Doctoral Thesis
박사 학위논문

MACHINE LEARNING-BASED SYSTEM MODELING FOR SPECTRUM SENSING & SEMANTIC SEGMENTATION

Jun Hee Kim (김 준 희 金 俊 熙)

Department of Information and Communication Engineering
정보통신융합전공

DGIST

2020

Doctoral Thesis
박사 학위논문

MACHINE LEARNING-BASED SYSTEM MODELING FOR SPECTRUM SENSING & SEMANTIC SEGMENTATION

Jun Hee Kim (김 준 희 金 俊 熙)

Department of Information and Communication Engineering
정보통신융합전공

DGIST

2020

MACHINE LEARNING-BASED SYSTEM MODELING FOR SPECTRUM SENSING & SEMANTIC SEGMENTATION

Advisor: Professor Jihwan P. Choi
Co-advisor: Professor Dooseok Lee

by

Jun Hee Kim
Department of Information and Communication Engineering
DGIST

A thesis submitted to the faculty of DGIST in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Information & Communication Engineering. The study was conducted in accordance with Code of Research Ethics[1]

12. 31. 2019

Approved by

Professor Jihwan P. Choi (signature)
(Advisor)

Professor Dooseok Lee (signature)
(Co-Advisor)

[1] Declaration of Ethical Conduct in Research**:** I, as a graduate student of DGIST, hereby declare that I have not committed any acts that may damage the credibility of my research. These include, but are not limited to: falsification, thesis written by someone else, distortion of research findings or plagiarism. I affirm that my thesis contains honest conclusions based on my own careful research under the guidance of my thesis advisor.

MACHINE LEARNING-BASED SYSTEM MODELING FOR SPECTRUM SENSING & SEMANTIC SEGMENTATION

Jun Hee Kim

Accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

12. 31. 2019

Head of Committee	Prof. Jae Youn Hwang (signature)
Committee Member	Prof. Dooseok Lee (signature)
Committee Member	Prof. Kyung-Joon Park (signature)
Committee Member	Prof. Sunghoon Im (signature)
Committee Member	Prof. Jihwan Choi (signature)

Ph.D/IC
201632002

김 준 희. Jun Hee Kim. MACHINE LEARNING-BASED SYSTEM MODELING FOR SPECTRUM SENSING & SEMANTIC SEGMENTATION. Department of Information & Communication Engineering. 2020. 79p. Advisor: Prof. Jihwan Choi, Co-Advisor: Prof. Dooseok Lee.

ABSTRACT

Machine-learning algorithms have attracted much attention in a wide range of areas. Because machine-learning algorithms can extract patterns from data automatically, it is possible to model the input-output function of a system using a machine-learning algorithm. In cognitive radio networks, machine-learning-based spectrum sensing schemes depend on the positioning of the nodes. During building extraction tasks based on high-resolution aerial images, large-scale datasets are required, with customized architecture necessary to process the datasets. Moreover, when there is a domain gap between the training and test data, trained models fail to segment objects for unseen images. In this thesis, the challenges facing machine-learning-based systems and solutions are discussed.

The first application is for spectrum sensing in cognitive radio networks. The hidden primary user (PU) problem, however, is a critical issue in cognitive radio networks because spectrum sensing nodes (SNs) can misclassify the degree of spectrum occupancy. To cope with this, machine-learning-based cooperative spectrum sensing schemes (CSSs) have been proposed. CSSs that do not consider node placement, however, continue to be affected by the hidden PU problem. In this paper, we present a method by which to place SNs to guarantee the performance of machine-learning-based CSSs. We verify that the hidden PU problem causes some overlap in the data distribution, which deteriorates of the spectrum sensing ability. Based on Kullback-Leibler divergence, analytical expressions for the

spectrum-sensing coverage of a single SN are derived. We then propose a strategy for placing a few SNs to cover the entire area of the PU and prove the feasibility of the proposed method experimentally.

The second application is related to deep-learning architecture for semantic segmentation from high-resolution aerial images. Extracting manufactured features such as buildings, roads, and water from aerial images is critical for urban planning, traffic management, and industrial development. Recently, convolutional neural networks (CNNs) have become a popular strategy to use to capture contextual features automatically.

We design a multi-object segmentation system and propose an algorithm that utilizes pyramid pooling layers to improve U-Net. Test results indicate that U-Net with pyramid pooling layers, referred to as UNetPPL, learns fine-grained classification maps and outperforms other algorithms, specifically FCN and U-Net, achieving a mean intersection of union (mIOU) value of 79.52 and pixel accuracy of 87.61% for four types of objects (buildings, roads, water, and background).

The final application is domain adaptation for building extraction. CNN-based semantic segmentation models garnered much attention in relation to remote sensing and achieved remarkable performance during the extraction of buildings from high -resolution aerial images. However, limited generalization for unseen images remains. When there is a domain gap between the training and test datasets, CNN-based segmentation models that are trained using a training dataset fail to segment buildings in the test dataset. In this paper, we propose domain-adaptive transfer attack (DATA)-based segmentation networks for building extraction from aerial images. The proposed system utilizes jointly both domain transfers and adversarial attacks. Based on the DATA scheme, the distribution of

input images can be shifted to that of target images while maintaining the semantic spaces. This reduces the domain gap and expands the generalization of the segmentation model. From test results with two different datasets (i.e., the Inria aerial image labeling dataset and the Massachusetts building dataset), it is verified that the performance when extracting buildings is improved to 0.16% and 7.12%, respectively.

Keywords: Deep learning, wireless cognitive radio networks, spectrum detection, semantic segmentation, domain adaptation.

CONTENTS

List of Figures

List of Tables

1. INTRODUCTION

1.1 MACHINE LEARNING ALGORITHM

Machine learning is defined as an automated process that extracts patterns from data [1]. When the system characteristics are unknown, machine-learning techniques are effective for estimating the input-output function of the system without complete knowledge of the system parameters [2]. Therefore, machine-learning algorithms are widely used in various areas, such as medical [3] and agricultural areas [4] as well as wireless networks [5]-[6] and aerial image processing [7]. Although machine-learning algorithms can effectively model the functions of the systems, they cannot solve all problems. In cognitive radio networks, spectrum sensing schemes depend on the positioning of nodes. They require large-scale datasets and customized architecture to process the datasets. To address these issues, node placement for machine-learning-based spectrum sensing schemes and deep-learning architectures for aerial image processing are discussed in this thesis.

1.2 BACKGROUND AND CHALLENGING ISSUES

1.2.1 COGNITIVE RADIO NETWORKS

In policies pertaining to spectrum allocation, while spectrum bands such as unlicensed industrial, scientific, and medical (ISM) bands and mobile cellular channels have high occupancy rates, whereas there exist spectrum bands that are underutilized. This imbalance reduces the overall spectrum utilization rate. To solve both the spectrum scarcity issue at specific bands and the deterioration of the overall spectrum utilization rate, cognitive radio (CR) technology has been proposed as a promising solution to the IoT demand issue of sensing and accessing radio spectrum [8], improving spectrum band usage by using unused portions of the spectrum. In previous research, an energy-detection method was widely used for better spectrum occupancy [9]. Energy detection does not require much a priori

knowledge of PU signals and can be implemented more simply than other methods (i.e., the channel state is determined by whether or not the received power exceeds a certain threshold) while the performance is affected by uncertain noise power. However, it is highly challenging to find a proper threshold, and the spectrum sensing performance depends on doing so. To find proper decision boundaries for spectrum sensing, machine-learning algorithms have been applied [10]. Because it is possible to find flexible decision boundaries, it is recommended to use machine-learning-based energy-detection methods for spectrum sensing.

Although machine-learning-based cooperative spectrum sensing schemes (CSSs) are applied to avoid the hidden PU problem, there is still an unsolved technical challenge. The spectrum sensing coverage of each spectrum sensing node (SN) depends on several parameters. CSSs without considering these parameters, however, are affected by the hidden PU problem. In this thesis, we verify that the hidden PU problem causes some overlap in the data distribution, which deteriorates of the spectrum sensing ability. Based on the received power distributions, analytical expressions for the spectrum sensing coverage of a single SN are derived, after which we propose a strategy that involves placing a few SNs to cover the entire area of the PU. We prove the feasibility of our proposal experimentally.

1.2.2 SEMANTIC SEGMENTATION FROM HIGH-RESOLUTION AERIAL IMAGES

Aerial images can provide valuable information about areas that are difficult for people to access or access non-intrusively. The information obtained using aerial images can be utilized in a variety of areas, including land inventory, vegetation monitoring, and environmental assessments. To utilize aerial images, object detection tasks related to buildings, roads and bodies of water are required. These tasks mainly rely on manual digitization [11], SVM [12], and artificial neural networks (ANNs) [13]. Although these methods require minimal human intervention, their results have generally been discouraging due to various adverse factors (e.g., image noise, shadows)

Recently, as deep-learning technology has evolved, it has been applied to many problems related to computer vision, such as image classification, object segmentation, and semantic segmentation.

Additionally, deep-learning techniques such as convolutional neural networks (CNNs) have attracted much attention in the segmentation of objects in satellite images. Although it is possible for deep-learning algorithms to learn contextual features automatically, there are several issues that must be addressed. First, a considerable amount of data is required to train them. Existing datasets cannot sufficiently train deep-learning algorithms. The performance of deep-learning algorithms depends on their architecture. In this thesis, geo data generation systems based on deep learning are discussed.

1.2.3 Domain Gap between Training and Test Dataset

CNN-based semantic segmentation models have gained much attention in relation to remote sensing and have achieved remarkable performance outcomes when used to extract buildings from high-resolution aerial images. However, limited generalization with regard to unseen images remains to be an issue. When there is a domain gap between the training and test datasets, CNN-based segmentation models trained with a training dataset fail to segment buildings in the test dataset. To solve this problem, a domain adaptation method has been proposed [14]-[15]. However, these approaches are not customized for aerial images. While driving datasets such as GTA5 [16] and Cityscapes [17] as discussed in the literature [15] have static class distributions for objects (e.g., the road is at the bottom and the sky is at the top in the input images), the buildings in aerial images are located randomly. In this thesis, a domain adaptation method customized for aerial images is proposed.

1.3 Overview of Thesis

The objective of this thesis is to make a contribution that solves the challenges of machine-learning-based systems. In Section 2, sensing coverage-based cooperative spectrum detection in cognitive radio networks is discussed. The deep-learning architecture for building and road extraction from high-resolution aerial images is proposed in Section 3. The domain gap between the training and test datasets is presented in Section 4. Finally, the proposed scheme is summarized in Section 5.

2. SENSING COVERAGE-BASED SPECTRUM DETECTION

2.1 INTRODUCTION

The Internet of Things (IoT) is expected as a promising solution to improve many existing industrial systems such as agriculture, food processing, environmental monitoring, etc. [18] By incorporate wireless sensor networks (WSNs) with the internet, the IoT enables the connected objects to communicate with others and users, mainly over the unlicensed industrial, scientific, and medical (ISM) band. While spectrum bands such as ISM and mobile cellular channels are highly occupied [19]-[20], there exist spectrum bands that are underutilized [21]-[23] and this imbalance reduces overall spectrum utilization. To solve both the spectrum scarcity in specific bands and the deterioration of overall spectrum utilization, the cognitive radio (CR) technology has been proposed as a promising solution of the IoT demand to sense and access radio spectrum [24], improving the spectrum band usage by using the unused portions of the spectrum.

Opportunistic spectrum access enables unlicensed users called secondary users (SUs) to utilize spectrum bands adaptively when primary users (PUs) do not access the spectrum. In order to access spectrum opportunistically, the spectrum sensing node (SN) must sense the activity of the PU. To utilize the spectrum opportunistically when the PU is inactive, the network requires a spectrum sensing scheme. In [9], three methods (matched filter detection, energy detection, and feature detection) are reviewed to detect the activity of PUs. Energy detection does not need a lot of priori knowledge of PU signals and can be implemented easier than the other methods (i.e., the channel state is determined by whether the received power is more than a certain threshold or not) while performance is affected by uncertain noise power. Therefore, we use the energy detection method for the opportunistic spectrum access, as commonly used to sense the spectrum in cognitive radio [25], [26].

When a single SN attempts to sense the spectrum bands, however, it cannot solve the hidden

PU problem in which the node can misclassify the PU's state mainly because of large and small scale fading. To handle this, cooperative spectrum sensing schemes (CSSs) that share the received power among nodes are proposed [27]-[29]. The distributed SNs receive signal power from PU and the results are reported to the fusion center (FC). By consolidating the reported results, the channel state is determined at the fusion center. In previous research, *k*-out-of-*n* rules have been used in CSS. That is, the fusion center declares that the spectrum bands are occupied if at least *k* out of *n* nodes sense the PU's activities [30]. There are several types of *k*-out-of-*n* rules. If $k = 1$ and $k = n$, it is named the OR and the AND rule, respectively [31]. However, it is very challenging to find a proper threshold that the spectrum sensing performance depends on. If the threshold is set too high, the FC tends to classify the channel state as vacant. On the contrary, with the low threshold, even if the spectrum is not occupied by PUs, the SUs will not access the spectrum. These problems get worse as the number of SNs is increased since the manual setting of large-scale threshold values is a non-trivial task.

When the system characteristics are unknown, learning techniques are effective for estimating the input-output function of the system without a complete knowledge of system parameters [2]. Thus, machine learning algorithms are widely used in wireless networks, such as modulation classification [32], channel estimation [5], anti-jamming [33], and spectrum sensing [34], [10]. In [34], a threshold or a decision boundary to determine channel states is changed depending on the transmit power levels of the PU. Under this environment, proper machine learning algorithms were analyzed and it was proved to outperform the conventional ones. Furthermore, the machine learning algorithm can add flexibility to finding decision boundaries. Therefore, it is recommendable to use the machine learning based energy detection method for spectrum sensing.

2.1.1 RELATED WORK

Recently, machine learning algorithm-based spectrum sensing schemes have been proposed to find a particular threshold [34], [10]. In [34], a threshold or a decision boundary to determine channel

states is changed depending on the transmit power level of PU. Under this environment, machine learning algorithms are analyzed in terms of the average training time and the sample classification delay and it was proved that the performance of the machine learning based method is better than the conventional ones. A new framework for spectrum sensing schemes is proposed in [10], consisting of the two phases of recognizing the transmit power level of PU and choosing the decision boundary for channel states. The first phase is conducted by K-means clustering and the second phase tasks to select the decision boundary that is trained by the support vector machine (SVM) with a Gaussian kernel function.

Although machine learning-based CSSs are applied to avoid the hidden PU problem, there is still an unsolved technical challenge. Because the spectrum sensing coverage of each SN depends on several parameters (which we discuss in Section II and III), PUs outside of the coverage of SNs cause SNs to misclassify spectrum occupancy. While the spectrum sensing coverage is critically related to the performance of CSS, it has not been considered in previous research.

In [35], methods of node placement for wireless sensor networks are presented. Each sensor has sensing distance r_s and communication distance r_c. Based on these distances, two possible methods for placing sensors are proposed. With the coverage-first placement method, neighboring sensors are separated by a distance of $\sqrt{3}r_s$. This is effective when $r_c \geq \sqrt{3}r_s$ by reducing the number of sensors and minimizing the overlapping coverage. The connectivity-first placement scheme separates neighboring sensors by a distance of r_c, which can meet the communication requirement efficiently when $r_c \leq \sqrt{3}r_s$. In the energy detection method, because SNs use the received power to sense the spectrum of the PU, the sensing distance is directly related to the received power. On the other hand, the communication distance depends on the signal-noise ratio (SNR). Even with low power received, communication signals can be decoded if the noise power is sufficiently low [36]. Therefore, the communication distance is longer than the sensing distance in general and the coverage-first placement method is proper in cooperative spectrum sensing. In order to apply the coverage-first placement to cognitive radio networks, however, the spectrum sensing coverage of SNs should be derived, which is

the main contribution of our research. In this paper, we derive the maximum coverage where a single SN can detect the activity of the PU and quantify the hidden PU problem. Then, experimental results are presented to show feasibility of the proposed scheme.

2.1.2 SUMMARY AND ORGANIZATION

The main contributions of this paper are summarized as follows:

- We derive the spectrum sensing coverage of the SN that guarantees the performance of spectrum sensing schemes.
- We demonstrate feasibility of the proposed schemes by conducting real-environment experiments.

The rest of this paper is organized as follows: In Section 2.2, system model and problem statements are presented. In Section 2.3, the spectrum sensing coverage of SN and the performance are analyzed. In Section 2.4, experiment results and analysis are discussed. Finally, our proposed schemes are concluded in Section 2.5. The notations used in this paper are summarized in TABLE I.

2.2 SYSTEM MODEL AND PROBLEM FORMULATION

Several parameters, such as the distance between the PU and SNs and channel states, affect the performance of machine learning algorithms in the hidden PU problem. In this section, we discuss how we can improve the performance of the machine learning based spectrum sensing schemes.

2.2.1 SYSTEM MODEL OF COGNITIVE RADIO NETWORKS

We consider a cognitive radio network that has more SUs than PUs. The transmitted signals from the PU are through multipath fading channels with path loss attenuation. The channel states can be categorized into busy or idle. The PU has multiple power levels selected from a transmit power set, $P_{PU} \in \{P_0, P_1, P_2, \dots, P_N\}$, in which P_0 is the power level of the idle state with zero power and P_i with $i = 1, \dots, N$ is that of the busy state (i.e., $P_0 < P_1 < P_2 < \cdots < P_N$). The coordinates of the PU and SNs are denoted as $S_{PU}(X, Y)$ and $S_{SN}(X, Y)$, respectively, where $(X, Y) \in \mathbb{R}^2$. The transmitted power from PU is attenuated by path

TABLE I

THE LIST OF PARAMETER NOTATIONS

Notation	Definition
λ	Wavelength
d_0	Reference distance
γ	Path loss exponent
ζ_{SN}	Multipath fading component of SN
ρ	Mean of noise power
${\sigma_n}^2$	Variance of noise power
N_{input}	Number of input layer nodes
N_{hidden}	Number of hidden layer nodes
N_{output}	Number of output layer nodes
λ_{ANN}	Regularization constant of ANN
f_{act}	Activation function of ANN
$1/\lambda_{SVM}$	Reciprocal number of regularization constant
σ_{SVM}	Bandwidth parameter
P_{PU}	Transmit power levels of PU
$d_{PU,SN}$	Distance between PU and SN
k	Distance between data distributions
${\sigma_{v,i}}^2$	Variance of received power
ΔP	Differential transmit power levels of PU
d_{real}	Real sensing coverage of SN
d_{init}	Computed sensing coverage of SN
σ^2	Variance of the Rayleigh distribution

loss, $P_L(d_{PU,SN})$ [dBm], and can be modeled as follows [37]:

$$P_L\left(d_{PU,SN}\right) = -20\log\left(\frac{\lambda}{4\pi d_0}\right) + 10\gamma\log\left(\frac{d_{PU,SN}}{d_0}\right), \quad (1)$$

where $d_{PU,SN}$ denotes the distance between PU and SN, λ is the wavelength, d_0 is the reference distance, and γ is the path loss exponent. SN receives the power of P_{SN} [W] which can be found as follows [34].

$$P_{SN} = P_{PU} \cdot 10^{-\left[\frac{P_L(d_{PU,SN})}{10}\right]} \cdot \zeta_h + P_n \ , \quad (2)$$

where ζ_h is the multipath fading component in channel h, and P_n denotes the noise power according to a normal distribution with mean ρ and variance ${\sigma_n}^2$, i.e., $P_n \sim \varkappa(\rho, {\sigma_n}^2)$.

2.2.2 Machine Learning-based Cooperative Spectrum Sensing

The distributed SNs report the received power to the FC. Based on the trained decision boundaries, the spectrum occupancy is determined at the FC. The spectrum sensing scheme in [10] includes a phase of *K*-means clustering, an unsupervised learning scheme. As supervised learning tends to outperform unsupervised [34], we use artificial neural networks (ANNs), illustrated in Fig. 1, to recognize the transmit power levels of the PU. Afterwards, the channel state is classified by the SVM with a Gaussian kernel function in the second phase (i.e., if the transmit power level is estimated as none, the

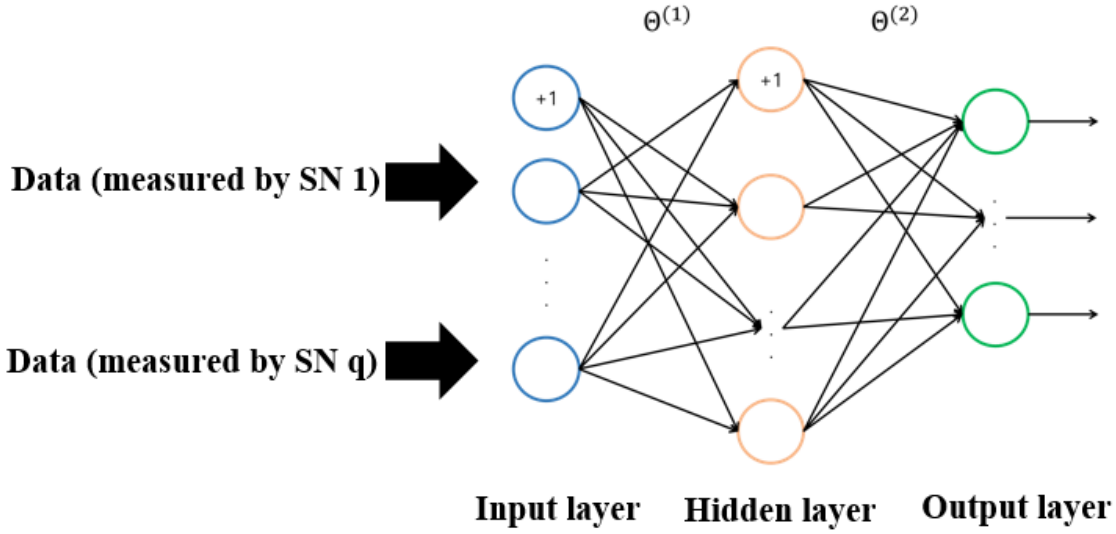


Fig. 1. An artificial neural network with three layers. In the input layer, the data measured by all SNs (from first to q-th SN) are inputted, and the results, which are multiplied with weight, $\theta^{(1)}$, are used as inputs in the hidden layer. The output of the hidden layer is also multiplied with weight, $\theta^{(2)}$, and the probability for each class is calculated in the output layer. Finally, the class that corresponds to the maximum value of the probability is selected.

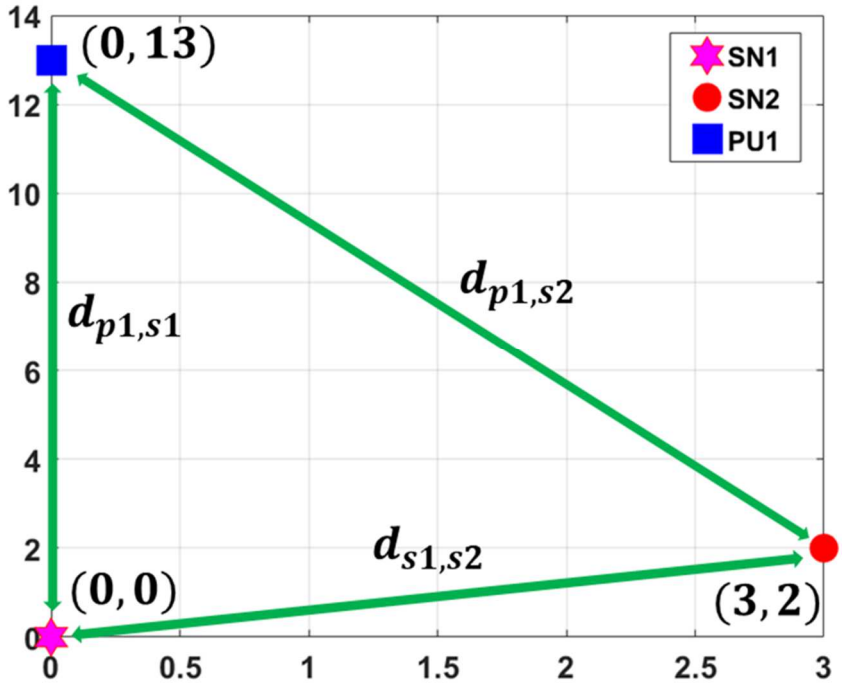


Fig. 2. Distribution of PU and SNs, with $S_{PU}(0, 13)$, $S_{SN1}(0, 0)$, $S_{SN2}(3, 2)$, $d_{p1s1} = 13\ m$, $d_{p1s2} = 11.40\ m$, $and\ d_{s1s2} = 5.66\ m$.

channel state is classified as idle directly). The SVM, a binary classifier in this case, can find the decision boundary that has the maximum margin. Note that the training data in this case can be labeled for supervised learning, by exploiting transmit power levels known from the protocol standard and specification of devices. In addition, except for zero power, the spectrum state can be regarded as occupied.

As a simple machine learning based CSS, we consider a scenario where two SNs attempt to sense the PU spectrum. The distributed SNs, as shown in Fig. 2, measure signal power from PU and the results are reported to the FC. Finally, the channel states are classified. Fig. 3 shows both the distributions of training data and the decision boundary for each power level. The PU is assumed to have multiple transmit power levels, given as $P_0 = 0$, $P_1 = 100$, $P_2 = 200$, $P_3 = 300$, and $P_4 = 400\ [mW]$. By assuming quasi-static multipath fading during the time of interest, the average value for the multipath fading components is used. Relationship between multipath fading components and

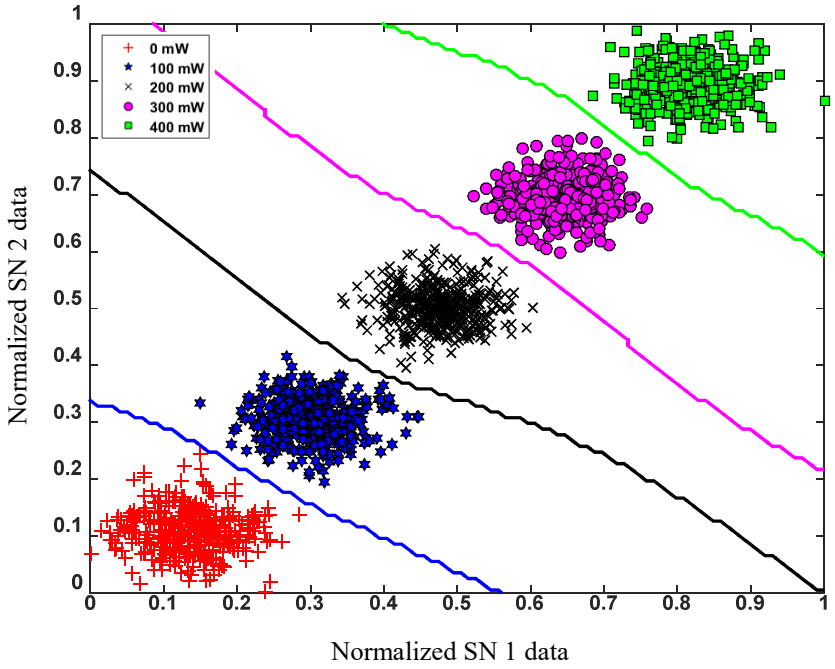


Fig. 3. Decision boundaries depending on transmit power levels, with λ = 0.375 m, d_0 = 1 m, γ = 2, ζ_{SN1} = 0.7, ζ_{SN2} = 0.8, ρ = 0.1 μW, $\sigma_n{}^2$ = -140 dBm, and a power gap = 100 mW.

the spectrum sensing scheme is discussed in Section III. In simulations, it is assumed that SUs can switch their modes with negligible switching delay. We measure 2,500 data, of which 80% are used for training the proposed sensing scheme and the remaining, 500 samples for each power level, are for tests. The decision boundary that is trained by the SVM for channel states depends on the PU transmit power levels, as shown in Fig. 3. The ANN estimates the PU power level in the first stage, from which a classifier is selected to determine the spectrum occupancy.

2.2.3 Hidden PU Problem and Kullback-Leibler Divergence

In the energy detection method, the performance of spectrum occupancy sensing is susceptible to the received power. For weak signals, a single SN can misclassify the spectrum occupancy, which is called the hidden PU problem. To mitigate this, CSSs which utilize several SNs have been used in previous research. However, the SN has a limited coverage for spectrum sensing, depending on PU transmit power levels, channel states, and the distance from the PU. For example, in a real environment,

due to limitation of battery capacity or regulation for electromagnetic compatibility, the PU does not have a large amount of transmit power. Fig. 4 illustrates the training results with these factors into account.

The overlap of data distributions leads to deteriorating the class separation capability. In the first phase of recognizing the PU transmit power levels, the accuracy of the ANN is reduced from 99.6% to 75.8%, while that of *K*-means clustering is from 99.6% to 70.6%. The performance of the SVM in the second phase for channel state classification also decreases from 100.0% to 84.2%. These results verify that the hidden PU problem is related to the interval of data distributions. Therefore, we can quantify performance in the hidden PU problem by using the Kullback–Leibler Divergence (KLD).

When the two probability density functions of a continuous random variable x are given as $q(x)$ and $r(x)$, respectively, the KLD is defined as follows [38]:

$$D_{KL}(q(x)||r(x)) = \int_{-\infty}^{+\infty} q(x) \log \frac{r(x)}{q(x)} dx. \quad (3)$$

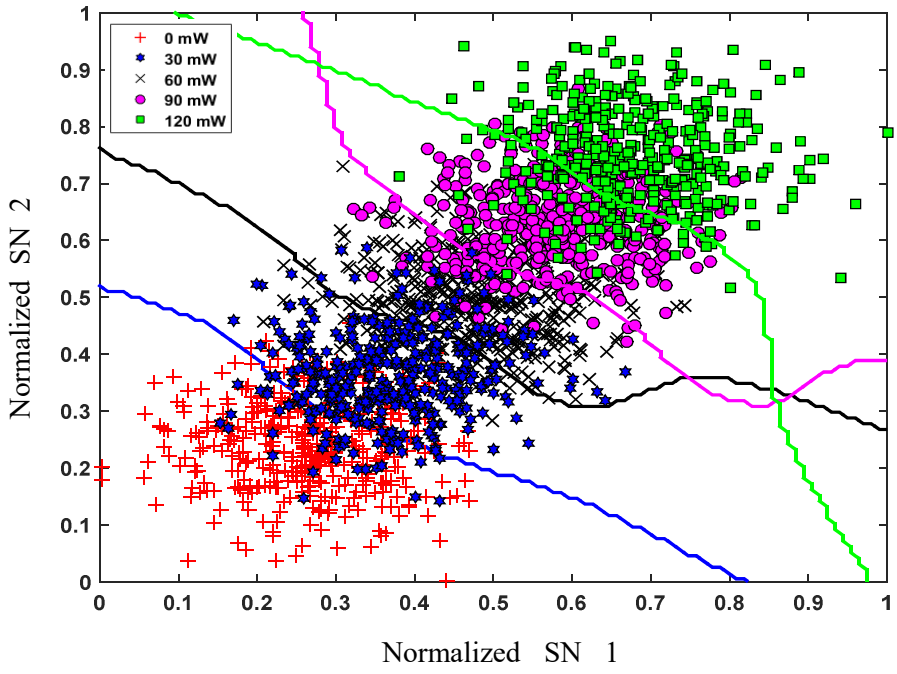


Fig. 4. Decision boundaries depending on transmit power levels, with λ = 0.375 m, d_0 = 1 m, γ = 2, ζ_{SN1} = 0.7, ζ_{SN2} = 0.8, ρ = 0.1 μW, $\sigma_n{}^2$ =-140 dBm, and a power gap = 30 mW.

If $q(x)$ and $r(x)$ are the Gaussian probability density functions with the variance of σ^2, the KLD in Eq. (3) can be rewritten by

$$D_{KL}(q(x)||r(x)) = \frac{(\mu_q-\mu_r)^2}{2\sigma^2} \quad [43], \tag{4}$$

where μ_q and μ_r are the mean values of $q(x)$ and $r(x)$, respectively. Because the noise component of the received power is drawn from a normal distribution, the received power can be modeled as the following probability density function:

$$v_i(p_{SN}|P_{PU} = P_i) = \frac{1}{\sigma_{v,i}\sqrt{2\pi}} e^{-\left[\frac{(p_{SN}-\mu_{v,i})^2}{2{\sigma_{v,i}}^2}\right]}, \tag{5}$$

where p_{SN} is the random variable of the received power, P_i is the i-th power level of the PU, ${\sigma_{v,i}}^2$ is the variance, and $\mu_{v,i}$ is the expectation of v_i. The interval between the adjacent distributions can be calculated by using the KLD:

$$\begin{aligned} D_{KL}[v_{i+1}||v_i] &= \frac{(\mu_{v,i+1}-\mu_{v,i})^2}{2{\sigma_{v,i}}^2} \\ &= Q \cdot \left\{(P_{i+1}-P_i) \cdot 10^{-\left[\frac{P_L(d_{PU,SN})}{10}\right]} \cdot \zeta_h\right\}^2, \end{aligned} \tag{6}$$

where Q is defined as $1/2{\sigma_{v,i}}^2$. Eq. (6) easily verifies that the main factors of the hidden PU problem are path loss attenuation, fading components, and gaps between transmit power levels. These factors change the KLD value, leading to an overlap of data distributions, and eventually the SN faces the hidden PU problem. The maximum distance between the SN and the PU is derived to avoid the overlap of data distributions in the following section.

2.3 ANALYSIS OF THE MAXIMUM DISTANCE

As the KLD value increases, data distributions move apart from each other, which makes it easy to discriminate channel states. Therefore, our approach is to compute the maximum distance at which the spectrum sensing ability is guaranteed and to place several SNs to cover a whole area where a PU can be placed. In this section, we derive the maximum distance at which SNs can be located from the PU while still sensing the spectrum, and then analyze performances according to parameter values.

2.3.1 MAXIMUM DISTANCE BETWEEN SN AND PU

The accuracy of spectrum sensing is determined by the distance between classes. Let *k* denote the minimum KLD value that can correctly classify the channel state. First, the maximum path loss that satisfies the interval of data distributions to be larger than *k* can be derived from the KLD equation:

$$
\begin{aligned}
D_{KL}[v_{i+1}||v_i] = \frac{\left(\mu_{v,i+1} - \mu_{v,i}\right)^2}{2{\sigma_{v,i}}^2} &\geq k \\
\Rightarrow \quad \mu_{v,i+1} - \mu_{v,i} &\geq \sqrt{2k{\sigma_{v,i}}^2} \\
\Rightarrow (P_{i+1} - P_{\ i}) \cdot \zeta_h &\geq \sqrt{2k{\sigma_{v,i}}^2} \cdot 10^{\left(\frac{P_L(d_{PU,SN})}{10}\right)} \\
\Rightarrow 10 \cdot \log\left[\frac{(P_{i+1} - P_{\ i}) \cdot \zeta_h}{\sqrt{2k{\sigma_{v,i}}^2}}\right] &\geq P_L\left(d_{PU,SN}\right) \\
\Rightarrow L_{max} \quad &\geq \quad P_L\left(d_{PU,SN}\right).
\end{aligned}
\tag{7}
$$

By holding the equality in (7) and substituting the path loss equation in (1), the maximum distance, $\left(d_{PU,SN}\right)_{max}$, can be derived as follows:

$$
\left(d_{PU,SN}\right)_{max} = d_0 \left(\frac{\Delta P \cdot \zeta_h \cdot \lambda^2}{(4\pi d_0)^2 \sqrt{2k{\sigma_{v,i}}^2}}\right)^{\frac{1}{\gamma}} , \tag{8}
$$

where $\Delta P = P_{i+1} - P_i$ is the differential of the transmit power levels.

2.3.2 COVERAGE AREA

From (8), we can see again that the maximum distance is affected by several parameters such as the multipath component ζ_h, the gap of transmission power levels ΔP, and the minimum interval between data distributions k. The maximum distance depending on k and ΔP is plotted in Fig. 5. With a large gap of power levels, the maximum distance increases as well because it can easily satisfy the inequality (7). On the other hand, a high value of k reduces the maximum distance. In addition, the maximum distance is related to multipath fading between the SN and the PU. Fig. 6 shows that the maximum distance and the channel gain have a positive correlation.

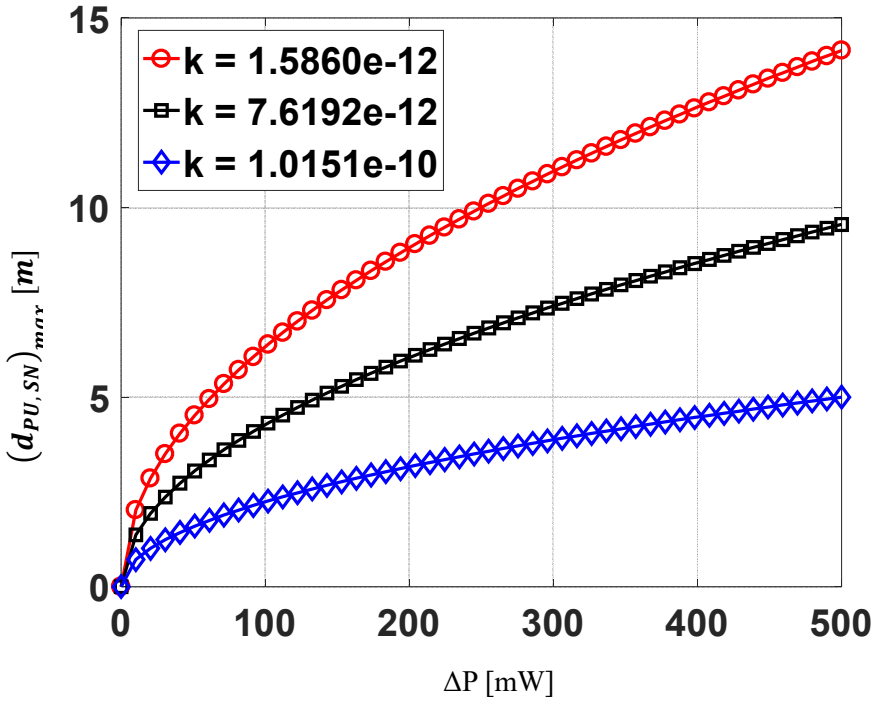


Fig. 5. The maximum distance depending on k and ΔP, with λ = 0.375 m, d_0 = 1 m, γ = 2, ${\sigma_{v,i}}^2 = 1$, and $\zeta_h = 0.8$.

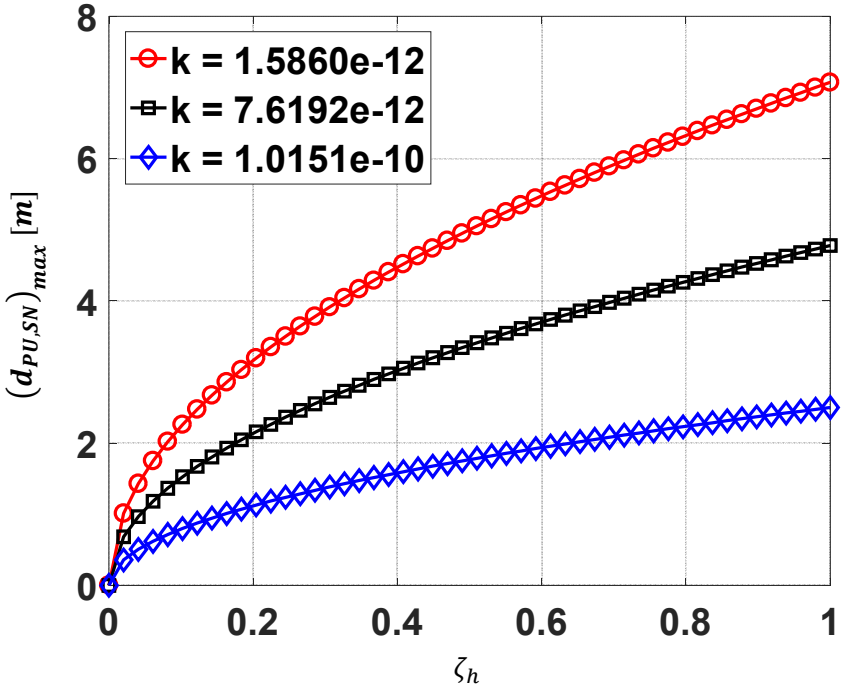


Fig. 6. The maximum distance depending on k and ζ_h, with λ = 0.375 m, d_0 = 1 m, γ = 2, ${\sigma_{v,i}}^2$ = 1, and $\Delta P = 100\ mW$.

2.3.3 Outage Probability for the Maximum Distance

In real environments, the multipath component, ζ_h, is random, affecting the calculation of the maximum distance. We denote the maximum distance with the estimated multipath fading as d_{init} and that with the real-time multipath component into account as d_{real}. Since the channel environment changes in real-time, the computed sensing coverage can be different from the real value. For example, if the estimated multipath component is larger than the real-time value, the estimated coverage is larger than the real coverage ($d_{real} < d_{init}$). If the PU is outside the real sensing coverage as shown in Fig. 7, the interval between data distributions is not guaranteed to be larger than k. Also, as the gap between the estimated and the real sensing coverage becomes large, the error for channel state classification increases rapidly. To measure a risk of losing the discriminatory capability, we analyze the outage probability, assuming that multipath fading is drawn from the Rayleigh distribution:

$$
\begin{aligned}
&\Pr\{d_{real} < d_{init}\} \\
&= \Pr\left\{ d_0 \left(\frac{\Delta P \cdot \zeta_h \cdot \lambda^2}{(4\pi d_0)^2 \sqrt{2k\sigma_{v,i}{}^2}} \right)^{\frac{1}{\gamma}} < d_{init} \right\} \\
&= \Pr\left\{ \zeta_h < \left(\frac{d_{init}}{d_0} \right)^{\gamma} \frac{(4\pi d_0)^2 \sqrt{2k\sigma_{v,i}{}^2}}{\Delta P \cdot \lambda^2} \right\} \\
&= \Pr\{\zeta_h < \alpha\} \\
&= \int_0^{\alpha} \frac{\zeta_h}{\sigma^2} e^{-\left(\frac{\zeta_h{}^2}{2\sigma^2}\right)} d\zeta_h \\
&= 1 - e^{-\left(\frac{\alpha^2}{2\sigma^2}\right)},
\end{aligned} \tag{9}
$$

where σ^2 is the variance of the Rayleigh distribution. The outage probability depending on d_{init} and k is plotted in Fig. 8. If large values of ζ_{init} and k are used to compute d_{init}, the outage probability grows rapidly because the probability of $d_{init} > d_{real}$ increases.

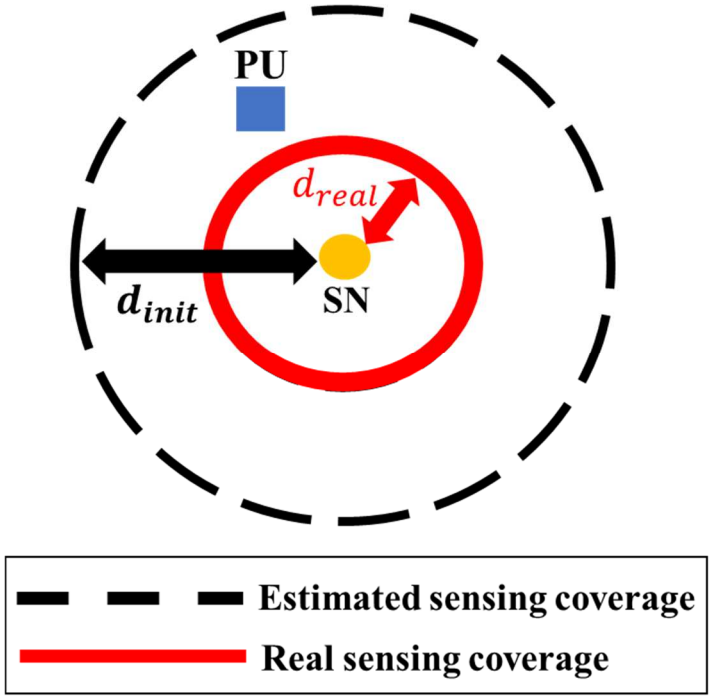


Fig. 7. The mismatched sensing coverage due to inaccurate channel estimation.

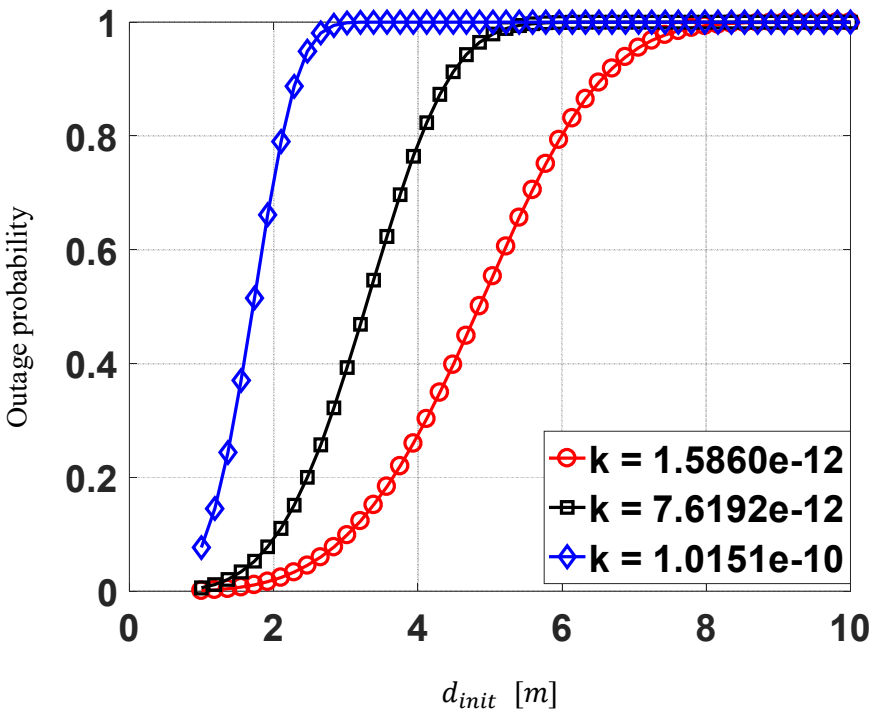


Fig. 8. The outage probability depending on k and d_{init}, with λ = 0.375 m, d_0 = 1 m, γ = 2, ${\sigma_{v.i}}^2 = 1$, ΔP = 100 mW, and σ = 0.4.

2.3.4 Positioning of Spectrum Sensing Nodes

Regardless of the position of the PU, several SNs are better to sense the spectrum for a particular area. A single SN can sense the spectrum within a circle with a radius of $\left(d_{PU,SN}\right)_{max}$. However, as placing several SNs, the sensing coverage is overlapped. In order to minimize the overlapped coverage while covering the whole area, the proper distance between SNs should be calculated. In previous research, two possible methods for placing the SN, coverage-first placement and connectivity-first placement [35] depend on sensing and communication distances. Because the sensing coverage is shorter than the communication distance, we place the SN by the coverage-first placement method as illustrated in Fig. 9. However, in practical situations, SNs near obstacles or blockages have shorter sensing distances than others. In this case, the sensing coverage should be re-computed by adjusting the multipath factor, and then, additional sensing nodes are required.

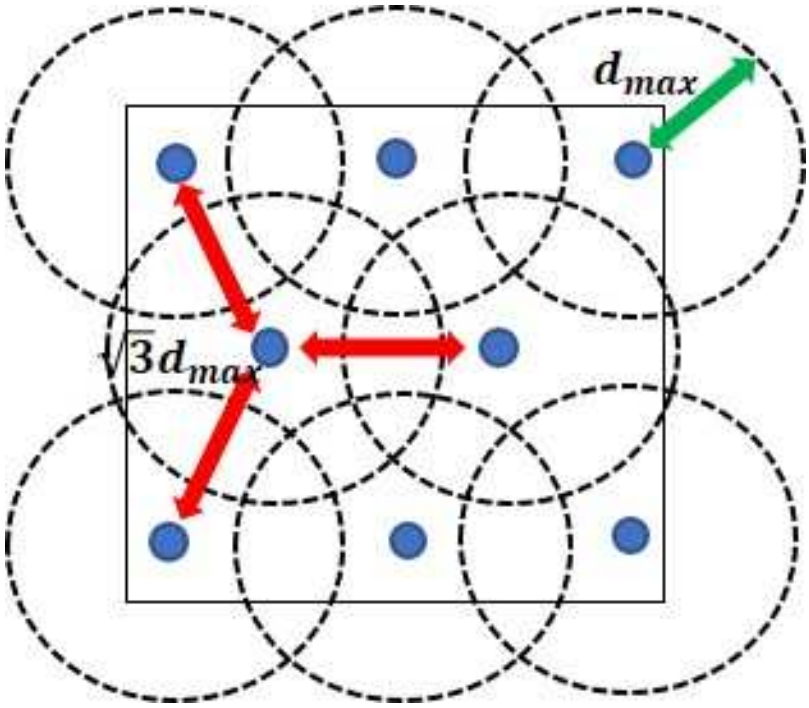
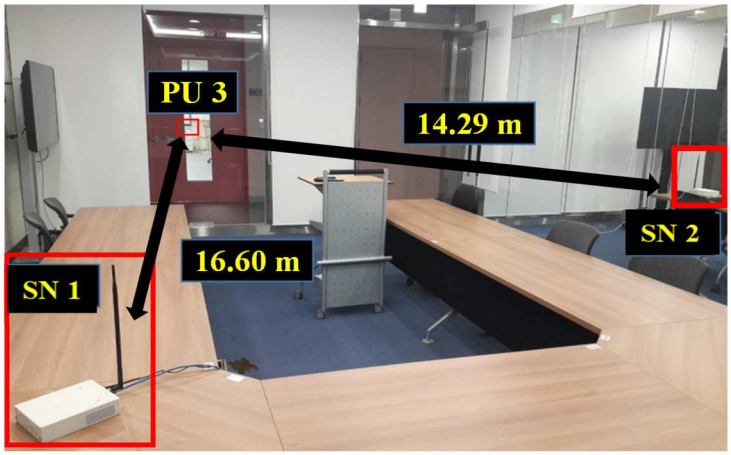


Fig. 9. The positioning of SNs based on the coverage-first placement method to reduce the overlapping coverage.

2.4 EXPERIMENT SETUP AND RESULTS

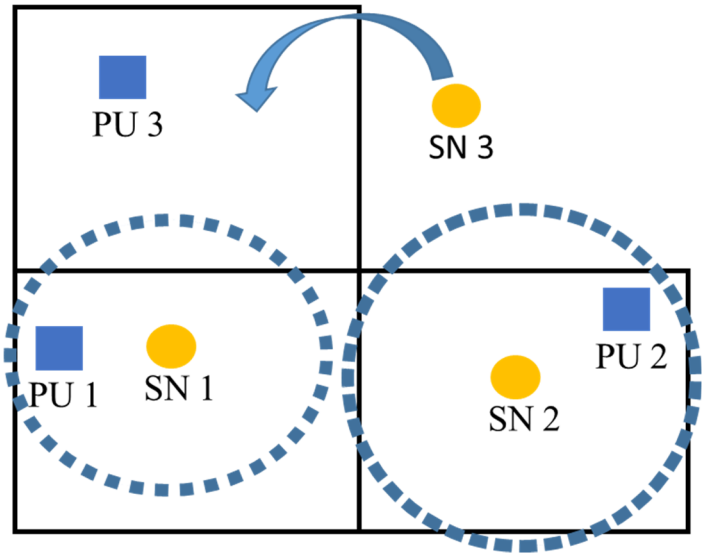
In this section, to show feasibility of the proposed schemes, experiment environments and results are discussed. We conduct experiments with and without the hidden PU problem, respectively. Fig. 10 shows the experimental setup for the hidden PU problem and our solution. All of SNs and the PU are connected to NI USRP-2921 and the received signals are processed using Simulink in MATLAB. PU 3 has the access authority for 2.45 GHz with the three power levels of 0, 50, and 100 mW. SN 1 and 2 are unlicensed users, located far from PU 3 at 16.60 m and 14.29 m, respectively. Because PU 1 and PU 2 are located within the sensing coverages of SN 1 and SN2, respectively, the channel state can be classified correctly. However, they cannot cover the whole area. If PUs move or new ones appear outside of the sensing coverage, this can cause the hidden PU problem. Fig. 11 (a) shows that when a new PU (PU 3) appears outside of the coverage for SN 1 and SN 2, their data distributions are overlapped.

Therefore, even if SN 1 and SN 2 may cooperate, it is still difficult to sense the spectrum of PU 3 correctly. To cover the whole area where the PU can be placed, we suggest to place several SNs. The sensing coverage of SN 3 is computed as 10.86 m by Eq. (8), using k = 1.5860e-12 from simulation results. However, due to difficulty in estimating channel states, the computed distance may misclassify the spectrum occupancy. We thus use 7.00 m to conservatively reduce the outage probability. The number of the data measured by each SN is 2,103 with the equal number of samples for each power level. 80 % of the total data are used to train and the others are to test ANNs and the SVM. In training, as preprocessing, min-max normalization is used to equalize the scale of features. In the results of ANNs under the hidden PU environment, the PU transmission power levels can be classified with 81.19 % accuracy and the SVM can classify channel states with 50.00 % and 86.07 % accuracy for the PU levels of 50 mW and 100 mW, respectively. We present confusion matrices for experiment results to analyze the accuracy of spectrum sensing in more detail. For the hidden PU problem, as shown in TABLE II (a), the spectrum sensing capability of the FC is deteriorated. The true positive rate (TPR) is 4.12% and the false positive rate (FPR) is 0.00%. Note that the FC classifies "busy" state for all cases. Although the TPR improves by increasing the power level of PU to 100mW (TPR = 43.98%, FPR = 1.46%), it is still low to detect the spectrum. However, if the PU is within the sensing coverage of SN, it is possible to detect the spectrum correctly. By introducing SN 3, the distribution of the received power has separations larger than the previous one as shown in Fig. 11(b), and the accuracies of ANNs increase to 98.80 % and 97.50 % with different power levels, respectively, and that of the SVM to 99.29 %. In this case, because the PU is in the sensing coverage of SN 3, the accuracy of spectrum sensing is very high (e.g., TPR = 95.86% and FPR = 1.48% for 50mW; TPR = 99.30% and FPR = 0.72% for 100mW). Our experimental results show that the hidden PU problem can be mitigated by deploying another SN.



(a)

(b)



(c)

Fig. 10. Experiment setup: (a) with the hidden PU problem (b) for CSS with a new SN (c) illustration for node placement.

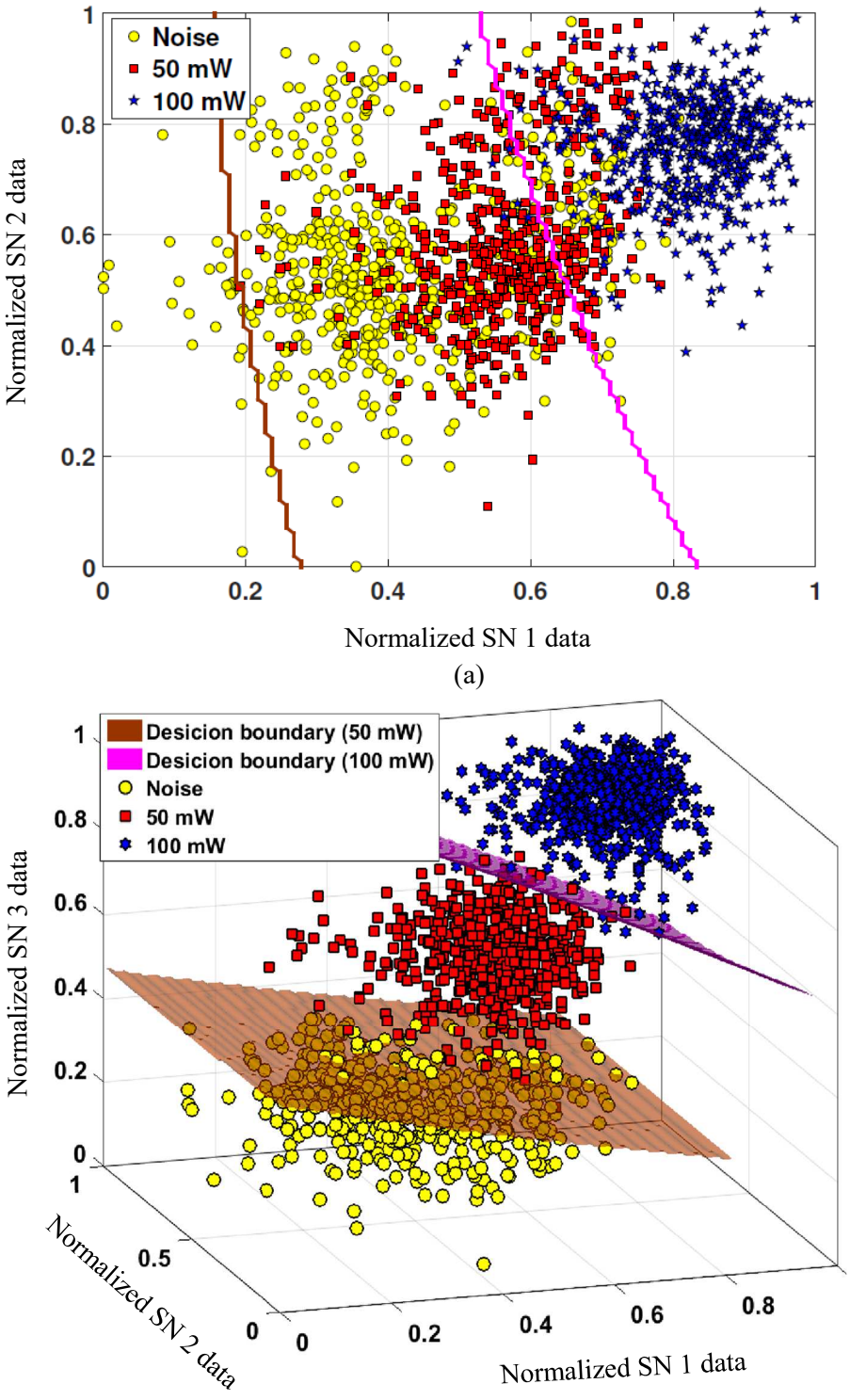


Fig. 11. Experiment results: (a) Overlap of data distribution (b) Results of spectrum sensing

TABLE II

CONFUSION MATRIX FOR EXPERIMENT RESULTS

P = 50 mW	Predict: Idle state	Predict: Busy state
Actual: Idle state	6 (True Positive)	139 (False Negative)
Actual: Busy state	0 (False Positive)	135 (True Negative)

P = 100 mW	Predict: Idle state	Predict: Busy state
Actual: Idle state	106 (True Positive)	37 (False Negative)
Actual: Busy state	2 (False Positive)	135 (True Negative)

(a)

P = 50 mW	Predict: Idle state	Predict: Busy state
Actual: Idle state	139 (True Positive)	6 (False Negative)
Actual: Busy state	2 (False Positive)	133 (True Negative)

P = 100 mW	Predict: Idle state	Predict: Busy state
Actual: Idle state	142 (True Positive)	1 (False Negative)
Actual: Busy state	1 (False Positive)	136 (True Negative)

(b)

TABLE II. Confusion matrices for spectrum sensing: (a) with (b) without the hidden PU problem.

Since the proposed node placement guarantees separation between data distributions, it can have the computational effect for machine learning algorithms such as K-nearest neighbors (K-NN) [44]. In K-NN, the class of test data is determined by a plurality vote of its K nearest neighbors among

TABLE III

THE RESULT OF K-NEAREST NEIGHBORS

	K=3	K=5	K=7	K=9
Test Accuracy (%)	98.57	98.75	99.10	99.10
Time for classification (s)	0.6222	0.6647	0.7108	0.7471

training points. In general, a sufficiently large value of K requires excessive computation time. However, large intervals between data distributions with the proposed node placement make it possible to classify the class of test data without a large value of K. The results of K-NN for the test set are summarized in TABLE III. Even with a small value of K, the accuracy is comparable to those of larger values.

2.5 CONCLUSION

In this section, we investigated a node placement scheme for machine learning based cooperative spectrum sensing in cognitive radio networks. It has been verified that the hidden PU problem causes the overlap of data distributions and reduces the accuracy of machine learning algorithms. Our approach was to find the conditions that secure sufficient distances between data distributions. From the conditions, we derived the sensing coverage of a single SN and proposed the method of placing several SNs to cover the whole area where the PU can be placed. Experiment results showed that the hidden PU problem can be mitigated by the proposed node placement.

The purpose of this section is to investigate the node placement method for cooperative spectrum. Relatively small cognitive networks have been considered for simulations and experiments. However, because the FC classifies the channel state by consolidating the received power from the distributed nodes, energy consumption and the complexity increase with the number of nodes, which are key issues in large-scale cognitive networks. In terms of energy consumption and the complexity for sensing the spectrum, it is inefficient for all the SNs to participate in cooperative spectrum sensing. To solve this challenge, several methods which utilize the desired detection alarm probability and false alarm

probability (e.g., sensor selection [41] and alternation between censoring and sleeping [42]) have been discussed. In this research, we propose how to find the conditions that secure sufficient distances between data distributions. Applying the suggested methods from previous research can achieve the constraints for detection and false alarm probability, leading to extension to large-scale cognitive networks.

3. SEMANTIC SEGMENTATION FOR GEO DATA GENERATION

3.1 INTRODUCTION

Aerial images can provide valuable information on areas that are difficult for people to access or access non-intrusively [43]. The information obtained using aerial images is used in a variety of industries including land inventory, vegetation monitoring, and environmental assessment [44]. In particular, extraction of manufactured features such as buildings, roads, railway lines, etc. or natural ones from aerial images is important in many applications that depend on geographic information systems (GISs). GISs are used to carry out tasks such as urban planning, traffic management, industrial development, and cartography as well as for emergency planning systems for evacuation and fire response [45].

The feature extraction method for creating or modifying a GIS is mainly to perform a visual analysis and manually digitize aerial images. It is still the main method to generate geospatial data but takes much labor and time to extract or identify features manually [46]. With the development of the optical sensor technology, it is possible to obtain higher resolution images, and more accurate information can be therefore obtained from their images. Nevertheless, it is difficult yet to generate all the geospatial data by the above method.

There are two main types of research for detecting objects from aerial images: semi-automatic and fully-automatic [47]. However, because the semi-automatic methods require prior knowledge of the extraction process, such as identifying areas of the road and buildings through human intervention, the automatic methods, which are modeled on the analysis and interpretation of a human operator, to segment objects from aerial images is widely used in segmenting roads [11], and buildings [48]. Also, traditional machine learning techniques such as support vector machine (SVM) and artificial neural networks (ANNs) have been studied in [12], [13]. The automatic methods are objective without human

intervention, but their results have generally been discouraging due to various adverse factors (image noise, shadows, etc.) [49].

Recently, as the deep learning technology has evolved, it has been applied to many problems related to computer vision fields such as image classification, object segmentation, and semantic segmentation. Additionally, deep learning techniques such as convolutional neural networks (CNNs) have attracted much attention to segment objects in satellite images [50]-[51]. To segment roads and buildings from satellite images, several semantic segmentation models based on deep learning techniques such as Fully Convolutional Networks [52]-[56], and U-Net [57] algorithms have been used.

Although it is possible for deep learning algorithms to learn contextual features automatically, there are several issues. First, it requires many data to train them. Existing open datasets are imperfect and not enough to train such deep learning algorithms. though there are the large-scale datasets for classification [58]-[59]. Second, due to the lack of multiclass datasets, binary data have been used in previous research [52]-[53]. To address these problems, we make a large scale of the dataset (i.e., the numbers of training and test data are 72,400 and 9,600, respectively), including multiple objects (e.g., building, road, water, and background). We also design a new semantic segmentation model, exploiting multi-scale features by using the pyramid pooling layer (PPL) [60] to extract information from various classes. The proposed model with our self-made dataset outperforms other algorithms, as shown in the test results.

The main contributions of this paper are summarized as follow:

- We integrated U-Net architecture with an eight-level PPL module to solve a problem that U-Net is not able to clearly distinguish the gap between buildings as well as to achieve the better shape prediction of the buildings mainly because the PPL provides a capability of global context aggregation.

• We developed object segmentation systems by improving architecture and compared performances with other algorithms. The proposed systems show the performance of pixel accuracy of 87.61% and the mean intersection of the union (mIOU) of 79.52 for the 9,600 test images.

• We made a large scale of datasets for multiple object segmentation from aerial images, including many complex obstacles in the metropolitan area of Seoul, Korea.

The rest of this paper is organized as follow: In Section 3.2, the related work for deep learning based objects segmentation is discussed. In Section 3.3, the proposed object segmentation systems are discussed. In Section 3.4 and 3.5, our dataset and test results are described. Finally, we conclude our contributions in Section 3.6.

3.2 DEEP LEARNING-BASED OBJECT SEGMENTATION SYSTEMS

There are several algorithms that are applied to object segmentation based on CNNs [61]-[63]. Because the performance of deep learning algorithms depends on their structures, it should be optimized to improve the performance by adjusting and fine-tuning. In this Section, CNN based segmentation algorithms are described as follow.

3.2.1 FULLY CONVOLUTIONAL NETWORKS

A Fully Convolutional Network (FCN) is a modified CNN to semantically segment images. The FCN predicts every pixel of images, and this can be trained end-to-end and pixels-to-pixels. A key idea of FCN is changing of the CNN model from classification to dense prediction by reinterpretation of fully connected layers of the classifier as a fully convolution layer [64]. The FCN consists of an encoder of input images and a decoder that up-samples encoded images by their original image size. The encoder part of FCN consists of visual geometry group network (VGGNet) [65] that is a famous CNN classification model and the decoder part consists of a deconvolution layer for up-sampling.

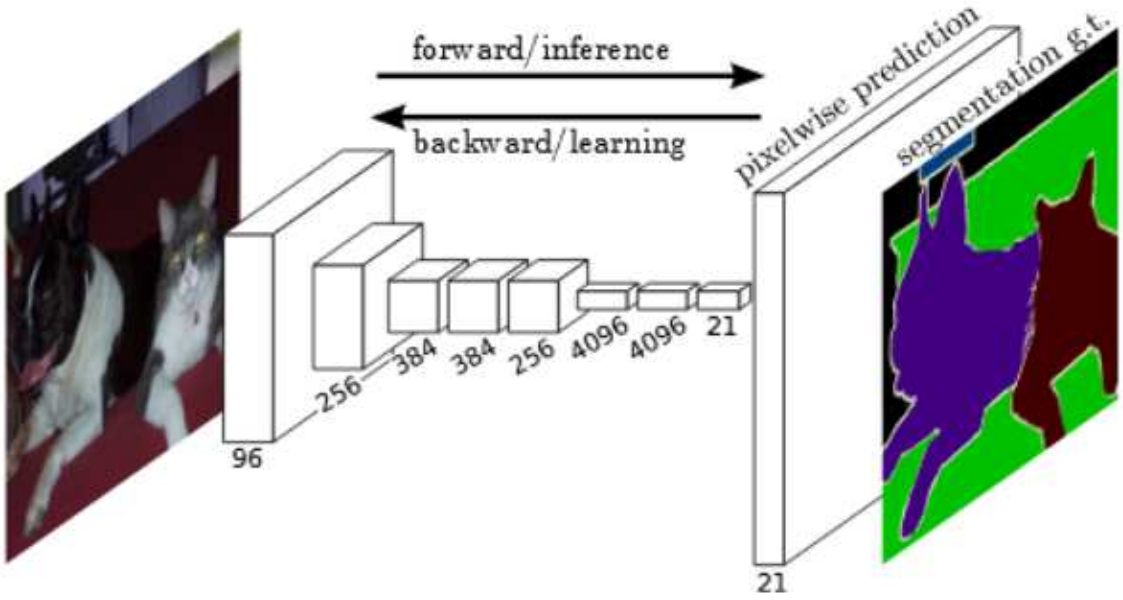


Fig. 12. Fully convolutional networks provided by [64].

3.2.2 U-NET

U-Net [57] is a modified FCN for yielding more precise segmentation. U-Net has two different architectures compared to FCN as illustrated in Fig. 13. First, U-Net uses high-resolution features to increase localization accuracy by combining the decoding layer and the encoding layer. Second, while FCN uses 1x1 convolution layer at the last layer of the encoder to use the pre-trained model, U-Net does not use 1x1 convolution at the encoding layer. Thus, U-Net shows outstanding performance not only for processing biomedical images but also for segmenting objects from satellite imagery [67]. In this section, we also exploit the U-Net architecture for multiclass object segmentation.

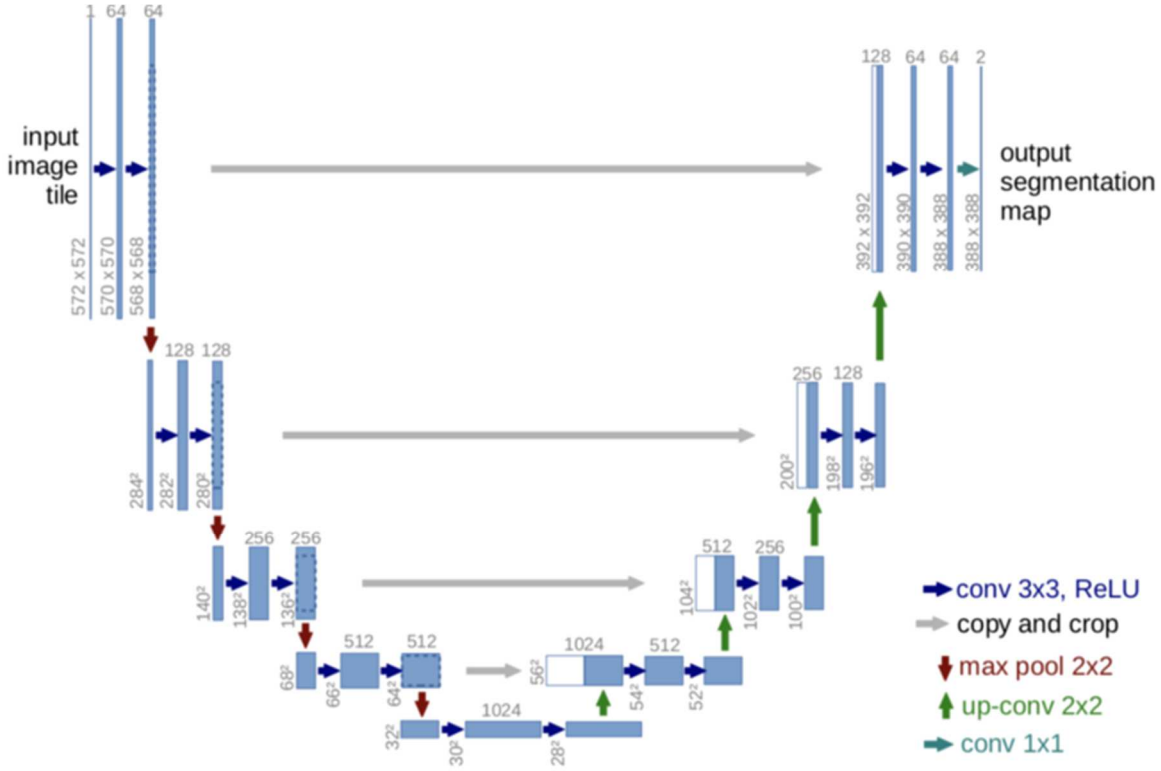


Fig. 13. U-Net architecture provided by [57].

3.3 U-NET WITH PYRAMID POOLING LAYERS

Feature maps in different sub-regions generated by Pyramid Pooling Layers (PPL) significantly enhanced segmentation of various classes. Thus, We re-designed the U-Net architecture by exploiting PPL. In the PPL, the output of each pyramid level is combined and up-sampled to the same resolution as the input via transposed convolution. Finally, the per-pixel prediction is presented.

The proposed algorithm consists of the encoder-decoder structure. In the encoder part, spatial dimension is reduced up to 14x14 by convolution filters (3x3, stride 1) and the max-pooling layer. The numbers of the filter channels are increased. In this process, we utilize batch normalization [67] to prevent the overfitting problem as shown in Fig. 14. At the end of the encoder, we use PPL, consisting of 8 layers. Because the pooling size is different (1, 3, 8, 9, 11, 12, 13, and 14), it is possible to extract features from several ranges. In the decoder part, to recover spatial dimension, up-convolution (or the

transposed convolution) is used.

Although using consecutive pooling layers allows to reduce parameters and extract long-range information, it can lose local information such as boundaries of objects. To overcome this problem, skip connection methods are used [57], [64]. The features extracted from PPL are up-sampled and concatenated with the output of the fourteenth rectified linear unit (ReLU). By transposed convolution, the spatial dimension of combined features is recovered from 14x14 to 28x28. As it combines the output of the eleventh, eight, fifth, and second ReLU and the transposed convolution is applied, the spatial dimension is recovered gradually, and the number of filters is regulated by convolution layers. Finally, the pixel-wise prediction is presented.

3.4 DATASET AND EXPERIMENTS

Most work to segment objects such as roads and buildings has been carried out using aerial images in rural areas [53]-[55]. Additionally, they have used many data from the OpenStreetMap (OSM) [68]. However, the OSM has several defects (e.g., buildings match in a different direction or some regions are not mapped) since the OSM is a free editable mapping platform, which might be easily contaminated by misinformation added by anonymous editors. The defects of OSM data might disturb training for CNN [69]. To overcome the aforementioned disadvantages, a fine-tuning scheme is proposed in [52], showing that the fine-tuned network outperforms the un-tuned one. After they trained CNN by using a large scale of raw OSM data for binary classification, a tiny piece of the manually labeled data was used to tune convolutional filters. On the other hand, we created a large scale of accurately labeled datasets for multiclass object segmentation and performed experiments on the self-made datasets that were based on RGB aerial images over the area of Seoul, Suwon, Anyang, Gwacheon, and Goyang, which are among the most complex cities in Korea. Our datasets are divided into training and test datasets, each of which is used for training and evaluating the proposed model. Note that contrasted from the binary labeled data in previous work, ours are labeled for multiple object classes.

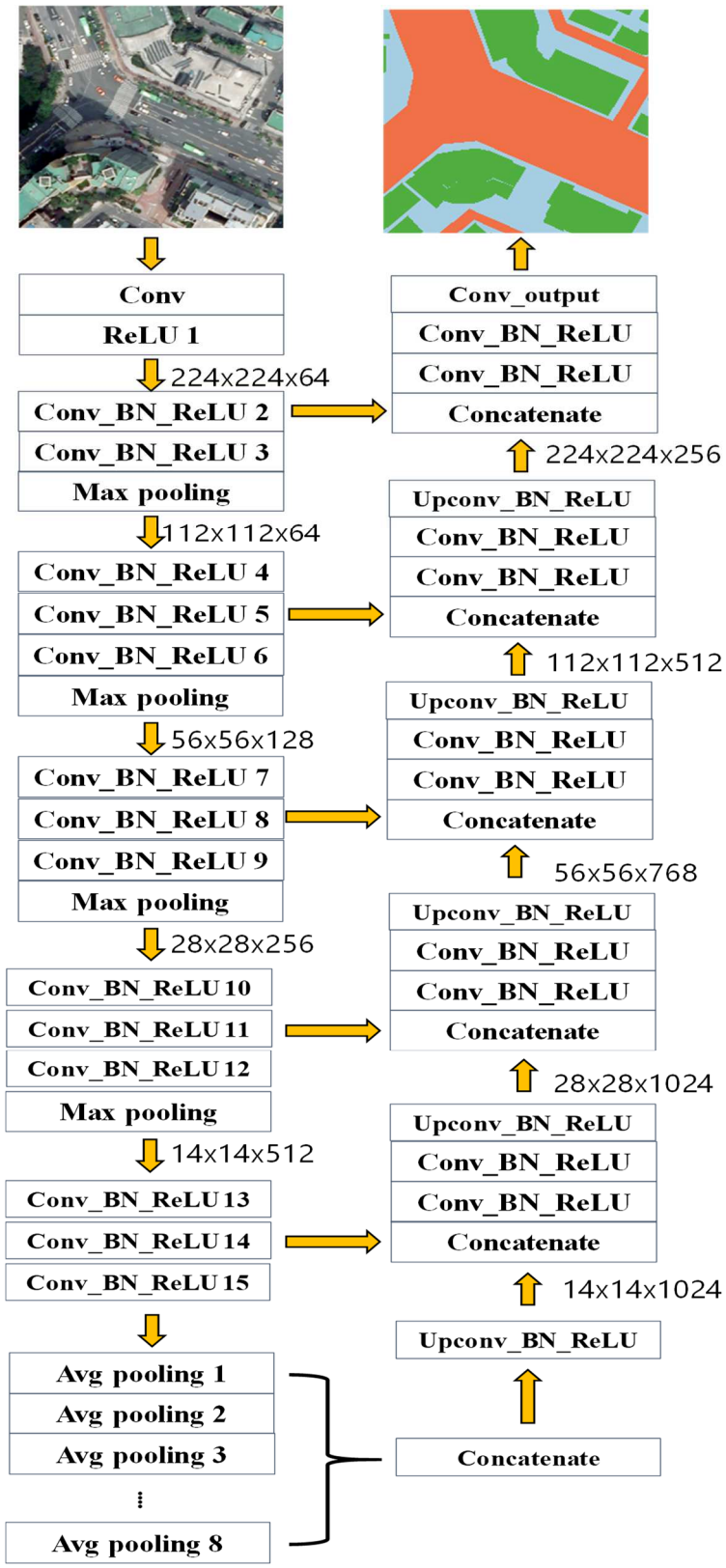


Fig. 14. A U-Net based object segmentation system. Note that Conv is the convolution layer, Upconv is the transposed convolution, and BN is batch normalization.

3.4.1 DATASET

We perform our experiments with images of all the areas of the five cities mentioned above. The ground-truth data were obtained by changing vector data provided by the government agency of National Geographic Information Institute (NGII) to images using Quantum Geographic Information System (QGIS), a free and open-source GIS application. Fig. 15 illustrates the process of making the ground truth. These data are more accurate than the ones from the OSM, because they have been made by experts for many years. As changing a data format of the ground-truth data, the data are labeled into four-classes: background, building, road, and water. Our dataset, as shown in Fig. 16, consists of pairs of RGB images with 0.51m spatial resolution and ground-truth images. The dataset covers an area of $551km^2$ and are randomly divided into an area of $486.5km^2$ for training and $64.5km^2$ for testing. All the data were divided into multiple images with the pixel size of 224x224, of which 72,400 images were assigned to the training set and 9,600 to the test set.

Our datasets have two advantages: First, the datasets consisted of images of the most complex cities. For applications, we here highlight that the proposed method performs well in complex urban areas as well as in rural areas, for which our datasets are appropriate. Second, our datasets were labeled

Fig. 15. The process of making the ground truth. The dataset covers an area of 551 km^2 and is randomly divided into an area of 486.5 km^2 for training and 64.5 km^2 for testing. All the data were divided into multiple images with the pixel size of 224 × 224, of which 72,400 images were assigned to the training set and 9,600 to the test set.

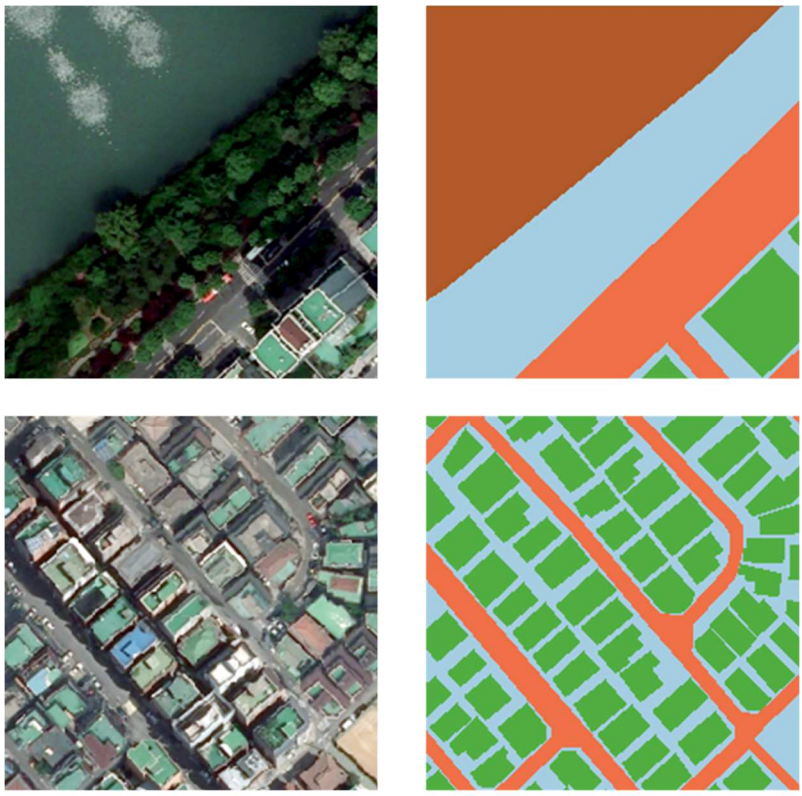


Fig. 16. Examples of our dataset. In the right side images, green pixels designate buildings, blue background, orange roads and brown water.

as multiple classes. Many problems have been investigated in binary classification, mainly road classification, and building segmentation due to the lack of multiclass datasets [52]-[53]. Moreover, classifying several classes at once can solve the problem of two or more classes being overlapped at one point when superposing several binary classification results.

3.4.2 TRAINING SETUP

The average pyramid levels were applied to the end of feature extract and the ReLU are used as the activation functions for the hidden layers. We here used the cross-entropy loss between prediction results and the ground truth for training the proposed model. The Adam optimizer [70] with the learning rate $\alpha = 10^{-4}$, the exponential decay rates for the moment of estimates $\beta_1 = 0.9$, $\beta_2 = 0.999$ and

the constant value $\varepsilon = 10^{-8}$ were used to train our deep neural network. The FCN with PPL were trained and evaluated by the self-made datasets containing 4 objects classes (i.e., building, water, road, and background). The number of 72,400 data were used for training and 9,600 for testing. The mIOU and pixel accuracy were utilized as performance indices. The experiments were implemented using the public platform Tensorflow [71] and run on an Intel core 6 i7-7820X CPU @ 3.6 GHz with 2 GPUs of Nvidia Geforce Titan XP (12 GB). The training setup is summarized in TABLE IV.

TABLE IV
TRAINING SETUP FOR DEEP LEARNING ALGORITHMS

Parameter	**Description**
Activation function	Rectified linear unit (ReLU)
Optimizer	Adam optimizer
Learning rate	10^{-4}
The exponential decay rate	$\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-4}$
Batch size	4
Image size	224 x 224 x 3
Number of classes	4 (building, road, water, background)
Training data	72,400
Test data	9,600
Hardware	GPU: Nvidia Geforce Titan XP (12 GB) * 2 ea CPU: Intel core 6 i7-7820X @ 3.6 GHz * 6 ea
Deep learning library	Tensorflow

3.5 RESULTS AND DISCUSSION

In this Section, we analyze the results of training for U-Net with PPL (UNetPPL) in detail. Fig. 17, including input images and the ground truth, illustrates the results of training for each model. Note that each object is colored differently (e.g., green pixels designate buildings, blue background, orange roads and brown water). To compare with each of deep learning architectures quantitatively, we compute mIOU, IOU per object and the average pixel accuracy of all ojbects as the results are summarized in TABLE V. Although the IoU of FCNPPL is lower than that of FCN in a building class, the number of false positive pixels in FCNPPL($\approx$ 20M pixel) is much lower than FCN($\approx$ 23M pixel). Also, as shown in Fig. 17. FCNPPL distinguishes buildings better and moreover represent the better-shaped building than FCN. Likewise in FCNPPL, the number of false positive pixels in UNetPPL($\approx$ 19M pixel) is much lower than that in U-Net($\approx$ 23M pixel). The main reason would be that the use of an eight-level pyramid pooling module allows to obtaining the effective global context information [60] and therefore produces a high-quality result on object segmentation.

In addition to FCN, FCNPPL, U-Net, and UNetPPL, we also conducted experiments on Pyramid Scene Parsing Network (PSPNet), which has proposed the pyramid pooling module. For most evaluation criteria, the proposed method provides higher scores than those of the others. It can achieve mIOU to 79.52, which is 0.50 higher than mIOU of U-Net. It is important to note that the compared U-Net is not the original architecture proposed in [57] but a highly calibrated model for the enhanced capability of object segmentation from satellite images [66]. Therefore, even a small gain of mIOU with UNetPPL is remarkable. In particular, UNetPPL outperforms other methods of segmenting roads and buildings, which are the most typical human-made objects as illustrated last example in Fig. 17. Compared with the worst performance of FCN, UNetPPL shows the IOU improvement of 2.92 and 1.97 for building and roads, respectively.

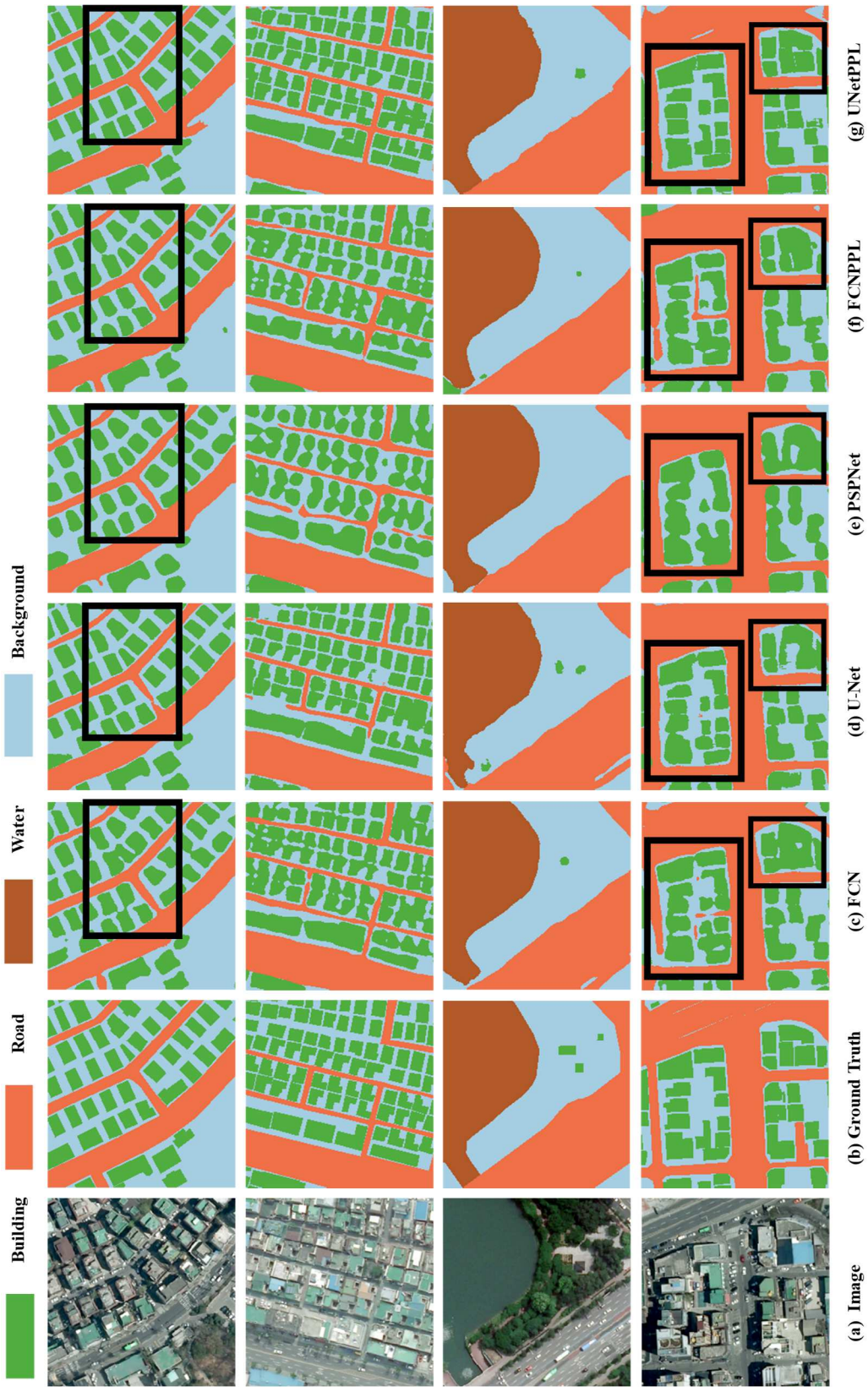


Fig. 17. Results on the test set for each model.

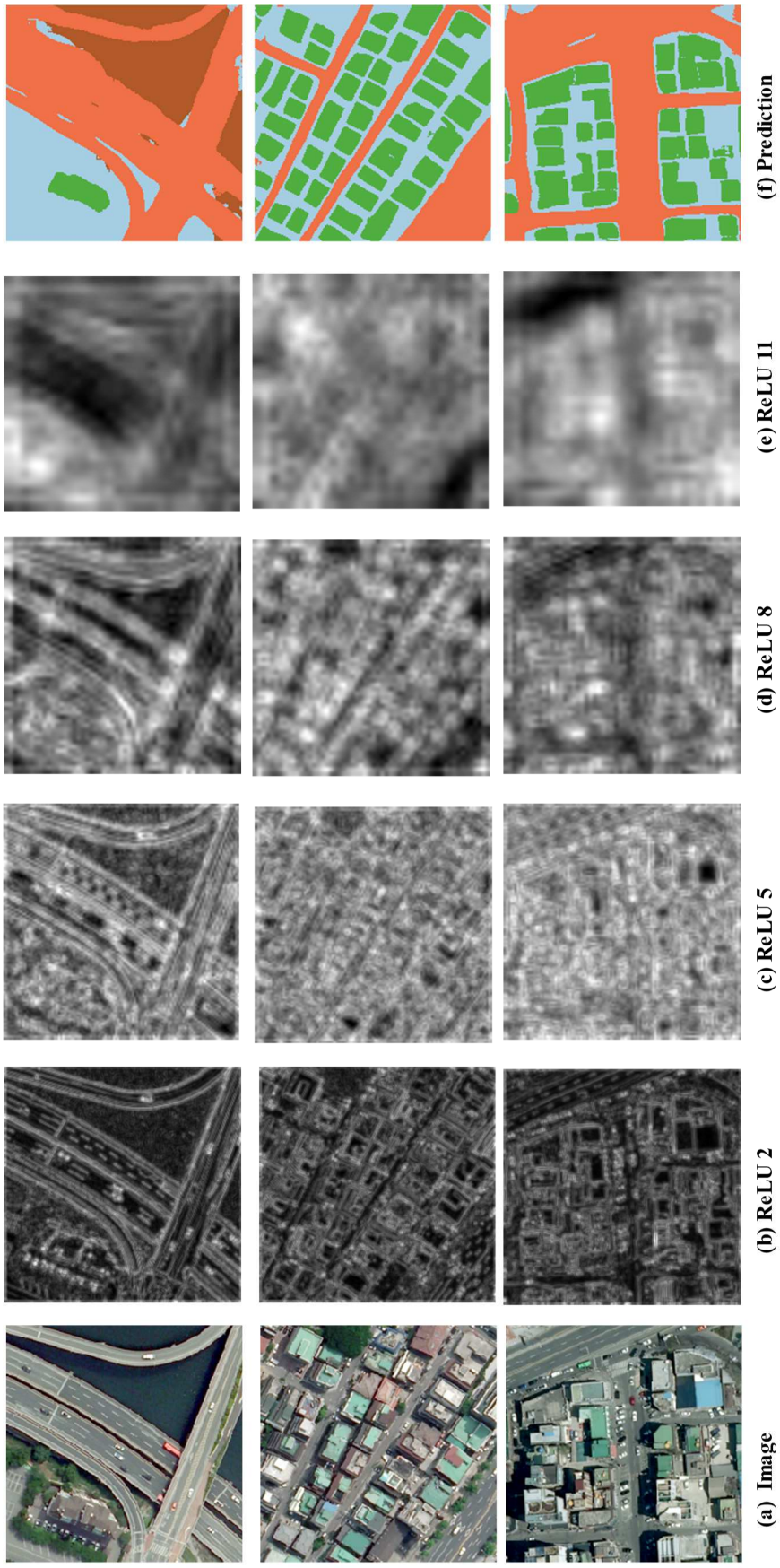


Fig. 18. Output of activation function.

TABLE V

TEST RESULTS FOR 9,600 DATA

	mIOU	Back.	Build.	Road	Water	Pixel.acc
FCN	77.88	78.16	67.42	73.64	92.30	85.90
U-Net	79.02	79.83	69.11	74.58	92.55	86.91
PSPNet	78.30	78.82	67.02	73.64	**93.71**	86.14
FCNPPL	78.03	78.56	67.05	74.12	92.39	86.08
UNetPPL	**79.52**	**80.16**	**70.34**	**75.61**	91.96	**87.61**

We additionally demonstrated through these experiments that our model shows better performance in aerial photographies than PSPNet, which had suggested the pyramid pooling module. General convolution is superior to dilated convolution used in PSPNet for feature extraction in aerial images.

3.6 CONCLUSION

In this section, we modeled the enhanced semantic segmentation model capable of better segmenting multiclass objects from aerial images by exploiting a state-of-the-art CNN-based algorithm. Different from previous research that has utilized a lot of raw OSM data and a tiny piece of the manually labeled dataset, we made a large scale of an accurately labeled dataset, including major cities in Korea. Our dataset, based on city centers, includes many complex obstacles, and thus, may be more practical than the binary labeled data in previous work.

Our proposed UNetPPL utilized pyramid pooling layers and skip connections widely to extract multi-scale features of objects. It showed better performance than other deep learning models in the segmentation of multiclass objects from aerial images. Test results showed that UNetPPL outperforms FCN, FCNPPL, U-Net and PSPNet regarding pixel accuracy and mIOU for four classes (building, road, water, and background). Moreover, UNetPPL has a capability of classifying objects in detail (e.g., the interval of buildings), which is significantly advantageous in the complex environments.

Our CNN-based multi-object segmentation with a large scale of dataset made to train algorithms only focused on the urban area of Korea. Data attributes of objects can be different from one area to another. Therefore, to enhance generalization of UNetPPL, supplementing other country data and the corresponding training should be required, which will be as addressed in our future work.

4. DOMAIN ADAPTATION FOR AERIAL IMAGES

4.1 INTRODUCTION

With the development of sensor technology, a considerable amount of remote sensing images are now available [72]-[73], and these images contain highly detailed information, including the shapes of buildings, vehicles, roads and other objects. To utilize aerial images, extracting the objects of interest from aerial images is necessary. However, images obtained under different conditions (e.g., locations, times, weather) make it difficult to extract even identical objects. This is especially true for buildings, which have large variations in the sizes, colors, and angles, frequently causing misclassifications [74]. Thus, building extraction from aerial images is mainly performed by a digitizing process done by hand [13], but this is laborious and time-consuming.

Recently, convolutional neural networks (CNNs) [75], replacing hand-crafted feature extraction with learnable filters, have shown remarkable performance in a wide range of areas, such as image classification [76]-[81] and object detection [82]-[85]. However, CNNs face structural limitations during the pixel-wise labeling of images. In the process of feature extraction, using consecutive pooling layers causes a loss of spatial information (e.g., the boundaries or edges of objects). Thus, most CNN architectures are good at recognizing objects but rarely succeed in localizing objects precisely [46]. During the pixel-wise labeling of aerial or satellite images, these problems become more challenging. Misclassifications of object boundaries hamper the overall performance of semantic segmentation systems. To achieve both recognition and localization, several types of architecture for aerial images have been studied [46] [54]. A refinement method for the coarse output of CNNs has been proposed [46]. A method that fuses multi-level feature maps to achieve accurate boundary inference outcomes has also been presented [86]. On the other hand, in other work [87], adversarial networks [88] were utilized to enforce the forms of high-order structural features learned from ground-truth label maps. In one study

[89], a Sobel detector [90] and a fully convolutional network were used together to rectify a semantic contour. Additionally, combining a pyramid pooling layer [60] with the U-Net architecture [57] was proposed [7].

Although these methods improve the performance of deep-learning models, trained models fail to segment objects for unseen images when there is a domain gap between the training and test data. To solve this problem, domain adaptation methods have been proposed [57] [91]. In one such study [57], convolutional neural networks were trained by means of a gradient reversal approach to adapt the test domain. The goal of the trained model is to ensure that the results of feature extraction are such that a discriminator cannot classify between the training and test domains. In another study [14], pixel-level domain adaptation in the output space is proposed, indicating that the performance of semantic segmentation is better than that of feature adaptation. On the other hand, an adversarial iterative scheme that use samples generated by an adversarial attack during the training procedure was also proposed [91]. However, these methods are not customized for aerial images. While driving datasets such as GTA5 [15] and Cityscapes [16] as discussed in the literature [14] have static class distributions for objects (e.g., the road is at the bottom and the sky is at the top in the input images), the buildings in aerial images are located randomly. Moreover, adversarial-attack-based data augmentation can be expanded with information for target distributions. In this paper, we suggest a domain-adaptation-based data augmentation scheme for aerial image segmentation. Because both domain adaptation and adversarial attacks are used jointly, it is possible to force the direction of transformation for input images. Thus, the generated images are outside of the capacity of trained models and close to the target distribution. With adversarial samples during the training procedure, it is possible to expand the capacity of the model to realize better extraction from aerial images.

The main contributions of this chapter are summarized below.

- The designed systems utilize both domain adaptation and adversarial attacks to force the direction of image generation. By using adversarial training with generated images, the capacity of the segmentation model can be expanded.
- The segmentation model trained by the generated images show the following performance outcomes: mean intersection of the union (mIoU) values of 80.00 and 63.78 for Inria [72] and Massachusetts [73] datasets, representing corresponding increments of 0.16% and 7.12% for these datasets.

The rest of this paper is organized as follows. In Section 4.2, semantic segmentation models for building extraction from high-resolution aerial images are studied. In Section 4.3, the relationship between the domain gap and the segmentation performance is described. In Section 4.4, the proposed augmentation systems are discussed. In Sections 4.5 and 4.6, training results with/without augmentation are described. Finally, we conclude the paper with its contributions in Section 4.7.

4.2 DEEP LEARNING-BASED SEMANTIC SEGMENTATION

The goal of semantic segmentation is to assign each pixel a semantic label (e.g., buildings and background) in images. Because the performance of semantic segmentation is related to the architecture of the models used, it should be designed to process input data properly. In this section, we suggest a semantic segmentation approach for building extraction from aerial images.

4.2.1 SEMANTIC SEGMENTATION SYSTEMS

One of the basic types of architecture is the fully convolutional network (FCN) [17], which is a modified type of CNN used to segment images. Due to the excellent performance of the FCN, recently designed segmentation models are mostly based on the architecture of the FCN. Because FCNs consist

of an encoder for feature extraction and a decoder that up-samples the extracted features to the original image size, in recently proposed models, more efficient basic encoder models are adopted to improve the performance. In addition, integrating the extracted features with certain independent modules such as a pyramid pooling layer [60] and a summation-based skip connection [92] can be utilized. We also suggest semantic segmentation architecture based on an efficient type of encoder architecture and an integrated feature module. Fig. 19 illustrates the overall architecture for building extraction. In the proposed model, residual dense blocks [93] for fusing multiple feature maps to compensate for the spatial loss occurring during the process of feature extraction are adopted. Also, to utilize the interdependencies between the channels of the convolutional features, squeeze-and-excitation (SE) blocks [94] are used at the end of the residual dense block, as shown in Fig. 19 (c). The designed model is trained using the Inria aerial image labeling dataset [72] and compared with other deep-learning models.

4.2.2 INRIA AERIAL IMAGE LABELING DATASET

The Inria aerial image labeling dataset [72] consists of 360 orthorectified aerial images for an area of 810 km^2. The size of each image is 5,000 x 5,000 pixels and the resolution is 0.3 m. The ground truth data has two semantic classes (buildings and background). The dataset covers ten different cities and incorporates various urban landscapes and settlements. It is divided into 155 images for training, 25 images for validation and the remaining 180 images for testing. The regions of the training set are Austin, Chicago, Kitsap County, Western Tyrol and Vienna and the regions of the testing set are Bellingham, Bloomington, Innsbruck, San Francisco and Eastern Tyrol. Inria dataset statistics for training and testing as done here are summarized in Table VI.

TABLE VI

INRIA DATASET STATISTICS [72]

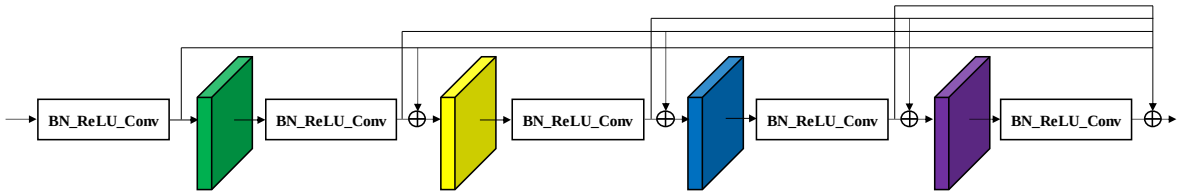
Train	Tiles	Total area
Austin, TX	36	81 km^2
Chicago, IL	36	81 km^2
Kitsap County, WA	36	81 km^2
Vienna, Austria	36	81 km^2
West Tyrol, Austria	36	81 km^2
Total	180	405 km^2

Test	Tiles	Total area
Bellingham, WA	36	81 km^2
San Francisco, CA	36	81 km^2
Bloomington, IN	36	81 km^2
Innsbruck, Austria	36	81 km^2
East Tyrol, Austria	36	81 km^2
Total	180	405 km^2

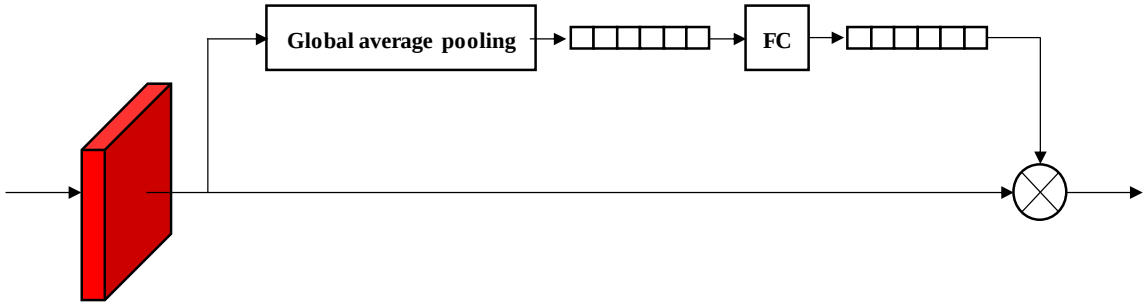
4.2.3 TRAINING SETUP

Our model was implemented based on TensorFlow. We used the Adam optimizer with a learning rate α = 1e-4 and the exponential decay rates for the moment of estimates of β_1= 0.9 and β_2= 0.999. The epsilon and momentum values of the batch normalization layer are 0.99 and 0.001, respectively. As the activation function for hidden layers, the rectified linear unit (ReLU) is used. The segmentation network is trained with binary cross-entropy loss and modified IoU loss between the prediction results and the ground truth. The network takes in patches 448 × 448 in size, and the patches are selected randomly. In this case, 155 images were used for training and 180 images were used for testing. We chose a batch size of 4. The data augmentation methods of rotation, flip, brightness changes, blur,

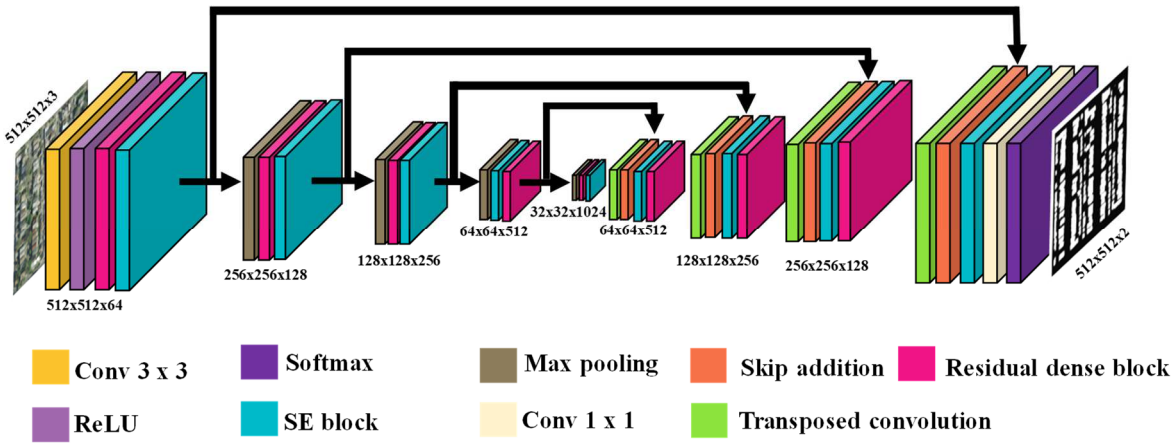
and Gaussian noise were used. The experiments were run on an Intel core 6 i7-7820X CPU at 3.6 GHz with 2 NVIDIA GeForce TITAN Xp GPUs (12 GB).



(a)



(b)



(c)

*BN: Batch normalization, Conv: Convolution, FC: Fully connected layer

Fig. 19. Overall architecture for semantic segmentation. (a) Residual dense block (b) Squeeze and excitation block (c) System architecture for building extraction.

We chose batch size of 4. Data augmentation methods such as rotation, flip, brightness change, blur, and Gaussian noise were used. The experiments were run on Intel core 6 i7-7820X CPU at 3.6 GHz with 2 GPUs of NVIDIA Geforce TITAN Xp (12 GB).

4.2.4 TEST RESULTS AND COMPARISON WITH OTHER ARCHITECTURES

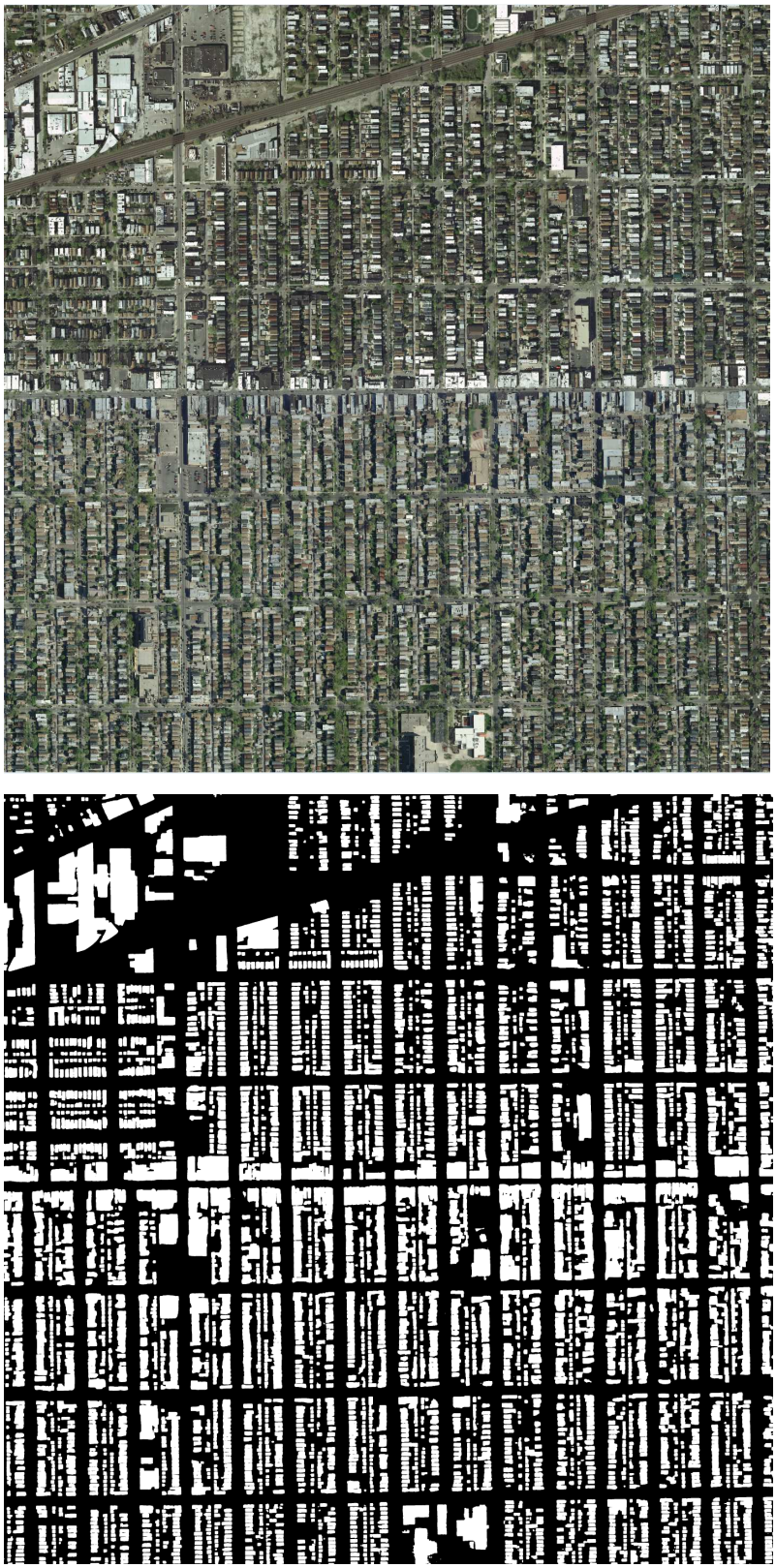
The test results for the Inria test dataset are summarized in Table VII. The designed model achieves an overall IoU score of 79.87. Specifically, it shows higher performance on the San Francisco and East Tyrol images (i.e., for these cities, the proposed model shows an IoU of more than 80.00). We also compare other architectures trained with the Inria labeling dataset and prove that the performance of the designed model is comparable to those of state-of-the-art models. Despite this, however, it has limited generalization power for other cities not included among the training images. Due to the domain gap between the training and test data, the designed model trained by Inria labeling fails to segment the Massachusetts building [80] dataset images not included the Inria training dataset. The relationship between the domain gap and segmentation performance is discussed in detail in the following section. Prediction results are illustrated in Fig. 20.

TABLE VII

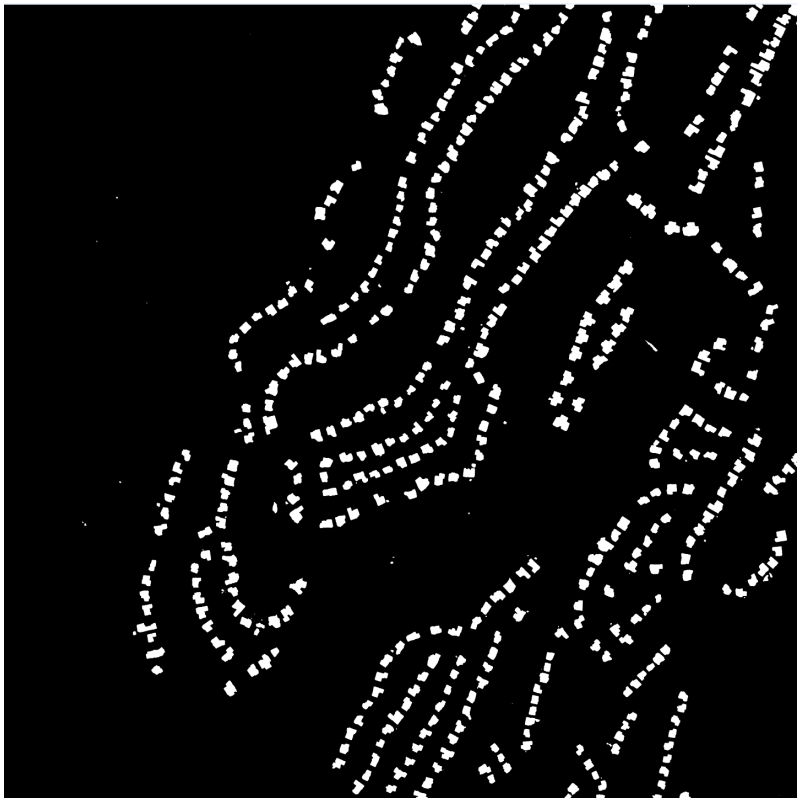
TEST RESULTS ON INRIA AERIAL IMAGE LABELLING TEST SET

Method	Bellingham	Bloomington	Innsbruck	San Francisco	East Tyrol	Overall IoU
Building-A-Net [87]	65.50	66.63	72.59	76.14	71.86	72.36
Dual-resolution U-Nets [95]	70.74	66.06	73.17	73.57	76.06	72.45
Sobel Heuristic Kernel [89]	70.73	69.98	76.74	76.73	79.09	75.33
Koki Takahashi [96]	74.15	75.55	78.62	80.65	80.80	78.80
Proposed model	**73.96**	**79.58**	**78.34**	**81.89**	**81.55**	**79.87**
ICT-Net [97]	74.63	80.80	79.50	81.85	81.71	80.32

(a)

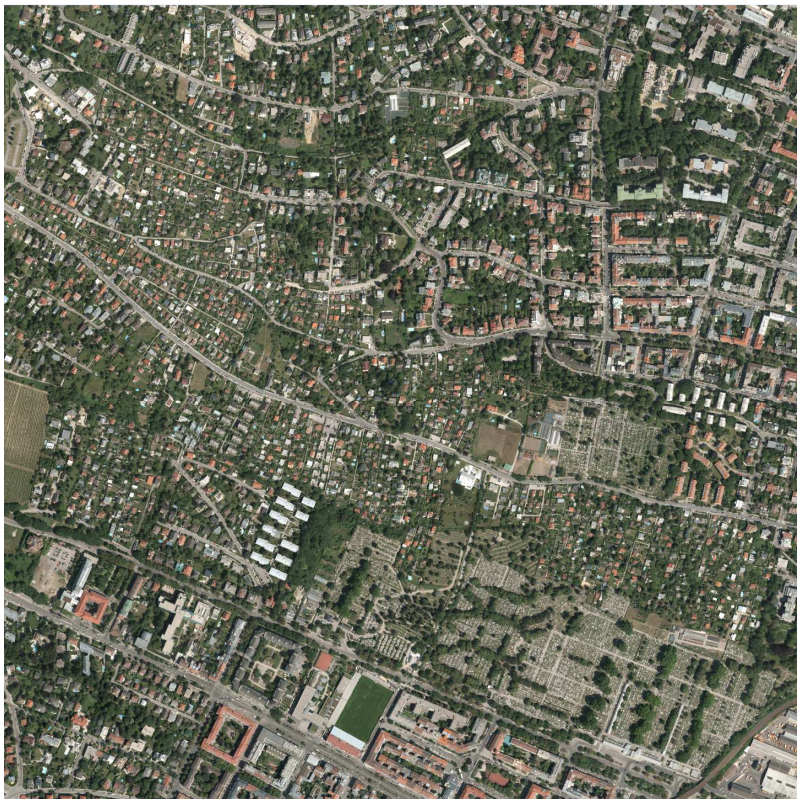


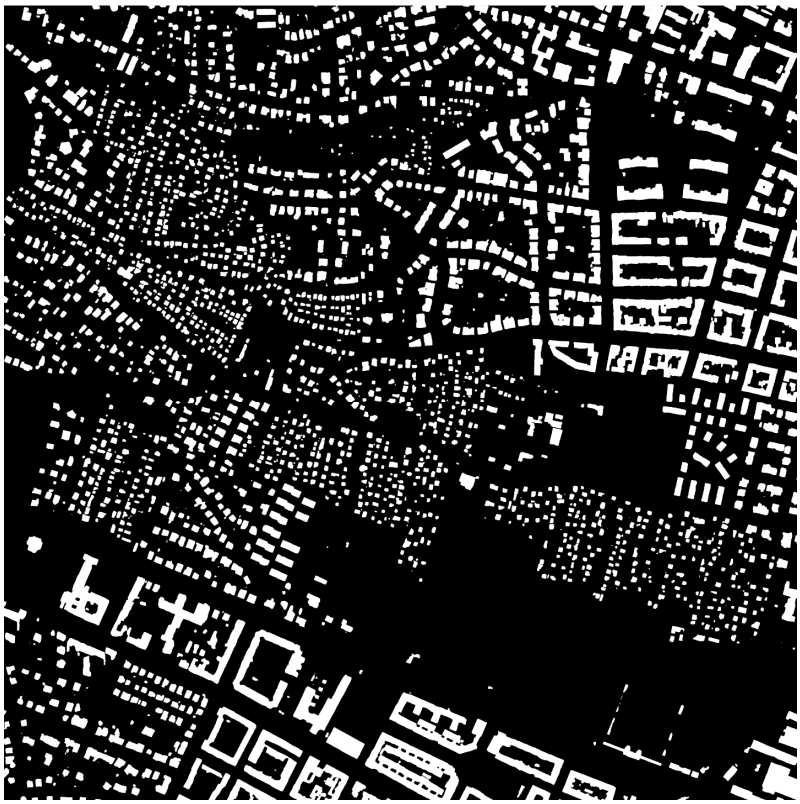
(b)



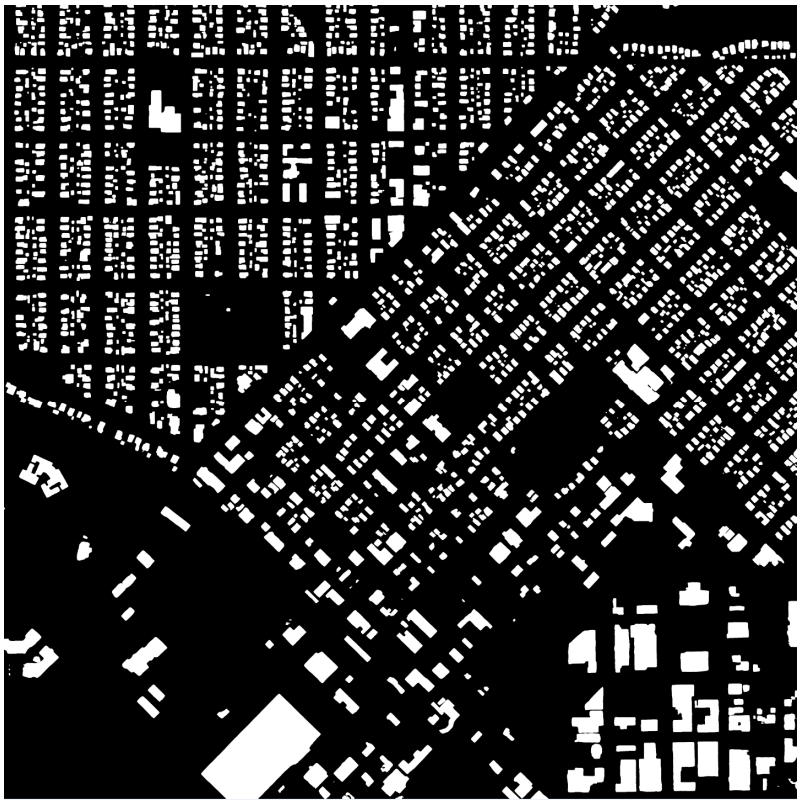
(c)

(d)

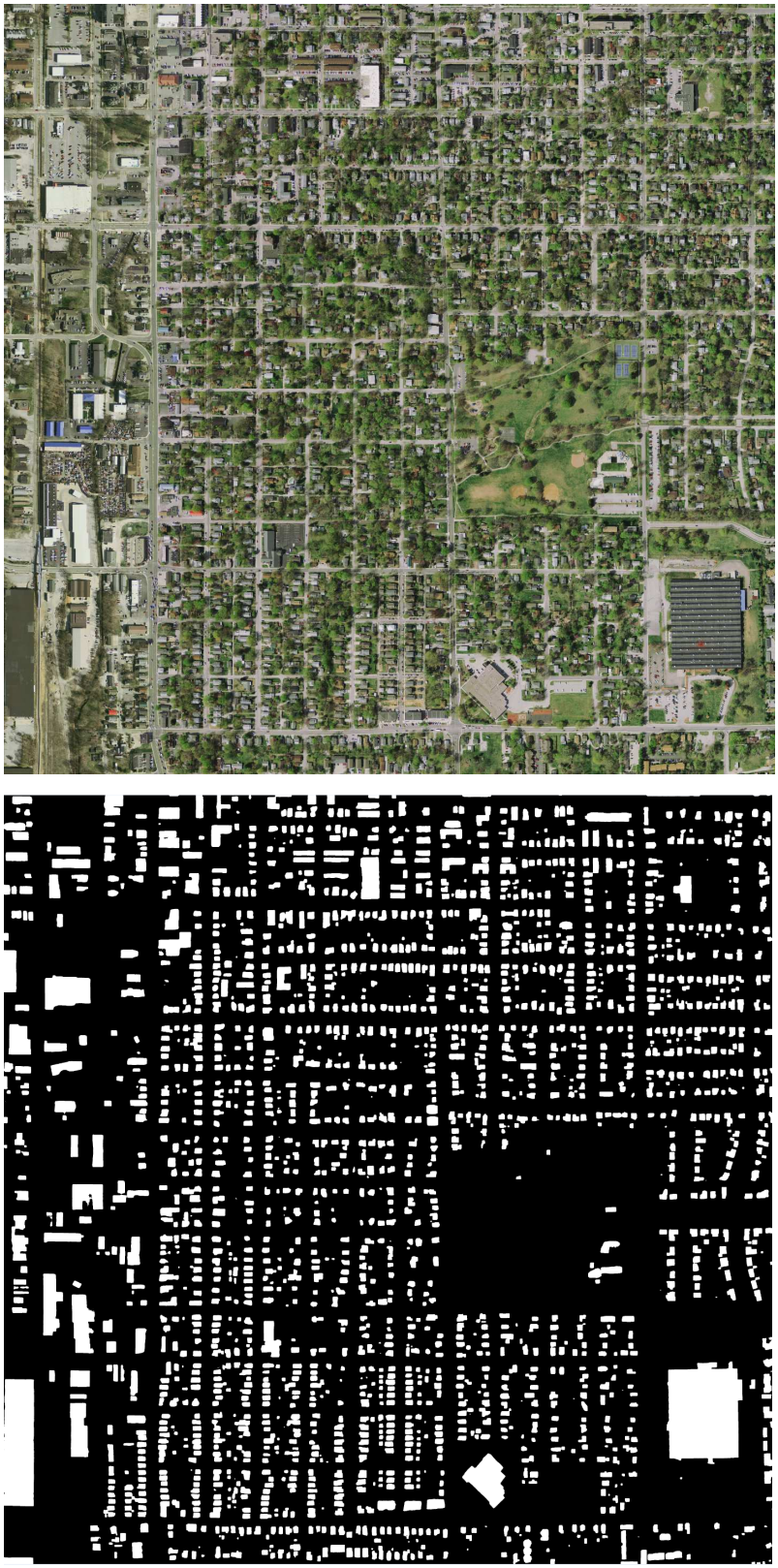




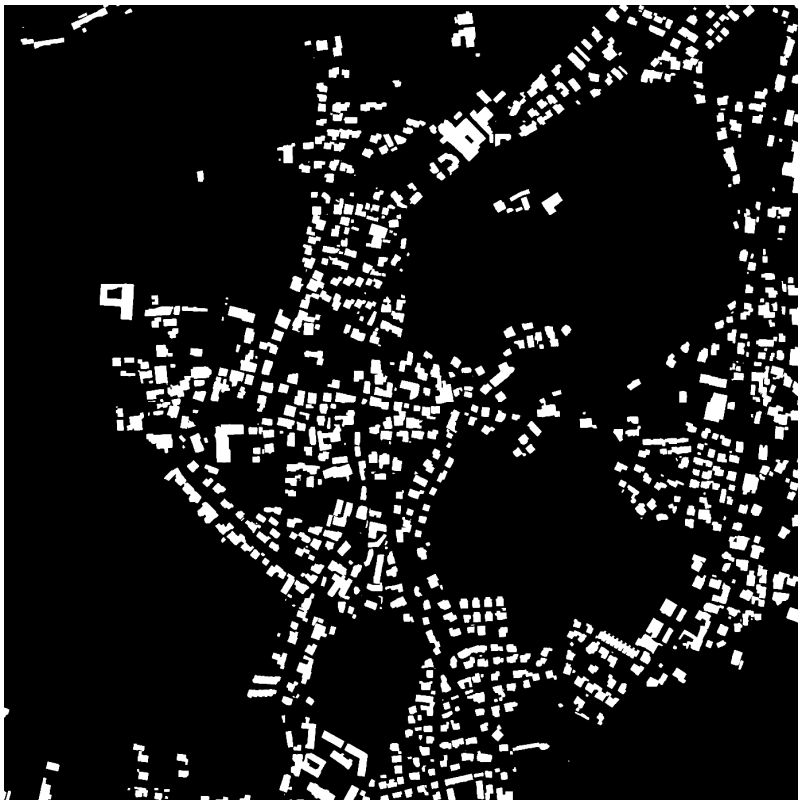
(e)



(f)



(g)



(h)

(i)

(j)

Fig. 20. Prediction results. (a) Austin (b) Chicago (c) Kitsap (d) West Tyrol (e) Vienna (f) Bellingham (g) Bloomington (h) Innsbruck (i) San Francisco (j) East Tyrol

4.3 DOMAIN GAP AND SEGMENTATION PERFORMANCE

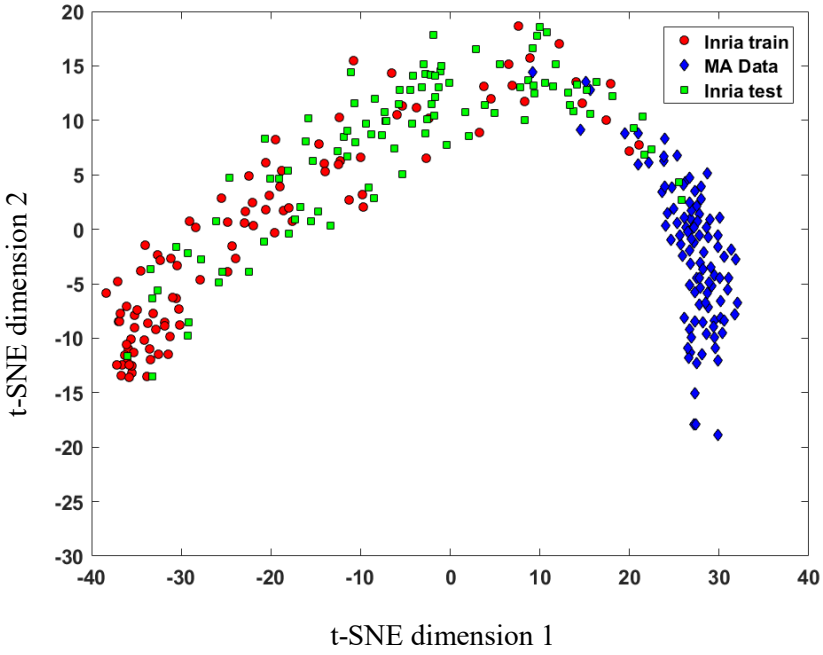
One of the main problems associated with semantic segmentation is the limited generalization, leading to segmentation failures for unseen image domains. This is caused by the domain gap between the training data (source) and the test data (target). For example, the appearance for buildings in aerial images in different cities can vary. Thus, a model trained for one city misclassifies buildings in other cities. In this section, we show the relationship between the domain gap and the segmentation performance for the Massachusetts building dataset.

4.3.1 MASSACHUSETTS BUILDING DATASET

The Massachusetts building dataset [73] consists of aerial images of the Boston area. The total number of images is 151 and the spatial resolution is 1.0 m. Each image has a size of 1,500 x 1,500 pixels for an area of 2.25 km^2. The dataset is split into a training set of 137 images, a test set of 10 images and a validation set of four images. Note that we used 104 training images that do not contain a white background to evaluate the network performance here.

4.3.2 DATA DISTRIBUTION FOR EACH DATASET

We visualize the data distribution for each dataset using t-stochastic neighbor embedding (t-SNE) [98]. Fig. 21 illustrates the data distribution for the Inria and Massachusetts datasets. The Inria training distribution (plotted as the red circle), while close to the Inria test distribution, is far from the Massachusetts dataset. We evaluate the deep-learning model, trained by the Inria training dataset discussed in Section 4.2, using the Massachusetts dataset. The test results are summarized in Table VIII. Note that we referred to the designed model in Section 4.2 as the base model. The test results verify that the building segmentation performance is related to the domain gap between the training and test data. Moreover, as the interval of the data distribution increases, the overall IoU score decreases. Although the IoU for the Inria test dataset, which is relatively close to the training distribution, is 79.87, it dropped



*MA: Massachusetts
Fig. 21. Visualization for each dataset.

to 59.54 for the Massachusetts dataset. To reduce the domain gap between the training and test data, we propose a domain-adaptive transfer attack (DATA) scheme.

4.4 Domain Adaptive Transfer Attack

In the previous section, we discussed the relationship between the domain gap and the performance of the segmentation model. Due to the domain gap, the deep-learning model fails to segment buildings from datasets that are not included in the training data. To solve this problem, we propose segmentation networks based on domain-adaptive transfer attack (DATA) in this section.

TABLE VIII
Test Results on Each Dataset

Method	Inria test	Massachusetts
Base model	79.87	59.54

4.4.1 OVERVIEW OF THE PROPOSED MODEL

The proposed system consists of three modules: an adversarial attack network ***G*** and a discriminator ***D***, and a segmentation network ***S***. There are two goals when training an adversarial attack network. The first is to convert the source to the target domain while maintaining the semantic space. The second is to manipulate the features of the source images to cause the segmentation model ***S*** to make a misclassification. To fulfill these tasks, we use jointly both domain adaptation and an adversarial attack. The sets of the source and target images are denoted as $\{\mathcal{I}_S\}$ and $\{\mathcal{I}_T\} \in \mathbb{R}^{H \times W \times 3}$, respectively, where *H* denotes the height of the images and *W* is the width of the images. The source image I_s is forwarded as the input of adversarial attack model ***G***. Note that the architecture of adversarial attack model ***G*** is identical to that of segmentation model ***S***. Then, segmentation model ***S*** trained by the Inria training dataset, uses the converted image $\boldsymbol{G}(I_s)$ to assign each pixel a semantic label. Because the goal of adversarial attack model ***G*** is to cause segmentation model ***S*** to make a mistake, the segmentation prediction $P_s \in \mathbb{R}^{H \times W \times 2}$, corresponding to $\boldsymbol{S}(\lambda_1 I_s + \lambda_2 \boldsymbol{G}(I_s))$, should be far from the ground truth ***Y***. Moreover, in order to adapt the domain from the source to the target while maintaining the semantic space of the source, we use the weighted summation of the input and converted image, $\lambda_1 I_s + \lambda_2 \boldsymbol{G}(I_s)$, and the target image I_t as the input of discriminator ***D*** to classify which input is from the target. Note that λ_1 and λ_2 are weighted values for the combination of the input and converted image. The overall system architectures are plotted in Fig. 22.

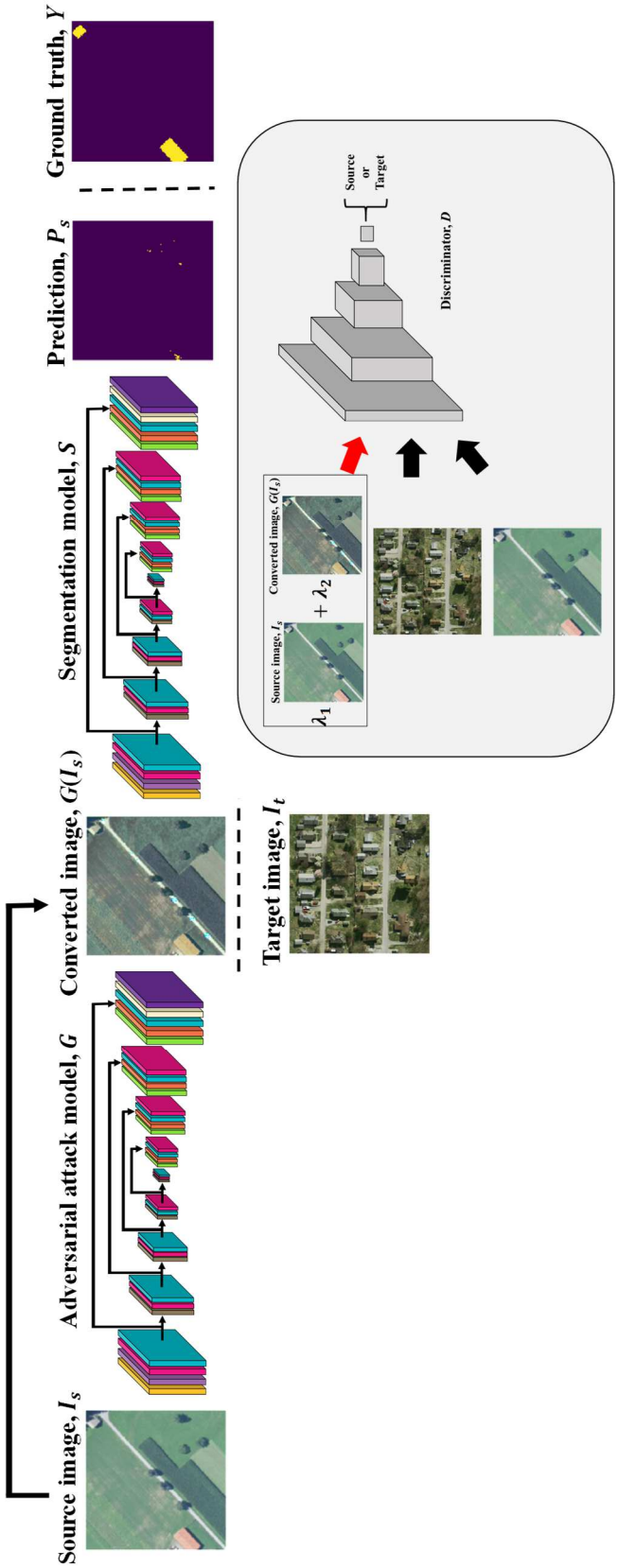


Fig. 22. Overall architecture for domain adaptive transfer attack based segmentation networks (DATA scheme).

4.4.2 OBJECTIVE FUNCTION FOR DOMAIN ADAPTATION

We formulate the objective function for domain adaptation to train the adversarial attack network $\boldsymbol{G}$ as follows,

$$\mathcal{L}_{inv}(I_s) = \sum_{h,w} \|I_s - (\lambda_1 I_s + \lambda_2 \boldsymbol{G}(I_s))\|, \tag{10}$$

where $\mathcal{L}_{inv}(I_s)$ is the invariance loss to prevent $\boldsymbol{G}$ from losing the semantic space, and h and w are correspondingly the height and width of I_s. $\sum_{h,w}\|\cdot\|$ is the L1 loss, and λ_1 and λ_2 are weighted values for the combination of the input and converted image. We constrain the weighted values via $\lambda_1 + \lambda_2 = 1.0$.

To adapt the distribution of the converted images to the target distribution, we introduce the following generative adversarial network:

$$\mathcal{L}_{GAN}(I_s) = \left\|\boldsymbol{D}\left(\lambda_1 I_s + \lambda_2 \boldsymbol{G}(I_s)\right) - 1\right\|^2. \tag{11}$$

Note that adversarial attack model $\boldsymbol{G}$ is trained by $\mathcal{L}_{inv}(I_s)$ and $\mathcal{L}_{GAN}(I_s)$.

4.4.3 OBJECTIVE FUNCTION FOR ADVERSARIAL ATTACK

To cause the segmentation model $\boldsymbol{S}$ to make a mistake, the segmentation prediction P_s should be far from the ground truth $\boldsymbol{Y}$. Thus, we model the adversarial attack function as follows,

$$\mathcal{L}_{atk}(P_s, Y) = -\mathcal{L}_{BCE}(P_s, Y) - \mathcal{L}_{IoU_{new}}(P_s, Y), \tag{12}$$

where $\mathcal{L}_{atk}(P_s, Y)$ is the adversarial attack loss, $\mathcal{L}_{BCE}(P_s, Y)$ is the cross-entropy loss, and $\mathcal{L}_{IoU_{new}}(P_s, Y)$ is the modified IoU loss. The cross-entropy loss $\mathcal{L}_{BCE}(P_s, Y)$ can be written as

$$\mathcal{L}_{BCE}(P_s, Y) = -\sum_{h,w}\sum_{c\in C} Y^{(h,w,c)} \cdot \log\left(P_s{}^{(h,w,c)}\right). \tag{13}$$

We modified the IoU loss [116] for semantic segmentation. When the number of true-positive (TP) pixels is zero (e.g., all pixels are background), the conventional IoU loss computes the IoU as 0 regardless of any false positives (FP) and false negatives (FN). Moreover, if there are relatively few object pixels in a batch, a misclassification for objects by a few pixels causes a large IoU loss. Thus, we simply modified the conventional IoU loss by multiplying the ratio of the union area, as presented below.

$$\begin{aligned}
&\mathcal{L}_{IoU_{new}}(P_s, Y) \\
&\quad = \mathcal{L}_{IoU_{old}}(P_s, Y) \cdot \frac{\sum(P_s + Y - P_s \cdot Y)}{N} \\
&\quad = \left(1 - \frac{\sum(P_s \cdot Y)}{\sum(P_s + Y - P_s \cdot Y)}\right) \cdot \frac{\sum(P_s + Y - P_s \cdot Y)}{N} \\
&\quad = \frac{\sum(P_s + Y - P_s \cdot Y) - \sum(P_s \cdot Y)}{N},
\end{aligned} \tag{14}$$

Here, $\mathcal{L}_{IoU_{old}}(P_s, Y)$ is the conventional IoU loss [99] and *N* denotes the number of pixels.

4.4.4 OBJECTIVE FUNCTION FOR DISCRIMINATOR

The goal of the discriminator ***D*** is to distinguish whether the input is from the source or the target domain. The objective function for the discriminator ***D*** can be formulated as follows:

$$\mathcal{L}_D(I_s, I_t) = \|\boldsymbol{D}(I_s)\|^2 + \|\boldsymbol{D}(I_t) - 1\|^2. \tag{15}$$

4.4.5 TRAINING ADVERSARIAL ATTACK MODEL & DISCRIMINATOR

The adversarial attack model ***G*** is trained to minimize the multi-task loss as given by

$$\mathcal{L}_G(I_s, P_s, Y) = \{(\mathcal{L}_{atk}(I_s, Y),\ \theta_G),\ (\mathcal{L}_{GAN}(I_s), \theta_G),\ (\mathcal{L}_{inv}(I_s), \theta_G)\}, \tag{6}$$

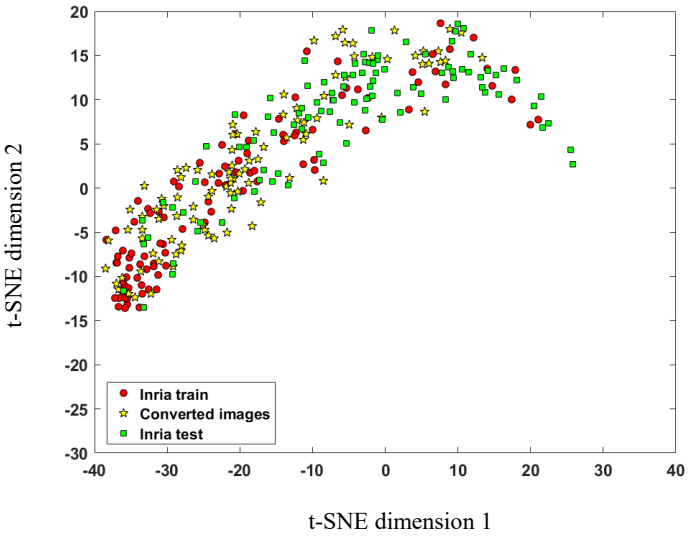
where p is the number of training and q is the switching parameter to determine the training loss for ***G***. Because holding the semantic space for original images is crucial in data augmentation, for segmentation, we set the weight values λ_1 and λ_2 to 0.9 and 0.1, respectively, during the training process. The trained ***G*** converts Inria training images to other cities without losing the semantic space. Fig. 23 shows sample images for the output of ***G***. There are a few differences between the original images and the converted images at the pixel level, as shown in Fig. 23. However, the effectiveness of the DATA

(a)

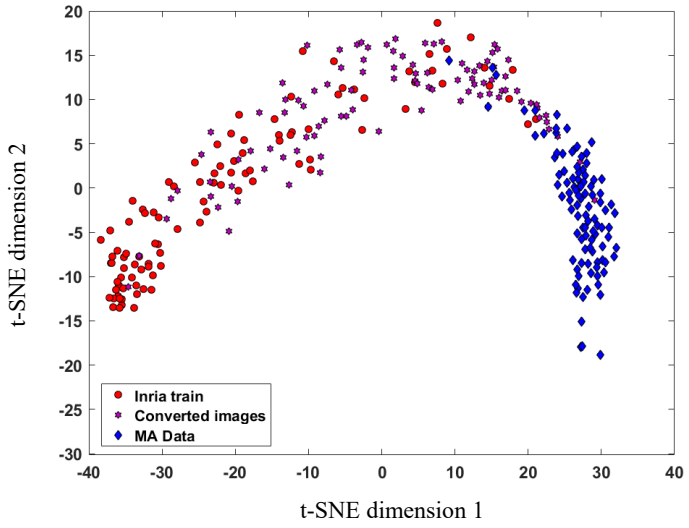
(b)

Fig. 23. The sample images for the output of adversarial attack model ***G***. (a) from Inria train (left) to Inria test (right) (b) from Inria train (left) to Massachusetts (right).

scheme can be verified by t-SNE rather than by the converted images at the pixel level. Data distributions for each dataset are plotted in Fig. 24. The distribution of the converted images for the Inria test dataset is slightly shifted to the Inria test distribution. On the other hand, the distributions of the converted images for the Massachusetts dataset move toward the test distribution. These outcomes indicate that the domain gap is reduced by the DATA scheme. We then utilize the converted images to expand the generalization power of the semantic segmentation network $\boldsymbol{S}$; i.e., both the training data and the converted data are used in training procedure. DATA-based adversary training is discussed in the following section in detail.



(a)



(b)

Fig. 24. Visualization for each dataset with converted images. (a) Inria test dataset (b) Massachusetts dataset

4.5 DATA-based Adversarial Training & Results

From the results of the previous section, the DATA scheme is proven to be able to reduce the domain gap between the training and test data as well as maintain the semantic space. In this section, to expand the generalization of segmentation networks, we utilize the adversary training method and discuss the results.

4.5.1 Adversary Training Setup

We use pre-trained weights for the segmentation model ***S*** as discussed in Section 4.2. The segmentation model ***S*** is trained by both the Inria training dataset and images converted by the generated DATA scheme. Moreover, the training setup (i.e., the learning rate, activation function for hidden layers, optimizers, etc.) is identical to that described in Section II. We introduce k as a switching parameter which determines whether adversarial attack model ***G*** operates or not. If k=5, before the number of training instances reaches 5, the segmentation model ***S*** is trained by Inria training images without adversarial attack model ***G***. When the number of training instances equals a multiple of k, the adversarial attack model ***G*** is operated and the converted images are used for training the segmentation model ***S***. Note that we set the value of k to 2 and 5 for the Inria test and Massachusetts data. The adversary training process is summarized in **Algorithm 1**.

4.5.2 Test Results for Adversary Training

The test results for the segmentation network ***S*** trained using Inria training and the converted images are summarized in Table IX. Because the converted images for each dataset are relatively close to the test distribution as compared to the Inria training data, the generalization of the segmentation network is expanded for unseen images. For the Inria test dataset, the performance of the base model with the DATA scheme is improved by 0.16% (i.e., from 79.87 to 80.00) compared to that without the DATA scheme. In addition, there are relative increments of IoU for the Massachusetts dataset. For the Massachusetts dataset, the base model with the DATA scheme achieves a score of 63.78, which is 7.12% more than the base model without the DATA scheme. From the test results, we conclude that the effectiveness of the DATA scheme is related to the gap between the training and test distribution. While it is improved by 0.16% in terms of the IoU for the Inria test dataset, it achieves a 7.12% improvement in the IoU outcome for the Massachusetts dataset. Fig. 25 compares the prediction results for the base model with and without the DATA scheme. Due to the narrow gap between the training dataset and the test dataset of Inria, the prediction results for the base model with/without the DATA scheme are similar to each other. For the Massachusetts data, however, that the effectiveness of the DATA scheme can easily be verified. While the base model fails to segment buildings in the Massachusetts dataset, the model trained by the DATA scheme assigns building pixels semantic labels.

TABLE IX

TEST RESULTS ON EACH DATASET

Method	Inria test	Massachusetts
Base model	79.87	59.54
Base+DATA	80.00(+0.16%)	63.78(+7.12%)

Algorithm 1 Adversary Training Process

01:**Initialize**: weights of the segmentation model ***S*** as learned weights θ_S trained by Inria training dataset

02: **for** j = 1, …, J **do**

Sample $I_s$uniformly from dataset $\{\mathcal{I}_S\}$ and corresponding ground truth ***Y***

03: **if** $j\%k == 0$ **do**

04: Operate the adversary attack model ***G*** and convert input image ***G***(I_s)

05: Generate the prediction label map P_s as ***S***(***G***(I_s))

06: **else**

07: Do not operate the adversary attack mode ***G***

08: Generate the prediction label map P_s as ***G***(I_s)

09: **end if**

10: **Compute** $\mathcal{L}_{BCE}(P_s, Y)$ and $\mathcal{L}_{IoUnew}(P_s, Y)$

11: **Update** the weights of segmentation model ***S*** with respect to $\mathcal{L}_{BCE}(P_s, Y)$ and $\mathcal{L}_{IoUnew}(P_s, Y)$

12: **end for**

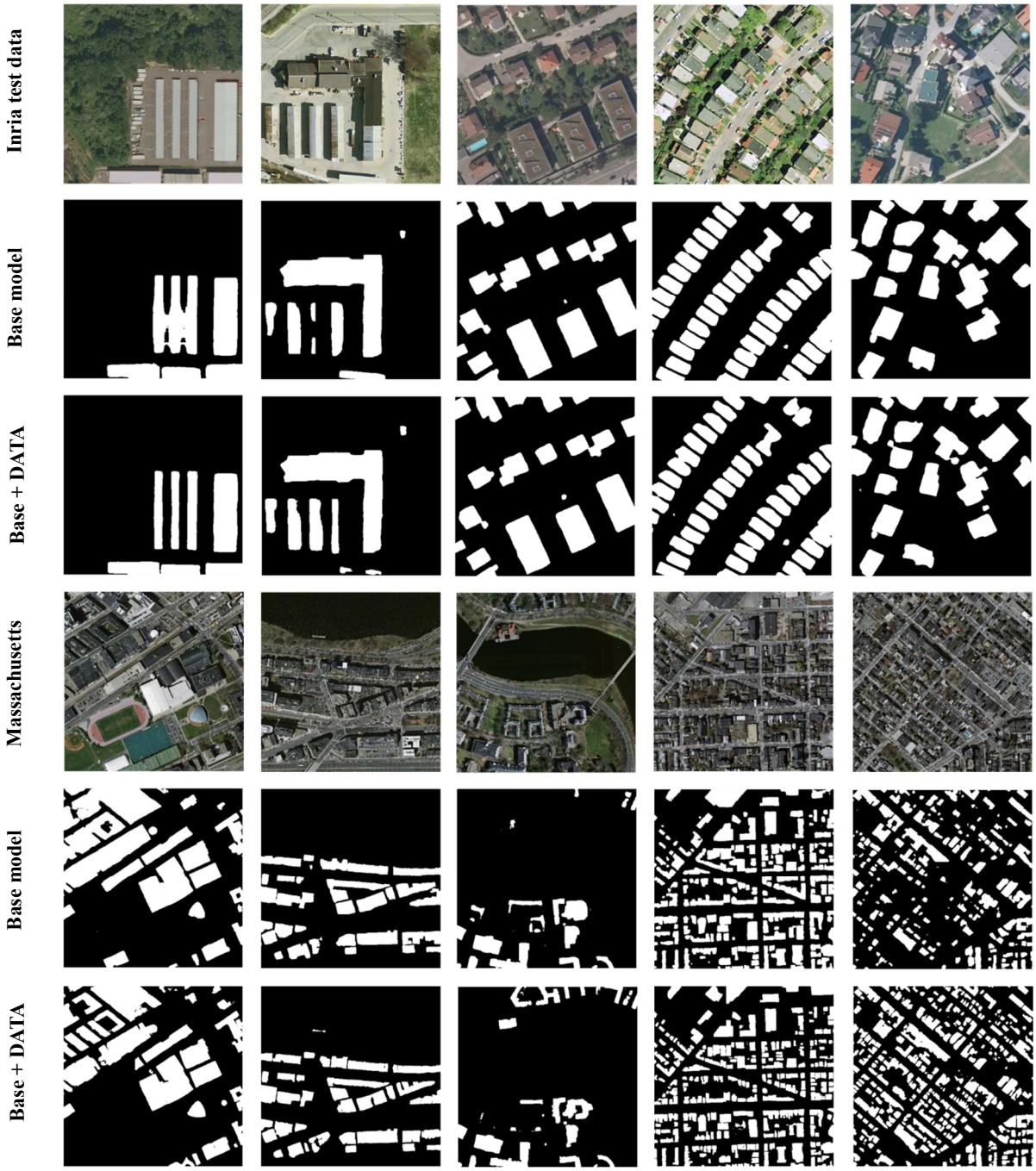


Fig. 25. The results of adversary training for each dataset. From first to third rows, it indicates Inria test dataset. The results for Massachusetts dataset is from fourth to sixth of rows.

4.6 Conclusion

In this section, we investigate the relationship between the domain gap and semantic segmentation. The designed deep learning model can extract buildings from high-resolution images with performance outcomes comparable to those of state-of-the-art methods. However, for other datasets, it fails to segment buildings due to the domain gap between the training and test datasets. We visualized data distributions for each dataset. The distribution of the Inria training dataset is close to that of the Inria test dataset. However, for the Massachusetts dataset, it is far from these datasets, meaning that the domain gap for the Massachusetts dataset is larger than for the Inria test dataset. Thus, the performance of the segmentation model trained by the Inria training dataset deteriorates with the Massachusetts dataset. To reduce the domain gap between the training and test data, we propose segmentation networks based on the domain-adaptive transfer attack (DATA) that jointly utilize both domain transfer and adversarial attack approaches. The proposed method reduces the domain gap by transferring the source images to the target images while maintaining the semantic space. The converted images can be used as a training dataset to expand the generalization power of the segmentation networks. Experimental results show that the generalization power for an unseen dataset is enhanced by the effectiveness of the DATA-scheme-based segmentation model.

5. CONCLUSION REMARKS

In this thesis, applications of machine-learning algorithms are discussed. The first application is wireless cognitive radio networks. We investigated a node placement scheme for machine-learning-based cooperative spectrum sensing in cognitive radio networks. It was verified that the hidden PU problem causes an overlap of the data distributions and reduces the accuracy of machine-learning algorithms. Our approach was to find the conditions that secure sufficient distances between the data distributions. From these conditions, we derived the sensing coverage of a single SN and proposed a method that involved placing several SNs to cover the entire area in which the PU can be placed. Experimental results showed that the hidden PU problem can be mitigated by the proposed node placement scheme.

The second application involved semantic segmentation systems for aerial images. We modeled an enhanced semantic segmentation model capable of better segmentation of multiclass objects from aerial images by exploiting a state-of-the-art CNN-based algorithm. The proposed UNetPPL utilized pyramid pooling layers and skip connections widely to extract multi-scale features of objects. It outperformed other deep-learning models during the segmentation of multiclass objects from aerial images. Test results showed that UNetPPL outperforms FCN, FCNPPL, U-Net and PSPNet in terms of pixel accuracy and the mIOU score for four classes (buildings, roads, water, and background). Moreover, UNetPPL is capable of classifying objects in greater detail (e.g., building intervals), which is significantly advantageous in complex environments.

The final application examined here is domain adaptation for building extraction. We investigated the relationship between the domain gap and semantic segmentation. The designed deep-learning model can extract buildings from high-resolution images with performance levels comparable to those of state-of-the-art methods. However, for other datasets, it fails to segment buildings due to the domain gap between the training and test datasets. We visualized data distributions for each dataset. The distri-

bution of the Inria training dataset is close to that of the Inria test dataset. However, for the Massachusetts dataset, it is far from these datasets, indicating that the domain gap for the Massachusetts data is larger than that for the Inria test dataset. Thus, the performance of the segmentation model trained by the Inria training dataset deteriorates with the Massachusetts dataset. To reduce the domain gap between the training and test data, we proposed segmentation networks based on the domain-adaptive transfer attack (DATA) approach that jointly utilize both domain transfer and adversarial attack techniques. The proposed method reduces the domain gap by transferring the source images to the target images while maintaining the semantic space. The Converted images can be used as a training dataset to expand the generalization power of the segmentation networks. Experimental results demonstrate that the generalization power for unseen datasets is enhanced by the effectiveness of the DATA-scheme-based segmentation model.

REFERENCES

[1] Kelleher, John D., Brian Mac Namee, and Aoife D'arcy. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT Press, 2015.

[2] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1136–1159, 3rd Quart., 2013.

[3] Shin, H. C., Tenenholtz, N. A., Rogers, J. K., Schwarz, C. G., Senjem, M. L., Gunter, J. L., ... & Michalski, M. (2018, September). *Medical image synthesis for data augmentation and anonymization using generative adversarial networks.* In International Workshop on Simulation and Synthesis in Medical Imaging (pp. 1-11). Springer, Cham.

[4] Kamilaris, Andreas, and Francesc X. Prenafeta-Boldu. "Deep learning in agriculture: A survey." *Computers and Electronics in Agriculture* 147 (2018): 70-90.

[5] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless 525 Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[6] J. H. Kim and J. P. Choi, "Sensing Coverage-based Cooperative Spectrum Detection in Cognitive Radio Networks," *IEEE Sensors Journal*, vol. 19, no. 13, pp. 5325-5332, July 1, 2019.

[7] J. H. Kim, H. Lee, S. J. Hong, S. Kim, J. Park, J. Y. Hwang, and J. P. Choi, "Objects Segmentation From High-Resolution Aerial Images Using U-Net With Pyramid Pooling Layers." *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 115–119, Jan. 2019.

[8] I. Akyildiz, B. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40-62, 2011.

[9] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag*., vol. 46, no. 4, Apr. 2008.

[10] K. Zhang, J. Li, and F. Gao. "Machine learning techniques for spectrum sensing when primary user has multiple transmit powers," *in Proc. IEEE Int. Conf. Commun. Sys. (ICCS)*, Macau, China, Nov. 2014, pp.137-141.

[11] Q. Zhang and I. Couloigner, "Accurate centerline detection and line width estimation of thick lines using the radon transform," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 310–316, 2007.

[12] X. Huang, L. Zhang, "Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines", *Int. J. Remote Sens.*, vol. 30, no. 8, pp. 1977-1987, 2009.

[13] M. Mokhtarzade and M. V. Zoej, "Road detection from high-resolution satellite images using artificial neural networks," *International journal of applied earth observation and geoinformation*, vol. 9, no. 1, pp. 32–40, 2007.

[14] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. S. Lempitsky, "Domain-adversarial training of neural networks", *J. Mach. Learn. Res.*, vol. 17, pp. 59:1-59:35, 2016.

[15] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, M. Chandraker, "Learning to adapt structured output space for semantic segmentation", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 7472-7481, Jun. 2018.

[16] S. R. Richter, V. Vineet, S. Roth, V. Koltun, "Playing for data: Ground truth from computer games", *Proc. Eur. Conf. Comput. Vis*, pp. 102-118, 2016.

[17] M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding", *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016.

[18] L. D. Xu, W. He, S. Li, "Internet of things in industries: A survey", *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2233-2243, Nov. 2014.

[19] A. Ahmad, S. Ahmad, M. H. Rehmani, and N. U. Hassan, "A survey on radio resource allocation in cognitive radio sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 888–917, 2nd Quart. 2015.

[20] M. Matinmikko, M. Mustonen, M. Hyhty, T. Rauma, H. Sarvanko, A. Mmmel, "Distributed and directional spectrum occupancy measurements in the 2.4 GHz ISM band", *Proc. 7th ISWCS*, pp. 676-980, 2010-Sep.

[21] R. Chiang, G. Rowe, K. Sowerby, "A Quantitative Analysis of Spectral Occupancy Measurements for Cognitive Radio", *Proc. IEEE 65th Vehicular Technology Conf. (VTC '07-Spring)*, pp. 3016-3020, Apr. 2007.

[22] M. H. Islam, C. L. Koh, S. W. Oh, X. Qing, Y. Y. Lai, C. Wang, Y.-C. Liang, B. E. Toh, F. Chin, G. L. Tan, W. Toh, "Spectrum survey in Singapore: Occupancy measurements and analysis", *Proc. 3rd Int. Conf. CROWNCOM*, pp. 1-7, 2008-May.

[23] V. Valenta, R. Marlek, G. Baudoin, M. Villegas, M. Suarez, F. Robert, "Survey on spectrum utilization in Europe: Measurements analyses and observations", *Proc. 5th Int. Conf. CROWNCOM*, pp. 1-5, 2010-Jun.

[24] T. Li, J. Yuan, Jin, M. Torlak, "Network Throughput Optimization for Random Access Narrowband Cognitive Radio Internet of Things (NBCR-IoT)," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1436–1448, Jun. 2018.

[25] A. Ebrahimzadeh, M. Najimi, S. M. H. Andargoli, A. Fallahi, "Sensor selection and optimal energy detection threshold for efficient cooperative spectrum sensing", *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1565-1577, Apr. 2015.

[26] P. Sofotasios et al., "Energy detection based spectrum sensing over κ−μ and κ−μ extreme fading channels ", *IEEE Trans. Veh. Technol.*, vol. 62, no. 3, pp. 1031-1040, Mar. 2013.

[27] W.-L. Chin, C.-W. Kao, Y. Qian, "Spectrum sensing of OFDM signals over multipath fading channels and practical considerations for cognitive radios", *IEEE Sensors J.*, vol. 16, no. 8, pp. 2349-2360, Apr. 2016.

[28] G. C. Deepak, and K. Navaie. "A low-latency zone-based cooperative spectrum sensing." *IEEE Sensors J.*, vol. 16, no. 15, pp. 6028-6042, Jun. 2016.

[29] G. Ding, "Spectrum sensing in opportunity-heterogeneous cognitive sensor networks: How to cooperate?", *IEEE Sensors J.*, vol. 13, no. 11, pp. 4247-4255, Nov. 2013.

[30] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: The cooperation-processing tradeoff," *Wireless Commun. And Mobile Comput.*, vol. 7, no. 9, pp. 1049–1060, Nov. 2007.

[31] S. Bokharaiee, H. H. Nguyen, E. Shwedyk, "Cooperative spectrum sensing in cognitive radio networks with noncoherent transmission", *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2476-2489, Jul. 2012.

[32] M. Petrova, P. Ma andho andnen, and A. Osuna, "Multi-class classification of analog and digital signals in cognitive radios using support vector machines," in *7th International Symposium on Wireless Communication Systems (ISWCS '10),* Sep. 2010, pp. 986 –990.

[33] Y. Ma, H. Cao, and J. Ma, ‘‘The intrusion detection method based on game theory in wireless sensor network,’’ *in Proc. 1st IEEE Int. Conf. Ubi-Media Comput*., Jul./Aug. 2008, pp. 326–331.

[34] K. M. Thilina, K. W. Choi, N. Saquib, E. Hossain, "Machine learning techniques for cooperative spectrum sensing in cognitive radio networks," *IEEE J. Sel. Areas in Commun*., vol. 31, no. 11, pp. 2209-2221, 2013.

[35] Y.-C. Wang, C.-C. Hu, Y.-C. Tseng, "Efficient placement and dispatch of sensors in a wireless sensor network", *IEEE Trans. Mobile Comput.*, vol. 7, no. 2, pp. 262-274, Feb. 2008.

[36] K. Huang, X. Zhou, "Cutting last wires for mobile communication by microwave power transfer", *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 86-93, Jun. 2015.

[37] Y. S. Cho, J. Kim, W. Y. Yang, C. G. Kang, *MIMO-OFDM Wireless Communications With MATLAB*, NJ, Hoboken:Wiley, Oct. 2010.

[38] J. Hershey and P. Olsen, "Approximating the Kullback Leibler Divergence Between Gaussian Mixture Models," in *Proc. IEEE Int. Conf. on Acous., Speech and Sig. Processing*, Apr. 2007, vol.4, pp. 317 - 320.

[39] F. Nielsen, V. Garcia, "Statistical exponential families: A digest with flash cards", *ArXiv.org:0911.4863*, 2009.

[40] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, San Diego, CA, USA:Academic, 2015.

[41] M. Monemian, M. Mahdavi, M. J. Omidi, "Optimum sensor selection based on energy constraints in cooperative spectrum sensing for cognitive radio sensor networks", *IEEE Sensors J.*, vol. 16, no. 6, pp. 1829-1841, Mar. 2016.

[42] S. Maleki, G. Leus, S. Chatzinotas, B. Ottersten, "To AND or to OR: On energy-efficient distributed spectrum sensing with combined censoring and sleeping", *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4508-4521, Aug. 2015.

[43] L P. P. Singh and R. Garg, “Automatic road extraction from high resolution satellite image using adaptive global thresholding and morphological operations,” *Journal of the Indian Society of Remote Sensing*, vol. 41, no. 3, pp. 631–640, 2013.

[44] B. C. Reed, J. F. Brown, D. VanderZee, T. R. Loveland, J. W. Merchant, D. O. Ohlen, "Measuring phenological variability from satellite imagery", *J. Veg. Sci.*, vol. 5, pp. 703-714, 1994.

[45] D. Chaudhuri, N. K. Kushwaha, A. Samal, R. C. Agarwal, "Automatic building detection from high-resolution satellite images based on morphology and internal gray variance", *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 5, pp. 1767-1779, May 2016.

[46] J. Blundell and D. Opitz, “Object recognition and feature extraction from imagery: The feature analyst® approach,” *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. 4, p. C42, 2006.

[47] A. P. Dal Poz, R. B. Zanin, and G. M. do Vale, “Automated extraction of road network from medium-and high-resolution images,” *Pattern Recognition and Image Analysis, vol*. 16, no. 2, pp. 239–248, 2006.

[48] A M. Izadi and P. Saeedi, “Automatic building detection in aerial images using a hierarchical feature based image segmentation,” *in Proc. IEEE 20th Int. Conf. Pattern Recognit*. IEEE, 2010, pp. 472–475.

[49] A. P. Dal Poz, R. A. Gallis, J. F. da Silva, and ´ E. F. Martins, “Object-space road extraction in rural areas using stereoscopic aerial images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 654–658, 2012.

[50] N. Audebert, B. Le Saux, and S. Lefèvre, "Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017.

[51] X. Zhu et al., "Deep learning in remote sensing: A review", *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8-36, 2017.

[52] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, 2017.

[53] Y. Wei, Z. Wang, and M. Xu, "Road structure refined cnn for road extraction in aerial image," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 709–713, 2017.

[54] E. Maggiori, G. Charpiat, Y. Tarabalka, and P. Alliez, "Recurrent neural networks to correct satellite image classification maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4962–4971, 2017.

[55] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features," *in Proc. IEEE IGARSS*, Jul. 2015, pp. 1873–1876.

[56] F. Hu, G. S. Xia, J. Hu, L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery", *Remote Sens.*, vol. 7, no. 11, pp. 14,680-14,707, 2015.

[57] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation", *Proc. Int. Conf. Medical Image Comput. Comput.-Assisted Intervention*, pp. 234-241, 2015.

[58] G.-S. Xia et al., "AID: A benchmark data set for performance evaluation of aerial scene classification", *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965-3981, Jul. 2017.

[59] G. Xia, X. Bai, J. Ding, Z. Zhu, S. J. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang. DOTA: A large-scale dataset for object detection in aerial images. *In Proc. CVPR*, Jun. 2018, pp. 3794-3983.

[60] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, "Pyramid scene parsing network", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017.

[61] M. Volpi, D. Tuia, "Dense semantic labeling of subdecimeter resolution images with convolutional neural networks", *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 881-893, Feb. 2017.

[62] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, U. Stilla, "Classification with an edge: Improving semantic image segmentation with boundary detection", *ISPRS J. Photogram. Remote Sensing*, vol. 135, pp. 158-172, 2018.

[63] D. Marcos, M. Volpi, B. Kellenberger, D. Tuia, "Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models", *ISPRS J. Int. Soc. Photogramm. Remote Sens.*, to be published. DOI: https://doi.org/10.1016/j.isprsjprs.2018.01.021, [Online].

[64] E. Shelhamer, J. Long, T. Darrell, "Fully convolutional networks for semantic segmentation", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640-651, Apr. 2017.

[65] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. *Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–14

[66] A. Nowaczynski. (2017). *Deep learning for satellite imagery via image segmentation*. Available: https://blog.deepsense.ai/deep-learning-for-satellite-imagery-via-image-segmentation/

[67] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *in Proc. Int. Conf. Mach. Learn*, 2015, pp. 448–456.

[68] M. Haklay and P. Weber, “Openstreetmap: User-generated street maps,” *IEEE Pervasive Comput.*, vol. 7, no. 4, pp. 12–18, 2008.

[69] A. J. Flanagin and M. J. Metzger, “The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information,” *New Media & Society*, vol. 9, no. 2, pp. 319–342, 2007.

[70] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[71] M. Abadi *et al.*, "TensorFlow: A System for Large-Scale Machine Learning," in *OSDI*, 2016, vol. 16, pp. 265-283.

[72] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, “Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark,” *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 3226–3229.

[73] V. Mnih, “Machine learning for aerial image labeling,” Ph.D. dissertation, Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, 2013.

[74] Guo, Yiyou, et al. "Global-Local Attention Network for Aerial Scene Classification." *IEEE Access* (2019).

[75] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[76] S. Chen, H. Wang, F. Xu, Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images", *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806-4817, Jun. 2016.

[77] F. Luus, B. Salmon, F. Van Den Bergh, B. Maharaj, "Multiview deep learning for land-use classification", *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2448-2452, Dec. 2015.

[78] Y. Chen, X. Zhao, X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network", *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381-2392, Jun. 2015.

[79] W. Li, H. Liu, Y. Wang, Z. Li, Y. Jia, G. Gui, "Deep learning-based classification methods for remote sensing images in urban built-up areas", *IEEE Access*, vol. 7, pp. 36274-36284, 2019.

[80] Y. Liu, Y. Zhong, Q. Qin, "Scene classification based on multiscale convolutional neural network", *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7109-7121, Dec. 2018.

[81] Q. Liu, R. Hang, H. Song, Z. Li, "Learning multiscale deep features for high-resolution satellite image scene classification", *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 117-126, Jan. 2017.

[82] W. Diao, X. Sun, X. Zheng, F. Dou, H. Wang, K. Fu, "Efficient saliency-based object detection in remote sensing images using deep belief networks", *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 137-141, Feb. 2016.

[83] X. Chen, S. Xiang, C.-L. Liu, C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks", *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797-1801, Oct. 2014.

[84] Z. Lin, K. Ji, X. Leng, G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images", *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751-755, May 2019

[85] Z. Deng, H. Sun, S. Zhou, and J. Zhao, “Learning deep ship detector in SAR images from scratch” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 4021–4039, Jun. 2019.

[86] L. Mou, X. X. Zhu, RiFCN: Recurrent network in fully convolutional network for semantic segmentation of high resolution remote sensing images, 2018, [Online]. Available: https://arxiv.org/abs/1805.02091.

[87] Li, Xiang, Xiaojing Yao, and Yi Fang. "Building-A-Nets: Robust Building Extraction From High-Resolution Remote Sensing Images With Adversarial Networks." *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3680-3687, Aug. 2018.

[88] I. Goodfellow et al., "Generative adversarial nets", *Proc. Adv. Neural Inf. Process. Syst.*, pp. 2672-2680, 2014.

[89] Hu, Tao, et al. "SOBEL heuristic kernel for aerial semantic segmentation." *2018 25th IEEE International Conference on Image Processing (ICIP). IEEE*, 2018.

[90] S. Gupta, S. G. Mazumdar, "Sobel edge detection algorithm", *Int. J. Comput. Sci. Manag. Res.*, vol. 2, no. 2, pp. 1578-1583, 2013

[91] Volpi, Riccardo, et al. "Generalizing to unseen domains via adversarial data augmentation." *Advances in Neural Information Processing Systems*. 2018.

[92] T. M. Quan, D. G. Hilderbrand, W.-K. Jeong, "FusionNet: A deep fully residual convolutional neural network for image segmentation in connectomics", *arXiv:1612.05360*, 2016.

[93] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, "Residual dense network for image super-resolution", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2472-2481, Jun. 2018.

[94] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, Sep. 2017, [online] Available: https://arxiv.org/abs/1709.01507.

[95] Lu, Kangkang, Ying Sun, and Sim-Heng Ong. "Dual-Resolution U-Net: Building Extraction from Aerial Images." *2018 24th International Conference on Pattern Recognition (ICPR). IEEE*, 2018.

[96] Inria Aerial Image Labeling Benchmark Leader Board. [Online]. Available: https://project.inria.fr/aerial-imagelabeling/leaderboard/

[97] Chatterjee, Bodhiswatta, and Charalambos Poullis. "On Building Classification from Remote Sensor Imagery Using Deep Neural Networks and the Relation Between Classification and Reconstruction Accuracy Using Border Localization as Proxy." *2019 16th Conference on Computer and Robot Vision (CRV). IEEE*, 2019.

[98] L. van der Maaten, G. Hinton, "Visualizing data using t-SNE", *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579-2605, 2008.

[99] M. A. Rahman, Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation", *Proc. 12th Int. Symp. I Adv. Vis. Comput. (ISVC)*, pp. 234-244, Dec. 2016.

요 약 문

기계학습 알고리즘을 이용한 스펙트럼 검출 및 영항분살 시스템 설계

기계학습 알고리즘은 데이터로부터 자동으로 패턴을 추출하는 과정으로 정의된다. 이러한 기계학습 알고리즘의 자동특징추출과정은 입력 및 출력에 대한 시스템이 모델링 되어 있지 않은 분야에서 입/출력 함수를 근사하는 방법으로써 뇌종양 진단을 위한 의료영상처리, 채널추정, 스펙트럼 검출, 항공영상처리 등 많은 분야에서 사용되고 있다. 하지만 기계학습 알고리즘을 적용하더라도 반드시 주어진 문제를 해결할 수 있는 것은 아니다. 인지 무선 네트워크에서 스펙트럼을 검출하기 위한 분류기는 스펙트럼 감지 노드들의 배치에 의존적이며 항공 이미지를 가공하는 분야에서는 주어진 데이터를 처리하는데 적합한 기계학습 알고리즘 구조를 설계해야한다. 본 논문은 주어진 문제를 기계학습 알고리즘을 이용하여 해결한 연구 사례 및 기존 딥러닝 알고리즘의 성능을 개선한 연구 사례를 제시한다.

첫 번째 적용 분야는 인지 무선 통신 기술 분야이다. 무선통신기술이 발달함에 따라 특정 주파수 대역은 포화가 되는 반면 수요가 적은 스펙트럼 대역은 제대로 활용되지 않는 상태이며 이러한 스펙트럼 불균형 문제를 해결하기 위한 방안으로 인지무선통신기술이 제안되었다. 인지무선통신에서 비면허 유저가 면허 유저에게 할당된 스펙트럼을 기회적으로 접속하기 위해 스펙트럼 감지기법이 요구되며 최근 들어 기계학습 알고리즘 기반의 스펙트럼 감지기법이 연구되고 있다. 또한 비면허 유저가 면허 유저의 스펙트럼 점유상태를 오판할 수 있는 은닉유저문제를 해결하기 위해 협력검출기법이 제안되었다. 하지만 협력검출기법을 적용하는데 있어서 여러 환경변수가 존재하며 기존의 연구에서는 이를 고려하지 않으며 기계학습기반 스펙트럼 검출기법 성능과 은닉유저문제 간 관계를 분석하여 은닉유저문제를 해결하기 위한 협력검출기법의 조건을 제시하고 시뮬레이션과 실험결과를 통해 이를 검증하였다.

두 번째 적용 분야는 딥러닝 알고리즘을 이용한 영상분할 시스템 설계이다. 인공위성 이미지는 도시계획, 교통량 관리 등 여러 분야에서 활용된다. 하지만 인공위성 이미지를 활용하기 위해서는 건물, 도로, 수로 등 인공물을 추출하는 과정이 요구된다. 최근 들어, 딥러닝 알고리즘을 이용하여 이미지로부터 특징을 추출하는 연구가 진행되었으며 그 성능이 뛰어남을 입증하였다. 하지만 딥러닝 알고리즘의 성능은 딥러닝 구조에 의존적이므로 목적에 맞는 구조를 설계해야한다. 따라서, 딥러닝 학습을 위한 많은 양의 학습데이터를 제작하였으며 지리정보생성 시스템의 목적에 맞게 딥러닝 구조를 설계하였다. 제안된 시스템은 겹침공간을 기준으로 다른 딥러닝 구조와 비교하여 그 성능이 뛰어남을 입증하였다.

세 번째 적용 분야는 영상분할 시스템의 일반화 능력을 향상시키기 위한 도메인 적응 분야이다. 일반적으로 딥러닝 알고리즘은 주어진 학습데이터와 유사한 테스트 데이터에 대해서는 뛰어난 성능을 보여주지만 학습데이터와 유사하지 않은 테스트 데이터에 대해서는 성능이 상당히 저하되게 된다. 이러한 현상은 도메인 격차로 발생하게 되며 이를 해결하기 위해 도메인 적응형 적대적 공격 기법 (Domain Adaptive Transfer Attack, DATA)을 제안하였다. 제안한 기법은 메사추세츠 건물 데이터를 이용하여 그 효과를 입증하였다.

핵심어: 딥러닝, 인지 무선 네트워크, 스펙트럼 검출, 영상 분할, 도메인 적응