Land Use and Land Cover Classification Using CNN, SVM, and Channel Squeeze & Spatial Excitation Block

H I Dewangkoro\(^1\) and A M Arymurthy\(^2\)

\(^{1,2}\) Faculty of Computer Science, Universitas Indonesia, Depok, Jawa Barat, Indonesia

Email: herdito.ibnu@ui.ac.id\(^1\), aniati@cs.ui.ac.id\(^2\)

Abstract. One of the materials essential for human life that must manage properly is the land. Land use and land cover (LULC) classification can help us how to manage land. The satellite can record images that can use as the data for LULC classification. This research aims to perform LULC classification using Convolutional Neural Network (CNN) on EuroSAT remote sensing image dataset taken from the Sentinel-2 satellite. CNN has become a well-known method to deal with image feature extraction. We used several CNN for feature extraction, such as VGG19, ResNet50, and InceptionV3. Then, we recalibrated the feature of CNN using Channel Squeeze & Spatial Excitation (sSE) block. We also used Support Vector Machine (SVM) and Twin SVM (TWSVM) as the classifier. VGG19 with sSE block and TWSVM achieved the highest experimental results with 94.57% accuracy, 94.40% precision, 94.40% recall, and 94.39% F1-score.

1. Introduction
The land is a necessary material related to human life. Because of the importance of land, people must manage it properly. One example of the beneficial activity related to land is land use and land cover (LULC) classification. Land use refers to land used by humans that requires natural environment management and modification. Then, the features or characteristics of the surface and subsurface of the Earth, the surface of the water, the structures of the human, topographic, and containing soil is called land cover [1]. LULC classification’s objective is to automatically include labels defining the kind and usage of physical land represented [2]. It can support multiple sectors such as environment, economics, and nature-related research [3].

Remote sensing application refers to observing and recording the Earth’s surface objects [1]. The development of technology can make doing some work more straightforward. One of the examples is remote sensing application has supported by the technology of satellite. The satellite can record the surface of Earth to produce remote sensing images.

Remote sensing images can be used as the data to perform LULC classification. Hence, the LULC classification can relate to image classification tasks. The image classification task aims to label the image based on their category [4].

In recent times, deep learning (DL) has become so popular for remote sensing image analysis. DL can surpass machine learning methods in various applications. There are many DL method use for remote sensing, such as Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Autoencoder (AE), and Generative Adversarial Network (GAN) [5].

CNN is one example of a DL method used for many tasks in image processing, such as image classification, semantic segmentation, and object detection [6]. CNN has the ability to do feature
extraction automatically [7]. Then, extract features with CNN can be easier instead of getting handcrafted features manually.

This research will study the classification of remote sensing images provided by the Sentinel-2 satellite. Several CNN are used to perform feature extraction. We also recalibrated CNN’s feature by the method called Channel Squeeze & Spatial Excitation (sSE) block [8]. Support Vector Machine (SVM) [9] and Twin SVM (TWSVM) [10] are also used as the classifier.

2. Related Work
There are many works did by researchers to do remote sensing image classification. Yang and Newsam used several methods to extract features from remote sensing images, such as Bag of Visual Words (BOVW), Spatial Co-occurrence Kernel (SCK), and Spatial Pyramid Match Kernel (SPMK). Then, they use SVM as the classifier [11]. It is one of the most popular classifiers for supervised learning [12].

With the development of Graphics Processing Unit (GPU) and supported by the availability of public large image repositories, many researchers used CNN to classify images [13]. It is also supported by the ability of CNN to extract features automatically [7]. The example of CNN used for remote sensing image classification are VGG19, ResNet50, and InceptionV3 [14].

CNN mostly uses the softmax activation function for the classification task [15]. The combination of CNN and SVM also can be applied for remote sensing image classification. CNN is used to extract features from remote sensing images and SVM is used as classifier. Cheng et al. combined several CNN, such as AlexNet, VGG16, and GoogleNet with SVM [16]. The other combination of CNN and SVM is also proposed by Sun et al. They combined seven-layer CNN with SVM [17].

Channel Squeeze and Spatial Excitation (sSE) block is proposed by Roy et al. to recalibrate the CNN’s feature map. The recalibration process is aimed to boost meaningful features and to suppress weak features. sSE block is initially proposed for image segmentation [8]. This research used sSE block on CNN for remote sensing image classification.

3. Methodology
This research used sSE block on CNN to classify remote sensing images. We used several CNN, such as ResNet50, InceptionV3, and VGG19. Then, we also replaced softmax on CNN with SVM and TWSVM as the classifier. CNN used as feature extraction when combined with SVM and TVSVM.

VGG19 is the influential Deep CNN proposed by Simonyan and Zisserman [14], [18]. It contains 3 x 3 convolution and 2 x 2 max pooling. It also has 19 layers deep with 16 convolutional layers and 3 fully connected layers [14].

Microsoft’s ResNet50 is a kind of CNN that uses the skip connection [19], [20]. It has 50 layers deep. ResNet50 uses skip connection to add extra information to the layer. The use of skip connection can get deeper model with minimizing overfitting [19]. The example of skip connection shows in Figure 1. Supposed we have x as input, the function of layer 1 and 2 is f(x), and the target function is g(x). Then, we can get g(x) = f(x) + x or f(x) = g(x) - x.

Google’s InceptionV3 uses the inception module on the network [21], [22]. It combines convolutional layers in the same module [23]. The inception module aims to get the representation with ideal sparse structure [24]. InceptionV3 has 48 layers deep [23].
\[ f(x) = g(x) - x \]

Skip connection

\[ g(x) \]

Figure 1. Skip connection.

Figure 2. Illustration of TWSVM.

Figure 3. EuroSAT sample dataset: (a) highway, (b) pasture, (c) residential, (d) annual crop, (e) sea lake, (f) forest, (g) industrial, (h) river, (i) permanent crop, and (j) herbaceous vegetation.

Softmax is the most used activation function on CNN for the classification task [15]. Softmax function with vector \( x \) as the input can be defined with (1) [25]. Here, we also replaced softmax and used Support Vector Machine (SVM) and Twin SVM (TWSVM) for the classifier. We get the features from the last dense layer of trained CNN. Then, we classified the features using SVM and TWSVM.

\[ f_i(x) = \frac{e^{x_i}}{\sum_{j} e^{x_j}} \]  \hspace{1cm} (1)

SVM uses one plane to classify the data. It is based on the margin maximization principle that can make it less sensitive to the overfitting problem [12]. SVM was designed to solve the binary classification task [26].

TWSVM is one of the developments of SVM. Same as with SVM, TWSVM was designed to solve the binary classification task. Instead of using one plane, TWSVM uses two nonparallel planes for
classification. Each plane is near one of the two labels and far from the other label. TWSVM is good at generalization [10]. The illustration of TWSVM with rectangle and circle as the labels can be seen in Figure 2.

We also combined CNN with sSE block to recalibrated the feature map on CNN [8]. sSE block contains convolutional layer with one filter, $1 \times 1$ kernel, and followed by sigmoid function [27]. Then, the result will be multiplied with previous inputs.

The dataset we used is EuroSAT RGB remote sensing images taken from Sentinel-2 satellite. The RGB channels is getting from three bands of satellite image: red (B04), green (B03), and blue (B02). EuroSAT contains 27,000 images with ten labels: highway, pasture, residential, annual crop, sea lake, forest, industrial, river, permanent crop, and herbaceous vegetation. It can be downloaded at https://github.com/phelber/EuroSAT [2]. Sample images of the EuroSAT dataset presented in Figure 3.

EuroSAT image dataset contains more than two labels, so we can’t use binary classification method to classify the images. To deal with this problem, we used one vs. one approach for SVM and TWSVM. If we have $n$ labels of the dataset, binary classifiers with two labels are created and the total number of binary classifiers can be find with (2) [28]. For example, if we have A, B, C, and D as the labels, we can divide them into six binary classifiers: A vs. B, A vs. C, A vs. D, B vs. C, B vs. D, and C vs. D. Then, the majority vote is used to vote the predicted labels of binary classifiers [28].

$$\text{number of classifiers} = \frac{n(n-1)}{2}$$

Table 1. Hyperparameters

| Method | Hyperparameter | Value                |
|--------|----------------|----------------------|
| CNN    | Optimizer      | Stochastic Gradient Descent with Nesterov |
|        | Momentum       | $9 \times 10^{-1}$    |
|        | Decay          | $1 \times 10^{-6}$    |
|        | Epoch          | 50                   |
|        | Batch size     | 32                   |
|        | Loss           | Categorical crossentropy |
| SVM    | Kernel         | Radial Basis Function |
|        | C              | 1                    |
|        | Gamma          | $1 \times 10^{-3}$    |
| TWSVM  | Kernel         | Radial Basis Function |
|        | C1             | 1                    |
|        | C2             | 2                    |
|        | Gamma          | $1 \times 10^{-3}$    |
To match with the output of the dataset, we replaced the last fully connected layer of CNN with global average pooling followed by dense layer with ten output and classifier. Then, we used sSE block on CNN before the last fully connected layer. Implementation of sSE block on CNN can be seen in Figure 4.

The steps of the research method can be seen in Figure 5. This research was implemented in Google Colab with several libraries. We resized the image of EuroSAT dataset to 75 × 75. Dataset is also divided by 70% as train and the rest as test. The hyperparameters used on CNN, SVM, and TWSVM can be seen in Table 1. Then, the learning rates used for ResNet50, InceptionV3, and VGG19 are $4 \times 10^{-6}$, $1 \times 10^{-3}$, and $1 \times 10^{-5}$, respectively.

### Table 2. Classification Results (%) of Various Methods

| Base         | Method          | Accuracy | Precision | Recall | F1-score |
|--------------|-----------------|----------|-----------|--------|----------|
| ResNet50     | ResNet50        | 60.23    | 58.31     | 59.34  | 58.28    |
|              | ResNet50 + SVM  | 62.09    | 60.11     | 61.16  | 60.34    |
|              | ResNet50 + TWSVM| 62.58    | 60.39     | 61.56  | 59.85    |
|              | ResNet50 + sSE  | 60.21    | 58.11     | 59.11  | 58.14    |
| Model                          | Train Loss | Validation Loss |
|-------------------------------|------------|-----------------|
| **ResNet50 + sSE** + SVM      | 64.32      | 62.36           |
|                                | 63.85      | 62.84           |
| **ResNet50 + sSE** + TWSVM    | 62.66      | 61.24           |
| **InceptionV3**               | 79.46      | 78.32           |
| **InceptionV3 + SVM**         | 80.36      | 79.47           |
| **InceptionV3 + TWSVM**       | 79.73      | 79.65           |
| **InceptionV3 + sSE**         | 79.31      | 78.97           |
| **InceptionV3 + sSE + SVM**   | 80.35      | 79.36           |
| **InceptionV3 + sSE + TWSVM** | 79.37      | 78.27           |
| **VGG19**                     | 92.94      | 92.82           |
| **VGG19 + SVM**               | 94.25      | 94.07           |
| **VGG19 + TWSVM**             | 94.26      | 94.08           |
| **VGG19 + sSE**               | 93.91      | 93.70           |
| **VGG19 + sSE + SVM**         | 94.51      | 94.33           |
| **VGG19 + sSE + TWSVM**       | 94.57      | 94.39           |

**Figure 6.** Train loss result.

**Figure 7.** Validation loss result.

**Figure 8.** Train accuracy result.

**Figure 9.** Validation accuracy result.
4. Results
Loss and accuracy during the training of the CNN and CNN with sSE are visualized through the graphs. Figure 6 and Figure 7 show the graph of train and validation loss, respectively. Then, Figure 8 and Figure 9 show the graph of train and validation accuracy, respectively.

Classification results of various methods are presented in Table 2. We provided the results with accuracy, macro-averaged precision, recall, and F1-score as evaluation metrics. The results show that on the ResNet50-based model, ResNet50 with sSE and SVM gives the highest results with 64.32% accuracy, 62.36% precision, 63.37% recall, and 62.66% F1-score. On the InceptionV3-based model, InceptionV3 with SVM gives the highest results with 80.36% accuracy, 79.35% precision, 79.65% recall, and 79.47% F1-score. Then, on the VGG19-based model, VGG19 with sSE and TWSVM gives the highest results with 94.57% accuracy, 94.40% precision, 94.40% recall, and 94.39% F1-score.

5. Conclusion and Future Works
This research combined CNN with sSE block for Land Use and Land Cover (LULC) classification. CNN we used are ResNet50, InceptionV3, and VGG19. Then, sSE block is used to recalibrate the feature of CNN. We used EuroSAT RGB remote sensing images for the data. We also replaced the softmax classifier on CNN with SVM and TWSVM.

From the experimental results, ResNet50 achieved the best results when combined with sSE and SVM. Then, InceptionV3 achieved the best results when combined with SVM. Finally, VGG19 with sSE block and TWSVM achieved the best results among all classification methods.

For future development, the researchers may try to perform classification with another type of remote sensing image dataset, such as remote sensing images with more than three channels instead of just three channels RGB images. The researchers may also try to perform hyperparameter optimization to find optimal hyperparameters. Another development can be combining CNN and sSE with other classifiers like K-Nearest Neighbor (KNN), Decision Tree, Naïve Bayes, or ensemble classifiers.

6. References
[1] N. A. Mahmon, N. Ya’Acob, and A. L. Yusof, “Differences of image classification techniques for land use and land cover classification,” in Proceedings - 2015 IEEE 11th International Colloquium on Signal Processing and Its Applications, CSPA 2015, Aug. 2015, pp. 90–94.
[2] P. Helber, B. Bischke, A. Dengel, and D. Borth, “Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 12, no. 7, pp. 2217–2226, Jul. 2019.
[3] M. Ahishali, S. Kiranyaz, T. Ince, and M. Gabbouj, “Dual and single polarized sar image classification using compact convolutional neural networks,” Remote Sens., 2019.
[4] Y. Huang, Z. Wu, L. Wang, and T. Tan, “Feature coding in image classification: A comprehensive study,” IEEE Trans. Pattern Anal. Mach. Intell., 2014.
[5] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, “Deep learning in remote sensing applications: A meta-analysis and review,” ISPRS Journal of Photogrammetry and Remote Sensing, 2019.
[6] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, “HybridSN: Exploring 3-D-2-D CNN Feature Hierarchy for Hyperspectral Image Classification,” IEEE Geosci. Remote Sens. Lett., 2020.
[7] Y. Wang, M. Liu, P. Zheng, H. Yang, and J. Zou, “A smart surface inspection system using faster R-CNN in cloud-edge computing environment,” Adv. Eng. Informatics, 2020.
[8] A. G. Roy, N. Navab, and C. Wachinger, “Concurrent spatial and channel ’squeeze & excitation’ in fully convolutional networks,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018.
[9] C. Cortes and V. Vapnik, “Support-Vector Networks,” Mach. Learn., 1995.
[10] Jayadeva, R. Khemchandani, and S. Chandra, “Twin support vector machines for pattern classification,” IEEE Trans. Pattern Anal. Mach. Intell., 2007.
[11] Y. Yang and S. Newsam, “Bag-of-visual-words and spatial extensions for land-use classification,” in GIS: Proceedings of the ACM International Symposium on Advances in
Geographic Information Systems, 2010.

[12] Y. Bazi and F. Melgani, “Convolutional SVM Networks for Object Detection in UAV Imagery,” IEEE Trans. Geosci. Remote Sens., 2018.

[13] H. Qassim, A. Verma, and D. Feinzimer, “Compressed residual-VGG16 CNN model for big data places image recognition,” in 2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018, Feb. 2018, vol. 2018-January, pp. 169–175.

[14] W. Li et al., “Classification of High-Spatial-Resolution Remote Sensing Scenes Method Using Transfer Learning and Deep Convolutional Neural Network,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 13, pp. 1986–1995, 2020.

[15] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, “Diabetic retinopathy detection through deep learning techniques: A review,” Informatics in Medicine Unlocked. 2020.

[16] G. Cheng, C. Ma, P. Zhou, X. Yao, and J. Han, “Scene classification of high resolution remote sensing images using convolutional neural networks,” in International Geoscience and Remote Sensing Symposium (IGARSS), Nov. 2016, vol. 2016- November, pp. 767–770.

[17] X. Sun, L. Liu, C. Li, J. Yin, J. Zhao, and W. Si, “Classification for Remote Sensing Data with Improved CNN-SVM Method,” IEEE Access, vol. 7, pp. 164507–164516, 2019.

[18] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 2015.

[19] S. Roy and S. Paul, “Land-Use Detection Using Residual Convolutional Neural Network,” in 1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019, May 2019.

[20] I. Konovalenko, P. Maruschak, J. Brezinová, J. Viňáš, and J. Brezina, “Steel Surface Defect Classification Using Deep Residual Neural Network,” Metals (Basel), vol. 10, no. 6, p. 846, Jun. 2020.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016.

[22] D. Vasan, M. Alazab, S. Wassan, H. Naeem, B. Safaei, and Q. Zheng, “IMCFN: Image-based malware classification using fine-tuned convolutional neural network architecture,” Comput. Networks, 2020.

[23] N. Papandrianos, E. I. Papageorgiou, and A. Anagnostis, “Development of Convolutional Neural Networks to identify bone metastasis for prostate cancer patients in bone scintigraphy,” Ann. Nucl. Med., 2020.

[24] Z. Xiong, Y. Yuan, and Q. Wang, “AI-Net: Attention inception neural networks for hyperspectral image classification,” in International Geoscience and Remote Sensing Symposium (IGARSS), 2018.

[25] X. Qi, T. Wang, and J. Liu, “Comparison of Support Vector Machine and Softmax Classifiers in Computer Vision,” in Proceedings - 2017 2nd International Conference on Mechanical, Control and Computer Engineering, ICMCCE 2017, 2018.

[26] H. Huang, X. Wei, and Y. Zhou, “Twin support vector machines: A survey,” Neurocomputing, 2018.

[27] J. Naranjo-Alcazar, S. Perez-Castanos, P. Zuccarello, and M. Cobos, “Acoustic Scene Classification with Squeeze-Excitation Residual Networks,” IEEE Access, 2020.

[28] A. R. Abdul Raziff, M. N. Sulaiman, N. Mustapha, and T. Perumal, “Gait identification using One-vs-one classifier model,” in ICOS 2016 - 2016 IEEE Conference on Open Systems, 2017.