks. *Optical Engineering*, 50(7):072001, 2011.

[18] I. Rodger, B. Connor, and N. M. Robertson. Classifying objects in LWIR imagery via CNNs. In *Proc. SPIE*, volume 9987, page 99870H, 2016.

[19] F. A. Sadjadi and A. Mahalanobis. Robust automatic target reconition in flir imagery. In *Proc. SPIE*, volume 8391, page 839105, 2012.

[20] F. Schwenger, P. Grossmann, and A. Malaplate. Validation of the thermal code of RadTherm-IR, IR-Workbench, and F-TOM. In *Proc. of SPIE*, volume 7300, page 73000J, 2009.

[21] J. Snoek, O. Rippely, K. Swerskyx, R. Kirosx, N. Satishz, N. Sundaramz, M. M. A. Patwaryz, Prabhat, and R. P. Adams. Scalable bayesian optimization using deep neural networks. In *International Conference on Machine Learning*, volume 37, 2015.

[22] A. Vedaldi and K. Lenc. MatConvNet: Convolutional neural networks for MATLAB. In *Proceedings of the 23rd ACM International Conference on Multimedia*, pages 689–692, 2015.

[23] X. Wang. Deep learning in object recognition, detection, and segmentation. *Foundations and Trends in Signal Processing*, 8(4):217–382, 2016.

[24] T. Zhan and N. Sang. Forward-looking infrared target recognition based on histograms of oriented gradients. In *Proc. SPIE*, volume 8003, page 80030S, 2011.

[25] F. Zhang, S. Liu, D. Wang, and W. Guan. Aircraft recognition in infrared image using wavelet moment invariants. *Image and Vision Computing*, 27(4):313–318, 2009.