PQA-CNN: Towards Perceptual Quality Assured Single-Image Super-Resolution in Remote Sensing

Yang Zhang, Xiangyu Dong, Md Tahmid Rashid, Lanyu Shang, Jun Han, Daniel Zhang, Dong Wang
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN, USA
{yzhang42, xdong2, mrashid, lshang, jhan5, yzhang40, dwang5}@nd.edu

***Abstract*—Recent advances in remote sensing open up unprecedented opportunities to obtain a rich set of visual features of objects on the earth's surface. In this paper, we focus on a single-image super-resolution (SISR) problem in remote sensing, where the objective is to generate a reconstructed satellite image of high quality (i.e., a high spatial resolution) from a satellite image of relatively low quality. This problem is motivated by the lack of high quality satellite images in many remote sensing applications (e.g., due to the cost of high resolution sensors, communication bandwidth constraints, and historic hardware limitations). Two important challenges exist in solving our problem: i) it is not a trivial task to reconstruct a satellite image of high quality that meets the human perceptual requirement from a single low quality image; ii) it is challenging to rigorously quantify the uncertainty of the results of an SISR scheme in the absence of ground truth data. To address the above challenges, we develop PQA-CNN, a perceptual quality-assured conventional neural network framework, to reconstruct a high quality satellite image from a low quality one by designing novel uncertainty-driven neural network architectures and integrating an uncertainty quantification model with the framework. We evaluate PQA-CNN on a real-world remote sensing application on land usage classifications. The results show that PQA-CNN significantly outperforms the state-of-the-art super-resolution baselines in terms of accurately reconstructing high-resolution satellite images under various evaluation scenarios.**

***Index Terms*—Super-Resolution, Perceptual Quality, Uncertainty-Aware, Convolutional Neural Network**

I. INTRODUCTION

With the advent of high precision optical and image processing technologies, satellite-based remote sensing has become a powerful sensing paradigm that can obtain abundant visual features of the objects residing on the earth's surface [1]. Examples of remote sensing applications include performing damage assessment during disaster scenarios [2], predicting the poverty in underdeveloped areas [3], detecting cholera outbreaks from water bodies [4], and monitoring refugee movements in armed-conflict zones [5]. In this paper, we focus on a *single-image super-resolution (SISR)* problem in remote sensing, where the objective is to generate a reconstructed satellite image with high quality (i.e., a high spatial resolution) from a single satellite image with a relatively low quality. One example of our application scenarios is the classification of diversified land usages in a city (e.g., urban areas, trees, lakes, and transportation) where the classification results can help address important urban and social questions (e.g., assessment of urban environmental impacts and potential anthropogenic activities involved on land) [6]. Figure 1 shows an example of a land usage classification scenario involving different geographical components in an area. We observe that different land classes can be easily messed up if the quality (resolution) of the satellite image is low. For example, with the high-resolution image in Figure 1(a), the lake is correctly classified. However, in the case of the low-resolution image in Figure 1(b), both the lake and some buildings are misclassified as trees.

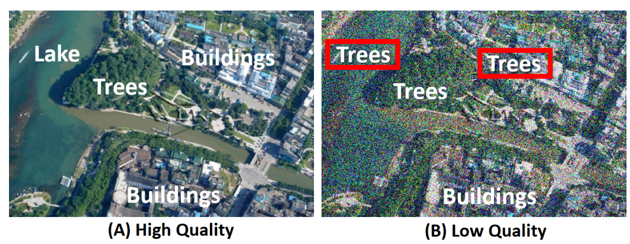


Figure 1. Classification of Diversified Land Usage Classes

While the high quality satellite images are normally more desirable as shown in the above example, they are not always available in remote sensing applications [7]. The reasons are multi-fold. First, high-resolution sensor packages are often quite expensive [8]. For example, a set of 8 high-resolution multi-spectral sensors kit required for a reasonable spatial resolution (e.g., 10 m$\times$10 m) costs more than 100,000 USD [9]. Second, many remote sensing applications need to utilize the historical satellite imagery data to study the spatial and temporal dynamics of an area or phenomenon (e.g., land cover changes [10], population migration [11]). Unfortunately, a large amount of historic satellite images are only available in relatively low-resolutions [12]. Third, it is hardly possible to have the 24/7 high-resolution image coverage of all objects on earth given the current satellite image updating frequency (i.e., from daily to yearly) and communication bandwidth constraints [13]. Therefore, there exists a strong motivation to develop an effective solution to accurately reconstruct high-resolution images from the low-resolution ones.

Efforts have been made to address the super-resolution problem in image processing, remote sensing, and deep learning [14]–[18]. Examples of those solutions include

978-1-7281-6887-6/20/$31.00 ©2020 IEEE

978-1-7281-6887-6/20/$31.00 ©2020 IEEE

regularization-based image interpolation [14], image-up-scaling using sub-pixel morphing [15], single-frame super-resolution through convolutional neural networks [16], and single-image upscaling using deep residual networks [18]. However, two important challenges have not been well addressed by current solutions. We elaborate them below.

Perceptual Quality Assurance. The first challenge lies in providing the desired perceptual quality assurance of the reconstructed satellite images from an SISR solution. The *perceptual quality* is a metric defined to describe the quality of a reconstructed satellite image as perceived by humans [19]. Previous efforts in remote sensing often failed to provide such perceptual quality assurance of the reconstructed images [14], [16], [18] due to two important limitations. First, current SISR solutions mainly focus on improving the pixel-wise estimation accuracy (e.g., peak signal-to-noise ratio (PSNR), structural similarity index (SSIM)) of the reconstructed images [14], [16], and ignore the actual perceptual quality [20]. Second, many current solutions utilize the deep neural networks to generate high-quality reconstructed images, which either introduce additional noise or ruin the structural integrity of images during the reconstruction process [16], [18]. Therefore, current super-resolution schemes often generate images that are sub-optimal to human perception, which can lead to inappropriate decision makings (e.g., inaccurate land usage classifications as shown in Figure 1).

Uncertainty-aware Super-resolution. The second challenge lies in the rigorous uncertainty quantification of the results (e.g., RGB values in reconstructed images) generated by an SISR scheme in the absence of ground truth data. For example, in an SISR based disaster damage assessment application, the uncertainty quantification of the assessment results (e.g., estimation confidence of an area being severely damaged in a reconstructed image) is critical to make life-saving decisions (e.g., where to dispatch the rescue teams) [21]–[23]. An important question that remains to be answered here is how to rigorously quantify the uncertainty of the results produced by SISR schemes without knowing the ground truth labels *a priori* and how to leverage the uncertainty quantification results to improve the quality of reconstructed images.

To address the above challenges, we develop an perceptual quality assured convolutional neural network (PQA-CNN) approach to solve the SISR problem in remote sensing applications. To address the first challenge, we design a duo-branch neural network that consists of two complementary convolutional neural architectures (i.e., Duo-CNN) that work collaboratively to achieve the desirable perceptual quality of the reconstructed images. To address the second challenge, we develop an uncertainty-driven ensemble model that integrates a maximum likelihood estimation approach with the Duo-CNN to accurately quantify the uncertainty of the estimation results. The uncertainty quantification is then used to guide the refinement of the reconstructed images. To the best of our knowledge, PQA-CNN is the first uncertainty-aware neural network approach to address the SISR problem in remote sensing. The perceptual quality-driven and uncertain-aware nature of our approach makes it possible to reconstruct a high resolution image with perceptual quality assurance from a single low resolution image. We evaluate PQA-CNN through a real-world remote sensing application where the satellite imagery dataset is collected from two different cities in Europe using Google Maps Platform. The results show that PQA-CNN significantly outperforms the state-of-the-art SISR baselines by reconstructing satellite images with higher perception quality under various evaluation scenarios.

II. Related Work

A. Remote Sensing

In recent times, remote sensing has received a significant amount of attention, enabling many applications that capture different phenomena occurring on the earth [1]. For example, Cervone *et al.* developed a machine learning based disaster damage assessment system by fusing satellite imagery with Twitter data [2]. Müller *et al.* utilized satellite imagery to assess the latent effects of human migration over the hydrological process of a river basin [5]. Zou *et al.* proposed a deep learning based feature selection for scene classification of satellite imagery [24]. Several important challenges prevail in current remote sensing applications. Examples include data irregularity, image obscurity, privacy concerns, and noise propagation [25]. The SISR task using low-resolution satellite imagery data remains to be an open and challenging problem in remote sensing. In this paper, we design a novel PQA-CNN framework to address this problem by developing novel convolutional neural network architectures and uncertainty quantification mechanisms.

B. Super-Resolution

Current solutions to the super-resolution problem can be classified into two categories: *conventional* and *deep learning* approaches [14]–[18]. *Conventional approaches*: Lukin *et al.* explored a regularization-based image interpolation method for image enhancement by using filtering and convergence techniques [14]. Yang *et al.* presented a morphing-based super-resolution method that leverages the complementary information contained in different sub-pixels among multiple low-resolution frames to construct a high-resolution image [15]. *Deep learning approaches*: Dong *et al.* proposed a conventional neural network approach to upscale low-resolution images to high-resolution ones through the bicubic interpolation [16]. Ledig *et al.* developed a generative adversarial network framework to generate high-resolution images from low-resolution ones through an optimization process regularized by adversarial and perceptual losses [17]. Lee *et al.* designed a deep residual network approach to improve the quality of the generated high-resolution images using a set of optimized residual blocks [18]. However, the above approaches often failed to provide the assured perceptual quality of the reconstructed high-quality satellite images in remote sensing because they either ignore the actual perceptual quality of the reconstructed satellite images or mainly focus on a single neural network design that does not address the noise and structural integrity issue well in the image reconstruction process [20]. In this paper, we develop a perceptual quality assured SISR scheme that integrates the uncertainty quantification model with the deep convolutional neural networks to provide high-resolution reconstructed images with perceptual quality assurance.

C. Uncertainty-Aware Estimation and Deep Learning

Our work is also related to the uncertainty-aware estimation and deep learning techniques, which have been studied in many areas such as reinforcement learning, computer vision, image generation, and Internet-of-Things [26]–[33]. For example, Wang *et al.* developed a set of uncertainty quantification schemes to rigorously quantify the quality of information in social sensing applications [26]. Houthooft *et al.* designed a curiosity-driven exploration strategy for high-dimensional deep reinforcement learning using Bayesian neural networks [30]. Yasarla *et al.* proposed a multi-scale residual learning framework based on cycle spinning that leverages the uncertainty of prediction for image de-raining tasks [31]. Tang *et al.* developed a multi-channel generative adversarial network that uses cascaded semantic uncertainty to improve the performance of the cross-view image translation [32]. However, a unique challenge in satellite-based remote imagery is the need for perceptual quality assurance, for which the current solutions on uncertainty quantification are not designed to address. In contrast, the PQA-CNN framework is the first work that aims to leverage the quantified uncertainty to reconstruct a high-resolution satellite image with high perceptual quality.

III. Problem Description

In this section, we formally define the perceptual quality assured single-image super-resolution problem in remote sensing. We first define a few key terms that will be used in the problem statement.

Definition 1: **Sensing Cell**: Given a studied area (e.g., a city) where we collect the satellite imagery data for the super-resolution task, we first divide the studied area into disjoint sensing cells. Each cell represents a subarea of interest (e.g., 250m $\times$ 250m as shown in Figure 2). In particular, we define N to be the number of sensing cells in the studied area and n to be the n^{th} sensing cell.

Definition 2: **Low-Resolution Satellite Image (L)**: we define L to be the satellite image (e.g., historical satellite imagery data) from each sensing cell collected in a specific remote sensing application. The low-resolution satellite image is usually in a relatively low spatial resolution (e.g., 112$\times$112 resolution for a sensing cell as shown in (A) of Figure 2). In particular, we define L^n to represent the low-resolution satellite image collected from the sensing cell n.

Definition 3: **High-Resolution Satellite Image (H)**: We define H to be the high-resolution satellite image for each sensing cell, which has a relatively high resolution (e.g., 224$\times$224 resolution for a sensing cell as shown in (B) of Figure 2). The high-resolution satellite images often provide more fine-grained details of the objects (e.g., clear building outlines and road edges). In particular, we define H^n to be the *actual* high-resolution satellite image of the sensing cell n.

Definition 4: **Reconstructed High-Resolution Satellite Image ($\widehat{H}$)**: We also define $\widehat{H}$ to be the *reconstructed* high-resolution satellite image, which is generated by our super-resolution scheme from the corresponding low-resolution satellite image L. In particular, we define $\widehat{H^n}$ to represent the *reconstructed* high-resolution satellite image for the sensing cell n and our goal is to ensure the reconstructed satellite image is as close to the *actual* high resolution satellite image H^n as possible.

Figure 2. *Low* and *High* Resolution Satellite Images

Definition 5: **Uncertainty Matrix ($\mathcal{U}$)**: Let us consider the error between the *actual* and *reconstructed* high resolution satellite images (i.e., H and $\widehat{H}$), where such an estimation error often follows a normal distribution [34]:

$$H - \widehat{H} \sim \mathcal{N}(\mathbf{0}, \mathcal{U}^2) \tag{1}$$

where $H - \widehat{H}$ is the matrix to represent the error of estimated RGB values at all pixels in the image. $\mathcal{U}$ is the *uncertainty matrix* that represents the standard deviation of the estimation errors. Such an uncertainty matrix is essential to refine the *reconstructed* satellite image $\widehat{H}$ to achieve the desired perceptual image quality, which will be discussed in detail in next section.

Definition 6: **Perceptual Quality**: To evaluate the quality of $\widehat{H}$, we use the state-of-the-art perceptual metric [20] to quantify the perceptual difference between the *actual* and *reconstructed* satellite images as follows:

$$\Phi(H, \widehat{H}) = \Gamma[\Theta(H) - \Theta(\widehat{H})] \tag{2}$$

where we set the $\Phi(\cdot)$ to represent the perceptual metric. $\Theta(H)$ and $\Theta(\widehat{H})$ represents the extracted deep feature vectors from the *actual* and *reconstructed* satellite images using ImageNet-trained deep convolutional neural networks (e.g., VGG [35]). $\Gamma(\cdot)$ is a function to calculate the difference between two deep feature vectors (e.g., Mean Squared Error (MSE), Mean Absolute Error (MAE)). This metric has been proven to be robust in capturing perceptual quality of images [19], [36].

The goal of the single-image super-resolution problem in remote sensing is to accurately reconstruct the high-resolution satellite image for each sensing cell from the collected low-resolution satellite image in that cell. Using the definitions above, our problem is formally defined as:

$$\arg\min_{\widehat{H^n}}(\Gamma[\Theta(H^n) - \Theta(\widehat{H^n})] \mid L^n), \quad \forall 1 \le n \le N \tag{3}$$

It is a challenging problem to reconstruct such a high-resolution satellite image with desired perceptual quality given the excessive fine-grained details in each satellite image, and the fuzzy and inadequate visual evidence provided by the input low-resolution satellite image. In this paper, we develop a PQA-CNN scheme to address these challenges, which is elaborated in the next section.

IV. Solution

A. Overview of PQA-CNN Framework

PQA-CNN is an uncertainty-aware convolutional neural network framework to address the SISR problem in remote sensing. The overview of the PQA-CNN framework is shown in Figure 3. It consists of two major components:

- *Uncertainty-aware Duo-CNN Architecture*: it constructs two effective yet complementary convolutional neural network architectures (i.e., *pre-upscaling* and *pos-upscaling* networks) to reconstruct the high-resolution satellite images and infer the uncertainty matrices.
- *Uncertainty-driven Satellite Imagery Ensemble*: it leverages the estimated uncertainty matrices from the Duo-CNN component to ensemble the satellite images generated by both *pre-upscaling* and *pos-upscaling* networks to further improve the perceptual quality of the reconstructed images.

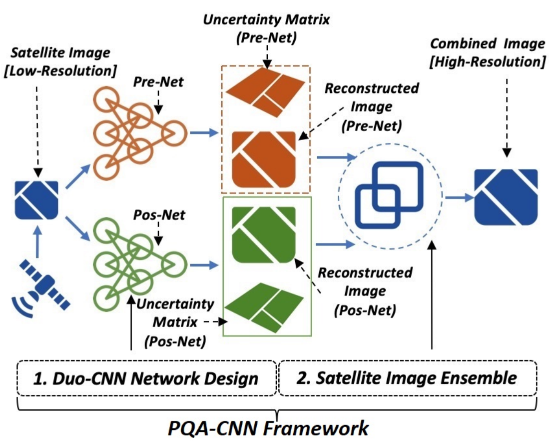


Figure 3. Overview of PQA-CNN framework

B. Uncertainty-Aware Duo-CNN Architecture

In this subsection, we present the Duo-CNN architecture design in our framework. The Duo-CNN consists of two convolutional neural network architectures to 1) reconstruct the high-resolution satellite images, and 2) infer the uncertainty matrices to quantify the accuracy of the estimated RGB values in the reconstructed images. In particular, we employ two different yet complementary neural network design strategies in Duo-CNN: *pre-upscaling* and *post-upscaling*. In *pre-upscaling*, it first scales the resolution of a low-resolution image to a high-resolution one (we refer to the process as *upscaling*) and then refines the generated high-resolution image to remove noise [37]. In *post-upscaling*, it first extracts and refines the semantic features from a low-resolution satellite image and then scales the refined semantic features to a high-resolution image [38]. The pre-upscaling can often effectively reduce the noise but is more likely to ruin the structure integrity (e.g., making building outlines fuzzier) in the reconstructed images. The post-upscaling often has an opposite effect on the images compared to the pre-scaling (i.e., successfully preserving the structure integrity while introducing the noise). Our Duo-CNN framework integrates both pre-upscaling and post-upscaling to reconstruct the satellite images to explore the benefits from both networks to improve the image quality. We define the two types of convolutional neural networks of our design as follows:

Definition 7: **Pre-upscaling Network (*Pre-Net*)**: We define *Pre-Net* to be a *pre-upscaling* convolutional neural network architecture to reconstruct the high-resolution image $\widehat{H}_{pre}$ and generate the corresponding uncertainty matrix $\mathcal{U}_{pre}$ as follows:

$$\langle \widehat{H}_{pre}, \mathcal{U}_{pre} \rangle = \textit{Pre-Net}(L) \tag{4}$$

where L is the low-resolution satellite image as the input to *Pre-Net*.

An example of the pre-upscaling network architecture and the associated model parameters are illustrated in Figure 4 In particular, it includes four different modules: a bicubic interpretation (BI) module, an image encoding module, an image decoding module, and an uncertainty matrix generation module. In the bicubic interpolation module, a bicubic interpolation operation [1] is applied to upscale a low-resolution image to a high-resolution one. The image encoding module contains a set of *ReflectionPad-Convolution-Relu* operations [40] to convert the upscaled satellite images to semantic feature representations and filters out the noise introduced by the bicubic interpolation process. Finally, the outputs of image encoding module are fed in parallel into both the image decoding and uncertainty matrix generation modules. The image decoding module converts the de-noised semantic feature representations to the reconstructed satellite images and the uncertainty matrix generation module generates the uncertainty matrix of the RGB values in the reconstructed images. Given the above pre-upscaling network architecture, our next question is how to define a loss function for our model to generate the high-resolution reconstructed images together with the uncertainty matrices.

To that end, we define the loss function $\mathcal{L}_{\text{pre}}$ for our Pre-Net that contains two sub-loss functions as follows:

$$\mathcal{L}_{\text{pre}} : \min\left(\mathcal{L}^{\text{pre}}_{\text{reconstruct}} + \mathcal{L}^{\text{pre}}_{\text{uncertain}}\right) \tag{5}$$

where $\mathcal{L}^{\text{pre}}_{\text{reconstruct}}$ is the first sub-loss function to ensure our Pre-Net generates the high quality reconstructed images $\widehat{H}_{pre}$, and

[1]Bicubic interpolation is a conventional interpolation operation for image upscaling that fills an empty pixel by leveraging the RGB values from its neighboring pixels [39].

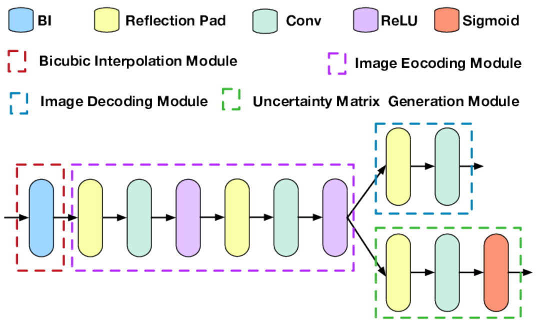


Figure 4. Illustration of Pre-upscaling Network (Pre-Net)

$\mathcal{L}^{\text{pre}}_{\text{uncertain}}$ is the second sub-loss function to ensure our Pre-Net derives accurate uncertainty matrix $\mathcal{U}_{pre}$. In particular, we first define the first sub-loss function $\mathcal{L}^{\text{pre}}_{\text{reconstruct}}$ as follows:

$$\mathcal{L}^{\text{pre}}_{\text{reconstruct}} : \min \left(\mathcal{L}_{\text{perceptual}}(H, \widehat{H}_{pre}) + \mathcal{L}_{\text{pixel}}(H, \widehat{H}_{pre}) \right) \quad (6)$$

where $\mathcal{L}_{\text{perceptual}}(H, \widehat{H}_{pre})$ is the perceptual loss [20] to quantify the perceptual difference between the actual and reconstructed images. $\mathcal{L}_{\text{pixel}}(H, \widehat{H}_{pre})$ is the Mean Squared Error (MSE) loss [41] to measure the pixel-wise RGB value difference between the actual and reconstructed images, which is used to reduce the pixel-wise noise in Pre-Net.

Next, we formulate a maximum likelihood estimation problem to derive the second sub-loss function $\mathcal{L}^{\text{pre}}_{\text{uncertain}}$. Our goal is to estimate the uncertainty matrix $\mathcal{U}_{pre}$ given the difference between the *actual* and *reconstructed* satellite images (i.e., (H - $\widehat{H}_{pre}$) as defined in Definition 5). By observing such an estimation discrepancy often follows a normal distribution [34], we derive the likelihood function of our estimation as follows:

$$\begin{aligned} &\mathbb{L}(\mathcal{U}_{pre} | H - \widehat{H}_{pre}) = \\ &(2\pi ||\mathcal{U}_{pre}||_2)^{-\frac{1}{2}} exp(-\frac{1}{2||\mathcal{U}_{pre}||_2} ||H - \widehat{H}_{pre}||_2) \end{aligned} \quad (7)$$

We can then derive the log-likelihood function accordingly:

$$\begin{aligned} &log\mathbb{L}(\mathcal{U}_{pre} | H - \widehat{H}_{pre}) = \\ &-\frac{1}{2} log 2\pi - \frac{1}{2} log ||\mathcal{U}_{pre}||_2 - \frac{1}{2||\mathcal{U}_{pre}||_2} ||H - \widehat{H}_{pre}||_2 \end{aligned} \quad (8)$$

Our goal is to maximize $log\mathbb{L}(\mathcal{U}_{pre} | H - \widehat{H}_{pre})$ to obtain an accurate uncertainty matrix estimation. This leads to the definition of the second sub-loss function $\mathcal{L}^{\text{pre}}_{\text{uncertain}}$ as the negation of $log\mathbb{L}(\mathcal{U}_{pre} | H - \widehat{H}_{pre})$:

$$\begin{aligned} &\mathcal{L}^{\text{pre}}_{\text{uncertain}} : \\ &\min \left(\frac{1}{2} log ||\mathcal{U}_{pre}||_2 + \frac{1}{2||\mathcal{U}_{pre}||_2} ||H - \widehat{H}_{pre}||_2 + \frac{1}{2} log 2\pi \right) \end{aligned} \quad (9)$$

By minimizing the loss function $\mathcal{L}^{\text{pre}}_{\text{uncertain}}$, we can obtain the optimal uncertainty matrix $\mathcal{U}_{pre}$ that maximizes the above likelihood function $\mathbb{L}(\mathcal{U}_{pre} | H - \widehat{H}_{pre})$.

Definition 8: **Post-upscaling Network (*Pos-Net*)**: We define *Pos-Net* to be a *post-upscaling* convolutional neural network architecture to reconstruct the high-resolution image $\widehat{H}_{pos}$ and generate the corresponding uncertainty matrix $\mathcal{U}_{pos}$ as follows:

$$\langle \widehat{H}_{pos}, \mathcal{U}_{pos} \rangle = \textit{Pos-Net}(L) \quad (10)$$

where L is the low-resolution satellite image as the input to *Pos-Net*.

An example of the post-upscaling network architecture and associated model parameters are illustrated in Figure 5. In particular, it also includes four different modules: an image encoding module, a residual block module, an image decoding module, and an uncertainty generation module. Different from the Pre-Net, the image encoding module directly takes the low-resolution image as the input and extracts the semantic feature representations from the images. This is done to ensure the structure integrity in the reconstructed satellite images. The residual block module has multiple residual blocks to segment individual objects of an image and apply augmented contents to improve the resolution of the identified objects. Similar to Pre-Net, the upscaled semantic feature representations of the image are simultaneously fed into two parallel output modules, i.e., image decoding and uncertainty matrix generation modules, where the outputs are the reconstructed satellite image and the corresponding uncertainty matrix.

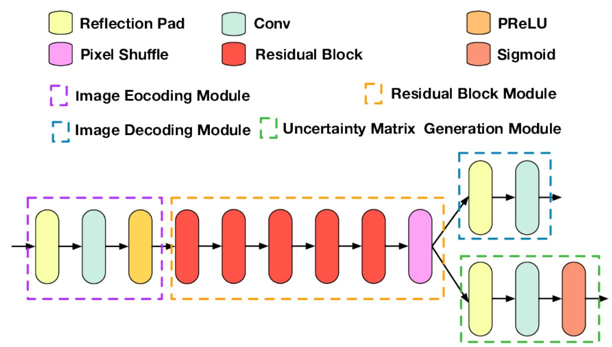


Figure 5. Illustration of Post-upscaling Network (Pos-Net)

Similar to Pre-Net, we define the loss function $\mathcal{L}_{\text{pos}}$ for our Pos-Net that contains two sub-loss functions to generate the reconstructed image $\widehat{H}_{pos}$ and the uncertainty matrix $\mathcal{U}_{pos}$ as:

$$\mathcal{L}_{\text{pos}} : \min \left(\mathcal{L}^{\text{pos}}_{\text{reconstruct}} + \mathcal{L}^{\text{pos}}_{\text{uncertain}} \right) \quad (11)$$

where $\mathcal{L}^{\text{pos}}_{\text{reconstruct}}$ is defined as:

$$\mathcal{L}^{\text{pos}}_{\text{reconstruct}} : \min \mathcal{L}_{\text{perceptual}}(H, \widehat{H}_{pos}) \quad (12)$$

Note that we only consider the perceptual loss in Pos-Net and ignore the pixel-wise MSE loss. This is done to enforce the Pos-Net to focus on generating images with high perceptual quality that preserves the structure integrity. In addition, we define the sub-loss function $\mathcal{L}^{\text{pos}}_{\text{uncertain}}$ in a similar way as the Pre-Net:

$$\begin{aligned} &\mathcal{L}^{\text{pos}}_{\text{uncertain}} : \\ &\min \left(\frac{1}{2} log ||\mathcal{U}_{pos}||_2 + \frac{1}{2||\mathcal{U}_{pos}||_2} ||H - \widehat{H}_{pos}||_2 + \frac{1}{2} log 2\pi \right) \end{aligned} \quad (13)$$

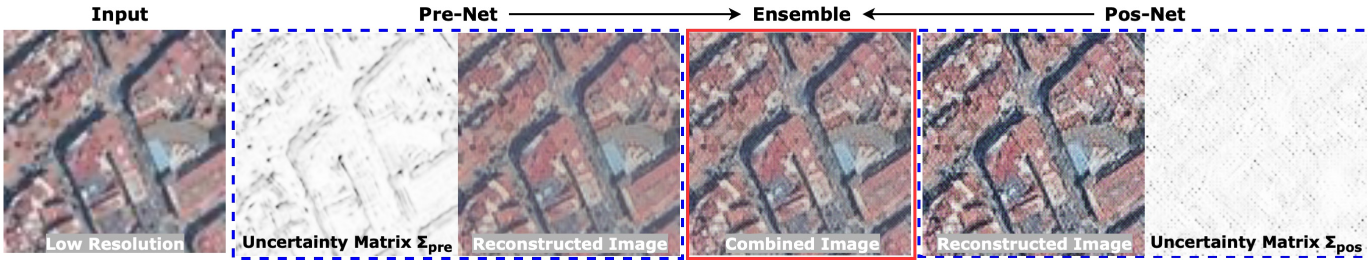


Figure 6. An Example of the Combined High-Resolution Image Generated by PQA-CNN

C. Uncertainty-driven Satellite Imagery Ensemble

In this subsection, we leverage the estimated uncertainty matrices ($\mathcal{U}_{pre}$ and $\mathcal{U}_{pos}$) output by the Duo-CNN networks to guide the ensemble of the satellite images generated by the pre-upscaling and post-upscaling networks (i.e., $\widehat{H}_{pre}$ and $\widehat{H}_{pos}$) to further improve the quality of the reconstructed images. We first define a key term in our ensemble mechanism as follows.

Definition 9: **Combined High-Resolution Satellite Image ($\widehat{H}_{combine}$):** We define $\widehat{H}_{combine}$ to be a high-resolution satellite image, where the RGB value at each pixel is a combination of the RGB values from the reconstructed satellite images ($\widehat{H}_{pre}$ and $\widehat{H}_{pos}$) generated from Pre-Net and Pos-Net as follows:

$$\widehat{H}_{combine} = (\mathbf{1} - \Lambda) \cdot \widehat{H}_{pre} + \Lambda \cdot \widehat{H}_{pos} \tag{14}$$

where Λ is a matrix to indicate the weights of each component at all pixels in the combined high-resolution image. $\mathbf{1}$ is a matrix with the same dimension as Λ, where all elements in $\mathbf{1}$ equal 1.

The key question now is how to derive the values in Λ to optimize the quality of the combined satellite image $\widehat{H}_{combine}$. To address this problem, we first consider the probabilistic model for the error between the actual and reconstructed satellite images generated by Duo-CNN as defined in Equation 1. We perform a random variable transformation to obtain the probabilistic models for the RGB values in the reconstructed images (i.e., $\widehat{H}_{pre} \sim \mathcal{N}(H, \mathcal{U}_{pre}^2)$ and $\widehat{H}_{pos} \sim \mathcal{N}(H, \mathcal{U}_{pos}^2)$). Using these models, we can derive the distribution of $\widehat{H}_{combine}$ in Equation (14) as follows:

$$\widehat{H}_{combine} \sim \mathcal{N}((\mathbf{1} - \Lambda) \cdot H, ((\mathbf{1} - \Lambda) \cdot \mathcal{U}_{pre})^2) + \mathcal{N}(\Lambda \cdot H, (\Lambda \cdot \mathcal{U}_{pos})^2) \tag{15}$$

We consider the ensemble mechanism to be optimized when the Pro-Net and Pos-Net share the maximum agreement in the estimation confidence/uncertainty of the pixel-wise RGB values in the reconstructed satellite image [42]. We enforce such an agreement by setting the variances of the two networks to be the same:

$$((\mathbf{1} - \Lambda) \cdot \mathcal{U}_{pre})^2 = (\Lambda \cdot \mathcal{U}_{pos})^2 \tag{16}$$

We can then derive the value of Λ as follows:

$$\Lambda = \frac{\mathcal{U}_{Pre}}{\mathcal{U}_{pre} + \mathcal{U}_{pos}} = \frac{\mathcal{U}_{pos}^{-1}}{\mathcal{U}_{pre}^{-1} + \mathcal{U}_{pos}^{-1}} \tag{17}$$

We plug the derived Λ value into Equation (14) as follows:

$$\widehat{H}_{combine} = \frac{\mathcal{U}_{pre}^{-1}}{\mathcal{U}_{pre}^{-1} + \mathcal{U}_{pos}^{-1}} \cdot \widehat{H}_{pre} + \frac{\mathcal{U}_{pos}^{-1}}{\mathcal{U}_{pre}^{-1} + \mathcal{U}_{pos}^{-1}} \cdot \widehat{H}_{pos} \tag{18}$$

where $\widehat{H}_{combine}$ is the final output of our PQA-CNN framework.

We further define a loss function $\mathcal{L}_{combine}$ to ensure the perceptual quality of the combined satellite image generated by the uncertainty-driven satellite imagery ensemble mechanism:

$$\mathcal{L}_{\text{combine}} : \min \mathcal{L}_{\text{perceptual}}(H, \widehat{H}_{combine}) \tag{19}$$

where $\mathcal{L}_{\text{perceptual}}(H, \widehat{H}_{combine})$ is the loss function to measure the perceptual difference between the actual and reconstructed satellite images as discussed in the previous subsection.

An example of the combined satellite image generated by our PQA-CNN framework is shown in Figure 6. First, we observe that the Pre-Net effectively reduces the noise from the input image but introduces a certain amount fuzziness into the reconstructed image. However, the fuzzy areas (e.g., building outlines) are accurately captured by the uncertainty matrix $\mathcal{U}_{pre}$ as shown in the figure [2]. Similarly, we observe that the Pos-Net successfully preserves the structure integrity but introduces a noticeable amount of noise (i.e., white dots in the figure). However, the noisy points are also accurately captured by the uncertainty matrix $\mathcal{U}_{post}$. Finally, we observe that the combined satellite image achieves an clearly improved perceptual quality compared to the input image as well as the reconstructed images from both Pre-Net and Post-Net.

Finally, we briefly summarize the optimization process of our PQA-CNN framework to learn the optimal parameters of Pre-Net and Pos-Net (i.e., Pre-Net* and Pos-Net*) based on the loss functions defined above. We first define an aggregated loss function $\mathcal{L}_{\text{overall}}$ for our PQA-CNN framework as:

$$\mathcal{L}_{\text{overall}} : \min \left(\mathcal{L}_{\text{pre}} + \mathcal{L}_{\text{pos}} + \mathcal{L}_{\text{combine}}\right) \tag{20}$$

The aggregated loss function combines the loss functions defined in each component of PQA-CNN: i.e., $\mathcal{L}_{\text{pre}}$ (Equation (5)), $\mathcal{L}_{\text{pos}}$ (Equation (11)), and $\mathcal{L}_{\text{combine}}$ (Equation (19)). By minimizing the aggregated loss, we ensure both Pre-Net and Pos-Net generate high quality reconstructed satellite images, which is used to generated the combined high-resolution satellite images. The loss function $\mathcal{L}_{\text{overall}}$ can be

[2] A darker color of a pixel in the uncertainty matrix graph indicates a higher degree of uncertainty for the generated RGB value of the corresponding pixel in the reconstructed image.

optimized using the Adaptive Moment Estimation (Adam) optimizer [43], which obtains the optimal parameters of both upscaling networks $PosNet^*$ and $PreNet^*$.

V. Evaluation

A. Dataset

In our experiment, we collect real-world satellite imagery datasets from two different cities in Spain (i.e., *Barcelona* and *Madrid*), a region well known for its diversified land features [44]. The collected satellite imagery data belongs to three diversified land usage classes (i.e., *urban fabric*, *forest and green land*, and *transportation* as shown in Figure 7). These classes have distinct visual and semantic characteristics (e.g., object layout and density, color distributions and complexity), which present a challenging evaluation scenario for the SISR problem we studied. We summarize the datasets as follows:

Google Maps Satellite Imagery Dataset: We collect the satellite imagery datasets from *Barcelona* and *Madrid* using Google Map Platform [3]. In our evaluation, each collected original satellite image is in 224×224 resolution with a 250m×250m ground coverage, which is considered as a *high* resolution satellite image in our evaluation as it provides sufficient visual information for our defined sensing cell [45]. In addition, we adopt the widely-used *bicubic interpretation* tool implemented in *scikit-image* package [4] to reduce the resolution of each original satellite image by 4 times as the *low* resolution satellite image in our experiment (i.e., each low-resolution satellite image is in 112×112 resolution as shown in Figure 2 (A)). Finally, we randomly select 1,200 *high* and *low* satellite images (i.e., 600 from each category) from the studied area for our experiments.

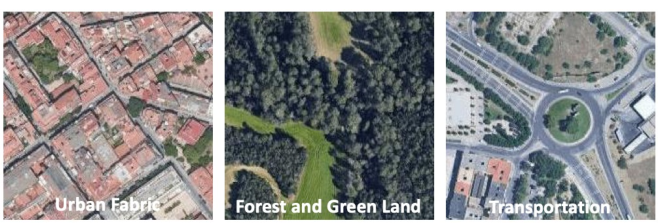


Figure 7. Examples of Satellite Imagery Data in Different Land Usage Classes

B. Baselines

We compare PQA-CNN with representative *conventional* and *deep learning* baselines that are used to solve the SISR problem.

1) Conventional Models
   - **Nearest-neighbour (NN)** [46]: it is a conventional satellite image upscaling scheme that fills each empty pixel with the same RGB value as the nearest available neighboring pixel.
   - **Bi-linear/quadratic/cubic** [39]: it is a set of representative satellite image super resolution schemes that leverage the bi-linear/quadratic/cubic interpolation techniques to generate an estimated RGB value for each empty pixel from its neighboring pixels.
2) Deep Learning Models
   - **SFSR18** [16]: it utilizes the bi-cubic interpolation and conventional neural networks to generate the high-resolution satellite image with a dedicated image refining process to improve the quality of reconstructed images.
   - **SRGAN17** [17]: it imposes a generative adversarial network architecture that utilizes an image generator network and an image discriminator network to refine the reconstructed images.
   - **SRResNet19** [18]: it is a deep convolutional neural network that leverages multiple residual blocks with skip-connection to capture the complex mapping between the low and high-resolution satellite images in the image reconstruction process.

C. Evaluation Metrics and Settings

To evaluate the performance of all compared schemes, we use the perceptual metric (discussed in Definition 6), which has proven to be an accurate metric that is close to human perception in the recent computer vision studies [19], [20], [36]. In particular, we use three commonly used deep features extracted before the 1^{st}, 2^{nd}, 3^{rd} convolutional layers of the 4^{th} convolutional block in VGG model (namely, VGG_{4-1}, VGG_{4-2}, VGG_{4-3}). In addition, we adopt two commonly used error measurement functions (i.e., $\Gamma(\cdot)$ in Definition 6): i) *Mean Absolute Error (MAE)* and ii) *Mean Squared Error (MSE)* to calculate the difference between the deep features extracted from the *actual* and *reconstructed* satellite images. Intuitively, a lower value in the error metric represents a higher perceptual quality and a better visual similarity between the *actual* and *reconstructed* satellite images, which indicates a better super-resolution performance. Note that we do not use pixel-wise evaluation metrics (e.g., PSNR, SSIM) in our evaluation as they are shown to be suboptimal in evaluating the actual perceptual quality of the images [19], [20].

In our experiment, we randomly sample 70% satellite images as training dataset and 10% satellite images as validation dataset to tune hyper-parameters of all compared algorithms. We then use the rest 20% satellite images as testing dataset to evaluate the performance of all compared algorithms. In addition, all hyper-parameters are optimized using the Adam optimizer [43]. In particular, we set the learning rate to be 10^{-4} and set the batch size to be 1 in our experiment. In addition, the model is trained over 500 epochs for all compared schemes.

D. Evaluation Results

Evaluation results on perceptual quality for *urban fabric*: In the first set of experiments, we study the performance of all compared schemes in *Barcelona* and *Madrid*, where

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

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

Table I
PERFORMANCE COMPARISONS (CLASS = *Urban Fabric*)

		Studied City = *Barcelona*						Studied City = *Madrid*					
		VGG_{4-1}		VGG_{4-2}		VGG_{4-3}		VGG_{4-1}		VGG_{4-2}		VGG_{4-3}	
Category	Algorithm	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*
Conv. Model	NN	1.1546	4.9464	0.6487	2.5650	0.4962	1.6637	1.2060	5.3422	0.6962	2.8903	0.5247	1.9278
	Bilinear	1.1396	5.0302	0.6442	2.6017	0.4891	1.6467	1.2458	5.9456	0.7152	3.1382	0.5293	1.9538
	Biquadratic	1.1091	4.7499	0.6222	2.4396	0.4703	1.5286	1.2084	5.5693	0.6893	2.9282	0.5088	1.8113
	Bicubic	1.1125	4.7782	0.6253	2.4584	0.4730	1.5450	1.2132	5.6149	0.6934	2.9572	0.5121	1.8341
Deep Model	SFSR18	1.1138	4.5774	0.6190	2.3757	0.4634	1.4894	1.1604	4.9477	0.6519	2.6021	0.4814	1.6420
	SRGAN17	1.0551	4.2401	0.5780	2.1161	0.4321	1.3126	1.1130	4.6791	0.6157	2.3641	0.4538	1.4706
	SRResNet19	1.0601	4.2572	0.5800	2.1250	0.4346	1.3255	1.1088	4.6512	0.6144	2.3512	0.4525	1.4598
Ours	**PQA-CNN**	**1.0195**	**3.9818**	**0.5561**	**1.9736**	**0.4155**	**1.2265**	**1.0701**	**4.3510**	**0.5901**	**2.1864**	**0.4335**	**1.3526**

Table II
PERFORMANCE COMPARISONS (CLASS = *Forrest and Green Land*)

		Studied City = *Barcelona*						Studied City = *Madrid*					
		VGG_{4-1}		VGG_{4-2}		VGG_{4-3}		VGG_{4-1}		VGG_{4-2}		VGG_{4-3}	
Category	Algorithm	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*
Conv. Model	NN	0.7527	2.4017	0.4763	1.2522	0.3819	0.9518	1.0164	4.0094	0.6208	2.2225	0.4795	1.5691
	Bilinear	0.6964	2.1695	0.4325	1.0525	0.3276	0.6925	0.9579	3.7372	0.5729	1.9716	0.4240	1.2430
	Biquadratic	0.6865	2.0991	0.4233	1.0136	0.3195	0.6589	0.9363	3.5478	0.5559	1.8635	0.4108	1.1667
	Bicubic	0.6890	2.1160	0.4261	1.0246	0.3219	0.6692	0.9394	3.5738	0.5590	1.8807	0.4133	1.1809
Deep Model	SFSR18	0.6943	2.1025	0.4280	1.0362	0.3168	0.6529	0.9445	3.4894	0.5549	1.8315	0.4037	1.1477
	SRGAN17	0.6193	1.7065	0.3690	0.7846	0.2712	0.4907	0.8396	2.8780	0.4802	1.4159	0.3458	0.8608
	SRResNet19	0.6180	1.7003	0.3677	0.7788	0.2701	0.4876	0.8483	2.9354	0.4847	1.4428	0.3487	0.8744
Ours	**PQA-CNN**	**0.5988**	**1.5986**	**0.3555**	**0.7340**	**0.2621**	**0.4601**	**0.8090**	**2.6894**	**0.4631**	**1.3224**	**0.3334**	**0.8067**

the land usage class of images is urban-fabric. The evaluation results are presented in Table I. We observed that the PQA-CNN scheme consistently outperforms all compared baselines across different deep features. For example, the performance gains of PQA-CNN over the best-performing baseline (i.e., SRGAN17) in Barcelona with the deep feature extracted by VGG_{4-1} on MAE and MSE are 3.49% and 6.48%, respectively. Such performance gains mainly come from the fact that PQA-CNN judiciously learns the uncertainty of the estimated RGB values in the reconstructed images through an integrated Duo-CNN and MLE hybrid design. The obtained uncertainty matrix is explicitly used to guide the reconstruction of the satellite image integrated from the ones generated by both *pre-upscaling* and *post-upscaling* networks.

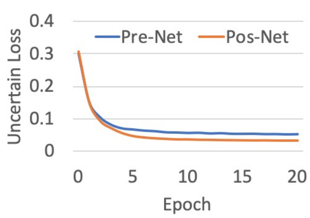
Evaluation results on perceptual quality for *forest and green land* and *transportation*: In addition to *urban fabric*, we also evaluate the performance of all schemes over the *forest and green land* and *transportation* land classes in both *Barcelona* and *Madrid*. Our objective here is to evaluate whether PQA-CNN and the baselines are capable of providing reliable super-resolution results across completely different land usage classes. The evaluation results are shown in Table II and Table III. We observe that PQA-CNN continues to outperform all baselines over both the *forest and green land* and *transportation* classes in the two cites. For example, the performance gains achieved by PQA-CNN compared to the best-performing baseline (i.e., SRGAN17) in the forest and green land class in Madrid with the deep feature extracted by VGG_{4-2} on MAE and MSE are 3.78% and 7.01%, respectively. Similar performance gains are also observed in the transportation class in both cities. Such consistent performance gains demonstrate the effectiveness and robustness of PQA-CNN in learning the accurate uncertainty matrices to guide the convolutional neural networks to reconstruct high-quality images across diversified classes of land usage. Additionally, we also observe that all compared schemes tend to have lower perception errors in the *forest and green land* class compared to the other two classes. This is mainly because that the forest and green land class often has much less complex object layouts and color distributions than other classes (as shown in Figure 7), making it an easier super-resolution task for all compared schemes.

Effectiveness of PQA-CNN on uncertainty estimation: In the third experiment, we study the effectiveness on uncertainty estimation for our PQA-CNN model by tacking the values of the uncertain loss functions for both Pre-Net and Pos-Net (defined in Equation (9) and Equation (13), respectively). The results are shown in Figure 8. We observe that the uncertain

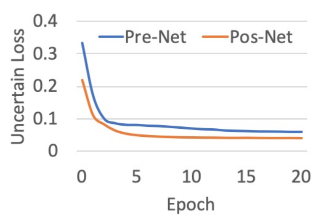
Table III
PERFORMANCE COMPARISONS (CLASS = *Transportation*)

		Studied City = *Barcelona*						Studied City = *Madrid*					
		VGG_{4-1}		VGG_{4-2}		VGG_{4-3}		VGG_{4-1}		VGG_{4-2}		VGG_{4-3}	
Category	Algorithm	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*	*MAE*	*MSE*
Conv. Model	NN	0.8189	3.0102	0.5070	1.6158	0.3964	1.1002	1.0101	4.1899	0.6157	2.2838	0.4717	1.5432
	Bilinear	0.7941	2.9851	0.4973	1.5824	0.3845	1.0659	0.9788	4.1157	0.5962	2.1983	0.4492	1.4317
	Biquadratic	0.7707	2.7832	0.4768	1.4610	0.3673	0.9733	0.9487	3.8388	0.5727	2.0376	0.4296	1.3143
	Bicubic	0.7736	2.8094	0.4801	1.4788	0.3701	0.9880	0.9530	3.8773	0.5769	2.0634	0.4331	1.3349
Deep Model	SFSR18	0.7759	2.7210	0.4761	1.4405	0.3657	0.9611	0.9601	3.7656	0.5725	2.0186	0.4286	1.3104
	SRGAN17	0.7690	2.6893	0.4645	1.3871	0.3552	0.9047	0.9099	3.4976	0.5330	1.7833	0.3962	1.1355
	SRResNet19	0.7683	2.6874	0.4639	1.3850	0.3543	0.8996	0.9054	3.4727	0.5304	1.7708	0.3940	1.1249
Ours	**PQA-CNN**	**0.7411**	**2.5159**	**0.4467**	**1.2943**	**0.3407**	**0.8417**	**0.8717**	**3.2378**	**0.5106**	**1.6537**	**0.3792**	**1.0520**

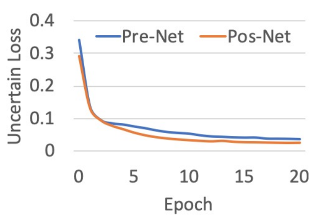
loss function values for both Pre-Net and Pos-Net converge to minimum values quickly and remain stable afterward in different settings, indicating that our PQA-CNN model is effective in terms of obtaining an accurate uncertainty estimation. In addition, such fast convergence rate also demonstrates the efficiency and scalability of our PQA-CNN scheme in real world remote sensing applications.



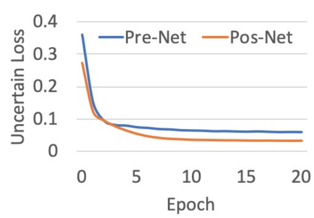
(a) Class = *Urban Fabric*, City = *Barcelona*



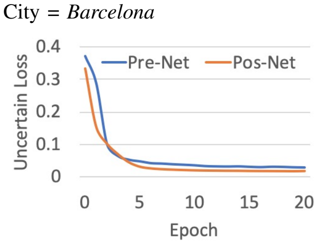
(b) Class = *Urban Fabric*, City = *Madrid*



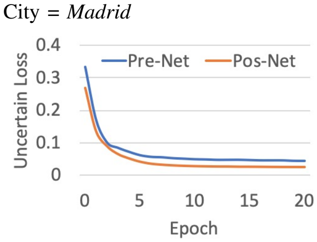
(c) Class = *Forrest and Green Land*, City = *Barcelona*



(d) Class = *Forrest and Green Land*, City = *Madrid*



(e) Class = *Transportation*, City = *Barcelona*



(f) Class = *Transportation*, City = *Madrid*

Figure 8. Effectiveness of PQA-CNN on Uncertainty Estimation

VI. CONCLUSION

In this paper, we develop a PQA-CNN approach to address the SISR problem in remote sensing applications. In particular, the PQA-CNN scheme addresses two intrinsic challenges (i.e., perceptual quality assurance and uncertainty-aware super resolution). The PQA-CNN scheme incorporates a hybrid duo-branch neural network design, namely Duo-CNN, to reconstruct the high-resolution satellite images with perceptual quality assurance from a low-resolution image. Our scheme also integrates an uncertainty quantification model with deep neural networks to further improve the quality of the reconstructed images. We evaluate PQA-CNN on a real-world remote sensing application involving land usage classification. The results demonstrate that our PQA-CNN scheme significantly outperforms state-of-the-art baselines in addressing the SISR problem. The results of this paper are important because they can directly contribute to a broad set of remote sensing applications that rely on the high quality satellite images that are not always available to the applications (e.g., disaster assessment, poverty prediction, disease outbreak detection).

ACKNOWLEDGMENT

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

REFERENCES

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

[2] G. Cervone, E. Schnebele, N. Waters, M. Moccaldi, and R. Sicignano, “Using social media and satellite data for damage assessment in urban areas during emergencies,” in *Seeing Cities Through Big Data*. Springer, 2017, pp. 443–457.

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

[4] B. Lobitz, L. Beck, A. Huq, B. Wood, G. Fuchs, A. Faruque, and R. Colwell, "Climate and infectious disease: use of remote sensing for detection of vibrio cholerae by indirect measurement," *Proceedings of the National Academy of Sciences*, vol. 97, no. 4, pp. 1438–1443, 2000.

[5] M. F. Müller, J. Yoon, S. M. Gorelick, N. Avisse, and A. Tilmant, "Impact of the syrian refugee crisis on land use and transboundary freshwater resources," *Proceedings of the national academy of sciences*, vol. 113, no. 52, pp. 14 932–14 937, 2016.

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

[7] Y. Zhang, R. Zong, J. Han, D. Zhang, T. Rashid, and D. Wang, "Transres: A deep transfer learning approach to migratable image super-resolution in remote urban sensing," in *2020 17th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2020.

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

[9] Dronenerds, "Micasense altum sensor." [Online]. Available: https://www.dronenerds.com/products/cameras-sensors/multispectral/micasense/

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

[11] J. Bartholy, R. Pongracz, Z. Barcza, and Z. Dezso, "Aspects of urban/rural population migration in the carpathian basin using satellite imagery," in *Environmental Change and its Implications for Population Migration*. Springer, 2004, pp. 289–313.

[12] B. L. Markham, J. C. Storey, D. L. Williams, and J. R. Irons, "Landsat sensor performance: history and current status," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 12, pp. 2691–2694, 2004.

[13] B. Yifang, P. Gong, and C. Gini, "Global land cover mapping using earth observation satellite data: Recent progresses and challenges," *ISPRS journal of photogrammetry and remote sensing (Print)*, vol. 103, no. 1, pp. 1–6, 2015.

[14] A. Lukin, A. S. Krylov, and A. Nasonov, "Image interpolation by super-resolution," in *Proceedings of GraphiCon*, vol. 2006. Citeseer, 2006, pp. 239–242.

[15] J. Yang and T. Huang, "Image super-resolution: Historical overview and future challenges," *Super-resolution imaging*, pp. 20–34, 2010.

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

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

[18] M. Kawulok, S. Piechaczek, K. Hrynczenko, P. Benecki, D. Kostrzewa, and J. Nalepa, "On training deep networks for satellite image super-resolution," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 3125–3128.

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

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

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

[22] D. Zhang, Y. Zhang, Q. Li, T. Plummer, and D. Wang, "Crowdlearn: A crowd-ai hybrid system for deep learning-based damage assessment applications," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2019, pp. 1221–1232.

[23] D. Zhang, Y. Huang, Y. Zhang, and D. Wang, "Crowd-assisted disaster scene assessment with human-ai interactive attention,," in *Thirty-Fourth AAAI Conference on Artificial Intelligence) (AAAI)*. AAAI, 2020, accepted.

[24] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2321–2325, 2015.

[25] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proceedings of the IEEE*, vol. 104, no. 11, pp. 2207–2219, 2016.

[26] D. Wang, L. Kaplan, T. Abdelzaher, and C. C. Aggarwal, "On credibility estimation tradeoffs in assured social sensing," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1026–1037, 2013.

[27] D. Zhang, D. Wang, N. Vance, Y. Zhang, and S. Mike, "On scalable and robust truth discovery in big data social media sensing applications," *IEEE Transactions on Big Data*, 2018.

[28] D. Wang, T. Abdelzaher, L. Kaplan, and C. C. Aggarwal, "On quantifying the accuracy of maximum likelihood estimation of participant reliability in social sensing," in *DMSN11: 8th International Workshop on Data Management for Sensor Networks*, 2011.

[29] D. Wang, T. Abdelzaher, H. Ahmadi, J. Pasternack, D. Roth, M. Gupta, J. Han, O. Fatemieh, H. Le, and C. C. Aggarwal, "On bayesian interpretation of fact-finding in information networks," in *14th International Conference on Information Fusion*. IEEE, 2011, pp. 1–8.

[30] R. Houthooft, X. Chen, Y. Duan, J. Schulman, F. De Turck, and P. Abbeel, "Curiosity-driven exploration in deep reinforcement learning via bayesian neural networks," *arXiv preprint arXiv:1605.09674*, 2016.

[31] R. Yasarla and V. M. Patel, "Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining," *arXiv preprint arXiv:1906.11129*, 2019.

[32] H. Tang, D. Xu, N. Sebe, Y. Wang, J. J. Corso, and Y. Yan, "Multi-channel attention selection gan with cascaded semantic guidance for cross-view image translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2417–2426.

[33] D. Wang, T. Abdelzaher, and L. Kaplan, *Social sensing: building reliable systems on unreliable data*. Morgan Kaufmann, 2015.

[34] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7482–7491.

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

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

[37] V. Dumoulin, J. Shlens, and M. Kudlur, "A learned representation for artistic style," *arXiv preprint arXiv:1610.07629*, 2016.

[38] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3147–3155.

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

[40] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

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

[42] "Uncertainty analysis: the variance of the variance," 2011.

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

[44] M. Stellmes, A. Röder, T. Udelhoven, and J. Hill, "Mapping syndromes of land change in spain with remote sensing time series, demographic and climatic data," *Land Use Policy*, vol. 30, no. 1, pp. 685–702, 2013.

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

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