TransLand: An Adversarial Transfer Learning Approach for Migratable Urban Land Usage Classification using Remote Sensing

Yang Zhang, Ruohan Zong, Jun Han, Hao Zheng, Qiuwen Lou, Daniel Zhang, Dong Wang
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
{yzhang42, rzong, jhan5, hzheng3, qlou, yzhang40, dwang5}@nd.edu

***Abstract*—Urban land usage classification is a critical task in big data based smart city applications that aim to understand the social-economic land functions and physical land attributes in urban environments. This paper focuses on a *migratable urban land usage classification* problem using remote sensing data (i.e., satellite images). Our goal is to accurately classify the land usage of locations in a *target city* where the ground truth land usage data is *not available* by leveraging a classification model from a *source city* where such data is available. This problem is motivated by the limitation of current solutions that primarily rely on a rich set of ground-truth data for accurate model training, which encounters high annotation costs. Two important challenges exist in solving our problem: i) the target and source cities often have different urban characteristics that prevent the direct application of a model learned from the source city to the target city; ii) the complex visual features in satellite images make it non-trivial to "translate" the images from the target city to the source city for an accurate classification. To address the above challenges, we develop *TransLand*, an adversarial transfer learning framework to translate the satellite images from the target city to the source city for accurate land usage classification. We evaluate our scheme on the real-world satellite imagery and land usage datasets collected from five different cities in Europe. The results show that TransLand significantly outperforms the state-of-the-art land usage classification baselines in classifying the land usage of locations in a city.**

1. Introduction

In this paper, we develop a principled adversarial transfer learning framework to address the urban land usage classification problem in big data based smart city applications. Urban land usage classification is the classification task to understand the social-economic land functions and physical land attributes in urban environments [1]. Examples of urban land usage classes include agriculture, urban-fabric, forest and green land, and transportation [2]. Accurate land usage classification is critical in various city planning and management tasks (e.g., urban planning, real-estate evaluation, road construction, and green lands protection) [3]. Recent advances in remote sensing provide a pervasive and scalable solution to obtain rich visual insights on urban land usage [4]. In particular, land usage information can be identified by analyzing the surface objects and layouts of the remote sensing data such as satellite images (e.g., an area of high-density buildings often indicates the urban-fabric class) [5].

A significant amount of efforts have been made to address the land usage classification problem in data mining, machine learning, and geographical information systems [5], [6], [7], [8], [9]. Those approaches primarily rely on a rich set of ground-truth data on land usage in the studied area as the training data to extract the relevant image features and build accurate classification models [6]. However, such ground-truth data is not always available due to the high cost of annotations and lack of government supports [10]. For example, less than 1% cities in Europe have publicly available land usage data, where the majority of them are major cities with a population larger than 100,000 [1]. In Austria, for instance, the land usage data are only available in 5 out of 77 cities. Therefore, the lack of ground-truth data presents a fundamental challenge to the urban land usage classification problem.

To address the above challenge, we focus on a *migratable land usage classification* problem using remote sensing data. Our goal is to accurately classify the land usage of locations in a *target city* where the ground-truth land usage data is *not available* by leveraging a classification model learned in a *source city* where such data is available. For example, consider two cities in Germany: *Berlin* and *Dortmund*, both of which are located in the northern Germany area with similar urban characteristics. However, Berlin has the ground-truth land usage data collected and published by European Environment Agency (EEA) but Dortmund does not. In this example, our goal is to learn the land usage classes of locations in Dortmund (i.e., target city) by leveraging the land usage classification model learned in Berlin (i.e., source city). Such a migratable urban land usage classification problem is non-trivial to solve due to several technical challenges we elaborated below.

Model Migration between Disparate Cities. A possible solution to address the migratable land usage classification

1. https://www.eea.europa.eu/data-and-maps/data/urban-atlas/

978-1-7281-0858-2/19/$31.00 ©2019 IEEE
978-1-7281-0858-2/19/$31.00 ©2019 IEEE

problem is to "transfer" (i.e., directly apply or modify) the classification model learned from the source city to classify the land usage of locations in the target city [6], [8]. However, two important limitations exist in current solutions. First, the target and source cities often have different urban characteristics (e.g., city landscape and layouts, architecture styles, and types of topography) that would prevent the direct application of a model learned from the source city to the target city. In particular, a previous study shows a significant classification performance drop when the classification model learned from one city is directly applied to another [6]. Second, a few solutions accommodate the disparity between different cities by re-training the classification model of the source city using a high quality labeled dataset from the target city [8]. However, such a training dataset with ground truth labels of the target city is not available in the migratable land usage classification problem we study in this paper.

Complex Satellite Imagery Data Translation. An alternative way to solve the migratable land usage classification problem is to judiciously "translate" the satellite images from the target city so that they can be better recognized by the classification model learned in the source city. For example, Figure 1 shows a set of satellite images of the same land usage class (i.e., urban-fabric) from two cities. We observe that the translated satellite image (C) shares several key visual features (e.g., high contrast ratio, red rooftop, and clear road outlines) with the image (B) in the source city while keeping the basic content of the original image (A) in the target city. Such visual features are critical for the accurate land usage classification task [11]. However, it remains a challenging task to automatically establish an effective image translation process as shown in Figure 1. In particular, it is not a trivial task for the translation scheme to accurately segment and translate all individual objects from a satellite image with the excessive fine-grained details and complex visual features [12].

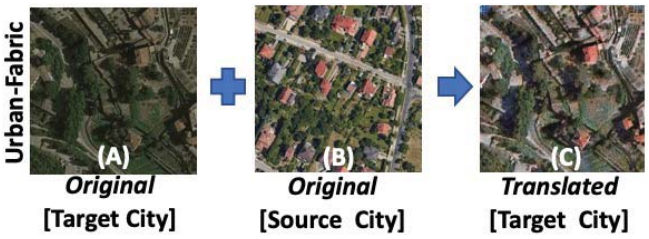


Figure 1. An Example of Satellite Image Translation

To address the above challenges, we develop TransLand, an adversarial transfer learning framework for migratable land usage classification using remote sensing data. TransLand is an *unsupervised* solution that does not require any labeled data from the studied area (i.e., target city) for the classification task. In particular, to address the first challenge, we develop a novel adversarial transfer learning approach to translate the satellite images from the target city to match the urban characteristics of the source city where a well-trained classification model has been built. To address the second challenge, we design a set of deep convolutional neural network architectures and a grouping mechanism to establish accurate object segmentation and reduce the noise introduced in the translated images. To the best of our knowledge, TransLand is the first adversarial transfer learning approach to address the migratable land usage classification problem using remote sensing data. The unsupervised nature of TransLand makes it applicable to the "data drought" problem in similar big data applications where the labeled data from the studied area is unavailable. We evaluate the TransLand framework using the real-world satellite imagery and land usage dataset collected from five different cities in Europe. The results show that TransLand significantly outperforms the state-of-the-art land usage classification baselines in terms of classification accuracy.

2. Related Work

2.1. Smart City Applications

Advanced information and communication technologies (e.g., Internet-of-Things (IoT), edge computing, 5G networks) are integrated to collect the sensing measurements about the urban environment for many smart city applications [13], [14]. Examples of such applications include intelligent traffic risk sensing systems to prevent road accidents [15], smart grid systems to reduce energy consumption and minimize the power shortage [16], urban crowd-sensing systems to report real-time air quality [17], and city-wide crime occurrences prediction to improve urban public safety [18]. Several important challenges exist in the smart city applications. Examples include data quality assurance [19], multi-modal data fusion [20], privacy protection [21], and platform heterogeneity [22]. However, the migratable urban land usage classification using satellite imagery data remains to be an open and challenging problem. In this paper, we develop a novel adversarial transfer learning framework to address this problem by translating the satellite images from the target city to match the urban characteristics of the source city.

2.2. Land Usage Classification

Efforts have been made towards addressing the land usage classification problem in data mining, machine learning, and geographical information systems [5], [6], [7], [8], [9]. For example, Castelluccio *et al.* tuned the deep convolutional network architectures for semantic classification of urban land usage [5]. Albert *et al.* developed a deep learning based approach to identify patterns in urban environments using the satellite imagery data and open-source survey data [6]. Liu *et al.* designed a spatial pyramid pooling framework to extract multi-scale deep features from high resolution satellite images for land use classification [7]. Hu *et al.* leveraged multiple pre-trained deep convolutional neural networks to learn efficient image representations for

land use classification [8]. Yang *et al.* proposed a bag-of-visual-words approach to learn the spatial pyramid match and co-occurrence kernels for land use classification of overhead imagery [9]. However, those approaches cannot be directly applied to our migratable land usage classification problem because they all rely on a rich set of ground-truth data on land usage in the studied area to extract the deep image features and build accurate classification models [10]. In contrast, we develop a novel adversarial transfer learning approach to classify the land usage classes in the areas where the ground-truth data is not available.

2.3. Generative Adversarial Learning

Our work is also related to the generative adversarial learning technique, which has been applied in many areas such as image generation, representation learning, intelligent transportation and remote sensing [23], [24], [25], [26]. For example, Denton *et al.* proposed a deep generative parametric image model to automatically create high quality image samples using the Laplacian pyramid of adversarial networks [23]. Radford *et al.* developed an unsupervised representation framework to learn the general representation of image features through deep convolutional generative adversarial networks [24]. Zhang *et al.* designed a deep transfer learning framework to provide reliable traffic risk estimation via generative adversarial networks [25]. Jia *et al.* proposed a cyclic domain adaptation approach to estimate the land covers under different weather conditions data [26]. To the best of our knowledge, the TransLand is the first adversarial transfer learning approach to solve the migratable urban land usage classification problem.

3. Problem Definition

In this section, we formally define the migratable urban land usage classification problem using the remote sensing data. We first define a few key terms that will be used in the problem statement.

Definition 1. **Source City** (A): We define a source city to be a city where the ground-truth urban land usage data is available.

Definition 2. **Target City** (B): We define a target city to be the studied city of interest where the ground-truth urban land usage data is not available.

Definition 3. **Sensing Cell**: Following a similar procedure in [6], we use the geographic grid to divide both the source and target cities into disjoint sensing cells where each cell represents a subarea of interest. In particular, we define I and J to be the number of cells in the source and target city, respectively. We denote i as the i^{th} grid cell in the source city, and j as the j^{th} grid cell in the target city.

Definition 4. **Satellite Image** (S): We define S to be the satellite images collected from the online map service (e.g., Google Map Platform). We define $S^A = \{S_1^A, S_2^A, ..., S_I^A\}$ and $S^B = \{S_1^B, S_2^B, ..., S_J^B\}$ to represent the set of satellite images collected from the sensing cells in the source and target city, respectively. Figure 1 shows examples of such images for both source and target city. In addition, we define S^{all} to be the complete set of satellite images from both source and target city (e.g., $S^{all} = S^A \cup S^B$).

Definition 5. **City-specific Urban Characteristics**: it refers to the visual features (e.g., contrast ratios, color distributions, surface objects) of the satellite images that are characteristic for a given city. For example, (A) and (B) in Figure 1 have clearly different urban characteristics in terms of contrast ratio (low vs. high) and building colors (dark vs. red building roof). Such differences are often originated from the architecture styles, population density, and topography types of different cities.

Definition 6. **Satellite Image Content**: it refers to the basic objects and their layouts (e.g., buildings, plants, roads, forests) in a satellite image.

Definition 7. **Translated Satellite Image** ($\overline{S}$): we define $\overline{S^B} = \{\overline{S_1^B}, \overline{S_2^B}, ..., \overline{S_J^B}\}$ to be the set of translated satellite images generated by our TransLand scheme for the target city. A translated image $\overline{S_j^B}$ matches the urban characteristics of the source city while preserving the intrinsic content of its original image S_j^B in the target city. Image (C) in Figure 1 is an example of the translated satellite image.

Definition 8. **Land Usage Class** (C): We use C to indicate the land usage class (e.g., agriculture, urban-fabric, forest and green land, and transportation) of a sensing cell in a city. In particular, we define $C^A = \{C_1^A, C_2^A, ..., C_I^A\}$ and $C^B = \{C_1^B, C_2^B, ..., C_J^B\}$ for the source and target city, respectively. In particular, C_i^A indicates the class of land usage in sensing cell i from the source city and C_j^B indicates the class of land usage in sensing cell j from the target city.

The goal of the migratable urban land usage classification problem is to correctly classify the land usage label of all sensing cells in the target city using the land usage classification model learned from the source city. Using the definitions above, our problem is formally defined as:

$$\arg\max_{\widehat{C_j^B}} \Pr(\widehat{C_j^B} = C_j^B \mid S^A, S^B, C^A), \quad \forall 1 \leq j \leq J \tag{1}$$

where $\widehat{C_j^B}$ is the *estimated* land usage class for the sensing cell j in the target city B. This problem is challenging due to the arbitrarily large and complex visual feature space in satellite imagery data and lack of training data of the land usage in the target city. To address this problem, we develop a TransLand scheme in the next section.

4. Solution

4.1. Overview of TransLand Framework

TransLand is an adversarial transfer learning framework that translates the satellite images from the target city to the source city for the land usage classification. It consists of three modules: i) *Cross-city Satellite Image Translation (CSIT)*, ii) *Grouping Mechanism for Noise Reduction (GMNR)*, and iii) *Land Usage Classification Module (LUCM)*. First, the *CSIT* module is the core of TransLand that translates the satellite images from the target city to match the urban characteristics of the source city without requiring any ground-truth land usage data from the target city. Second, the *GMNR* module divides the locations from both source and target city into different subsets based on the learned correlation between the complex visual features from satellite imagery data to further boost the classification accuracy. Finally, the *LUCM* module learns the estimated land usage class of each sensing cell in the target city by feeding the translated satellite images into a classification model trained in the source city. The overview of TransLand is shown in Figure 2.

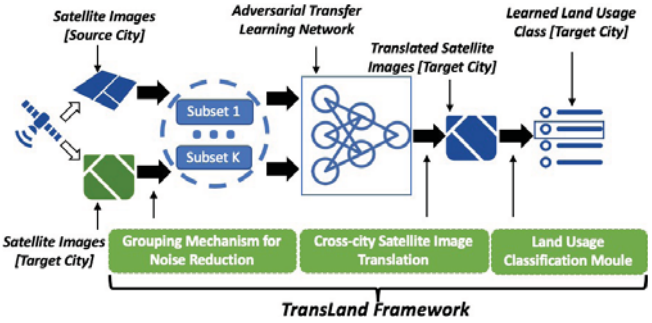


Figure 2. Overview of TransLand framework

4.2. Cross-city Satellite Image Translation (CSIT)

We develop a *cross-city satellite image translation* module that translates the satellite images from the target city to the source city using a novel adversarial transfer learning network. We first define two types of network architectures used in this module:

Definition 9. **Translator Network** (T): we define T_{BA} as a translator network whose function is to translate the satellite images from city B (target city) to match the urban characteristics of city A (source city). In particular, we characterize the function of the translator network below:

$$T_{BA} : S_j^B \in B \rightarrow \overline{S_j^B} \in A \tag{2}$$

where S_j^B and $\overline{S_j^B}$ are the original and translated satellite images in cell j from the target city, respectively.

An example of the translator network is shown in Figure 3. It consists of three components: an image encoder, a residual block component, and an image decoder. The image encoder component has multiple convolutional layers and instance normalization layers. In particular, the convolutional layers extract the content from a satellite image and the instance normalization layers normalize the parameters in the convolutional layers to stabilize the content extraction process. The residual block component has multiple residual blocks [27] to handle the complex task of segmenting individual objects in a satellite image and applying correct urban characteristics translation for the identified objects. The image decoder component has multiple deconvolutional layers and instance normalization layers. In particular, the deconvolutional layers convert the latent representations of translated images generated by the residual blocks to the actual images used for land usage classification.

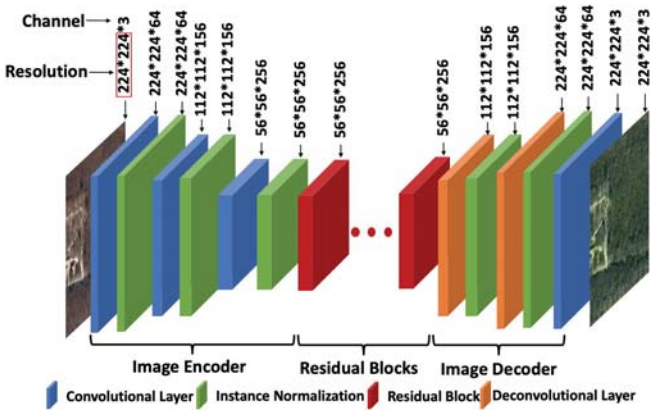


Figure 3. Illustration of Translator Network[2]

Definition 10. **Discriminator Network** (D): We define D_A as a discriminator network to exam whether a satellite image S_x matches the urban characteristics of the satellite images in a given city A or not:

$$D_A : \begin{cases} \mathbf{1} : S_x \in A \\ \mathbf{0} : S_x \notin A \end{cases} \tag{3}$$

where D_A returns true (i.e., "1") if S_x matches the urban characteristics of city A and false (i.e., "0") otherwise.

An example of the discriminator network is shown in Figure 4. It consists of two components: 1) a down-sampling component that reduces the resolution of the satellite image representation (e.g., 224*224 -> 28*28) to identify a critical set of visual features that best represent the urban characteristics of a given city; 2) an output layer component that includes a convolutional layer and an average pooling layer to output the classification results using the learned low-dimensional satellite image representation extracted by the down-sampling component.

2. *Channel* refers to the number of dimensions to represent the semantic information embedded in each pixel (e.g., we have three channels (i.e., RGB) in an image). *Resolution* refers to the size of the matrix that represents the visual information preserved in each layer.

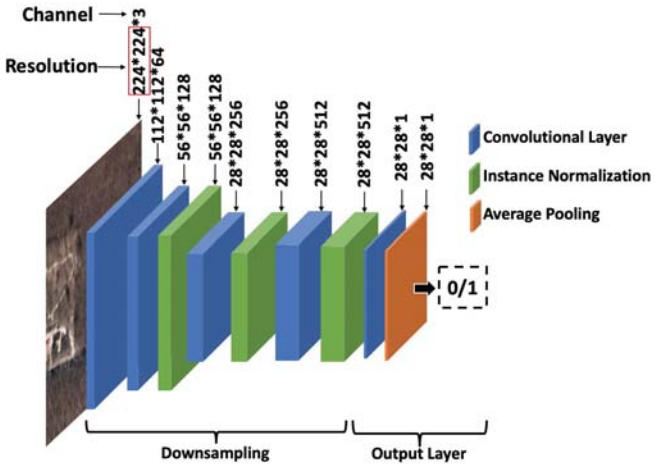


Figure 4. Illustration of Discriminator Network

We define our loss functions for the translator network T_{BA} and the discriminator network D_A to ensure the desired quality of translated satellite images generated from the target city as follows:

$$\begin{aligned}\mathcal{L}_{T_{BA}} &: \min_{T_{BA}} \sum_{S_j^B \in S^B} ||\mathbf{1} - D_A(T_{BA}(S_j^B))||_2 \\ \mathcal{L}_{D_A} &: \min_{T_{D_A}} \sum_{S_j^B \in S^B} ||\mathbf{0} - D_A(T_{BA}(S_j^B))||_2 \\ &+ \sum_{S_i^A \in S^A} ||\mathbf{1} - D_A(S_i^A)||_2\end{aligned} \tag{4}$$

where $\mathcal{L}_{T_{BA}}$ and $\mathcal{L}_{D_A}$ are the loss functions for the translator network T_{BA} and discriminator network D_A, respectively. $||\cdot||_2$ denotes the L2-norm of a given matrix. The rationale of the above loss functions is to let T_{BA} and D_A compete with each other so that T_{BA} can generate high quality translated satellite images from the target city B that match well with the urban characteristics of the source city A. For example, the translator network T_{BA} is effective if $T_{BA}(S_j^B)$ is confirmed positively by D_A (i.e., returning true). On the other hand, the discriminator network D^A is required to identify the poorly translated satellite images from T_{BA} that do not match the urban characteristics of city A. Similarly, we also define the loss functions for the translator network T_{AB} and the discriminator network D_B for target city B (i.e., $\mathcal{L}_{T_{AB}}$ and $\mathcal{L}_{D_B}$). These loss functions are designed to convert the translated image ($T_{BA}(S_j^B)$) back to its original version to verify the quality of the translated satellite images ($T_{AB}(T_{BA}(S_j^B)) \rightarrow S_j^B$). In particular, $\mathcal{L}_{T_{AB}}$ and $\mathcal{L}_{D_B}$ are defined as follows:

$$\begin{aligned}\mathcal{L}_{T_{AB}} &: \min_{T_{AB}} \sum_{S_i^A \in S^A} ||\mathbf{1} - D_B(T_{AB}(S_i^A))||_2 \\ \mathcal{L}_{D_B} &: \min_{D_B} \sum_{S_i^A \in S^A} ||\mathbf{0} - D_B(T_{AB}(S_i^A))||_2 \\ &+ \sum_{S_j^B \in S^B} ||\mathbf{1} - D_B(S_j^B)||_2\end{aligned} \tag{5}$$

While the above adversarial loss functions ensure reasonable satellite image translation from the target city to the source city, it may also lead to a mode collapse problem during the translation process: the translator network may map different satellite images to an identical translated image where the contents of the original images are lost [28]. The collapse problem often happens when the source and target cities differ so significantly in their urban characteristics that the translator network fails to make accurate translations for the images in the target city. To address this problem, we adopt the cycle-consistent loss function [29] that leverages the translation consistency to regularize the image translation process. In particular, we combine the cycle-consistent loss function with the adversarial loss functions defined above (i.e., $\mathcal{L}_{AB}$ and $\mathcal{L}_{BA}$) to derive the final objectives for the translator networks as follows:

$$\begin{aligned}\mathcal{L}_{T_{AB},T_{BA}} : \min_{T_{AB},T_{BA}} &\sum_{S_i^A \in S^A} ||T_{BA}(T_{AB}(S_i^A)) - S_i^A||_1 \\ &+ \sum_{S_j^B \in S^B} ||T_{AB}(T_{BA}(S_j^B)) - S_j^B||_1 \\ &+ \sum_{S_i^A \in S^A} ||\mathbf{1} - D_B(T_{AB}(S_i^A))||_2 \\ &+ \sum_{S_j^B \in S^B} ||\mathbf{1} - D_A(T_{BA}(S_j^B))||_2\end{aligned} \tag{6}$$

where $\mathcal{L}_{T_{AB},T_{BA}}$ indicates the final objective for the translator network T_{AB} and T_{BA}. $||\cdot||_1$ indicates the L1-norm of a given matrix. The idea is to ensure the translator networks can translate the satellite images S_i^A and S_j^B back to the original version after the translation process (i.e.,$T_{BA}(T_{AB}(S_i^A)) \rightarrow S_i^A$ and $T_{AB}(T_{BA}(S_j^B)) \rightarrow S_j^B$). This process ensures the content from the original version of a translated image is preserved to its maximum extent.

Similarly, we combine the loss function $\mathcal{L}_{D_A}$ and $\mathcal{L}_{D_B}$ to obtain the final objective $\mathcal{L}_{D_A,D_B}$ for the discriminator networks as follows:

$$\begin{aligned}\mathcal{L}_{D_A,D_B} : \min_{D_A,D_B} &\sum_{S_j^B \in S^B} ||\mathbf{0} - D_A(T_{BA}(S_j^B))||_2 \\ &+ \sum_{S_i^A \in S^A} ||\mathbf{1} - D_A(S_i^A)||_2 \\ &+ \sum_{S_i^A \in S^A} ||\mathbf{0} - D_B(T_{AB}(S_i^A))||_2 \\ &+ \sum_{S_j^B \in S^B} ||\mathbf{1} - D_B(S_j^B)||_2\end{aligned} \tag{7}$$

Given the final objectives $\mathcal{L}_{T_{AB},T_{BA}}$ and $\mathcal{L}_{D_A,D_B}$, the optimal instances (i.e., $T_{AB}^*, T_{BA}^*, D_A^*, D_B^*$) of all networks can be learned using the Adaptive Moment Estimation (Adam) optimizer [30]. Finally, we leverage T_{BA}^* to generate the translated satellite images for the target city as follows:

$$\overline{S_j^B} = T_{BA}^*(S_j^B), \forall S_j^B \in S^B \tag{8}$$

where S_j^B and $\overline{S_j^B}$ are the original and translated satellite image for sensing cell j in the target city, respectively.

4.3. Grouping Mechanism for Noise Reduction (GMNR)

While the adversarial transfer learning network can establish effective satellite image translations, it can also introduce unexpected noise during the translation process. Figure 5 shows an example of such a case. We observe that the translated image (D) does capture the characteristics of the *forest and green land* class in image (B) from the source city (i.e., changing the dominating color from green to brown). However, the characteristics of *urban-fabric* class in the image (C) from the source city is also mistakenly introduced into image (D) (e.g., the green trees are translated to the red building roofs). This is because the translated images can learn the urban characteristics from any land usage class in source city. Such noise could significantly degrade the classification performance (e.g., classifying (D) as the *urban-fabric* in the above example).

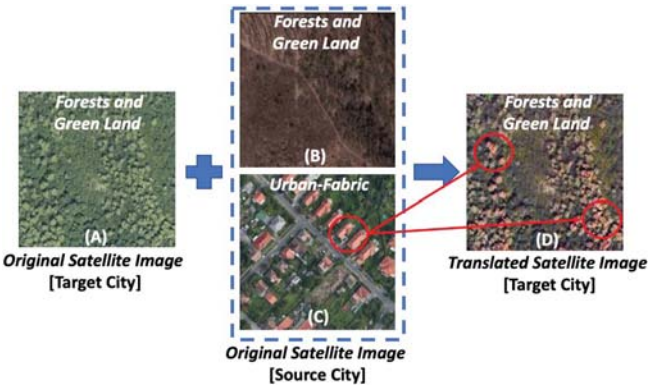


Figure 5. Illustration of Satellite Image Translation Affected by Noise

To address the above problem, we develop an unsupervised grouping mechanism that groups the satellite images with high visual similarities from the source and target city to avoid the noise in the translation process. In particular, we first define the translation subset as follows.

Definition 11. **Translation Subset** (S^{sub}): we divide the complete set of satellite images S^{all} into different subsets (i.e., $S^{sub} = \{S_1^{sub}, S_2^{sub}, ..., S_K^{sub}\}$), where each subset contains the satellite images from both source and target city that share high similarities in terms of the visual features (e.g., dominant color, texture). Such visual features can be extracted through deep neural network based feature extraction methods (e.g., DenseNet [31]).

The above process ensures the satellite images that share similar visual similarities are divided into the same translation subset. Finally, we apply the adversarial transfer learning network introduced in the previous subsection to preform the image translation task for each translation subset. The translated satellite images will be used for the final image classification task discussed in the next subsection.

4.4. Land Usage Classification Module (LUCM)

After obtaining the translated satellite image for each sensing cell in the target city, TransLand performs a multi-class image classification task to classify the land usage for each translated satellite image by leveraging the classification model learned in the source city. Rather than re-inventing the wheel, we adopt the state-of-the-art ImageNet-based conventional neural network tool (i.e., VGG [32]) trained in the source city for the classification task. In particular, we consider the label smoothing during the training process to improve the robustness of the learned classification model [33]. An important parameter in the label smoothing process is the smoothing parameter λ, which controls the trade-off between the model convergence and noise tolerance for the learned classification model. In the evaluation section, we show the performance of Transland in terms of various λ values.

Finally, we summarize the TransLand framework in Algorithm 1. The inputs to the framework are i) the collected satellite images S^A and S^B from the source and target city, respectively, and ii) the land usage data from source city C^A. The outputs are estimated land usage classes for all sensing cells in the target city $\widehat{C^B}$.

Algorithm 1 Summary of the TransLand Framework

input: S^A,S^B, C^A
output:$\widehat{C^B}$
▷ *grouping phase*
set $S^{all} = S^A \cup S^B$
generate S^{sub} (Def. 11) from S^{all} (GMNR in Sec. 4.3)
▷ *translation phase*
for each S_k^{sub} in S^{sub} **do**
  optimize $\mathcal{L}_{T_{AB},T_{BA}}$ (Equation 6) and $\mathcal{L}_{D_A,D_B}$ (Equation 7) to obtain T_{BA}^*
  for each S_j^B in S_k^{sub} **do**
    generate $\overline{S_j^B}$ using T_{BA}^* (Equ. 8)
    add $\overline{S_j^B}$ to $\overline{S^B}$
  end for
end for
▷ *classification phase*
obtain classification model using S^A and C^A (LUCM in Sec. 4.4)
for each $\overline{S_j^B}$ in $\overline{S^B}$ **do**
  apply classification model to generate $\widehat{C_j^B}$
  add $\widehat{C_j^B}$ to $\widehat{C^B}$
end for
output $\widehat{C^B}$

5. Evaluation

In this section, we evaluate the performance of the TransLand scheme using the real-word satellite imagery and land usage datasets collected from five different cities in Europe. We compare the performance of TransLand with the state-of-the-art deep learning frameworks for land usage classification. The results show that TransLand significantly outperforms the baselines in terms of classification accuracy.

5.1. Dataset

We collect real-world satellite imagery and land usage datasets from five different cities in Europe: *Budapest*

(Hungary), *Berlin (Germany)*, *Madrid (Spain)*, *Barcelona (Spain)*, and *Athens (Greece)*. These cities have diversified cultural and architectural characteristics (e.g., architecture styles, urban layouts, and population density), which presents a challenging scenario for the migratable land usage classification problem. We summarize the datasets as follows:

- **Satellite Imagery Dataset**: we collect the publicly available satellite imagery dataset from the five European cities through Google Maps Platform [3]. In particular, the collected images are of 224×224 resolution that provides sufficient visual information for a sensing cell of a city for our application [6].
- **Ground-truth Land Usage Dataset**: we use the *Urban Atlas* dataset published by the European Environment Agency [4] to obtain the ground truth labels of land usage class in the studied cities. Please note that the above ground-truth dataset is used for the purpose of evaluation only and is not always available due to the high cost of annotations and lack of government supports [10].

In our evaluation, we randomly sample 800 satellite images that belong to four widely adopted urban land usage classes [5] for our experiments: *agriculture*, *urban-fabric*, *forest and green land*, and *transportation* as shown in Figure 6.

Figure 6. Examples of Land Usage Classes

5.2. Baselines and Metrics

We compare TransLand with several state-of-the-art deep learning frameworks that are widely used in the previous literature for land usage classification. To ensure the fairness of comparison, the inputs to all compared schemes are set to be same (i.e., satellite imagery data from both source and target city and the training data on land usage from the source city only.)

- **InceptionResNetV2 [34]**: it is a popular deep neural network framework that combines inception architectures with residual connections to accelerate the training process of the land usage classification task.
- **InceptionV3 [35]**: a scalable deep learning architecture that leverages convolutions factorization and regularization to ensure the effectiveness of the learned classification model.
- **DenseNet [31]**: a densely connected convolutional network that imposes a novel concatenation mechanism to ensure the classification accuracy while assuaging the vanishing-gradient problem. In particular, we consider all three variations of DenseNet with different network depths in our experiments (i.e., DenseNet121, DenseNet169, DenseNet201).
- **VGG [32]**: it is a very deep convolutional network architecture for large-scale image classification tasks by employing multiple deep convolutional layers to improve the classification accuracy.
- **MobileNet [36]**: a lightweight deep neural network that utilizes multiplier operations to improve the computational efficiency in its classification model.
- **NasNet [37]**: a transfer learning based neural network image classification model that dynamically adapts convolutional architectures to achieve the desired classification accuracy across different image classification tasks.

In our experiments, all baselines are pre-trained on ImageNet [38] and fine-tuned on the ground-truth land usage dataset. In our evaluation, we also consider the baseline *Grouping* that leverages the translation subsets generated by the grouping mechanism (discussed in Section 4.3) for land usage classification (i.e., the land usage class of a satellite image from the target city is considered to be the same as the class of majority satellite images from the source city in that subset). In addition, the *Random* baseline estimates the land usage class of a sensing cell by randomly picking a class label from all possible land usage classes.

To evaluate the performance of all compared schemes, we adopt four representative metrics that are commonly used for multi-class classification problem. In particular, we use i) *Mirco-F1* and *Marco-F1* [39]; ii) *Cohen's kappa Score ($\mathcal{K}$-Score)* [40]; iii) *Matthews Correlation Coefficient (MCC)* [41]. Intuitively, a higher *Micro-F1*, *Marco-F1*, *MCC*, or *$\mathcal{K}$-Score* indicates a better classification performance. In our experiments, all hyper-parameters are optimized using the Adam optimizer [30]. All reported results are averaged over 50 experiments.

5.3. Evaluation Results

In the first set of experiments, we evaluate the performance of all compared schemes over different target cities for a given source city. In particular, we select the source city as *Athens* and vary the target cities to be *Madrid*, *Berlin*, and *Barcelona*. The three target cities are observed to have distinct urban characteristics (e.g., architecture styles, city layouts, and topography types). The evaluation results are presented in Table 1. We observe that the TransLand scheme consistently outperforms the baselines on all metrics across different target cities. For example, the performance gains of TransLand over the best-performing baseline (i.e., VGG) with *Madrid* as the target city on Micro-F1, Marco-F1, $\mathcal{K}$-Score, and MCC are 6.52%, 6.67%, 8.71%, and 8.43%, respectively. Such performance gain mainly comes

3. https://developers.google.com/maps/documentation/javascript/get-api-key

4. https://www.eea.europa.eu/data-and-maps/data/urban-atlas/

5. https://land.copernicus.eu/user-corner/technical-library/urban-atlas-mapping-guide

Table 1. EVALUATION RESULTS BY VARYING TARGET CITIES

	Athens-> Madrid				Athens -> Berlin				Athens -> Barcelona			
Algorithm	Mirco-F1	Macro-F1	$\mathcal{K}$-Score	MCC	Mirco-F1	Macro-F1	$\mathcal{K}$-Score	MCC	Mirco-F1	Macro-F1	$\mathcal{K}$-Score	MCC
Random	0.2505	0.2493	0.0007	0.0004	0.2522	0.2511	0.0029	0.0028	0.2520	0.2508	0.0027	0.0025
Grouping	0.3750	0.3580	0.1667	0.1735	0.4063	0.4015	0.2083	0.2099	0.3375	0.3134	0.1167	0.1211
InceptionResNetV2	0.4233	0.3538	0.2311	0.2743	0.4836	0.4539	0.3115	0.3294	0.4071	0.3610	0.2094	0.2383
InceptionV3	0.4635	0.3806	0.2846	0.3329	0.4997	0.4519	0.3329	0.3632	0.4437	0.3943	0.2583	0.2997
DenseNet121	0.4067	0.2951	0.2089	0.2878	0.4740	0.4084	0.2986	0.3797	0.3952	0.3373	0.1936	0.2441
DenseNet169	0.4057	0.3037	0.2076	0.3007	0.4436	0.3481	0.2581	0.3402	0.3562	0.2809	0.1416	0.2190
DenseNet201	0.2744	0.1491	0.0325	0.1070	0.2970	0.1965	0.0627	0.1519	0.2760	0.1579	0.0347	0.1052
VGG	0.4880	0.4829	0.3172	0.3219	0.4063	0.4027	0.2085	0.2109	0.3720	0.3804	0.1626	0.1627
MobileNet	0.2849	0.2052	0.0456	0.0774	0.4034	0.3092	0.2045	0.2773	0.3186	0.2367	0.0915	0.1391
NASNet	0.4395	0.3922	0.2527	0.2734	0.4376	0.3854	0.2502	0.2793	0.4208	0.3814	0.2277	0.2426
TransLand	**0.5532**	**0.5496**	**0.4043**	**0.4062**	**0.5178**	**0.4698**	**0.3571**	**0.3712**	**0.4797**	**0.4487**	**0.3063**	**0.3156**

Table 2. EVALUATION RESULTS BY VARYING SOURCE CITIES

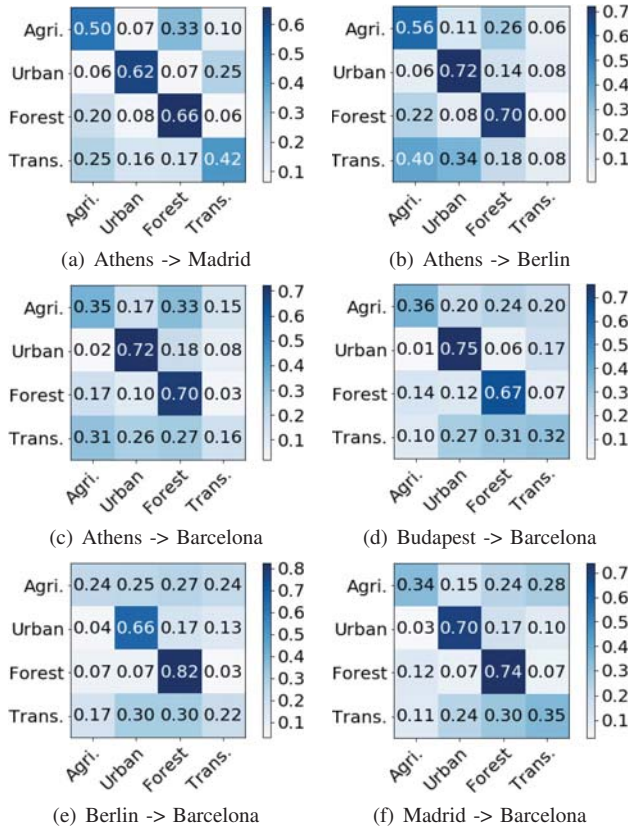
	Budapest -> Barcelona				Berlin-> Barcelona				Madrid -> Barcelona			
Algorithm	Mirco-F1	Macro-F1	$\mathcal{K}$-Score	MCC	Mirco-F1	Macro-F1	$\mathcal{K}$-Score	MCC	Mirco-F1	Macro-F1	$\mathcal{K}$-Score	MCC
Random	0.2508	0.2496	0.0011	0.0018	0.2513	0.2501	0.0018	0.0010	0.2523	0.2512	0.0031	0.0014
Grouping	0.3999	0.3941	0.1999	0.2021	0.4438	0.4075	0.2583	0.2666	0.3188	0.2302	0.0917	0.1071
InceptionResNetV2	0.4251	0.3861	0.2334	0.2481	0.4496	0.4129	0.2661	0.2844	0.4717	0.4501	0.2956	0.3029
InceptionV3	0.4177	0.3994	0.2236	0.2294	0.4584	0.4061	0.2779	0.2979	0.4178	0.3703	0.2238	0.2414
DenseNet121	0.3645	0.3467	0.1526	0.1674	0.3735	0.3450	0.1646	0.1728	0.3577	0.3294	0.1436	0.1713
DenseNet169	0.3560	0.3305	0.1413	0.1531	0.3715	0.3311	0.1620	0.1728	0.3852	0.3672	0.1803	0.2116
DenseNet201	0.2976	0.2523	0.0635	0.0701	0.2931	0.2440	0.0575	0.0667	0.2821	0.2041	0.0428	0.0817
VGG	0.4340	0.4212	0.2453	0.2483	0.4586	0.4415	0.2781	0.2808	0.4633	0.4360	0.2845	0.2905
MobileNet	0.3170	0.2387	0.0893	0.1132	0.3580	0.2902	0.1440	0.1662	0.3390	0.3143	0.1187	0.1280
NASNet	0.3774	0.3631	0.1698	0.1812	0.4555	0.4211	0.2740	0.2900	0.4209	0.3709	0.2278	0.2484
TransLand	**0.5218**	**0.5063**	**0.3625**	**0.3686**	**0.4826**	**0.4471**	**0.3101**	**0.3209**	**0.5278**	**0.5122**	**0.3705**	**0.3784**

from the fact that TransLand judiciously learns the urban characteristics of the source city and translates the satellite images from the target city through a principled adversarial transfer learning network. In particular, the translated satellite images successfully capture the key visual features of the satellite images from the source city without losing the basic content of the original images.

In the second set of experiments, we evaluate the performance of all schemes over different source cities for given a target city. In particular, we set the target city as *Barcelona* and vary the source cities to be *Budapest*, *Berlin*, and *Madrid*. Our objective here is to evaluate whether TransLand and the baselines are capable of providing reliable land usage classification results across the classification models learned from different source cities. The evaluation results are presented in Table 2. We observe that TransLand continues to outperform all baselines across different source cities. For example, the performance gains achieved by TransLand compared to the best-performing baseline with *Budapest* as the source city on Micro-F1, Marco-F1, $\mathcal{K}$-Score, and MCC are 8.78%, 8.51%, 11.72% and 12.03%, respectively. Such performance gains demonstrate the effectiveness of TransLand in translating the satellite images to provide accurate land usage classification services by leveraging classification models trained in different source cities.

In the third set of experiments, we study the per-class land usage classification performance of the TransLand scheme by presenting the classification confusion matrix for each source and target city combination in the above experiments. The results are shown in Figure 7. We observe that our scheme is able to achieve high accuracy when classifying the *urban-fabric* and *forest and green land* across different source and target city combinations. For example, the classification accuracy achieved by TransLand are 0.75 for urban-fabric in Budapest -> Barcelona and 0.82 for forest and green land in Berlin -> Barcelona. Such a high classification accuracy in some critical urban land usage classes demonstrates the promising use cases of TransLand in smart city applications (e.g., effective urban planning, green land protection). We also observe that the *agriculture* class is often mis-classified as the *forest and green land* class in several source and target city combinations. This is mainly because those two classes share some similar visual characteristics (e.g., dominant color, object texture),

which provides fuzzy and insufficient visual evidences for TransLand to perform accurate classification tasks.



(a) Athens -> Madrid (b) Athens -> Berlin

(c) Athens -> Barcelona (d) Budapest -> Barcelona

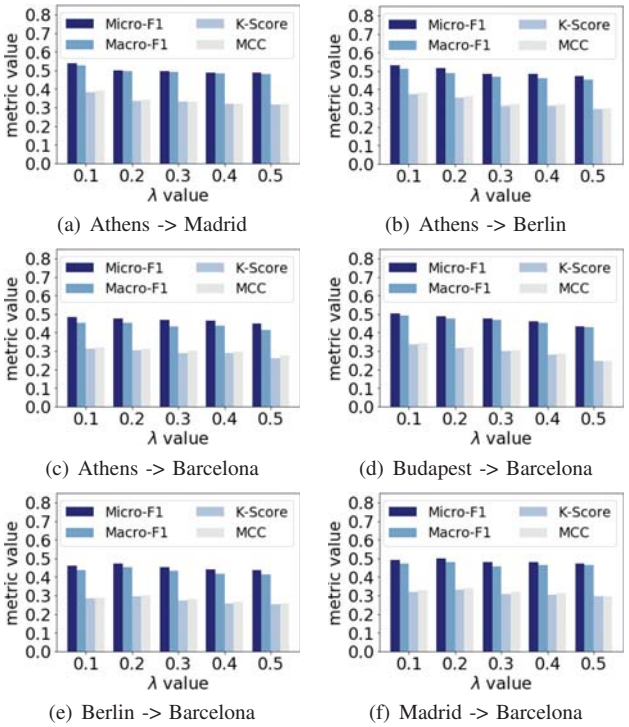
(e) Berlin -> Barcelona (f) Madrid -> Barcelona

Figure 7. Confusion Matrices of TransLand Scheme

Finally, we study the robustness of TransLand model by varying the values of the parameters in our model. An important parameter in our model is the smoothing parameter λ (defined in Section 4.4). This parameter controls the trade-off between the model convergence and noise tolerance in TransLand model. The results are shown in Figure 8. We observe that the performance of our scheme is relatively stable as the value of λ changes across different source and target city combinations.

6. Conclusion

This paper develops a TransLand framework to solve the migratable land usage classification problem using remote sensing data. Our framework addresses two critical challenges, namely *model migration between disparate cities* and *complex satellite imagery data translation*. In particular, we develop a principled adversarial transfer learning framework to effectively translate the satellite images from the target city to the source city for the land usage classification. The evaluation results on the real world case studies from five different cities in Europe demonstrate that TransLand achieves significant performance gains compared to the state-of-the-art land usage classification baselines in accurately classifying the land usage of locations in a city. The authors believe TransLand provides useful insights to address similar "data drought" problems in other big data applications.



(a) Athens -> Madrid (b) Athens -> Berlin

(c) Athens -> Barcelona (d) Budapest -> Barcelona

(e) Berlin -> Barcelona (f) Madrid -> Barcelona

Figure 8. Robustness of TransLand Scheme

Acknowledgement

This research is supported in part by the National Science Foundation under Grant No. CNS-1845639, CNS-1831669, Army Research Office under Grant W911NF-17-1-0409. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] D. Lu and Q. Weng, "Use of impervious surface in urban land-use classification," *Remote Sensing of Environment*, vol. 102, no. 1-2, pp. 146–160, 2006.

[2] E. Montero, J. Van Wolvelaer, and A. Garzón, "The european urban atlas," in *Land use and land cover mapping in Europe*. Springer, 2014, pp. 115–124.

[3] A. O'sullivan, *Urban economics*. McGraw-Hill/Irwin, 2007.

[4] J. Dash and B. O. Ogutu, "Recent advances in space-borne optical remote sensing systems for monitoring global terrestrial ecosystems," *Progress in Physical Geography*, vol. 40, no. 2, pp. 322–351, 2016.

[5] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," *arXiv preprint arXiv:1508.00092*, 2015.

[6] A. Albert, J. Kaur, and M. C. Gonzalez, "Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2017, pp. 1357–1366.

[7] Q. Liu, R. Hang, H. Song, and Z. Li, "Learning multi-scale deep features for high-resolution satellite image classification," *arXiv preprint arXiv:1611.03591*, 2016.

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

[9] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. ACM, 2010, pp. 270–279.

[10] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, 2016.

[11] Z. Murez, S. Kolouri, D. Kriegman, R. Ramamoorthi, and K. Kim, "Image to image translation for domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4500–4509.

[12] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2849–2857.

[13] P. Neirotti, A. De Marco, A. C. Cagliano, G. Mangano, and F. Scorrano, "Current trends in smart city initiatives: Some stylised facts," *Cities*, vol. 38, pp. 25–36, 2014.

[14] D. Wang, B. K. Szymanski, T. Abdelzaher, H. Ji, and L. Kaplan, "The age of social sensing," *Computer*, vol. 52, no. 1, pp. 36–45, 2019.

[15] Y. Zhang, Y. Lu, D. Zhang, L. Shang, and D. Wang, "Risksens: A multi-view learning approach to identifying risky traffic locations in intelligent transportation systems using social and remote sensing," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 1544–1553.

[16] B. Morvaj, L. Lugaric, and S. Krajcar, "Demonstrating smart buildings and smart grid features in a smart energy city," in *Proceedings of the 2011 3rd international youth conference on energetics (IYCE)*. IEEE, 2011, pp. 1–8.

[17] Y. Zhang, D. Zhang, N. Vance, and D. Wang, "Optimizing online task allocation for multi-attribute social sensing," in *2018 27th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2018, pp. 1–9.

[18] H.-W. Kang and H.-B. Kang, "Prediction of crime occurrence from multi-modal data using deep learning," *PloS one*, vol. 12, no. 4, p. e0176244, 2017.

[19] D. Wang, M. T. Amin, S. Li, T. Abdelzaher, L. Kaplan, S. Gu, C. Pan, H. Liu, C. C. Aggarwal, R. Ganti *et al.*, "Using humans as sensors: an estimation-theoretic perspective," in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*. IEEE, 2014, pp. 35–46.

[20] Y. Zhang, H. Wang, D. Zhang, and D. Wang, "Riskcast: Social sensing based traffic risk forecasting via inductive multi-view learning," in *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2019.

[21] N. Vance, D. Y. Zhang, Y. Zhang, and D. Wang, "Privacy-aware edge computing in social sensing applications using ring signatures," in *2018 IEEE 24th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 2018, pp. 755–762.

[22] D. Y. Zhang, T. Rashid, X. Li, N. Vance, and D. Wang, "Heteroedge: Taming the heterogeneity of edge computing system in social sensing," in *Proceedings of the International Conference on Internet of Things Design and Implementation*. ACM, 2019, pp. 37–48.

[23] E. L. Denton, S. Chintala, R. Fergus *et al.*, "Deep generative image models using aï¿$\frac{1}{4}$ laplacian pyramid of adversarial networks," in *Advances in neural information processing systems*, 2015, pp. 1486–1494.

[24] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[25] Y. Zhang, H. Wang, D. Zhang, and D. Wang, "Deeprisk: A deep transfer learning approach to migratable traffic risk estimation in intelligent transportation using social sensing," in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2019, pp. 123–130.

[26] X. Jia, G. Nayak, A. Khandelwal, A. Karpatne, and V. Kumar, "Classifying heterogeneous sequential data by cyclic domain adaptation: An application in land cover detection," in *Proceedings of the 2019 SIAM International Conference on Data Mining*. SIAM, 2019, pp. 540–548.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[29] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

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

[31] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks." in *CVPR*, vol. 1, no. 2, 2017, p. 3.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[33] R. J. Muller, S. Kornblith, and G. E. Hinton, "When does label smoothing help?" *ArXiv*, vol. abs/1906.02629, 2019.

[34] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[35] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[36] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[37] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.

[38] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[39] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.

[40] R. Artstein and M. Poesio, "Inter-coder agreement for computational linguistics," *Computational Linguistics*, vol. 34, no. 4, pp. 555–596, 2008.

[41] G. Jurman, S. Riccadonna, and C. Furlanello, "A comparison of mcc and cen error measures in multi-class prediction," *PloS one*, vol. 7, no. 8, p. e41882, 2012.