On Coupling Classification and Super-Resolution in Remote Urban Sensing: An Integrated Deep Learning Approach

Yang Zhang, *Student Member, IEEE*, Ruohan Zong, *Student Member, IEEE*, Lanyu Shang, *Student Member, IEEE*, and Dong Wang, *Member, IEEE*

***Abstract*— Motivated by the state-of-the-art optical sensing and image processing technologies, remote urban sensing (RUS) has emerged as a powerful sensing paradigm to capture abundant visual information about the urban environment for intelligent city monitoring, planning, and management. In this article, we focus on a classification and super-resolution coupling (CSC) problem in RUS applications, where the goal is to explore the interdependence between two critical tasks (i.e., *classification* and *super-resolution*) to concurrently boost the performance of both the tasks. Two fundamental challenges exist in solving our problem: 1) it is challenging to obtain accurate classification results and generate high-quality reconstructed images without knowing either of them *a priori* and 2) the noise embedded in the image data could be amplified infinitely by the complex interdependence and coupling between the two tasks. To address these challenges, we develop *SCLearn*, a novel deep convolutional neural network architecture, to couple the classification task with the super-resolution task in an integrated learning framework to *concurrently* boost the performance of both the tasks. The evaluation results on a real-world RUS application over two different cities in Europe (Barcelona and Berlin) show that SCLearn consistently outperforms the state-of-the-art baselines by simultaneously achieving better land usage classification accuracy and higher reconstructed image quality under various application scenarios.**

***Index Terms*— Classification, integrated deep learning, smart urban sensing, super-resolution.**

I. Introduction

REMOTE urban sensing (RUS) has emerged as a powerful and scalable sensing paradigm to capture abundant visual information about the urban environment by leveraging high-quality images from satellites and unmanned aerial vehicles (UAVs) [1]. Examples of RUS applications include urban infrastructure health monitoring for smart urban management [2], anomaly event detection and investigation for emergency response [3], remote crop growth sensing for precise agriculture [4], and city-wide land usage classification for intelligent urban planning [5]. In this article, we focus on two critical tasks—*classification* and *super-resolution*—in RUS applications. In particular, the classification task in RUS refers to the process of learning the class label (e.g., infrastructure health, damage severity, crop condition, and land usage) of a given sensing image. In contrast, the super-resolution task targets at improving the spatial resolution of an input sensing image. On one hand, while classification is effective in categorizing the sensing images that share similar visual characteristics, its performance often depends on the resolution of the input images [6]. On the other hand, super-resolution is dedicated to improving the image quality by refining its visual details, but it often requires a good amount of training data with similar visual features (e.g., images from the same class) to learn a specific super-resolution model [7]. In this article, we study a new classification and super-resolution coupling (CSC) problem, where the goal is to explore the interdependence between the classification and super-resolution tasks to concurrently improve the performance of both the tasks.

An example RUS application of our CSC problem is the classification of diversified land usages using aerial images in urban areas as shown in Fig. 1. The land usage classification results are essential to address various important social and urban challenges in a city (e.g., urban planning and management, natural resource, and environment protection) [8]. In Fig. 1, we observe that the waterbody and recreation area are misclassified as green lands in the image in Fig. 1(a) due to insufficient image resolution. This problem can be addressed by exploring the interdependence between the classification and super-resolution tasks. On one hand, the land usage classification scheme could leverage the high-resolution (HR) images reconstructed by the super-resolution scheme [e.g., Fig. 1(b)] for a more accurate land usage classification result. On the other hand, the super-resolution scheme is also able to leverage the accurate land usage labels learned by the classification scheme to build a more specific and refined super-resolution model for high-quality image reconstruction (e.g., improving the image quality of the red box areas in Fig. 1(a) using the visual details of the waterbody and recreation areas instead of green lands).

Manuscript received December 10, 2021; revised February 25, 2022 and March 24, 2022; accepted April 11, 2022. Date of publication April 22, 2022; date of current version May 9, 2022. This work was supported in part by the National Science Foundation under Grant IIS-2008228, Grant CNS-1845639, and Grant CNS-1831669; and in part by the Army Research Office under Grant W911NF-17-1-0409. *(Corresponding author: Dong Wang.)*

Yang Zhang is with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: yzhang42@nd.edu).

Ruohan Zong, Lanyu Shang, and Dong Wang are with the School of Information Sciences, University of Illinois Urbana-Champaign, Champaign, IL 61820 USA (e-mail: rzong2@illinois.edu; lshang3@illinois.edu; dwang24@illinois.edu).

Digital Object Identifier 10.1109/TGRS.2022.3169703

Fig. 1. Example of land usage classification. (a) Low resolution. (b) High resolution.

A good amount of efforts in RUS focus on solving the classification and super-resolution problems separately. In particular, examples of classification solutions include land usage classification using Siamese-prototype network [9], satellite image texture recognition using convolutional neural networks (CNNs) [10], and remote sensing image scene classification using the best representation branch model [11]. Examples of super-resolution solutions include remote sensing image enhancement using reference-based generative adversarial network (GAN) [12], image upscaling using CNNs [13], and satellite video super-resolution using multiscale deformable convolution networks [14]. There also exist efforts that use super-resolution models to improve the performance of the downstream tasks such as classification and segmentation [15]–[17]. However, those efforts largely ignore the opportunity to leverage the estimated classification labels or segmentation results to also improve the image resolution, which can be used to further increase the classification accuracy. To the best of our knowledge, no existing work explicitly explores the interdependence between classification and super-resolution to *simultaneously* boost the performance of *both* the tasks. In contrast, this article focuses on a CSC problem, where the goal is to integrate the classification and super-resolution tasks into a holistic framework to systematically explore the interdependence between the two tasks and improve both of their performance. However, solving such a CSC problem is a nontrivial task due to two intrinsic challenges, which are elaborated as follows.

A. Complex Interdependence

The first challenge of integrating the classification and super-resolution tasks lies in the complex interdependence between the two tasks. In particular, there exists a "chicken-and-egg" issue in our CSC problem where the two tasks depend on the results generated by each other. Specifically, it is challenging to concurrently obtain accurate classification results and generate high-quality reconstructed images without knowing either of them *a priori*. A straightforward solution to address this problem is to perform one task first and then the other. However, a major problem of this approach is its ignorance of the interdependence between the two tasks and the corresponding suboptimal results. Consider the case where we first apply a super-resolution scheme to reconstruct a high-resolution image and then perform the classification task based on the reconstructed image. In this case, the classification result could be completely wrong if we use a low-quality reconstructed image generated by a *one-size-fits-all* super-resolution model trained by data samples from *all* possible classes [18]. Therefore, it is difficult to design an effective integration solution that addresses the complex interdependence between the classification and super-resolution tasks in RUS applications.

B. Noise Amplification

The second challenge of integrating the classification and super-resolution tasks lies in how to avoid the potential noise amplification between the two coupled tasks. In particular, an effective way to address the complex interdependence between the classification and super-resolution tasks is to develop a solution that iteratively leverages the output from one task to improve the result of the other. However, there exists a potential "vicious circle" problem for such an iterative solution where the noise embedded in the data (e.g., fuzziness of the images, incorrect class labels) could be amplified infinitely in the iterations between the two tasks. For example, a poorly reconstructed sensing image from the super-resolution task (e.g., due to low-quality training data or inappropriate models) could easily lead to inaccurate classification labels, which could further reduce the quality of the reconstructed image. Therefore, the integration model has to effectively reduce the noise in the iterations between the two tasks to offer the desirable classification accuracy and reconstructed image quality for the CSC problem.

To address the above challenges, we develop *SCLearn*, an integrated deep learning framework to solve the CSC problem in RUS applications. In particular, we develop an integrated deep CNN architecture that integrates both the classification and super-resolution tasks into a holistic learning framework to *concurrently* boost the performance of both the tasks. Furthermore, we develop a noise-sensitive deep refinement framework to effectively improve the inaccurate class labels and poorly reconstructed images by reducing the noise propagation between the two tasks. To the best of our knowledge, SCLearn is the first solution that effectively integrates the interdependent classification and super-resolution tasks into a holistic solution to boost the performance of both the tasks in RUS applications. We evaluate the SCLearn through a real-world urban land usage classification application over two different cities in Europe (Barcelona and Berlin). The results show that SCLearn consistently outperforms the state-of-the-art baselines by simultaneously achieving better land usage classification accuracy and higher reconstructed image quality under various application scenarios.

A preliminary version of this work was accepted in [19]. We refer to the scheme developed in the conference paper as the SuperClass scheme. The journal article is a significant extension of the previous work in the following aspects. First, we identify two new intrinsic challenges (i.e., complex interdependence and noise amplification) in solving the CSC problem and explicitly discuss how our scheme addresses those two challenges in the introduction (Section I). Second, we extend the class-aware perception-quality refinement (CPR) module in SuperClass by adding a new ensemble learning mechanism to optimize the reconstructed image

quality for the super-resolution task (Section IV). Third, we study the performance of all the compared schemes over two different cities in Europe (i.e., Barcelona and Berlin), while we only studied the performance of SuperClass over Barcelona in the conference paper. We also consider two different low-resolution satellite image evaluation settings (i.e., 56×56 and 112×112), while we only studied a single setting (i.e., 56×56) in the conference paper (Section V). Fourth, we compare SCLearn with four additional recent classification and super-resolution schemes including SuperClass and demonstrate the performance gains achieved by SCLearn compared with all the baselines (Section V). Fifth, we add a new ablation study to evaluate the effect of each component of SCLearn in terms of their contributions to both the classification and super-resolution tasks in addressing the CSC problem (Section V). Finally, we extend the related work by adding discussions on the recent progress in RUS and integrated machine learning, respectively (Section II).

II. Related Work

A. Remote Urban Sensing

Motivated by the state-of-the-art optical sensing and image processing technologies, RUS has emerged as a powerful sensing paradigm to capture a rich set of visual information of the urban environments at an unprecedented scale [1], [20]. Examples of RUS applications include obtaining the structural health conditions of the city bridges using traffic camera for smart urban management [2], detecting severely damaged areas using satellite and UAV images for efficient disaster response [21], monitoring photovoltaic array efficiency in solar power plant using surveillance camera for intelligent manufacturing [22], and sensing city-wide air quality using aerial panoramic images for pollution detection and prevention [23]. Several key challenges exist in the current RUS applications. Examples include data sparsity, image obscurity, noise propagation, and privacy protection [24], [25]. However, the CSC problem remains to be a challenging and unresolved problem in RUS applications. In this article, we develop the SCLearn scheme to address this problem by designing a novel integrated deep learning framework to concurrently improve the performance of both the tasks.

B. Classification and Super-Resolution

Previous efforts in RUS often focus on solving the classification and super-resolution problems separately. For example, Albert *et al.* [5] proposed a deep learning land usage classification framework that uses high-resolution satellite images to learn land usage classes by fine-tuning the CNNs. Vetrivel *et al.* [26] leveraged the damage-specific visual features extracted by the deep neural network to segment and classify the disaster damage severity using incremental learning. Behrendt *et al.* [27] proposed an end-to-end deep learning framework to detect and classify the traffic light for autonomous driving in urban environments using CNNs. Similar examples also exist in super-resolution literature. For example, Tuna *et al.* [13] proposed a deep learning approach that leverages a set of convolutional operations to refine the reconstructed aerial images generated by bicubic interpolation for single image super-resolution. Kawulok *et al.* [28] presented a deep super-resolution framework that implies multiple residual blocks with skip connection to boost the quality of reconstructed high-resolution hyperspectral images. Wang *et al.* [29] designed a novel deep neural network framework that applies the cycle-consistent convolutional network design to capture the complex mapping between low- and high-resolution satellite images in the image reconstruction process. There also exist a couple of initial explorations that use the super-resolution models to help with the classification task [15]–[17]. However, those efforts mainly focus on improving the classification accuracy using reconstructed images from the super-resolution task but ignore the interdependence between the classification and super-resolution tasks. We also note that there exist current works that leverage the synthetic bands to improve the land usage classification performance [30]. Those approaches acquire additional spectral information from channels other than RGB [e.g., near-infrared response (NIR), light detection and ranging (LiDAR)] to facilitate identification of land usage classes through 3-D CNNs. In particular, those approaches leverage the *Extended Multiattribute Profiles* to generate augmented image bands from the limited number of input channels, which can be used by different classifiers [e.g., CNN, joint sparse representation (JSR), support vector machine (SVM)] to effectively identify class-specific visual features for desirable land usage classification performance. However, those efforts primarily focus on improving the classification accuracy using multiband remote sensing data but do not leverage the estimated classification labels to further improve the image resolution by exploring the interdependence between the classification and super-resolution tasks. In contrast, this article develops a novel SCLearn framework to integrate the classification and super-resolution tasks into a holistic framework to improve the performance of both the tasks simultaneously.

C. Integrated Machine Learning

Our work is also related to the integrated machine learning technique that is designed to concurrently improve the performance of interdependent tasks. In particular, integrated machine learning has been applied in domains such as data mining, medical image processing, information retrieval, and computer vision [31]–[34]. For example, Sun *et al.* [31] designed a hybrid information network to jointly improve the performance of ranking and clustering in heterogeneous information network analysis. Girard *et al.* [32] developed a CNN to incorporate blood vessel segmentation with the arteries and veins' classification using fundus images. Müller *et al.* [33] proposed an ontology-driven text-mining framework to integrate information retrieval and extraction in a unified framework to formulate semantic queries for scientific literature search. Yang *et al.* [34] proposed a deep metric learning framework to improve both image retrieval and classification performance in effective image understanding. To our knowledge, SCLearn is the first integrated deep learning approach to solve the CSC problem in RUS applications. In particular, our scheme explicitly addresses the complex interdependence

between the classification and super-resolution tasks by effectively reducing the amplified noise between the two tasks and offers the desirable classification accuracy and reconstructed image quality for RUS applications. Meanwhile, we observe that our work is also related to the techniques from multitask learning, which aims at jointly solving multiple homogeneous learning tasks and exploring the correlations across different tasks to improve the overall task performance [35]. However, multitask learning primarily focuses on jointly solving multiple *homogeneous* learning tasks [35]. For example, multitask learning schemes often focus on jointly solving multiple related classification tasks on the same dataset [36]. In particular, the multitask learning schemes leverage the visual features extracted from each task to supervise the model convergence process of other classification tasks to reduce the classification bias in each task. Therefore, the multitask learning techniques cannot be applied to the *heterogeneous* learning tasks (i.e., super-resolution and classification) that we study in this article. In contrast, we develop an integrated deep learning framework to explicitly explore the interdependence between the two heterogeneous learning tasks to improve the performance of both the tasks.

III. Problem Description

In this section, we formally define the CSC problem in RUS applications. In our CSC problem, we focus on integrating super-resolution with *scene classification* to improve the performance of both the tasks. In particular, the studied land usage classification application focuses on examining surface object variations and fine-grained details of an image to infer its land usage class as perceived by humans instead of identifying low-level pixels or physical objects (e.g., trees and buildings). The scene classification focuses on the visual details of a studied image to infer the class label of the scene captured in the image, which matches well with our application objective [37], [38]. Note that our CSC problem does not study pixel or object classification problems that focus on identifying lower level physical objects (e.g., trees and buildings). This is because the identified physical objects often provide insufficient evidence in classifying the land usage class of a sensing cell. For example, an identified object of a tree cannot help us differentiate the land usage classes between the *urban fabric* and *forest and green land* as they both often contain trees. We first define a few key terms used in problem statement.

Definition 1 (Sensing Cell): Given a studied area (e.g., a city) where we collect the imagery data for the classification and super-resolution integration task, we first divide the studied area into disjoint sensing cells. In particular, a sensing cell represents a subarea of interest. In addition, we define N to be the number of sensing cells from the studied area and n to be the nth sensing cell.

Definition 2 [Low-Resolution Image (L)]: We define L to be the low-resolution image from each sensing cell collected in a specific RUS application. The low-resolution image is usually in a relatively low spatial resolution, which often does not provide sufficient fine-grained details for the classification tasks (e.g., land usage classification). For example, the satellite image in a sensing cell with a spatial resolution of 56×56 is shown in (a) of Fig. 2. In particular, we define L_n to represent the low-resolution image collected from cell n.

(a) (b)

Fig. 2. *Low*- and *high*-resolution images in RUS. (a) Low resolution. (b) High resolution.

Definition 3 [High-Resolution Image (H)]: We define H to be the high-resolution image for each sensing cell with a relatively high resolution. For example, the satellite image in a sensing cell with a spatial resolution of 224×224 is shown in (b) of Fig. 2. The high-resolution image often presents more fine-grained details of the surface objects (e.g., clear building shapes and road layouts), which provides more clear visual evidence for the classification tasks. In particular, we define H_n to be the *actual* high-resolution image of cell n.

Definition 4 [Reconstructed High-Resolution Image ($\widehat{H}$)]: We define $\widehat{H}$ to be the *reconstructed* high-resolution image for each sensing cell. The reconstructed high-resolution image is expected to have the same resolution as the actual high-resolution image H. In particular, we define $\widehat{H_n}$ to be the *reconstructed* high-resolution image for the sensing cell n.

Definition 5 [Class Label (C)]: We define $C = \{C_1, C_2, \ldots, C_N\}$ to represent the set of class labels for all the sensing cells in a specific classification task in an RUS application. In particular, we define C_n to be the class label in sensing cell n. For example, in an urban land usage classification application, we define C_n to be the land usage class (e.g., agriculture) of a sensing cell.

Definition 6 (Class Set): We define $\{1, 2, \ldots, K\}$ to represent the set of all possible classes in a CSC application, where k represents the kth class. In particular, we have the class label C_n of sensing cell n belonging to one of the classes in $\{1, 2, \ldots, K\}$.

Definition 7 [Estimated Class Label ($\widehat{C}$)]: We define $\widehat{C} = \{\widehat{C}_1, \widehat{C}_2, \ldots, \widehat{C}_N\}$ to be the set of *estimated* class labels for all the sensing cells learned by the CSC scheme. In particular, we define $\widehat{C}_n$ to indicate the estimated class label in cell n.

Definition 8 (Perception Error): To evaluate the quality of the reconstructed image $\widehat{H_n}$, we adopt the state-of-the-art perception error metric [39] to measure the perception difference between the *actual* and *reconstructed* images as

$$\mathcal{P}\left(H_n, \widehat{H_n}\right) = \mathcal{F}\left(\mathcal{E}(H_n) - \mathcal{E}\left(\widehat{H_n}\right)\right) \tag{1}$$

where $\mathcal{P}(\cdot)$ represents the perception error metric. $\mathcal{E}(H_n)$ and $\mathcal{E}(\widehat{H_n})$ are the deep features extracted from the *actual* and *reconstructed* images, respectively, using ImageNet-trained deep convolutional networks (e.g., Visual Geometry Group (VGG) [40]). $\mathcal{F}(\cdot)$ represents the error measurement function

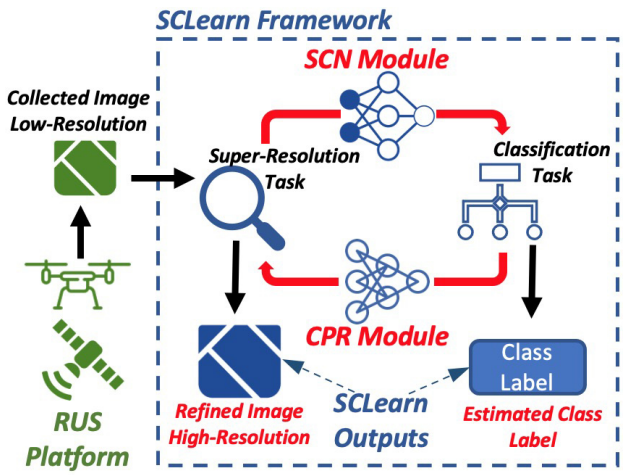


Fig. 3. Overview of the SCLearn framework.

[e.g., mean absolute error (MAE)] to measure the difference between the extracted deep features. This metric has been proven to be robust to capture the perception quality of images [41].

Given the above definitions, the goal of our CSC problem is to integrate both the classification and super-resolution tasks into an integrated learning framework to concurrently boost the performance of both the tasks. In particular, our goal is to correctly learn the class labels while accurately generating high-resolution images of all the sensing cells from the corresponding low-resolution ones. Our problem is formally defined as

$$\arg\max_{\widehat{C_n}} \Pr\left(\widehat{C_n} = C_n \mid L_n\right) \quad \forall 1 \leq n \leq N$$

while

$$\arg\min_{\widehat{H_n}} \left(\mathcal{P}\left(H_n, \widehat{H_n}\right) \middle| L_n\right) \quad \forall 1 \leq n \leq N. \tag{2}$$

IV. Solution

SCLearn is a deep CNN framework that couples the classification task with the super-resolution task into an integrated learning framework to *concurrently* boost the performance of both the tasks. An overview of SCLearn is shown in Fig. 3. It consists of two major modules: 1) *Super-Resolution-Assisted Classification Network (SCN)* and 2) *CPR*. We elaborate on how SCN and CPR work collaboratively to ensure that the classification and super-resolution tasks can achieve mutual promotion as follows.

1) *SCN*: it designs a holistic CNN architecture that focuses on using the super-resolution task to improve the performance of the classification task. In particular, the SCN module designs an SCN that explicitly augments enhanced visual details to optimize intraclass similarity and interclass dissimilarity to boost the classification accuracy. In particular, the SCN contains an upscaling subnetwork and a classification subnetwork that are sequentially concatenated and simultaneously optimized by the joint super-resolution and classification loss function design. As a result, the upscaling subnetwork in SCN primarily focuses on augmenting the visual features that are essential to the classification task instead of improving the resolution for all the objects in the input images, which could bring unexpected noise during the image reconstruction process.
2) *CPR*: it develops a noise-sensitive deep refinement model that focuses on leveraging the classification task to boost the performance of the super-resolution task. In particular, the CPR module introduces a set of paralleled Class-Aware Refinement Networks (CRNs) that work collaboratively to refine the object details in the reconstructed high-resolution images by augmenting the class-specific visual features using their specific class labels generated by the SCN module. In addition, our SCLearn includes a novel aggregated loss function design that jointly optimizes all the CPR networks to collaboratively refine the reconstructed high-resolution images under a probabilistic learning framework to ensure the desirable perceptual quality of the reconstructed images.

A. Super-Resolution-Assisted Classification Network

In this section, we present the super-resolution-assisted classification network (CN) architecture in SCLearn to explicitly reconstruct high-resolution images with enhanced visual details to boost the classification accuracy. In particular, the SCN architecture incorporates two components: an upscaling network (UN) and a CN. In particular, *UN* first generates the reconstructed high-resolution images with enhanced visual details to maximize intraclass similarity and interclass dissimilarity. Then, *CN* is used to learn the class label of each sensing cell using enhanced visual details generated by *UN*. In particular, we formally define *UN* and *CN* as follows:

Definition 9 [Upscaling Network]: We define *UN* as a generative network that reconstructs a high-resolution image $\widehat{H}$ from a corresponding low-resolution image L with enhanced visual details as follows:

$$\widehat{H} = \mathrm{UN}(L). \tag{3}$$

We show an example of *UN* in (a) of Fig. 4. In particular, it consists of a set of residual blocks to carefully segment each individual object (e.g., tree, road, and building) in an image and apply enhanced visual details that maximize intraclass similarity and interclass dissimilarity to each identified object (e.g., a layout of a road often indicates a land usage of transportation).

Definition 10 [Classification Network]: We define *CN* as a CN that estimates the class label of each sensing cell using enhanced visual details generated by *UN* as

$$\widehat{C} = \mathrm{CN}(\mathrm{UN}(L)) \tag{4}$$

where $\widehat{C}$ is the estimated class label. We show an example of *CN* in (b) of Fig. 4. It consists of an ImageNet pretrained deep CNN (i.e., VGG) with multiple trainable convolutional layers for visual feature extraction. This is done to ensure that the *CN* can accurately identify the key visual features enhanced by *UN* for accurate classification task. Then, the *CN* consists

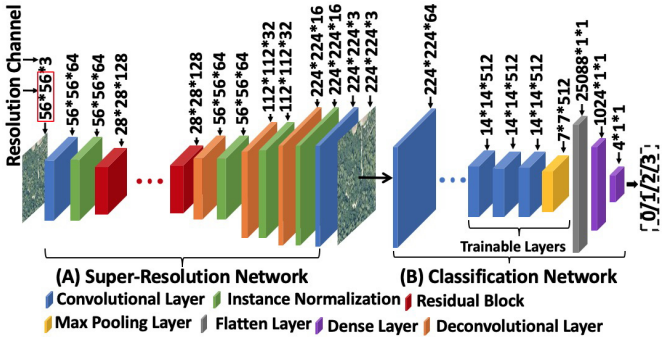


Fig. 4. Illustrations of network architectures in SCN.

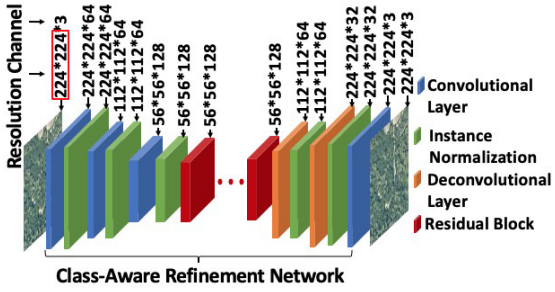


Fig. 5. Illustrations of network architectures in CPR.

of a flatten layer and multiple dense layers, which are used to learn the accurate class labels using the extracted key visual features.

Given the two network architectures above, our next question is how to learn the optimal instances of all the networks that reconstruct high-resolution images with enhanced visual details to maximize the classification accuracy. To address this question, we define two sets of loss functions in our SCN module. In particular, we first define the classification loss for the *UN* and *CN* as follows:

$$\mathcal{L}^{\text{CL}}_{\text{UN,CN}}\colon \mathcal{L}_{\text{cross-entropy}}(C, \text{CN}(\text{UN}(L))) \tag{5}$$

where C indicates the ground-truth class label for each sensing cell. $\mathcal{L}_{\text{cross-entropy}}$ indicates the cross entropy loss [42] that measures the difference between the ground-truth and estimated class labels for each sensing cell. Intuitively, the design of $\mathcal{L}^{\text{CL}}_{\text{UN,CN}}$ is to ensure that the UN can explicitly enhance visual details so that the CN can effectively learn the class label using enhanced image details. In addition to $\mathcal{L}^{\text{CL}}_{\text{UN,CN}}$, we also consider an upscaling loss $\mathcal{L}^{\text{UP}}_{\text{UN}}$ for the UN to further validate the quality of the reconstructed images as

$$\mathcal{L}^{\text{UP}}_{UN}\colon \mathcal{L}_{\text{pixel-loss}}(H, UN(L)) \tag{6}$$

where H and L indicate the actual high-resolution and collected low-resolution images, respectively. $\mathcal{L}_{\text{pixel-loss}}$ represents the pixel-wise RGB value difference between the actual and reconstructed high-resolution images (e.g., mean square error (MSE) loss [43]). Intuitively, $\mathcal{L}^{\text{UP}}_{UN}$ is designed to ensure the stable performance of UN of generating the reconstructed high-resolution images that are close to the actual ones.

Finally, we combine the above two sets of loss functions to derive the final loss $\mathcal{L}^{\text{Final}}_{\text{SCN}}$ for the SCN module as follows:

$$\mathcal{L}^{\text{Final}}_{\text{SCN}}\colon \mathcal{L}^{\text{CL}}_{\text{UN,CN}} + \mathcal{L}^{\text{UP}}_{\text{UN}}. \tag{7}$$

Using the above loss function, we can learn the optimal instances (i.e., UN*, CN*) of all the networks using the adaptive moment estimation (ADAM) optimizer [44]. Finally, we use UN* and CN* to estimate the class label for each sensing cell using the collected low-resolution image L as follows:

$$\widehat{C} = \text{CN}^*\big(\text{UN}^*(L)\big). \tag{8}$$

B. Class-Aware Perception-Quality Refinement

In Section IV-A, we present the SCN module that reconstructs high-resolution images with enhanced visual details to improve the classification performance. However, the reconstructed images from the SCN module are often noisy since they are generated by a one-size-fits-all super-resolution model that is trained with images from all possible classes [18]. Therefore, our next question here is that can we further improve the quality of the reconstructed high-resolution images by leveraging the class labels output by the SCN module? To address this question, we design a set of CRNs that judiciously refine the reconstructed high-resolution images by leveraging their specific class labels. In particular, we first formally define a CRN as follows.

Definition 11 [Class-Aware Refinement Network]: We define CRN^k as a refinement network to refine the reconstructed high-resolution images for a specific class k to improve the reconstructed image quality as follows:

$$\widehat{H}^k_{\text{refine}} = \text{CRN}^k\left(\widehat{H}^k\right) \tag{9}$$

where $\widehat{H}^k$ is the reconstructed high-resolution image generated by the UN in SCN (defined in Definition 9) of a specific class k. $\widehat{H}^k_{\text{refine}}$ is the refined image generated by CRN^k with an improved image quality compared with $\widehat{H}^k$. We show an example of *CRN* in Fig. 5. In particular, it uses multiple residual blocks to ensure the desired depth of the *CRN*, making it sensitive to the noise in the reconstructed images and capable of refining the visual details in the reconstructed images.

Given the CRN architecture above, our next question is how to learn the optimal instance of the CRN^k for each class k to maximize the reconstructed image quality. To address this question, we define the refinement loss function for the CRN^k of class k as follows:

$$\mathcal{L}^{\text{RF}}_{\text{CRN}^k}\colon \mathcal{L}_{\text{pixel-loss}}\left(H^k, \text{CRN}^k\left(\widehat{H}^k\right)\right) + \mathcal{L}_{\text{perc-loss}}\left(H^k, \text{CRN}^k\left(\widehat{H}^k\right)\right) \tag{10}$$

where $\mathcal{L}^{\text{RF}}_{\text{CRN}^k}$ indicates the refinement loss function for CRN^k. H^k indicates the actual high-resolution image of a specific class k. $\widehat{H}^k$ is the reconstructed high-resolution image generated by *UN* for class k. $L_{\text{pixel-loss}}(\cdot)$ is the pixel-wise MSE loss as defined in (6). $\mathcal{L}_{\text{perc-loss}}(\cdot)$ is the perception loss [39]

to quantify the perceptual difference between the actual and refined images. Intuitively, $\mathcal{L}^{\text{RF}}_{\text{CRN}^k}$ is designed for each class k to ensure CRN^k can effectively refine the reconstructed high-resolution images for better quality.

Our next question is how can we best leverage the class-aware refinement network above to refine reconstructed image using the class labels output by the SCN module? We first define a few key terms below.

Definition 12 [Refinement Network Set (RS)]: We define $\text{RS} = \{\text{CRN}^1, \text{CRN}^2, \ldots, \text{CRN}^K\}$ to represent a set of CRNs for all available classes in a CSC application.

Definition 13 [Refinement Image Candidates (RCs)]: Given a reconstructed image $\widehat{H}$ generated by the SCN module, we define $\text{RC}(\widehat{H}) = \{\text{CRN}^1(\widehat{H}), \text{CRN}^2(\widehat{H}), \ldots, \text{CRN}^K(\widehat{H})\}$ to represent a set of refined images generated by all the refinement networks in RS with the input reconstructed image $\widehat{H}$.

Given the RCs, a straightforward way is to directly select the $\text{CRN}^k(\widehat{H})$ from the RC as the refined image for $\widehat{H}$ if the class label of $\widehat{H}$ is estimated to be k (i.e., $\widehat{C} = k$) in the *SCN* module. However, a critical problem of such a deterministic solution is that it is not robust to the noisy estimated class labels generated by the *SCN* module. In particular, any inaccurate class labels from the *SCN* module will directly lead the *CPR* module to select the RC generated by a refinement network of a different class. As a result, the refinement network could impair the refined image quality by adding noise from a different class with very different visual details. To address this problem, we propose a probabilistic solution to handle noisy class labels. We first define a key term in our solution.

Definition 14 [Ensembled High-Resolution Image ($\widehat{H}_{ensemble}$)]: We define $\overline{H}_{\text{ensemble}}$ to be a high-resolution image, where the RGB value at each pixel is a combination of the RGB values from RCs in RC as follows:

$$\widehat{H}_{\text{ensemble}} = \sum_{k=1}^{K} \text{CRN}^k\left(\widehat{H}\right) \cdot \mathcal{W}^k\left(\widehat{H}\right) \tag{11}$$

where $\mathcal{W}^k(\widehat{H})$ indicates the weight of the RGB values of each RC $\text{CRN}^k(\widehat{H})$ in the ensembled high-resolution image. The key question now is how to derive the value of each $\mathcal{W}^k(\widehat{H})$ to optimize the quality of the ensembled image $\widehat{H}_{\text{ensemble}}$. To that end, we set the weight as follows:

$$\mathcal{W}^k\left(\widehat{H}\right) = \Pr\left(\text{CN}\left(\widehat{H}\right) = k\right) \tag{12}$$

where $\Pr(\text{CN}(\widehat{H}) = k)$ is the probability of the CN (defined in Definition 10) estimates $\widehat{H}$ to be a specific class k. Such a probability of $\Pr(\text{CN}(\widehat{H}) = k)$ can be obtained using the deep features generated by the CN after the final softmax activation function [45]. Intuitively, such a probabilistic design ensures $\widehat{H}_{\text{ensemble}}$ can still use the RGB values from the correct RCs even when the estimated class label from the CN is noisy.

Next, we define a loss function $\mathcal{L}_{\text{ensemble}}$ to validate the perceptual quality of the ensembled image as follows:

$$\mathcal{L}_{\text{ensemble}}\colon \mathcal{L}_{\text{perc-loss}}\left(H, \widehat{H}_{\text{ensemble}}\right) \tag{13}$$

where $\mathcal{L}_{\text{perc-loss}}(H, \widehat{H}_{\text{ensemble}})$ is the loss function to measure the perceptual difference between the actual and ensembled images.

Finally, we briefly discuss the optimization process of the CPR module to learn the optimal parameters of all the CRNs (i.e., $\text{CRN}^{1*}, \text{CRN}^{2*}, \ldots, \text{CRN}^{K*}$) based on the loss functions defined above. We first define an aggregated loss function for our CPR module as

$$\mathcal{L}^{\text{Final}}_{\text{CPR}} = \sum_{k=1}^{K} \mathcal{L}^{\text{RF}}_{\text{CRN}^k} + \mathcal{L}_{\text{ensemble}}. \tag{14}$$

The aggregated loss function combines all the loss functions defined in the CPR module, i.e., $\mathcal{L}^{\text{RF}}_{\text{CRN}^k}$ [defined in (10)] and $\mathcal{L}_{\text{ensemble}}$ [defined in (13)]. The minimization of the aggregated loss ensures all the CRNs generate high-quality ensembled images. The loss function $\mathcal{L}^{\text{Final}}_{\text{CPR}}$ can be optimized using the ADAM optimizer, which obtains the optimal parameters of all the class-aware-refinement networks.

While our current CPR module requires a CRN for each class in the studied application, the overall time and space overhead of our CPR module is still manageable for two reasons. First, each CRN in our CPR module has a reasonable time and space overhead. In particular, each CRN consists of a constant number of convolutional and deconvolutional layers. Each layer in CRN has a time and space overhead of $O(n)$ for CNNs where n indicates the resolution $\times$ channels of input features at each layer [46]. Second, we observe that the total number of classes in many RUS applications is limited (e.g., a land usage classification application often contains less than ten land usage classes and a disaster damage assessment application often contains three to five different damage severity levels). As a result, our SCLearn often only needs a small number of CPR networks to refine and improve the reconstructed image quality. We studied the time and space overhead of our CRN in the evaluation section below.

C. Summary of SCLearn Framework

The SCLearn is summarized in Algorithm 1. The input to SCLearn is the low-resolution image L for each sensing cell. The output is the estimated class label $\widehat{C}$ and the ensembled high-resolution image $\widehat{H}_{\text{ensemble}}$ for each cell.

V. Evaluation

In this section, we conduct extensive experiments on a real-world RUS application: urban land usage classification using satellite images to answer the following research questions.

1) Q1: Can SCLearn achieve a better classification accuracy than the state-of-the-art classification baselines in RUS applications?
2) Q2: Can SCLearn concurrently achieve a better reconstructed image quality compared with the state-of-the-art super-resolution baselines?
3) Q3: How does each component of SCLearn design contribute to its overall performance?
4) Q4: How does SCLearn perform in terms of both classification and super-resolution on different land usage classes?

Algorithm 1 SCLearn Framework Summary

input: L
output:$\widehat{C}$, $\widehat{H}_{ensemble}$
▷ *SCN phase*
initialize UN
initialize CN
for each epoch **do**
  for each batch **do**
    optimize UN and CN
  end for
end for
obtain UN^*, CN^*
generate $\widehat{H}$ using UN^*
generate $\widehat{C}$ using CN^*
▷ *CPR phase*
initialize $CRN^1, CRN^2, \ldots, CRN^K$
obtain $\{\mathcal{W}^1(\widehat{H}), \mathcal{W}^2(\widehat{H}), \ldots, \mathcal{W}^K(\widehat{H})\}$ from CN^*
for each epoch **do**
  for each batch **do**
    optimize $CRN^1, CRN^2, \ldots, CRN^K$
  end for
end for
obtain $CRN^{1*}, CRN^{2*}, \ldots, CRN^{K*}$
generate $\widehat{H}_{ensemble}$ using $CRN^{1*}, CRN^{2*}, \ldots, CRN^{K*}$
output $\widehat{C}$, $\widehat{H}_{ensemble}$

5) Q5: What are the impacts of image noise on the overall performance of SCLearn?

A. Dataset

We evaluate SCLearn on a real-world land usage classification application. In particular, we use the land usage datasets collected from two different cities in Europe (Barcelona, Spain, and Berlin, Germany). The datasets consist of four different land usage classes (*urban fabric*, *transportation*, *forest and green land*, and *agriculture* as shown in Fig. 6). We summarize the datasets as follows[1].

1) *Google Maps Satellite Image:* We collect the satellite imagery datasets from *Barcelona* and *Berlin* using publicly available Google Maps application programming interface (API).[2] In particular, we first divide a city into disjoint sensing cells (Definition 1). Each collected satellite image is in a 224 × 224 resolution with a 250 m × 250-m ground coverage for each sensing cell, which is considered as the *high-resolution* image in our evaluation. We then follow the standard process that is widely used in the state-of-the-art super-resolution literature to generate the *low-resolution* satellite images in our experiments [47], [48]. In particular, we adopt the widely used *bicubic interpretation* tool from the *scikit-image* package[3] to reduce the resolution of a high-resolution satellite image as the *low-resolution* satellite image. In particular, we consider two types of low-resolution satellite image settings in our evaluation: 1) *scale factor* = 2 × 2: we reduce the resolution of a collected satellite image by 2 × 2 times to generate a *low-resolution* image with a 112 × 112 resolution; 2) *scale factor* = 4 × 4: we reduce the resolution of a collected satellite image by 4 × 4 times to generate a *low-resolution* image with a 56 × 56 resolution. Finally, we randomly select 2400 images (i.e., 800 from high resolution, 800 from low resolution of each scale factor) from the studied area for our experiments. We then follow the standard deep learning training procedure by randomly sampling 70% of images as the training data and 30% of images as the testing data.
2) *Urban Atlas Land Usage Data:* Following the procedures in [5], we obtain the ground-truth labels of land usage for each sensing cell in our studied cities from the publicly available land usage dataset (i.e., *Urban Atlas* dataset) published by the European Environment Agency.[4] In addition, we adopt the commonly used mapping method in [5] to establish a one-to-one match between the satellite imagery data and land usage label of each sensing cell.

Fig. 6. Examples of land usage from Barcelona and Berlin.

B. Baseline

We compare SCLearn with a rich set of state-of-the-art baselines that are widely used in the previous literature for both the land usage classification and satellite imagery super-resolution tasks. In our experiments, we keep the same inputs to all the compared schemes for a fair comparison. In particular, the inputs to a scheme include 1) the studied low-resolution satellite imagery data and 2) the high-resolution satellite imagery data and land usage class labels in the training dataset.

1) Land Usage Classification:

1) *InceptionResNet [49]:* a recent land usage classification model that integrates the inception architectures with residual block design in a holistic deep neural network architecture for effective land usage classification.
2) *DenseNet [50]:* a deep land usage classification model that achieves dense connections between convolutional layers by incorporating a feed-forward mechanism.
3) *Neural search architecture network (NASNet) [51]:* a dynamic deep convolutional network that adjusts its convolutional network architecture for an optimized land usage classification performance.
4) *VGG [5]:* a popular deep convolutional network architecture that uses intensive sequential convolutional operations to boost land usage classification accuracy.

[1]We will make our datasets and codes publicly available on Github upon the acceptance of the article.

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

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

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

5) *Radius-adaptive convolutional neural network (RACNN) [15]:* a joint super-resolution and classification framework that leverages convolutional image super-resolution operations to improve classification results.
6) *EfficentNetV2 [52]:* a recent deep classification framework that uses a lightweight neural architecture search and scaling design to achieve fast model training and optimized classification accuracy.
7) *ResNeSt [53]:* a recent convolutional classification framework that leverages a split-attention network design to ensure a desirable classification accuracy.

In addition, we also consider the *Random* baseline, which estimates the land usage class of a sensing cell by randomly choosing a land usage class from all available class candidates.

2) Super-Resolution:

1) *Nearest-Neighbor [54]:* a conventional satellite image super-resolution model that applies the RGB values from the nearest neighboring pixels to improve the spatial resolution of the input satellite image.
2) *Bilinear/Bicubic [55]:* a set of widely used super-resolution solutions that apply the bilinear/bicubic upscaling operations to refine the visual details of the image.
3) *SFSR18 [13]:* a recent deep super-resolution solution that uses a set of recursive convolutional operations to refine the reconstructed satellite images.
4) *SRRES19 [28]:* a powerful deep super-resolution framework that uses multiple residual blocks with skip connection to boost the quality of reconstructed images.
5) *CycleCNN19 [29]:* a novel deep learning framework that applies the cycle-consistent neural network design to improve the reconstructed high-resolution image quality.
6) *Enhanced super-resolution generative adversarial network (ESRGAN) [56]:* a state-of-the-art GAN framework that uses residual-in-residual dense block and relativistic GAN to ensure desirable reconstructed image quality.
7) *Super-resolution residual convolutional generative adversarial network (SRResCGAN) [57]:* a new generative adversarial learning approach that introduces a deep cyclic generative adversarial residual network design to capture the complex mapping between low- and high-resolution images in the image reconstruction process.
8) *Unfolding super-resolution network (USRNet) [58]:* a recent deep-learning-based super-resolution framework that uses a deep unfolding network design to ensure a good quality of the reconstructed image.
9) *Blind image super-resolution generative adversarial network (BSRGAN) [59]:* a deep degradation super-resolution model that integrates shuffled blur, downsampling, and noise degradation to improve the visual details of the reconstructed image.

C. Evaluation Metrics

1) Classification Metrics: To evaluate the land usage classification performance, we adopt four representative metrics for multiclass classification problem in our evaluation: 1) *micro-F1*, 2) *macro-F1*, 3) *Cohen's kappa score (K-score)* [60], and 4) *Matthews correlation coefficient (MCC)* [61]. Intuitively, a higher value of *micro-F1*, *macro-F1*, *K-score*, and *MCC* indicates a better classification performance.

2) Super-Resolution Metrics: The land usage classification application studied in this article focuses on examining the surface object variations and fine-grained details of an image (i.e., the scene classification task) to infer its land usage class as perceived by humans instead of identifying low-level pixels or physical objects (e.g., trees and buildings) in the image. Hence, we select the perceptual metric (Definition 8), which has been proven to be close to human perception in evaluating the performance of super-resolution schemes [39], [41]. Note that we do not use the pixel-wise evaluation metrics [e.g., peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM)] in our evaluation because they have been proven to be inappropriate to measure the actual perceptual quality of the reconstructed images [43]. In particular, following [39], [41], we select four commonly used deep features [i.e., $\mathcal{E}(\cdot)$ in (1)] extracted by the 13th to 16th convolutional layers in the VGG19 model. We refer to them as *deep feature 1 (DF_1)* to *deep feature 4 (DF_4)* in the evaluation. We adopt the commonly used error measurement function [i.e., MAE as the $\mathcal{F}(\cdot)$ in (1)] to calculate the difference between the deep features extracted from the *actual* and *reconstructed* images. A lower value in the error metrics indicates a better super-resolution performance.

D. Evaluation Results

1) Q1 (Performance Comparisons on Land Usage Classification): In the first set of experiments, we evaluate the performance of all the compared schemes in estimating the land usage classes in the studied area. The evaluation results are presented in Tables I and II. We observe that the SCLearn scheme consistently outperforms all compared baselines in all the studied cities with different scale factors. For example, the performance gains of SCLearn over the best performing baseline (i.e., EfficientNetV2) in Barcelona at scale factor $= 4 \times 4$ on micro-F1, macro-F1, K-score, and MCC are 4.65%, 4.96%, 5.48%, and 6.72%, respectively. Such performance gains mainly come from the fact that SCLearn integrates both the super-resolution and classification tasks into an integrated learning framework to improve the accuracy of land usage classification. In particular, our scheme incorporates a set of collaborative upscaling and classification convolutional layers that are dedicated to reconstructing high-resolution images with refined visual details to boost the classification accuracy. We also note that the hybrid baselines (i.e., RACNN) that leverage the reconstructed HR images for classification tasks do not always improve the classification accuracy. This is because the classification result could be wrong when the reconstructed HR images generated by a one-size-fits-all super-resolution model are of low quality. We also present the visual results of our SCLearn compared with the best performing classification baseline (i.e., RACNN) in Fig. 7. We observe that our SCLearn can accurately identify a rich set of land usage class labels that RACNN misclassifies. The visual results further demonstrate the capability of our

TABLE I
LAND USAGE CLASSIFICATION PERFORMANCE COMPARISONS (CITY = BARCELONA)

		Scale Factor = 2×2				Scale Factor = 4×4			
Class	Algorithm	Mirco-F1	Marco-F1	K-Score	MCC	Mirco-F1	Marco-F1	K-Score	MCC
Random	Random	0.2833	0.2431	0.0444	0.0472	0.2750	0.2362	0.0332	0.0354
Deep Learning	InceptionResNet	0.6666	0.6450	0.5599	0.5725	0.6416	0.5727	0.5259	0.5576
	DenseNet	0.6500	0.5994	0.5371	0.5556	0.5333	0.4971	0.3811	0.4164
	NASNet	0.7083	0.6915	0.6144	0.6230	0.5750	0.5626	0.4380	0.4480
	VGG	0.7333	0.7354	0.6365	0.6467	0.7583	0.7525	0.6697	0.6848
	RACNN	0.7666	0.7683	0.6888	0.6906	0.7416	0.7435	0.6555	0.6585
	EfficientNetV2	0.7768	0.7753	0.7024	0.7032	0.7618	0.7572	0.6805	0.6813
	ResNeSt	0.7500	0.7429	0.6666	0.6682	0.7250	0.7126	0.6333	0.6439
Our Model	**SCLearn**	**0.7833**	**0.7860**	**0.7134**	**0.7113**	**0.8083**	**0.8068**	**0.7353**	**0.7485**

TABLE II
LAND USAGE CLASSIFICATION PERFORMANCE COMPARISONS (CITY = BERLIN)

		Scale Factor = 2×2				Scale Factor = 4×4			
Class	Algorithm	Mirco-F1	Marco-F1	K-Score	MCC	Mirco-F1	Marco-F1	K-Score	MCC
Random	Random	0.2731	0.2677	0.0413	0.0202	0.2518	0.2668	0.0431	0.0306
Deep Learning	InceptionResNet	0.5570	0.5537	0.4043	0.4094	0.6442	0.6611	0.5489	0.5658
	DenseNet	0.5076	0.5455	0.3935	0.4207	0.5209	0.5785	0.4371	0.4780
	NASNet	0.5289	0.4909	0.3722	0.3887	0.5310	0.5867	0.4503	0.4810
	VGG	0.7291	0.7272	0.6364	0.6380	0.7773	0.7768	0.7025	0.7033
	RACNN	0.7414	0.7416	0.6555	0.6562	0.8083	0.8084	0.7444	0.7479
	EfficientNetV2	0.6776	0.6772	0.5703	0.5709	0.7768	0.7753	0.7024	0.7032
	ResNeSt	0.6583	0.6620	0.5444	0.5462	0.7916	0.7886	0.7222	0.7268
Our Model	**SCLearn**	**0.7656**	**0.7667**	**0.6889**	**0.6913**	**0.8392**	**0.8416**	**0.7889**	**0.7898**

TABLE III
SUPER-RESOLUTION PERFORMANCE COMPARISONS (STUDIED CITY = BARCELONA)

		Scale Factor = 2×2				Scale Factor = 4×4			
		Mean Absolute Error (MAE)				Mean Absolute Error (MAE)			
Class	Algorithm	DF_1	DF_2	DF_3	DF_4	DF_1	DF_2	DF_3	DF_4
Conventional	Nearest-neighbor	4.0274	3.0621	1.8714	0.4705	7.2994	5.6831	3.5184	0.8331
	Bilinear	4.6253	3.3179	1.9870	0.4963	7.4280	5.4063	3.2417	0.7784
	Bicubic	4.4053	3.1702	1.9050	0.4763	7.4624	5.3640	3.2286	0.7764
Deep Learning	SFSR18	4.0150	2.8870	1.7447	0.4525	6.8228	4.9113	2.9848	0.7620
	SRRES19	3.7892	2.7491	1.6726	0.4366	7.1054	5.1125	3.1034	0.7653
	CycleCNN19	3.9348	2.8753	1.7424	0.4442	6.9456	5.0434	3.0643	0.7565
	ESRGAN	3.3573	2.5564	1.6179	0.4238	6.1567	4.8142	3.0435	0.7428
	SRResCGAN	4.5091	3.2442	1.9515	0.4912	7.4267	5.4057	3.2417	0.7784
	USRNet	3.9065	2.7850	1.6622	0.4305	7.1183	5.1084	3.0717	0.7704
	BSRGAN	5.2022	3.9941	2.5307	0.6516	6.7797	5.2530	3.3259	0.8457
Our Model	**SCLearn**	**3.2164**	**2.4356**	**1.5418**	**0.4060**	**5.5823**	**4.2390**	**2.6779**	**0.6909**

SCLearn in accurately identifying the land usage class of the studied area.

2) Q2 (Performance Comparisons on Super-Resolution): In the second set of experiments, we further evaluate the performance of all the compared schemes in accomplishing the super-resolution task. The evaluation results are shown in Tables III and IV. We observe that SCLearn consistently outperforms all the compared super-resolution baselines over different city and scale factor settings. For example, the performance gains achieved by SCLearn compared with the

TABLE IV
SUPER-RESOLUTION PERFORMANCE COMPARISONS (STUDIED CITY = BERLIN)

Class	Algorithm	Scale Factor = 2 × 2 Mean Absolute Error (MAE) DF_1	DF_2	DF_3	DF_4	Scale Factor = 4 × 4 Mean Absolute Error (MAE) DF_1	DF_2	DF_3	DF_4
Conventional	Nearest-neighbor	4.2879	3.2687	1.9815	0.4873	6.8480	5.3954	3.3505	0.7404
	Bilinear	4.9799	3.6157	2.1473	0.5220	7.3520	5.3193	3.1757	0.7171
	Bicubic	4.7649	3.4582	2.0560	0.5027	7.4354	5.3163	3.1714	0.7181
Deep Learning	SFSR18	4.2883	3.0873	1.8387	0.4694	6.8612	4.8864	2.9235	0.7013
	SRRES19	4.0856	2.9561	1.7786	0.4537	7.0642	5.0701	3.0426	0.7077
	CycleCNN19	4.3078	3.1583	1.8903	0.4742	6.8663	4.9911	2.9962	0.7025
	ESRGAN	3.5506	2.7306	1.7312	0.4433	5.8152	4.5496	2.8698	0.6804
	SRResCGAN	4.8667	3.5310	2.1019	0.5153	7.3520	5.3196	3.1764	0.7172
	USRNet	4.3587	3.1065	1.8507	0.4565	6.8958	4.9493	2.9680	0.6941
	BSRGAN	5.0305	3.8281	2.3978	0.5844	6.2882	4.8575	3.0519	0.7236
Our Model	**SCLearn**	**3.4737**	**2.6437**	**1.6651**	**0.4279**	**5.4805**	**4.1851**	**2.6302**	**0.6385**

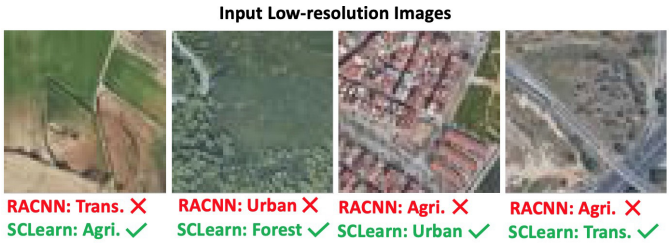


Fig. 7. Examples of classification results for SCLearn.

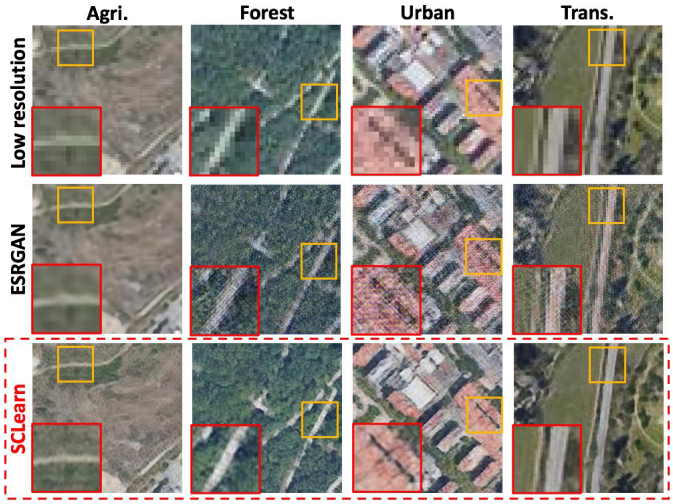


Fig. 8. Examples of super-resolution results for SCLearn.

best performing baseline (i.e., ESRGAN) for Barcelona at scale factor = 4 × 4 with DF_1, DF_2, DF_3, DF_4 (i.e., deep features extracted by the 13th to 16th convolutional layers in VGG19) are 10.28%, 13.57%, 13.65%, and 7.51%, respectively. Such consistent performance gains over various scenarios demonstrate the effectiveness of the CPR network design in SCLearn. We observe that the reconstructed images from the ESRGAN model are suboptimal compared with SCLearn because the reconstructed images are generated by a one-size-fits-all ESRGAN model that is trained with images from all possible land usage classes. As a result, ESRGAN could impair the refined image quality and introduce undesirable noise by adding fine-grained visual details from other land usage classes with very different visual characteristics. In contrast, our SCLearn designs a set of principled CRNs in the CPR module that works collaboratively to refine the object details in the reconstructed high-resolution images by augmenting class-specific visual features using their specific class labels generated by the SCN module. In addition, our SCLearn includes a novel aggregated loss function design [i.e., (14)] that jointly optimizes all the CPR networks to collaboratively refine the reconstructed high-resolution image under a probabilistic learning framework. We also observe that the size of each CRN in our SCLearn is only 25.74 MB and it only requires an average of 11.50 s/epoch to train our CRNs on a single NVIDIA Quadro RTX 6000 GPU, which is a reasonable space and time overhead of deep learning models for super-resolution and classification tasks in RUS applications [62]. In addition, we present the reconstructed images of our SCLearn compared with the best performing super-resolution baseline (i.e., ESRGAN) in Fig. 8. We observe that our SCLearn achieves an improved reconstructed image quality by successfully augmenting fine-grained details while avoiding unexpected noisy points. Such a visual quality improvement further demonstrates the effectiveness of our CPR network that effectively refines the reconstructed images to boost the perceptual quality of reconstructed images. The land usage classification application studied in this article focuses on examining the surface object variations and fine-grained details of an image (i.e., the scene classification task) to infer its land usage class as perceived by humans instead of identifying low-level pixels or physical objects (e.g., trees and buildings) in the image. We further include the evaluation results using the traditional pixel-wise evaluation metrics (i.e., PSNR and SSIM). The

TABLE V
PERFORMANCE COMPARISONS ON PIXEL-WISE EVALUATION METRICS (STUDIED CITY = BARCELONA)

	Scale Factor = 2 × 2		Scale Factor = 4 × 4	
Algorithm	PSNR	SSIM	PSNR	SSIM
Nearest-neighbor	23.8841	0.6620	21.2279	0.3861
Bilinear	24.4417	0.6632	21.7376	0.3952
Bicubic	24.7768	0.6955	21.9307	0.4257
SFSR18	25.0374	0.7123	22.1994	0.4520
SRRES19	25.0633	0.7137	22.0501	0.4423
CycleCNN19	24.8876	0.7045	22.0192	0.4405
ESRGAN	18.6723	0.3816	16.8636	0.1476
SRResCGAN	24.4763	0.6653	21.7380	0.3953
USRNet	25.2332	0.7197	21.4553	0.3764
BSRGAN	21.1487	0.3686	18.8488	0.1773
SCLearn	24.1398	0.6595	21.4154	0.3950

TABLE VI
PERFORMANCE COMPARISONS ON PIXEL-WISE EVALUATION METRICS (STUDIED CITY = BERLIN)

	Scale Factor = 2 × 2		Scale Factor = 4 × 4	
Algorithm	PSNR	SSIM	PSNR	SSIM
Nearest-neighbor	23.7178	0.6281	22.4703	0.4160
Bilinear	24.1072	0.6234	22.8791	0.4251
Bicubic	24.3578	0.6529	22.9841	0.4451
SFSR18	24.7151	0.6760	23.1526	0.4616
SRRES19	24.5594	0.6692	22.9682	0.4478
CycleCNN19	24.4758	0.6656	22.9396	0.4468
ESRGAN	18.3435	0.3396	17.0213	0.1566
SRResCGAN	24.1387	0.6261	22.8801	0.4254
USRNet	25.3394	0.6726	22.7299	0.4417
BSRGAN	22.0482	0.4045	20.2937	0.2582
SCLearn	23.6796	0.6149	22.1967	0.3978

evaluation results are shown in Tables V and VI. In addition, we show the visual result comparison between our model and the best performing baseline in terms of PSNR and SSIM (i.e., SFSF18) in Fig. 9. We observe that our SCLearn achieves a clearly better perceptual quality of the reconstructed images (e.g., avoiding making the image blurry) than SFSF18 despite the fact that SFSF18 has better PSNR and SSIM performance. Our observation also matches with the findings in current super-resolution literature that reports the pixel-wise evaluation metrics are *inappropriate* to measure the actual *perceptual quality* of the reconstructed images [39], [41].

3) Q3 (Ablation Study of SCLearn Scheme): In the third set of experiments, we perform a comprehensive ablation study to evaluate whether the key designs in our SCLearn can effectively explore the interdependence between the classification and super-resolution tasks to achieve the mutual promotion for both the tasks. First, we present the classification results by removing the UN (i.e., w/o UN) in SCN, where we use

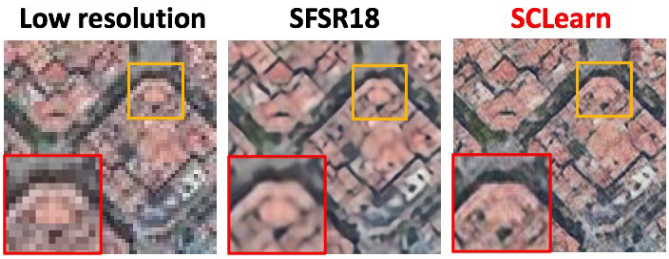


Fig. 9. Visual comparisons between SFSR18 and SCLearn.

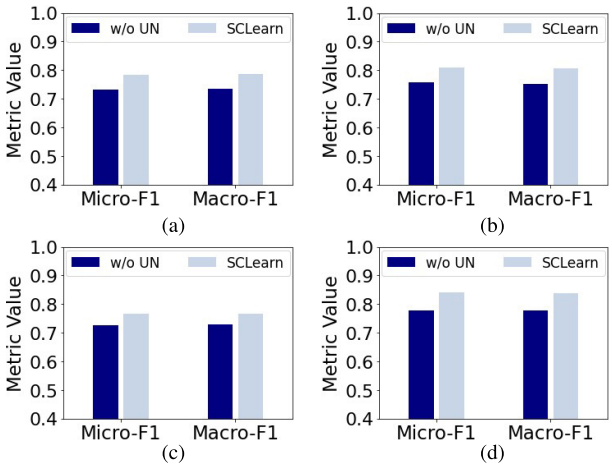


Fig. 10. Ablation study of SCLearn on classification. (a) Barcelona Scale Factor = 2 × 2. (b) Barcelona Scale Factor = 4 × 4. (c) Berlin Scale Factor = 2 × 2. (d) Berlin Scale Factor = 4 × 4.

low-resolution images for classification tasks. The results are shown in Fig. 10. We observe that the UN design makes essential contributions in improving the classification performance, which indicates the effectiveness of our SCLearn in explicitly leveraging the super-resolution task to improve the classification performance. Second, we present the super-resolution results by removing the CRN (i.e., w/o CRN) in CPR, where we do not use the classification results to improve the super-resolution tasks and output the high-resolution images reconstructed by the *one-size-fits-all* UN in SCN. The results are shown in Fig. 11. We observe that the CRN component effectively reduces the perceptual errors of the reconstructed high-resolution images, which indicates the effectiveness of our SCLearn in leveraging the classification tasks to improve the super-resolution performance in a closed-loop learning framework design.[5]

4) Q4 (Per-Class Performance of SCLearn Scheme): In the fourth set of experiments, we study the *per-class* performance of the SCLearn scheme by plotting the confusion matrix for the classification task and the class-wise perceptual error for the super-resolution task under various evaluation settings (i.e., different cities and scale factors). First, the confusion matrices for the classification task are presented in Fig. 12. We observe that SCLearn achieves a high classification accuracy in *forest and green land* and *urban fabric* classes in

[5]Due to the space limit, we only present the results on a subset of metrics (micro/macro-F1, DF_1/2). The results on other metrics are similar.

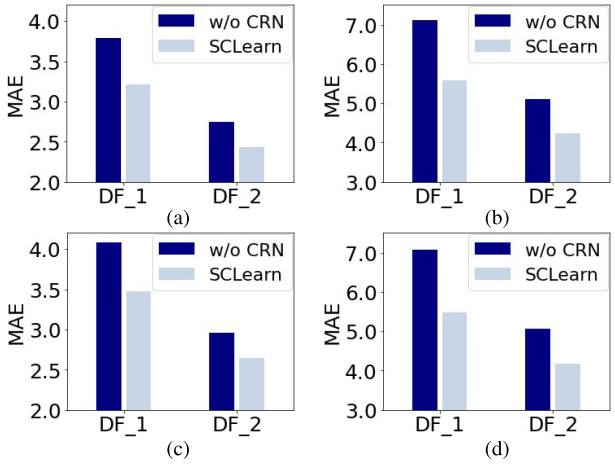


Fig. 11. Ablation study of SCLearn on super-resolution. (a) Barcelona Scale Factor = 2 × 2. (b) Barcelona Scale Factor = 4 × 4. (c) Berlin Scale Factor = 2 × 2. (d) Berlin Scale Factor = 4 × 4.

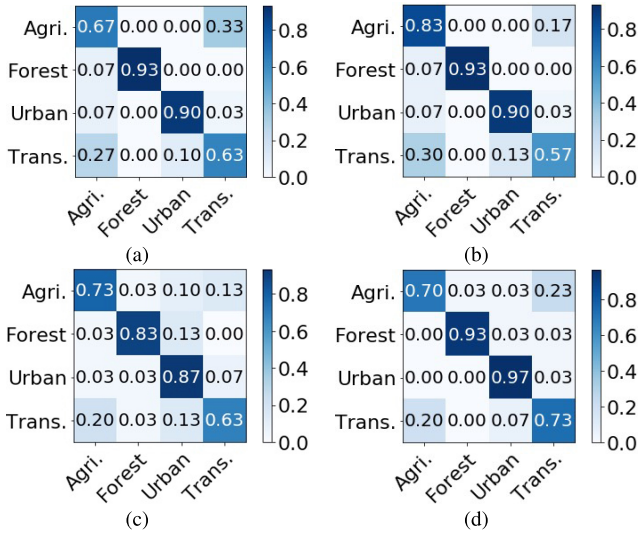


Fig. 12. Confusion matrices of the SCLearn scheme. (a) Barcelona Scale Factor = 2 × 2. (b) Barcelona Scale Factor = 4 × 4. (c) Berlin Scale Factor = 2 × 2. (d) Berlin Scale Factor = 4 × 4.

all evaluation settings. We also note that the *transportation* and *agriculture* classes are sometimes misclassified with each other. This is because these two classes sometimes share similar object shapes and layouts (e.g., the shape of highway in *transportation* class looks similar to the squared farmland in *agriculture* class as shown in Fig. 6). Such ambiguity leads to misclassifications between the two classes. Second, we show the class-wise perceptual error in Fig. 13. We observe that SCLearn achieves a low perceptual error in *forest and green Land* and *agriculture* classes in all evaluation settings. We also note that SCLearn has a higher perceptual error on *urban fabric* and *transportation* classes. This is because these two classes often contain more complex object layouts and fine-grained details compared with the other two classes, which presents a more challenging super-resolution task to our SCLearn framework.

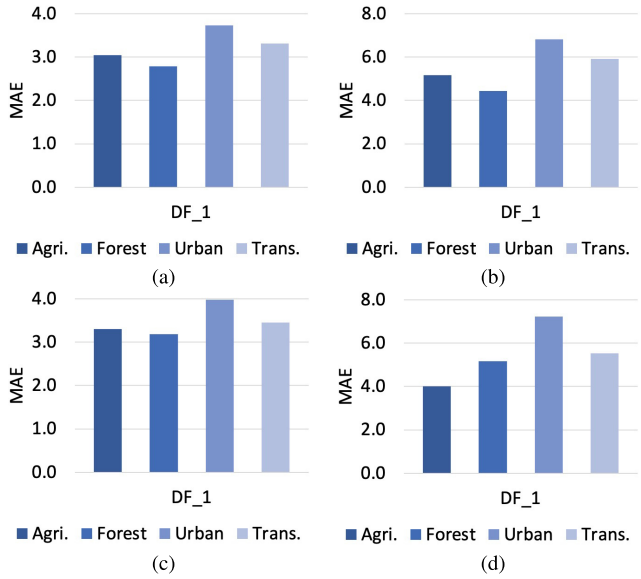


Fig. 13. Class-wise perceptual error of the SCLearn scheme. (a) Barcelona Scale Factor = 2 × 2. (b) Barcelona Scale Factor = 4 × 4. (c) Berlin Scale Factor = 2 × 2. (d) Berlin Scale Factor = 4 × 4.

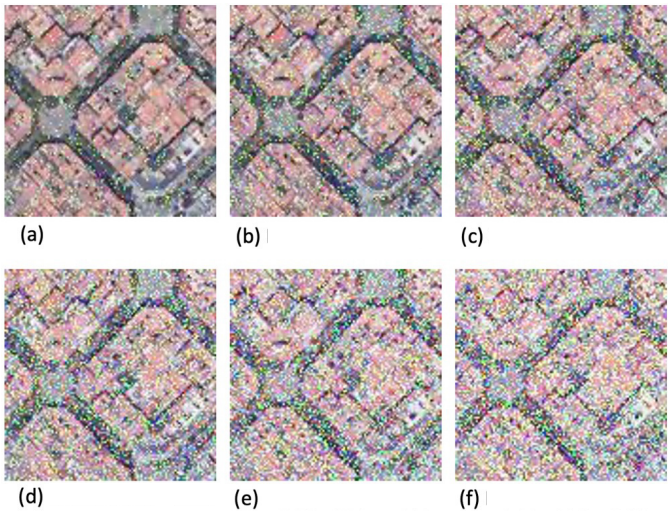


Fig. 14. Examples of low-resolution images under different noise ratios. (a) Noise ratio = 5%. (b) Noise ratio = 10%. (c) Noise ratio = 15%. (d) Noise ratio = 20%. (e) Noise ratio = 25%. (f) Noise ratio = 30%.

5) Q5 (Impact of Image Noise on SCLearn Scheme): In the last set of experiments, we study the impact of image noise on the performance of our SCLearn scheme on both the classification and super-resolution tasks. In particular, we adopt the widely used *random_noise* tool from the *scikit-image* package[6] to add random noise to the studied low-resolution images. In particular, the *random_noise* tool changes the RGB values to random values for a specific percentage of pixels within an image (we refer to such percentage as *noise ratio*). In our evaluation, we vary the noise ratio from 5% to 30% (examples of low-resolution images under different noise ratios are shown in Fig. 14). The performance of our SCLearn on classification

[6]https://scikit-image.org/docs/stable/api/skimage.util.html#random-noise

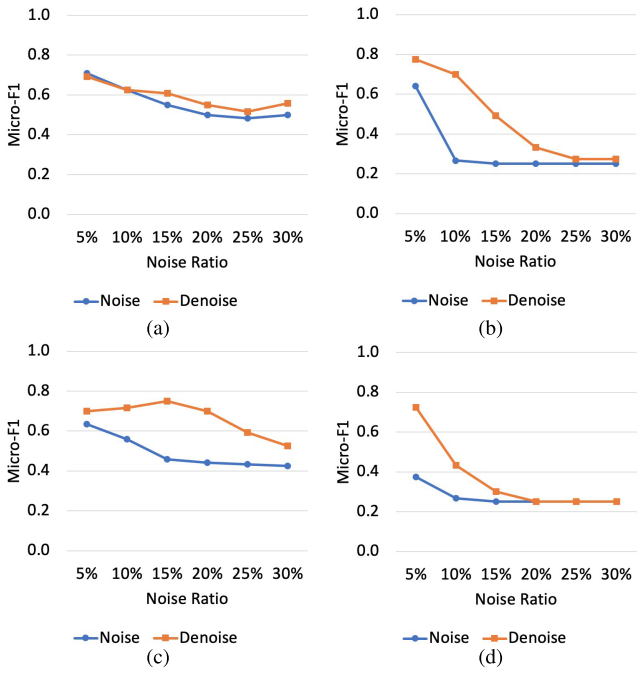


Fig. 15. Impact of image noise to SCLearn on classification. (a) Barcelona Scale Factor = 2 × 2. (b) Barcelona Scale Factor = 4 × 4. (c) Berlin Scale Factor = 2 × 2. (d) Berlin Scale Factor = 4 × 4.

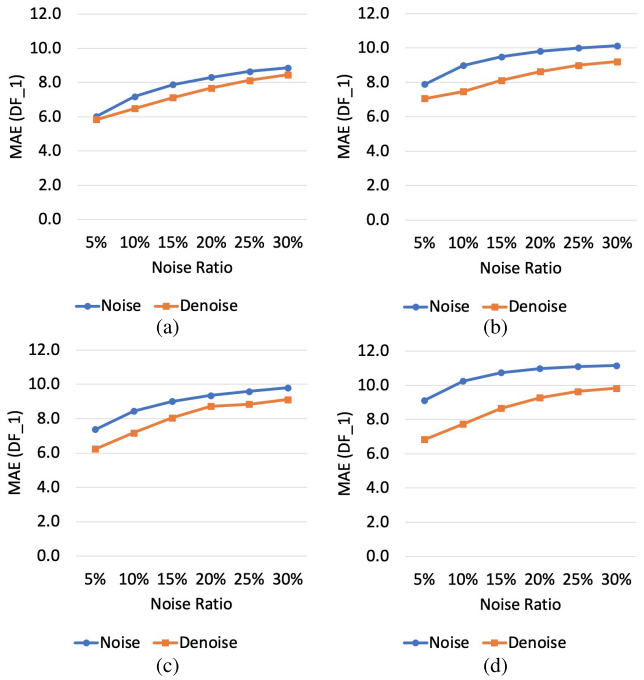


Fig. 16. Impact of image noise to SCLearn on super-resolution. (a) Barcelona Scale Factor = 2 × 2. (b) Barcelona Scale Factor = 4 × 4. (c) Berlin Scale Factor = 2 × 2. (d) Berlin Scale Factor = 4 × 4.

and super-resolution is shown in Figs. 15 and 16, respectively. We observe that the classification accuracy of SCLearn decreases when the noise ratio increases (marked as "Noise" in Fig. 15). This is because noise could confuse our model by recognizing incorrect object layouts and color distributions for inaccurate land usage classification. In addition, we also observe that the perceptual error of our SCLearn increases when the noise ratio increases (marked as "Noise" in Fig. 16). This is because the added noise could mislead our SCLearn to add inaccurate fine-grained details that result in suboptimal reconstructed images. To mitigate the negative effect of noise, we noted that our SCLearn can be coupled with the current image denoising approaches (e.g., median filtering [63]) to preprocess the input images to reduce noise. We observe that our SCLearn can clearly achieve improved classification accuracy and better reconstructed image quality using denoised images generated by the image denoising tool (marked as "Denoise" in both Figs. 15 and 16).

VI. Discussion

The land usage classification performance gains achieved by our SCLearn clearly demonstrate that improving the spatial resolution of the satellite image can help improve the classification accuracy. In particular, our SCLearn designs a holistic super-resolution-assisted convolutional classification (SCN) network to explicitly reconstruct high-resolution images with enhanced global features (e.g., color distributions, object layouts, contrast ratios) and local details (e.g., local object contours, shapes, and textures) that optimize intraclass similarity and interclass dissimilarity to improve the classification accuracy. In general, we observe that the global features and local details enhanced by our SCLearn model are complementary to each other and work collaboratively to improve the land usage classification accuracy. On one hand, the enhanced global features provide global visual evidence to describe the land usage of an image but it can be limited by object occlusion and intraclass variation. For example, the global features can help us clearly distinguish the *forest and green land* class from the *urban fabric* class as those two classes have clearly different color distributions and object layouts. However, global feature alone does not provide sufficient visual evidence to generate accurate classification results. For example, global features are insufficient to distinguish the images of the *agriculture* class from the ones of the *forest and green land* class as they share similar global visual features (e.g., dominant color, object texture). On the other hand, local details provide fine-grained visual evidence of local objects but different land usage classes can share the same type of local objects. For example, enhanced contour and texture of crops can help distinguish the *agriculture* class from the *forest and green land* class. However, the texture and shape of a tree can occur in both *forest and green land* class and *urban fabric* class. Therefore, our SCLearn jointly uses both the enhanced global features and local details to effectively boost the land usage classification accuracy. .

In our SCLearn design, we focus on integrating super-resolution with *scene classification* to improve the performance of both the tasks. However, our SCLearn framework can be further extended to integrate the super-resolution with the pixel or object-based classification under an end-to-end learning framework for object-driven RUS applications (e.g., measuring traffic flow from satellite images, segmenting

damaged areas after a disaster using remote sensing). First, we can replace the scene CN in our SCN module with a pixel or object-based CN [64] to build an integrated CNN architecture. Similar to our SCN, the new integrated convolutional network can be used to explicitly reconstruct high-resolution images with enhanced visual details that optimize the intraobject similarity and interobject dissimilarity to improve the object- or pixel-based classification accuracy. Second, we can modify the CRN in the CPR module to an object-aware refinement network. In particular, we can leverage the dense network connection design to connect the object-oriented convolutional layers with our current CPR module so that it can specifically focus on refining the image quality of each identified object. Each object-aware refinement network can then be used to refine a subset of closely related objects identified by the updated SCN model given the fact that the closely-related objects often share similar visual characteristics (e.g., similar shape and layouts). There also exist several open-ended questions that can be further explored in integrating the super-resolution task with the pixel- or object-based classification task. For example, how to effectively scale up the proposed solutions when the satellite images contain a rich set of surface objects with diversified visual characteristics? How to establish a robustness model when we only have sparse training labels of surface objects (e.g., buildings, trees, and infrastructures)?

In our experiments, we observe that noise within satellite images can affect the performance of our SCLearn model. In particular, a noisy satellite image could lead to a poorly reconstructed satellite image from the super-resolution task. The poorly reconstructed satellite image could lead to inaccurate classification labels, which could further reduce the quality of the reconstructed image during the iteratively super-resolution and classification tasks. To address this challenge, we could further extend our SCN to a discriminative learning-based deep network by introducing a noise-sensitive discriminator network to explicitly identify the noise within each image. In addition, we can introduce an adversarial loss function design to supervise our SCN network to effectively reduce noise within satellite images so that the discriminator network cannot detect any noise within each image. In general, such a design aims to effectively reduce noise within satellite images before reconstructing high-resolution images to boost the classification accuracy.

VII. Conclusion

This article presents a SCLearn framework to tackle a new CSC problem in RUS applications. SCLearn addresses two challenges, namely, *complex interdependence* and *noise amplification*. In particular, we develop a novel integrated deep learning framework to integrate the super-resolution and classification tasks to concurrently boost the performance of both the tasks. The results on a real-world RUS application show that SCLearn significantly outperforms both the state-of-the-art classification and super-resolution baselines. We believe SCLearn will provide useful insights to integrate important tasks with similar complex interdependence in other domains.

Acknowledgment

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," *Comput. J.*, vol. 52, pp. 90–100, Jan. 2009.

[2] R. Hou, S. Jeong, Y. Wang, K. H. Law, and J. P. Lynch, "Camera-based triggering of bridge structural health monitoring systems using a cyber-physical system framework," *Struct. Health Monitor*, pp. 3139–3146, Sep. 2017.

[3] M. T. Rashid and D. Wang, "Unravel: An anomalistic crowd investigation framework using social airborne sensing," in *Proc. IEEE Int. Perform., Comput., Commun. Conf. (IPCCC)*, Oct. 2021, pp. 1–10.

[4] W. An, D. Wu, S. Ci, H. Luo, V. Adamchuk, and Z. Xu, "Agriculture cyber-physical systems," in *Cyber-Physical Systems*. Amsterdam, The Netherlands: Elsevier, Jan. 2017, pp. 399–417.

[5] 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 *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2017, pp. 1357–1366.

[6] R. R. Vatsavai, E. Bright, C. Varun, B. Budhendra, A. Cheriyadat, and J. Grasser, "Machine learning approaches for high-resolution urban land cover classification: A comparative study," in *Proc. 2nd Int. Conf. Comput. Geospatial Res. Appl.*, May 2011, pp 1–10.

[7] Y. Zhang *et al.*, "PQA-CNN: Towards perceptual quality assured single-image super-resolution in remote sensing," in *Proc. IEEE/ACM 28th Int. Symp. Quality Service (IWQOS)*, Jun. 2020, pp. 1–10.

[8] D. Lu and Q. Weng, "Use of impervious surface in urban land-use classification," *Remote Sens. Environ.*, vol. 102, nos. 1–2, pp. 146–160, May 2006.

[9] G. Cheng *et al.*, "SPNet: Siamese-prototype network for few-shot remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2022.

[10] R. M. Anwer, F. S. Khan, J. van de Weijer, M. Molinier, and J. Laaksonen, "Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 138, pp. 74–85, Apr. 2018.

[11] X. Zhang, W. An, J. Sun, H. Wu, W. Zhang, and Y. Du, "Best representation branch model for remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9768–9780, Sep. 2021.

[12] R. Dong, L. Zhang, and H. Fu, "RRSGAN: Reference-based super-resolution for remote sensing image," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2021.

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

[14] Y. Xiao, X. Su, Q. Yuan, D. Liu, H. Shen, and L. Zhang, "Satellite video super-resolution via multiscale deformable convolution alignment and temporal grouping projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2021, Art. no. 5610819.

[15] D. Cai, K. Chen, Y. Qian, and J.-K. Kämäräinen, "Convolutional low-resolution fine-grained classification," *Pattern Recognit. Lett.*, vol. 119, pp. 166–171, Mar. 2017.

[16] S. Hao, W. Wang, Y. Ye, E. Li, and L. Bruzzone, "A deep network architecture for super-resolution-aided hyperspectral image classification with classwise loss," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4650–4663, Aug. 2018.

[17] Y. Pang, J. Cao, J. Wang, and J. Han, "JCS-Net: Joint classification and super-resolution network for small-scale pedestrian detection in surveillance images," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 12, pp. 3322–3331, Dec. 2019.

[18] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, "Image super-resolution by neural texture transfer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7982–7991.

[19] Y. Zhang, R. Zong, L. Shang, M. T. Rashid, and D. Wang, "SuperClass: A deep duo-task learning approach to improving QoS in image-driven smart urban sensing applications," in *Proc. IEEE/ACM 29th Int. Symp. Quality Service (IWQOS)*, Jun. 2021, pp. 1–6.

[20] 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 *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 1544–1553.

[21] M. T. Rashid, D. Y. Zhang, and D. Wang, "SocialDrone: An integrated social media and drone sensing system for reliable disaster response," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Jul. 2020, pp. 218–227.

[22] S. Rao *et al.*, "A cyber-physical system approach for photovoltaic array monitoring and control," in *Proc. 8th Int. Conf. Inf., Intell., Syst. Appl. (IISA)*, Aug. 2017, pp. 1–6.

[23] T. Evariste, W. Kasakula, J. Rwigema, and R. Datta, "Pollution context-aware representation in vehicular Internet of Things for smart cities," in *Proc. Int. Workshop Distrib. Comput. Emerg. Smart Netw.* Cham, Switzerland: Springer, 2020, pp. 23–39.

[24] E. K. Wang, F. Wang, R. Sun, and X. Liu, "A new privacy attack network for remote sensing images classification with small training samples," *Math. Biosci. Eng.*, vol. 16, no. 5, pp. 4456–4476, 2019.

[25] L. Shang, Y. Zhang, C. Youn, and D. Wang, "SAT-GEO: A social sensing based content-only approach to geolocating abnormal traffic events using syntax-based probabilistic learning," *Inf. Process. Manage.*, vol. 59, no. 2, Mar. 2022, Art. no. 102807.

[26] A. Vetrivel, N. Kerle, M. Gerke, F. Nex, and G. Vosselman, "Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning," Univ. Twente Fac. Geo-Inf. Earth Observ. (ITC), Hengelosestraat, AE Enschede, Netherlands, Tech. Rep. Proc. GEOBIA, pp. 1–5, 2016.

[27] K. Behrendt, L. Novak, and R. Botros, "A deep learning approach to traffic lights: Detection, tracking, and classification," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 1370–1377.

[28] M. Kawulok, S. Piechaczek, K. Hrynczenko, P. Benecki, D. Kostrzewa, and J. Nalepa, "On training deep networks for satellite image super-resolution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 3125–3128.

[29] P. Wang, H. Zhang, F. Zhou, and Z. Jiang, "Unsupervised remote sensing image super-resolution using cycle CNN," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 3125–3128.

[30] C. Kwan *et al.*, "Deep learning for land cover classification using only a few bands," *Remote Sens.*, vol. 12, no. 12, p. 2000, Jun. 2020.

[31] Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu, "Rankclus: Integrating clustering with ranking for heterogeneous information network analysis," in *Proc. 12th Int. Conf. Extending Database Technol. Adv. Database Technol.*, Mar. 2009, pp. 565–576.

[32] F. Girard, C. Kavalec, and F. Cheriet, "Joint segmentation and classification of retinal arteries/veins from fundus images," *Artif. Intell. Med.*, vol. 94, pp. 96–109, Mar. 2019.

[33] H.-M. Müller, E. E. Kenny, and P. W. Sternberg, "Textpresso: An ontology-based information retrieval and extraction system for biological literature," *PLoS Biol.*, vol. 2, no. 11, p. e309, Nov. 2004.

[34] J. Yang, D. She, Y.-K. Lai, and M.-H. Yang, "Retrieving and classifying affective images via deep metric learning," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.

[35] Y. Zhang and Q. Yang, "A survey on multi-task learning," 2017, *arXiv:1707.08114*.

[36] M.-L. Zhang and Z.-H. Zhou, "A review on multi-label learning algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1819–1837, Mar. 2013.

[37] C. Chen, B. Zhang, H. Su, W. Li, and L. Wang, "Land-use scene classification using multi-scale completed local binary patterns," *Signal Image Video Process.*, vol. 10, no. 4, pp. 745–752, 2016.

[38] Y. Zhang *et al.*, "Transland: An adversarial transfer learning approach for migratable urban land usage classification using remote sensing," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 1567–1576.

[39] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 586–595.

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

[41] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 1–22.

[42] W. Liu, Y. Wen, Z. Yu, and M. Yang, "Large-margin softmax loss for convolutional neural networks," in *Proc. ICML*, Jun. 2016, vol. 2, no. 3, pp. 1–10.

[43] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4491–4500.

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

[45] A. Martins and R. Astudillo, "From softmax to sparsemax: A sparse model of attention and multi-label classification," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1614–1623.

[46] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5353–5360.

[47] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.

[48] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4681–4690.

[49] A. Rakhlin, A. Davydow, and S. Nikolenko, "Land cover classification from satellite imagery with U-Net and Lovász-softmax loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 262–266.

[50] J. Zhang, C. Lu, X. Li, H.-J. Kim, and J. Wang, "A full convolutional network based on DenseNet for remote sensing scene classification," *Math. Biosci. Eng.*, vol. 16, no. 5, pp. 3345–3367, 2019.

[51] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.

[52] M. Tan and Q. Le, "EfficientNetV2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106.

[53] H. Zhang *et al.*, "ResNeSt: Split-attention networks," 2020, *arXiv:2004.08955*.

[54] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5197–5206.

[55] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 723–731.

[56] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1905–1914.

[57] R. M. Umer, G. L. Foresti, and C. Micheloni, "Deep generative adversarial residual convolutional networks for real-world super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2020, pp. 438–439.

[58] K. Zhang, L. Van Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3217–3226.

[59] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 4791–4800.

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

[61] 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, Aug. 2012, Art. no. e41882.

[62] X. Hu, L. Chu, J. Pei, W. Liu, and J. Bian, "Model complexity of deep learning: A survey," 2021, *arXiv:2103.05127*.

[63] M. C. Motwani, M. C. Gadiya, R. C. Motwani, and F. C. Harris, "Survey of image denoising techniques," in *Proc. GSPX*, vol. 27, 2004, pp. 27–30.

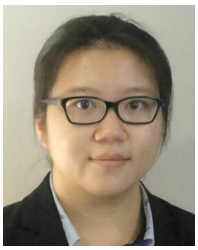
[64] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," 2017, *arXiv:1704.06857*.

Yang Zhang (Student Member, IEEE) received the B.S. degree in software engineering from Wuhan University, Wuhan, China, in 2013, and the M.S. degree in data science from Indiana University-Bloomington, Bloomington, IN, USA, in 2017. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN.

His research interests include social sensing, deep learning, and human-centered artificial intelligence.

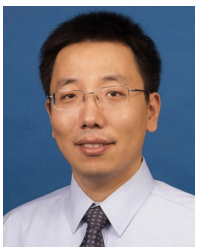
Ruohan Zong (Student Member, IEEE) received the B.E. degree in computer science and technology from Sichuan University, Chengdu, China, in 2020, and the M.S. degree in computer science from Columbia University, New York, NY, USA, in 2022. She is currently pursuing the Ph.D. degree with the School of Information Sciences, University of Illinois Urbana-Champaign (UIUC), Champaign, IL, USA.

Her primary research interests are human-centered AI and AI for social good.



Lanyu Shang (Student Member, IEEE) received the B.S. degree in applied mathematics from the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, in 2014, and the M.S. degree in data science from New York University, New York, NY, USA, in 2017. She is currently pursuing the Ph.D. degree with the School of Information Sciences, University of Illinois Urbana-Champaign (UIUC), Champaign, IL, USA.

Her research interest primarily lies in online misinformation detection using social media data.



Dong Wang (Member, IEEE) received the Ph.D. degree in computer science from the University of Illinois Urbana-Champaign (UIUC), Champaign, IL, USA, in 2012.

He is currently an Associate Professor with the School of Information Sciences, UIUC. His research interests lie in the area of reliable social sensing, human-centric AI, cyber-physical computing, and smart city applications.

Dr. Wang is a member of ACM. He was a recipient of the NSF CAREER Award, Google Faculty Research Award, Army Research Office Young Investigator Program (YIP) Award, the Wing-Kai Cheng Fellowship from the University of Illinois, and the Best Paper Award of IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS).