TransRes: A Deep Transfer Learning Approach to Migratable Image Super-Resolution in Remote Urban Sensing

Yang Zhang, Ruohan Zong, Jun Han, Daniel Zhang, Tahmid Rashid, Dong Wang
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
{yzhang42, rzong, jhan5, yzhang42, mrashid, dwang5}@nd.edu

***Abstract*—Recent advances in remote sensing provide a powerful and scalable sensing paradigm to capture abundant visual information about the urban environments. We refer to such a sensing paradigm as *remote urban sensing*. In this paper, we focus on a *migratable satellite image super-resolution* problem in remote urban sensing applications. Our goal is to reconstruct satellite images of a high resolution in a *target area* where the high-resolution training data is *not available* by transferring a super-resolution model learned in a source area where such data is available. This problem is motivated by the limitation of current solutions that primarily rely on a rich set of high-resolution satellite images in the studied area that are not always available. Two important challenges exist in solving our problem: i) the target and source areas often have very different urban characteristics that prevent the direct application of a super-resolution model learned from the source area to the target area; ii) it is not a trivial task to ensure effective model migration with desirable quality without sufficient high quality training data. To address the above challenges, we develop *TransRes*, a deep adversarial transfer learning framework, to effectively reconstruct high-resolution satellite images without requiring any ground-truth training data from the studied area. We evaluate the TransRes framework using the real-world satellite imagery data collected from three different cities in Europe. The results show that TransRes consistently outperforms the state-of-the-art baselines by achieving the lowest perception errors under various application scenarios.**

***Index Terms*—Urban Sensing, Remote Sensing, Migratable Image Super-Resolution, Transfer Learning**

I. Introduction

In this paper, we develop a principled deep adversarial transfer learning framework to address the migratable satellite image super-resolution problem in remote urban sensing applications. Recent advances in remote sensing (e.g., leveraging high resolution images from satellites and drones) provide a powerful and scalable sensing paradigm to capture abundant visual information (e.g., streets, traffic, land usage, disaster damage) about the urban environments [1]. We refer to this sensing paradigm as *remote urban sensing*. Examples of remote urban sensing applications include urban land usage classification [2], city-wide traffic risk detection [3], and real-time disaster situation awareness [4]. In this paper, we focus on a *migratable satellite image super-resolution* problem where our goal is to generate a reconstructed satellite image of a high resolution from a low resolution one in an area of interest where the high-resolution training data is *not available*.

A good amount of efforts have been made to address the satellite image super-resolution problem in image processing, machine learning, and remote sensing [5]–[9]. The current solutions primarily rely on a rich set of high-resolution satellite images in the studied area as the training data to learn an effective super-resolution model [10]. However, such a high-quality training dataset is not always available to the remote urban sensing applications due to the high cost of data acquisition and government/legal regulations [11]. For example, the high-resolution imagery data collected by DigitalGlobe for the urban land usage classification applications is often quite expensive (e.g, USD 1,750 per 100 sq.km.) [12]. Furthermore, high-resolution satellite imagery data collected by some advanced commercial satellites (e.g., *WorldView* satellite) is generally not available for cities outside US due to government regulations [13]. Additionally, open satellite imagery platforms (e.g., Google Maps) provide publicly available satellite imagery data with global coverage. However, the spatial resolution of such "free data" is usually too low in developing countries to be useful for many remote urban sensing applications [14]. For example, the satellite images from Google Maps are reported to provide insufficient resolutions to detect the harmful algal blooms that are highly correlated with cholera outbreaks in major Bangladesh cities [15]. Therefore, the lack of high-resolution training data presents a fundamental challenge to the satellite image super-resolution problem in remote urban sensing applications.

To address the above challenge, this paper develops a deep adversarial transfer learning solution to reconstruct the high-resolution satellite images in a *target area* where the training data is *not available* by taking advantage of a super-resolution model learned in a source area where such data is available. For example, consider two cities: city A and city B that have satellite images of different resolutions. In particular, city A has the high-resolution satellite images published by Google Maps but city B does not [16]. In this example, our goal is to

978-1-7281-6630-8/20/$31.00 2020

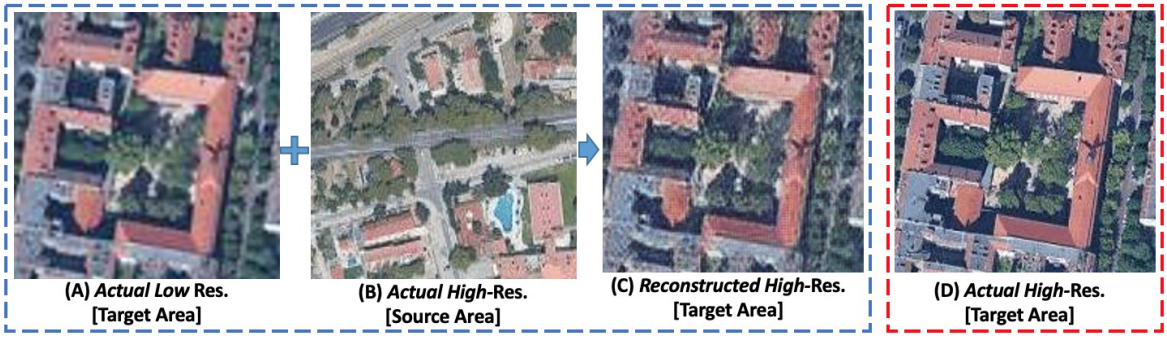


Figure 1. Example of Discrepancy between Source and Target Area

reconstruct the high-resolution satellite images in city B (i.e., target area) by leveraging the super-resolution model learned in city A (i.e., source area). Such a migratable satellite image super-resolution problem is non-trivial to solve due to two technical challenges we elaborated below.

Discrepancy on Area-specific Urban Characteristics: A simple solution to solve the migratable satellite image super-resolution problem is to directly apply the super-resolution model learned from the source area to reconstruct the high-resolution satellite image in the target area [7], [17]. However, a major issue of this solution is that the source and target area may have completely different area-specific urban characteristics (e.g, color distributions, architectural styles, and object layouts as shown in images (B) and (D) in Figure 1). Such a discrepancy would prevent the direct application of a super-resolution model learned from the source area to the target area (e.g, the quality of the reconstructed image (C) from (A) and (B) is not as good as the actual image (D) in Figure 1). Therefore, the migratable super-resolution model has to carefully accommodate the discrepancy on area-specific urban characteristics between the source and target areas to ensure the desirable quality of the reconstructed images.

Lack of Ground Truth Data in the Target Area: An alternative way to solve the migratable satellite image super-resolution problem is to carefully modify the super-resolution model learned from the source area to reconstruct the satellite image for the target area. However, one important limitation exists: current transfer learning solutions have established an effective model migration process from the source to target area by re-training the super-resolution model of the source area using a high-resolution training dataset from the target area [5], [6]. However, such a training dataset is not available in the unsupervised super-resolution problem we study in this paper. Therefore, it is not a trivial task to ensure effective model migration with the desired quality assurance given the lack of ground-truth training data in the target area.

To address the above challenges, we develop TransRes, a deep transfer learning framework for migratable satellite image super-resolution in remote urban sensing applications. TransRes is an *unsupervised* super-resolution solution that does not require any high-resolution training data from the studied area (i.e., target area). In particular, to address the first challenge, we transfer the super-resolution model learned from the source area by accommodating the discrepancy on area-specific urban characteristics through a deep transfer learning network design. To address the second challenge, we design a set of adversarial and cycle-consistent neural network architectures to improve the resolution and quality of the reconstructed images without requiring the direct matching between the high and low resolution training data from the same area. To the best of our knowledge, TransRes is the first deep adversarial transfer learning approach to address the migratable satellite image super-resolution problem in remote urban sensing applications. The unsupervised nature of TransRes makes it applicable to address similar data scarcity problems in many urban sensing applications (e.g., disaster damage assessment, traffic risk detection, urban land classification) where the ground-truth training data is not always available. We evaluate the TransRes framework using the real-world satellite imagery data collected from three different cities in Europe. The results show that TransRes consistently outperforms the state-of-the-art baselines by achieving the lowest perception errors in reconstructing high-resolution satellite images in the targe area under various application scenarios.

II. Related Work

A. Remote Urban Sensing

Motivated by the advent of modern optical and image processing technologies, remote urban sensing has emerged as a powerful and scalable sensing paradigm to capture abundant visual information of the urban environments [1]. Examples of remote urban sensing applications include urban land usage classification [2], city-wide traffic risk detection [18], and real-time disaster situation awareness [19]. A few key challenges exist in current remote urban sensing applications. Examples include data scarcity [20], spatial coverage [21], privacy preservation [22], and image obscurity [23]. However, the unsupervised migratable image super-resolution problem remains to be an open and challenging problem in remote urban sensing applications. In this paper, we develop a novel TransRes framework to address the problem by designing a novel deep adversarial transfer learning framework that utilizes a set of adversarial and cycle-consistent neural network design to ensure the desired reconstructed image quality.

B. Satellite Image Super-Resolution

Efforts have been made to address the satellite image super-resolution problem in image processing, machine learning, and remote sensing [5]–[9]. For example, Lin *et al.* proposed a domain transfer learning approach for hyper-spectral image enhancement by utilizing the cross-correlation between the low-resolution hyper-spectral image and the corresponding multi-spectral image [5]. Zhang *et al.* proposed a neural texture transfer framework to improve the spatial resolution of an image by leveraging the high-resolution contents learned from a set of reference images [6]. Tuna *et al.* presented a deep convolutional framework for the single frame satellite image super resolution problem by leveraging the conventional neural networks to refine the reconstructed high-resolution images [7]. Wang *et al.* proposed a cycle convolutional neural network framework to generate high-resolution images from the low-resolution ones using cycle-consistent network design [8]. However, those approaches cannot be directly applied to our migratable satellite image super-resolution problem because they primarily rely on a rich set of high-resolution satellite imagery data from the studied area to build an effective super-resolution model, which is not available in our problem setting. In contrast, we develop a deep adversarial transfer learning approach to output high-resolution satellite images with desired perception quality in the areas where high-quality training data is not available.

C. Generative Adversarial Learning

Our work is also related to the generative adversarial learning technique, which has been applied in many areas such as nature language processing, recommender systems, intelligent transportation, and image generation [24]–[27]. For example, Pascual *et al.* designed an end-to-end speech enhancement framework to provide fast raw audio quality improvement via generative adversarial networks [24]. Kang *et al.* proposed a visually-aware fashion recommendation system to provide personalized cloth recommendations using generative image models [25]. Zhang *et al.* developed a traffic risk forecasting scheme to provide reliable traffic accident rate prediction using an adversarial transfer learning network [26]. Ledig *et al.* proposed a generative adversarial network framework to generate high-quality natural sense images with an augmented resolution by utilizing an image generator and image discriminator network design [27]. To the best of our knowledge, the TransRes is the first adversarial transfer learning approach to solve the migratable image super-resolution in remote urban sensing by addressing challenges of the discrepancy on area-specific urban characteristics between the source and target area and the lack of ground truth data in the target area.

III. Problem Description

In this section, we formally define the migratable satellite image super-resolution problem in remote urban sensing. We first define the key terms used in the problem statement.

Definition 1: **Source Area** (S): We define a source area to be an area where the high-resolution satellite imagery data is available.

Definition 2: **Target Area** (T): We define a target area to be the studied area of interest where the high-resolution satellite imagery data is *not* available.

Definition 3: **Sensing Cell**: Following a similar procedure in [28], we first divide the source and target area into disjoint sensing cells. Each cell represents a subarea of interest. In particular, we define M to be the number of sensing cells in the source area and m to be the m^{th} sensing cell, and N to be the number of sensing cells in the target area and n to be the n^{th} sensing cell.

Definition 4: **High-Resolution Image in Source Area** (S^H): We define S^H to be a set of high-resolution satellite images that are collected from the source area with a relatively high resolution. In particular, we define S_m^H to be the high-resolution image of the sensing cell m in the source area.

Definition 5: **Low-Resolution Image in Target Area** (T^L): we define T^L to be a set of the satellite images that are collected from the target area with a relatively low resolution. In particular, we define T_n^L to represent the low-resolution image of the sensing cell n in the target area.

Definition 6: **Reconstructed High-Resolution Image in Target Area** ($\widehat{T^H}$): We define $\widehat{T^H}$ to be the set of *reconstructed* high-resolution satellite images in target area. The reconstructed high-resolution satellite images are expected to have the same resolution as the high-resolution satellite images in the source area. In particular, we define $\widehat{T_n^H}$ as the *reconstructed* high-resolution satellite image for the sensing cell n in the target area.

Definition 7: **Area-specific Urban Characteristics**: it refers to the specific visual features (e.g, color distributions, architectural styles, and object layouts) of the satellite images that are characteristic in a given area. In particular, the source and target area often have very different area-specific urban characteristics as shown in Figure 1.

Definition 8: **Perception Quality**: To evaluate the quality of $\widehat{T^H}$, we use the state-of-the-art perception metric [29] to quantify the perception difference between the *actual* and *reconstructed* satellite images as follows:

$$perc(T_n^H, \widehat{T_n^H}) = \Omega\left(\mathcal{D}(T_n^H) - \mathcal{D}(\widehat{T_n^H})\right) \tag{1}$$

where $perc(\cdot)$ represents the perception metric. $\mathcal{D}(T_n^H)$ and $\mathcal{D}(\widehat{T_n^H})$ represent the extracted deep features from the *actual* and *reconstructed* satellite images using ImageNet-trained deep convolutional neural networks (e.g., VGG [30]). $\Omega(\cdot)$ is a function to calculate the difference between two deep feature vectors (e.g., Mean Squared Error (MSE), Mean Absolute Error (MAE)). This metric has been proven to be robust in capturing perception quality of images [31].

The goal of our migratable super-resolution problem is to accurately reconstruct the high-resolution satellite image for each sensing cell in the target area from its corresponding low-resolution satellite image by leveraging the super-resolution

model learned from the source area. Using the definitions above, our problem is formally defined as:

$$\arg\min_{\widehat{T_n^H}} \left(\Omega\left(\mathcal{D}(T_n^H) - \mathcal{D}(\widehat{T_n^H})\right) \mid S^H, T^L\right), \quad \forall 1 \leq n \leq N \tag{2}$$

This problem is challenging considering the discrepancy on area-specific urban characteristics between the source and target areas and the lack of ground truth data in the target area. In this paper, we develop a TransRes scheme to address these challenges, which is elaborated in the next section.

IV. Solution

A. Overview of TransRes Framework

TransRes is a deep adversarial transfer learning approach to address the migratable image super-resolution problem in remote urban sensing. The overview of the TransRes framework is shown in Figure 2. It consists of two modules: 1) *migratable image super-resolution networks (MISN)* and 2) *perception quality-aware optimization process (PQOP)*. First, in the MISN module, we present the adversarial transfer learning network design in TransRes that enables the effective super-resolution model migration between the source and target areas to accommodate the discrepancy on area-specific urban characteristics. Second, in the *PQOP* module, we present the optimization process of TransRes to learn the optimal instances of all neural networks in the MISN module to achieve the desired perception quality of the reconstructed satellite images.

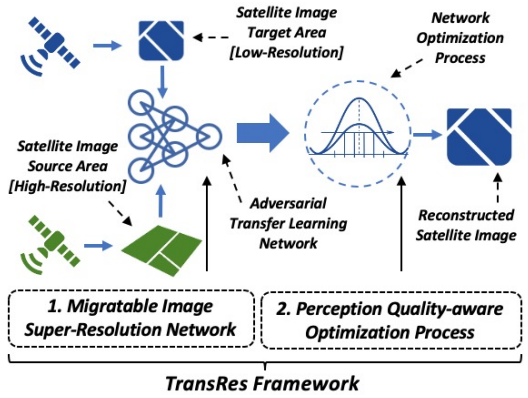


Figure 2. Overview of TransRes framework

B. Migratable Image Super-Resolution Networks (MISN)

In this subsection, we present the *adversarial transfer learning network architecture* design in MISN. An overall architecture of the adversarial transfer learning network is shown in Figure 3. The adversarial transfer learning network design consists of three neural network architectures: an upscaling network (UN), a downscaling network (DN), and an examination network (EN). The upscaling and downscaling networks work collaboratively to learn a migratable super-resolution model without requiring any ground-truth data for the target area. In particular, the upscaling network first generates the reconstructed images in the target area with the same resolution as the high-resolution ones in the source area. The downscaling network then converts the reconstructed images back to the low-resolution ones in the target area and compared with the their original low-resolution counterparts. The goal of such a design is to verify if the area-specific urban characteristics of the target area can be successfully preserved during the image reconstruction process. The examination network is used to exam the reconstructed high-resolution satellite image to ensure the desired image quality. Intuitively, the examination network is used to regulate the upscaling network to ensure the reconstructed images generated by the upscaling network have the same resolution and image quality as the high-resolution images in the source area. We first formally define the three network architectures as follows.

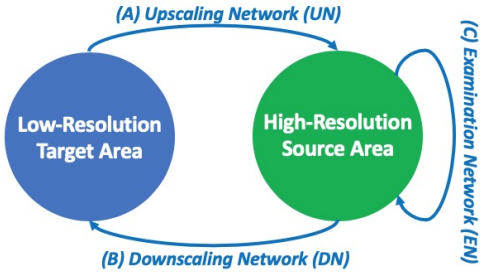


Figure 3. Overall of Network Architectures

Definition 9: **Upscaling Network (*UN*)**: we define *UN* as a generative network that reconstructs a high-resolution satellite image $\widehat{X^H}$ from a corresponding low-resolution satellite image X^L as follows:

$$\widehat{X^H} = UN(X^L) \tag{3}$$

An example of the upscaling network is shown in (A) of Figure 4. It consists of three components: an image encoder, a set of residual blocks, and an image decoder. The image encoder contains a set of convolutional and instance normalization layers, which are used to extract the semantic feature representations of the contents in low-resolution images. The residual block component contains multiple residual blocks [32] that handle the complex task of segmenting individual objects of an image and applying augmented contents (e.g., more fine-grained object details) to each identified object to improve the image resolution. The image decoder has multiple deconvolutional and instance normalization layers to convert the augmented semantic feature representations generated by the residual block component into the reconstructed high-resolution satellite images.

Definition 10: **Downscaling Network (*DN*)**: we define *DN* as an additional generative network that transforms the reconstructed high-resolution satellite image $\widehat{X^H}$ back to its original low-resolution satellite image X^L as follows:

$$X^L = DN(\widehat{X^H}) \tag{4}$$

An example of the downscaling network is shown in (B) of Figure 4. Similar to the upscaling network, the downscaling network also consists of three components: an image encoder,

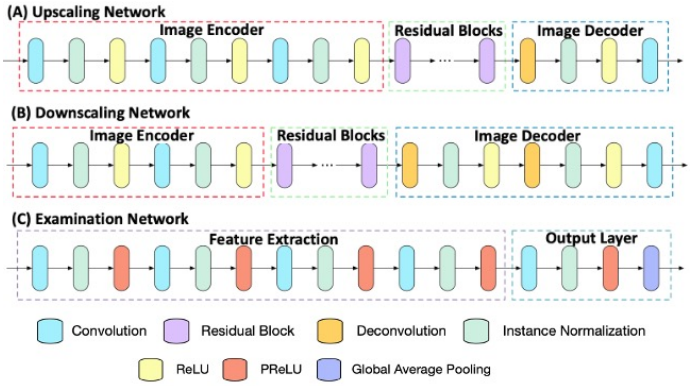


Figure 4. Illustrations of Network Architectures

a set of residual blocks, and an image decoder. The only difference here is that the residual blocks in the downscaling network are used to remove the fine-grained object details in reconstructed images to reduce the resolution.

Definition 11: **Examination Network (*EN*)**: We define *EN* as an examination network to exam whether a high-resolution satellite image X^H matches the same image quality (i.e., fine-graininess and image clarity) of a reference high-resolution satellite image Y^H or not:

$$EN : \begin{cases} \mathbf{1} : X^H \in Y^H \\ \mathbf{0} : X^H \notin Y^H \end{cases} \tag{5}$$

where *EN* returns true (i.e., "1") if X^H matches the same image quality of the reference satellite image Y^H and false (i.e., "0") otherwise.

An example of the examination network is shown in (C) of Figure 4. It consists of two components: a feature extractor and an output layer. The feature extractor contains a set of convolutional layers, which are used to identify a critical set of visual features that best represent the content of a reconstructed satellite image. The output layer includes a convolutional layer and an average pooling layer, which are used to evaluate the reconstructed image quality and output the classification results using the extracted visual features generated by the feature extractor.

C. Perception Quality-aware Optimization Process (PQOP)

Given the three network architectures above, our next question is how to learn the optimal instances of all networks to maximize the quality of the reconstructed satellite image in the target area. To that end, we define three sets of loss functions in our TransRes framework, which are elaborated below.

In the first set of loss functions, we consider the stability loss, which is used to ensure the stable performance of the upscaling network UN using the high resolution satellite images collected from the source area S^H as follows:

$$\mathcal{L}_{UN}^{ST} : \left(\mathcal{L}_{\text{perc}}(S^H, UN(\overline{S^L})) + L_{\text{pix}}(S^H, UN(\overline{S^L}))\right) \tag{6}$$

where $\mathcal{L}_{UN}^{ST}$ represents the stability loss function for the upscaling network UN. $\overline{S^L}$ indicates the low-resolution image in source area, which is generated from S^H through the bi-cubic down-sampling operation [33]. $\mathcal{L}_{\text{perc}}(\cdot)$ is the perception loss [29] to quantify the perceptual difference between the actual and reconstructed images (i.e., X^H and $\widehat{X^H}$). $L_{\text{pix}}(\cdot)$ is the MSE loss [34] to measure the pixel-wise RGB value difference between the actual and reconstructed images. Intuitively, $\mathcal{L}_{UN}^{ST}$ is designed to ensure the stable performance of UN. Similarly, we have the stability loss function $\mathcal{L}_{DN}^{ST}$ for the downscaling network DN as follows:

$$\mathcal{L}_{DN}^{ST} : \left(\mathcal{L}_{\text{perc}}(T^L, DN(\overline{T^H})) + L_{\text{pix}}(T^L, DN(\overline{T^H}))\right) \tag{7}$$

where $\overline{T^H}$ is the high-resolution image for the target area, which is generated from T^L through the bi-cubic up-sampling operation [33]. Intuitively, $\mathcal{L}_{DN}^{ST}$ is designed to ensure the stable performance of DN.

After defining the loss functions for the upscaling and downscaling networks, our next question is: how to ensure the two networks work collaboratively to maximize the quality of the reconstructed satellite images in the target area. To address this question, we define another set of transformation loss function as follows:

$$\mathcal{L}_{UN,DN}^{TR} : \mathcal{L}_{\text{pix}}(T^L, DN(UN(T^L))) + \mathcal{L}_{\text{pix}}(S^H, UN(DN(S^H))) \tag{8}$$

where $\mathcal{L}_{UN,DN}^{TR}$ represents the transformation loss for the UN and DN. The idea is to ensure the upscaling and downscaling networks can translate the satellite images T^L and S^H to the original versions after the image upscaling and downscaling process (i.e., $DN(UN(T^L)) \rightarrow T^L$ and $UN(DN(S^H)) \rightarrow S^H$). In particular, the reconstructed images (i.e., $UN(T^L)$) are translated back to the previous resolution and compared with the original images (i.e., compare $DN(UN(T^L))$ with T^L) to ensure the area-specific urban characteristics of the target area are preserved in the reconstructed satellite images.

Finally, we consider the adversarial loss for the examination network EN to exam the quality of reconstructed high-resolution satellite images generated by the upscaling network UN. In particular, we define the adversarial loss for the examination network EN as follows:

$$\mathcal{L}_{EN}^{AD} : \left(||\mathbf{0} - EN(\widehat{T^H})||_2 + ||\mathbf{1} - EN(S^H)||_2\right) \tag{9}$$

where $\mathcal{L}_{EN}^{AD}$ is the adversarial loss function for the examination network EN. $||\cdot||_2$ donates the L2-norm of a given matrix. Intuitively, the examination network EN is used to identify the poorly reconstructed high-resolution satellite images $\widehat{T^H}$ that do not match the same level of image quality as the actual high-resolution satellite images in the source area. Similarly, we also define the adversarial loss for the upscaling network UN as follows:

$$\mathcal{L}_{UN}^{AD} : \left(||\mathbf{1} - EN(\widehat{T^H})||_2\right) \tag{10}$$

$\mathcal{L}_{UN}^{AD}$ is used to ensure that UN can generate reconstructed satellite image $\widehat{T^H}$ with desired image quality verified by EN (i.e., returning 1). Intuitively, $\mathcal{L}_{UN}^{AD}$ and $\mathcal{L}_{EN}^{AD}$ are designed to

enforce the competition between UN and EN so that UN can generate high quality reconstructed high-resolution satellite images with the highest image quality possible.

Finally, we combine the above three set of loss functions to derive the final loss $\mathcal{L}_{UN,DN}$ for the generative networks (i.e, UN and DN) and the final loss $\mathcal{L}_{EN}$ for the examination network (i.e., EN) to jointly optimize the objectives of our adversarial transfer learning networks (i.e., loss functions defined above) as follows:

$$\begin{aligned} &\mathcal{L}_{UN,DN} : \left(\mathcal{L}_{UN}^{ST} + \mathcal{L}_{DN}^{ST} + \mathcal{L}_{UN,DN}^{TR} + \mathcal{L}_{UN}^{AD}\right) \\ &\mathcal{L}_{EN} : \mathcal{L}_{EN}^{AD} \end{aligned} \tag{11}$$

Given the above loss functions, the optimal instances (i.e., UN^*, DN^*, and EN^*) of all networks can be learned using the Adaptive Moment Estimation (ADAM) optimizer [35]. We then use UN^* to generate the reconstructed satellite images for the target area $\widehat{T^H}$ as the output of TransRes as follows:

$$\widehat{T^H} = UN^*(T^L) \tag{12}$$

V. Evaluation

In this section, we conduct extensive experiments on real-world datasets to answer the following questions:

- Q1: Can TransRes achieve a better reconstructed satellite image quality than the state-of-the-art baselines for different areas in a city with distinct area-specific urban characteristics?
- Q2: Can TransRes consistently outperform other baselines in more challenging application scenarios when the source and target area come of different cities with completely different urban environments?
- Q3: How do the different choices of model parameters (e.g., the depth and width of neural networks) affect the performance of TransRes?

A. Dataset

In our experiment, we collect real-world satellite imagery datasets from three different cities in Europe (i.e, *Barcelona (Spain)*, *Athens (Greece)*, and *Berlin (Germany)*) with two land usage classes (i.e., *urban fabric* and *transportation* as shown in Figure 5). The two different land usage classes in different cities have clearly different area-specific urban characteristics, which create a challenging evaluation scenario for the migratable super-resolution problem we studied in this paper. We summarize the datasets as follows.

Urban Imagery Dataset from Google Maps: We collect the urban imagery datasets from *Barcelona*, *Athens*, and *Berlin* using Google Map Platform [1]. In our evaluation, each collected satellite image is in 224×224 resolution with a 250m×250m ground coverage, which is considered as the *high* resolution image in our evaluation. We adopt the widely-used *bicubic interpretation* tool implemented in *scikit-image* package [2] to reduce the resolution of a collected satellite image by 4 times as the *low* resolution image in our experiment (i.e., each low-resolution satellite image is in 112×112 resolution as shown in Figure 1). Finally, we randomly select 600 *high* and *low* resolution images (i.e., 300 from each category) from the studied area for our experiments.

Figure 5. Examples of Two Classes of Urban Images from Three Different Cities in Europe

B. Baselines

We compare TransRes with the state-of-the-art *conventional* and *deep learning* baselines. To ensure the fairness of comparison, the inputs to all compared schemes are set to be the same (i.e., the low-resolution images from the target area and the high-resolution images from the source area).

1) Conventional Models
   - **Nearest-neighbour (NN)** [36]: it is a conventional super-resolution scheme that augments the satellite image contents by utilizing the RGB value from the nearest available neighboring pixels.
   - **Bilinear** [37]: it is a super-resolution scheme that utilizes bilinear upsampling operations to upscale the image resolution.
   - **Bicubic** [38]: it is a popular super-resolution scheme that leverages the bicubic interpolation technique to estimate the RGB values of all empty pixels in the reconstructed high-resolution satellite images.
2) Deep Learning Models
   - **SFSR18** [7]: it utilizes a set of deep convolutional operations to scale the low-resolution satellite image to a high-resolution one and refine the reconstructed high-resolution satellite images.
   - **SRGAN17** [27]: it leverages a generative adversarial network design that utilizes an image generator and an image discriminator to improve the quality of reconstructed high-resolution satellite images.
   - **CycleCNN19** [8]: it is a deep transfer learning framework that utilizes the cycle-consistent design to capture the complex association of fine-grained object details with the low-resolution objects in the image reconstruction process.

[1] https://developers.google.com/maps/documentation/

[2] https://scikit-image.org/docs/dev/api/skimage.transform.html\#skimage.transform.resize

Table I
PERFORMANCE COMPARISONS (*Same* CITY *Different* CLASSES)

		Barcelona_U→Barcelona_T				Barcelona_T→Barcelona_U			
		Deep Feature 1		Deep Feature 2		Deep Feature 1		Deep Feature 2	
Class	Algorithm	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Conventional	Nearest-neighbor	0.5179	1.6749	0.4043	1.1325	0.6420	2.4936	0.4949	1.6365
	Bilinear	0.5092	1.6544	0.3948	1.1117	0.6168	2.3764	0.4712	1.5237
	Bicubic	0.4929	1.5539	0.3809	1.0359	0.5993	2.2503	0.4559	1.4282
Deep Learning	SFSR18	0.4966	1.5539	0.3816	1.0346	0.6001	2.2342	0.4517	1.4023
	SRGAN17	0.4893	1.5129	0.3752	0.9935	0.5919	2.1930	0.4443	1.3681
	CycleCNN19	0.5658	1.9136	0.4369	1.2966	0.5661	2.0206	0.4375	1.2837
Our Model	**TransRes**	**0.4611**	**1.3426**	**0.3532**	**0.8949**	**0.5416**	**1.8659**	**0.4185**	**1.1819**

C. Evaluation Metrics

To ensure the rigorous evaluation of the performance for all compared schemes, we use the perceptual metric (Definition 8) in our experiments, which has been proven to be a metric that is close to human perception in the recent computer vision studies [29], [31], [39]. In particular, following the literature in [29], [31], we select two commonly used deep features (i.e., $\mathcal{D}(\cdot)$ in Equation (1)) extracted by the 2^{nd}, 3^{rd} convolutional layers of the 4^{th} convolutional block in VGG model (namely, VGG_{4-2}, VGG_{4-3}.) We refer to them as *deep feature 1* and *deep feature 2* in the evaluation. In addition, we adopt two commonly used error measurement functions (i.e., *Mean Absolute Error (MAE)* and *Mean Squared Error (MSE)* as the $\Omega(\cdot)$ in Equation (1)) to calculate the difference between the deep features extracted from the *actual* and *reconstructed* satellite images. Intuitively, a lower value in the error metric represents a higher perception quality of the *reconstructed* satellite images, and hence a better super-resolution performance.

D. Evaluation Results

1) Q1: Performance Comparison across Different Land Classes in a City: In the first set of experiments, we evaluate the performance of all compared schemes by setting the source and target area to locations of different land usage classes in the same city. For example, we consider two land usage classes in Barcelona (i.e., *urban fabric* and *transportation*, which are referred to as *Barcelona_U* and *Barcelona_T*, respectively). We set locations of one class to be the source area and the locations of the other class to be the target area. For example, Barcelona_U→Barcelona_T represents that we set the locations of the urban fabric class in Barcelona (*Barcelona_U*) as the source area and the locations of the transportation class in Barcelona (*Barcelona_T)* as the target area). The evaluation results are presented in Table I. We observe that the TransRes scheme consistently outperforms all compared baselines. For example, the performance gains of TransRes over the best-performing baseline (i.e., SRGAN17) in Barcelona_U→ Barcelona_T with the deep feature 1 (i.e., deep feature extracted by VGG_{4-2}) on MAE and MSE are 6.12% and 12.68%, respectively. Such performance gains mainly come from the fact that TransRes reconstructs the high-quality images for the target area by transferring the super-resolution model learned from the source area. In particular, the adversarial transfer learning network design in TransRes enables the effective super-resolution model migration between the source and target areas to accommodate the discrepancy on the area-specific urban characteristics between the two areas.

2) Q2: Performance Comparison across Different Cities: In this experiment, we evaluate the performance of all compared schemes in a more challenging evaluation scenario when the source and target areas are from different cities. In particular, we consider two evaluation settings: 1) source and target areas are from two cities with the same land usage class (we refer to it as *different cities same class*) and 2) source and target areas are from two cities with different land usage classes (we refer to it as *different cities different classes*). The evaluation results are shown in Table II and Table III. We observe that TransRes consistently outperforms all baselines over different source and target area settings. For example, the performance gains achieved by TransRes compared to the best-performing baseline (i.e., CycleCNN19) in Barcelona_U→Athens_U with the deep feature 2 (i.e., deep feature extracted by VGG_{4-3}) on MAE and MSE are 5.11% and 10.02%, respectively. Such consistent performance gains over various scenarios demonstrate the effectiveness of the adversarial and transfer-consistent neural network design in our model.

3) Q3: Robustness Study of TransRes Scheme: Two key parameters in the upscaling network (Definition 9) and down-scaling network (Definition 10) are essential for our TransRes framework: 1) the number of residual blocks (N) that are used to control the depth of our networks, and 2) the channel transformation ratio (C) that is used to control the width of our networks. In this set of experiments, we examine how these two parameters affect the performance of our TransRes. Results are presented in Figure 6. Given the space limit, we only present the MAE results for three source and target area pairs (e.g,. Barcelona_U→Barcelona_T category in Table I). Performance in other scenarios are similar. We observe that the performance of TransRes is relatively stable as the number of residual blocks and channel transformation ratio change,

Table II
PERFORMANCE COMPARISONS (*Different* CITIES *Same* CLASS)

		Barcelona_U→Athens_U				Barcelona_T→Athens_T			
		Deep Feature 1		Deep Feature 2		Deep Feature 1		Deep Feature 2	
Class	Algorithm	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Conventional	Nearest-neighbor	0.6556	2.5838	0.5041	1.7083	0.5359	1.8301	0.4187	1.2036
	Bilinear	0.6550	2.6670	0.5003	1.6967	0.5187	1.7554	0.4015	1.1408
	Bicubic	0.6346	2.5162	0.4843	1.5945	0.5036	1.6582	0.3884	1.0685
Deep Learning	SFSR18	0.6247	2.3912	0.4739	1.5294	0.5001	1.6206	0.3822	1.0358
	SRGAN17	0.6199	2.3803	0.4714	1.5154	0.5094	1.6990	0.3927	1.0952
	CycleCNN19	0.5550	2.0250	0.4261	1.2541	0.5052	1.6929	0.3919	1.0746
Our Model	**TransRes**	**0.5307**	**1.8511**	**0.4054**	**1.1398**	**0.4912**	**1.5999**	**0.3809**	**1.0208**

Table III
PERFORMANCE COMPARISONS (*Different* CITIES *Different* CLASSES)

		Barcelona_U→Berlin_T				Barcelona_T→Berlin_U			
		Deep Feature 1		Deep Feature 2		Deep Feature 1		Deep Feature 2	
Class	Algorithm	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Conventional	Nearest-neighbor	0.5006	1.5740	0.3884	1.0756	0.6919	2.9329	0.5354	1.9279
	Bilinear	0.5141	1.6866	0.3967	1.1393	0.6833	2.9462	0.5244	1.8841
	Bicubic	0.4993	1.5949	0.3841	1.0714	0.6660	2.8021	0.5096	1.7786
Deep Learning	SFSR18	0.4984	1.5649	0.3824	1.0528	0.6572	2.7022	0.5030	1.7312
	SRGAN17	0.4984	1.5822	0.3824	1.0528	0.6808	2.9452	0.5200	1.8714
	CycleCNN19	0.5022	1.3985	0.3782	0.8727	0.6855	3.0137	0.5285	1.9442
Our Model	**TransRes**	**0.3481**	**0.7187**	**0.2629**	**0.4688**	**0.6482**	**2.6819**	**0.4976**	**1.7145**

demonstrating the robustness of our scheme over these key parameters of our model.

VI. CONCLUSION

This paper develops a TransRes framework to solve the migratable image super-resolution problem in remote urban sensing. In particular, we develop a novel deep adversarial transfer learning framework to effectively reconstruct high-resolution satellite images without requiring any ground-truth training data from the studied area. The evaluation results demonstrate that TransRes achieves non-trivial performance gains compared to the state-of-the-art super-resolution baselines in reconstructing high-resolution satellite images with the desirable quality. We believe TransRes will provide useful insights to address similar data scarcity problems in other urban sensing applications.

ACKNOWLEDGMENT

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] H. Shen, M. K. Ng, P. Li, and L. Zhang, "Super-resolution reconstruction algorithm to modis remote sensing images," *The Computer Journal*, vol. 52, no. 1, pp. 90–100, 2007.

[2] Y. Zhang, R. Zong, J. Han, H. Zheng, Q. Lou, D. Zhang, and D. Wang, "Transland: An adversarial transfer learning approach for migratable urban land usage classification using remote sensing," in *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 2019, pp. 1567–1576.

[3] 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.

[4] D. Zhang, D. Wang, N. Vance, Y. Zhang, and S. Mike, "On scalable and robust truth discovery in big data social media sensing applications," *IEEE Transactions on Big Data*, vol. 5, no. 2, pp. 195–208, 2018.

[5] X. Li, L. Zhang, and J. You, "Domain transfer learning for hyperspectral image super-resolution," *Remote Sensing*, vol. 11, no. 6, p. 694, 2019.

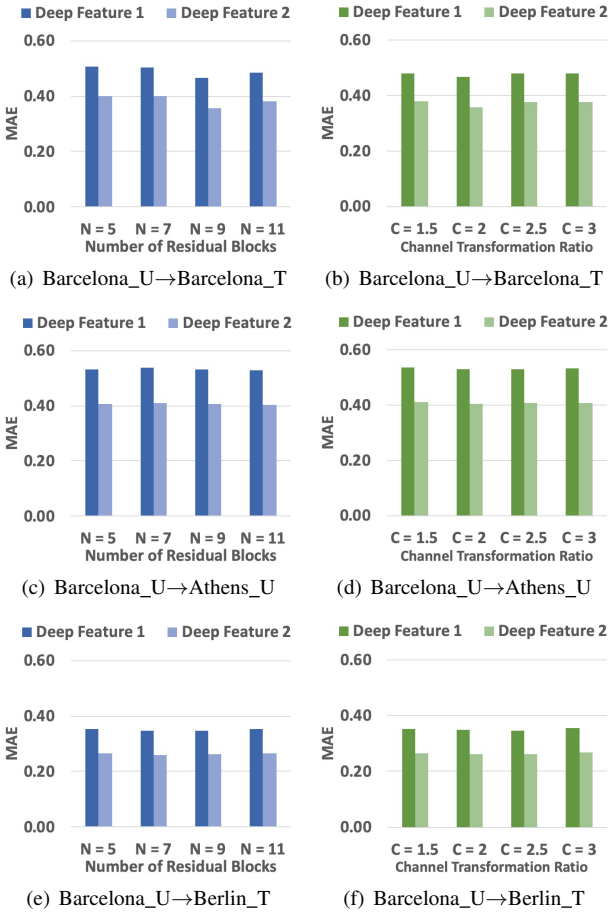
[6] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, "Image super-resolution by neural texture transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7982–7991.

[7] C. Tuna, G. Unal, and E. Sertel, "Single-frame super resolution of remote-sensing images by convolutional neural networks," *International journal of remote sensing*, vol. 39, no. 8, pp. 2463–2479, 2018.

[8] P. Wang, H. Zhang, F. Zhou, and Z. Jiang, "Unsupervised remote sensing image super-resolution using cycle cnn," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 3117–3120.

[9] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[10] L. Liebel and M. Körner, "Single-image super resolution for multispectral remote sensing data using convolutional neural networks," *ISPRS-*



(a) Barcelona_U→Barcelona_T (b) Barcelona_U→Barcelona_T

(c) Barcelona_U→Athens_U (d) Barcelona_U→Athens_U

(e) Barcelona_U→Berlin_T (f) Barcelona_U→Berlin_T

Figure 6. Robustness Study of TransRes Scheme

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 41, pp. 883–890, 2016.

[11] M. J. Falkowski, M. A. Wulder, J. C. White, and M. D. Gillis, "Supporting large-area, sample-based forest inventories with very high spatial resolution satellite imagery," *Progress in Physical Geography*, vol. 33, no. 3, pp. 403–423, 2009.

[12] S. A. Boyle, C. M. Kennedy, J. Torres, K. Colman, P. E. Pérez-Estigarribia, and U. Noé, "High-resolution satellite imagery is an important yet underutilized resource in conservation biology," *PLoS One*, vol. 9, no. 1, p. e86908, 2014.

[13] G. S. Page, "Worldview 4." [Online]. Available: https://space.skyrocket.de/doc_sdat/worldview-4.htm

[14] E. O. System, "Satellite data: What spatial resolution is enough for you?" [Online]. Available: https://eos.com/blog/satellite-data-what-spatial-resolution-is-enough-for-you/

[15] L. Shen, H. Xu, and X. Guo, "Satellite remote sensing of harmful algal blooms (habs) and a potential synthesized framework," *Sensors*, vol. 12, no. 6, pp. 7778–7803, 2012.

[16] Google, "Google map." [Online]. Available: https://www.google.com/maps

[17] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[18] Y. Zhang, H. Wang, D. Zhang, Y. Lu, and D. Wang, "Riskcast: social sensing based traffic risk forecasting via inductive multi-view learning," in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2019, pp. 154–157.

[19] D. Y. Zhang, C. Zheng, D. Wang, D. Thain, X. Mu, G. Madey, and C. Huang, "Towards scalable and dynamic social sensing using a distributed computing framework," in *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*. IEEE, 2017, pp. 966–976.

[20] M. T. Rashid, D. Zhang, Z. Liu, H. Lin, and D. Wang, "Collabdrone: A collaborative spatiotemporal-aware drone sensing system driven by social sensing signals," in *2019 28th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2019, pp. 1–9.

[21] M. T. Rashid, D. Zhang, and D. Wang, "Socialdrone: An integrated social media and drone sensing system for reliable disaster response," in *INFOCOM 2020-IEEE Conference on Computer Communications, IEEE*. IEEE, 2020.

[22] E. K. Wang, F. Wang, R. Sun, and X. Liu, "A new privacy attack network for remote sensing images classification with small training samples." *Mathematical biosciences and engineering: MBE*, vol. 16, no. 5, pp. 4456–4476, 2019.

[23] G. Q. Collins, M. J. Heaton, and L. Hu, "Physically constrained spatiotemporal modeling: generating clear-sky constructions of land surface temperature from sparse, remotely sensed satellite data," *Journal of Applied Statistics*, pp. 1–21, 2019.

[24] S. Pascual, A. Bonafonte, and J. Serra, "Segan: Speech enhancement generative adversarial network," *arXiv preprint arXiv:1703.09452*, 2017.

[25] W.-C. Kang, C. Fang, Z. Wang, and J. McAuley, "Visually-aware fashion recommendation and design with generative image models," in *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2017, pp. 207–216.

[26] 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.

[27] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[28] 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.

[29] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 586–595.

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

[31] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 pirm challenge on perceptual image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[32] 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.

[33] P. Xia, T. Tahara, T. Kakue, Y. Awatsuji, K. Nishio, S. Ura, T. Kubota, and O. Matoba, "Performance comparison of bilinear interpolation, bicubic interpolation, and b-spline interpolation in parallel phase-shifting digital holography," *Optical review*, vol. 20, no. 2, pp. 193–197, 2013.

[34] M. S. Sajjadi, B. Scholkopf, and M. Hirsch, "Enhancenet: Single image super-resolution through automated texture synthesis," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4491–4500.

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

[36] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5197–5206.

[37] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 723–731.

[38] P. Rasti, H. Taşmaz, M. Daneshmand, R. Kiefer, C. Ozcinar, and G. Anbarjafari, "Satellite image enhancement: systematic approach for denoising and resolution enhancement." *DYNA-Ingeniería e Industria*, vol. 91, no. 3, 2016.

[39] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.