UA-CNN: An Uncertainty-Aware Convolutional Neural Network Approach to Single-Image Super-Resolution in Remote Sensing

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 a high spatial resolution from a satellite image of a relatively low resolution. This problem is motivated by the lack of high-resolution 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 resolution that meets the human perceptual requirement from a single low-resolution 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 UA-CNN, an uncertainly-aware convolutional neural network framework, to reconstruct a high-quality satellite image from a low-resolution image by designing novel neural network architectures and integrating an uncertainty quantification model with the framework. We evaluate UA-CNN on a real-world remote sensing application on land usage classifications. The results show that UA-CNN significantly outperforms the state-of-the-art super-resolution baselines in terms of accurately reconstructing high-resolution satellite images under various evaluation scenarios.

KEYWORDS

Remote Sensing, Super-Resolution, Uncertainty-Aware, Convolutional Neural Network, Maximum Likelihood Estimation

1 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 [36]. Examples of remote sensing applications include performing damage assessment during disaster scenarios [8], predicting the poverty in underdeveloped areas [22], detecting cholera outbreaks from water bodies [28], and monitoring refugee movements in armed-conflict zones [34]. Due to its non-intrusive nature (i.e., not requiring any physical contact), remote sensing is increasingly exploited in scenarios where the detailed analysis of an area cannot be simply performed by modeling or field observations [42].

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 a high spatial resolution from a single satellite image with a relatively low resolution. The SISR problem is much more challenging than the traditional super-resolution problems that focus on reconstructing a high-resolution image from multiple low-resolution images of the same scene [45]. Our problem is motivated by the observation that the information extraction at a fine-grained scale of an object in remote sensing often requires a set of high-resolution satellite images [36]. One example of such applications 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) [29]. 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 resolution of the satellite image is not sufficiently high. 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-resolution satellite images are normally more desirable as shown in the above example, they are not always available in remote sensing applications [15]. The reasons are multi-fold. First, high-resolution sensor packages are often quite expensive [36]. For example, a set of 8 high-resolution multi-spectral sensors kit required for a reasonable spatial resolution (e.g., 10m×10m) costs more than 100,000 USD [13]. 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., the assessment of land cover changes over time [3], the study of population migration due to geological changes [5]). Such applications often require the access to a long duration of imagery data (e.g., more than 10 years). Unfortunately, the historic satellite images are often only available in relatively low-resolutions, e.g., the satellite

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

images before 2010 are primarily collected by Landsat 1-7 satellites that only provide low-resolution images (e.g., 30m×30m) [31]. 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 [49]. Therefore, there exists a strong demand 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 [11, 27, 30, 36, 41, 44]. Examples of those solutions include regularization-based image interpolation [30], image-up-scaling using sub-pixel morphing [44], single-frame super-resolution through convolutional neural networks [41], and single-image upscaling using deep residual networks [27]. 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 [6]. Previous efforts in remote sensing often failed to provide such perceptual quality of the reconstructed images [11, 36, 41] due to two important limitations. First, current SISR solutions in remote sensing 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 [11, 41], and ignore the actual perceptual quality, which is recently shown to be more appropriate in assessing the performance of SISR solutions [6, 50]. Second, existing solutions either introduce excessive noise or accidentally ruin the structural integrity (e.g., making building outlines fuzzier) in the reconstructed satellite images [36, 41]. Therefore, current super-resolution schemes often generate images that are ambiguous to human perception and lead to inappropriate decision makings (e.g., inaccurate land usage classifications as shown in Figure 1). The perceptual quality assurance of the reconstructed images remains to be a challenging problem in the SISR research of remote sensing.

Uncertainty-aware Super-resolution. The second challenge lies in the rigorous uncertainty quantification of the results (i.e., RGB values in reconstructed images) generated by an SISR scheme in the absence of ground truth data. Current super-resolution solutions (especially the deep learning based ones) mainly focus on improving the visual quality of the reconstructed images by imposing complex neural architectures or inference models but ignore an important aspect of their results: uncertainty quantification. 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 satellite image) is critical to make life-saving decisions (e.g., when and where to dispatch the rescue teams) [12]. An important question that remains to be answered in this challenge 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 uncertainty-aware convolutional neural network (UA-CNN) approach to solve the SISR problem in remote sensing applications. In particular, to address the first challenge, we develop a duo-branch neural network design that consists of two effective yet complementary convolutional neural networks (i.e., Duo-CNN). The Duo-CNN reconstructs satellite images with high perceptual quality by designing a new hybrid upscaling neural network architecture that effectively reduces the noises and keeps the structural integrity in the reconstruction process. To address the second challenge, we develop an uncertainty-driven ensemble model that integrates the uncertainty quantification with the deep convolutional neural networks in Duo-CNN to improve the quality of the reconstructed satellite images. To the best of our knowledge, UA-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 framework makes it possible to reconstruct a high resolution image with perceptual quality assurance from a single low resolution image. We evaluate UA-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 UA-CNN significantly outperforms the state-of-the-art SISR baselines by reconstructing satellite images with higher perception quality under various evaluation scenarios.

The main contributions of this paper are summarized below:

- We develop an UA-CNN framework to address the single-image super-resolution problem in remote sensing.
- We develop a hybrid duo-branch neural network architecture (Duo-CNN) in UA-CNN to reconstruct the images with a high perceptual quality.
- We integrate an uncertainty quantification model with deep neural networks in UA-CNN to improve the quality of the reconstructed images.
- We perform extensive experiments to evaluate the UA-CNN through a real-world case study and the results demonstrate significant performance gains of our scheme compared to state-of-the-art SISR baselines.

2 RELATED WORK

2.1 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 [10]. Several studies have leveraged remote sensing in the realms of precision agriculture [33], traffic risk identification [51], and urban behavior observation [38]. For example, Cervone *et al.* developed a machine learning based disaster damage assessment by fusing satellite imagery with Twitter data [8]. Müller *et al.* utilized satellite imagery to assess the latent effects of conflict and human migration over the hydrological process of a river basin [34]. Zou *et al.* proposed a deep learning based feature selection for scene classification of satellite imagery [52]. Several important challenges prevail in current remote sensing applications. Examples include data irregularity [7], image obscurity due to cloud cover [9], privacy concerns [43], and noise propagation [2]. The single-image super-resolution 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 UA-CNN framework

to address this problem by developing novel convolutional neural network architectures and uncertainty quantification mechanisms.

2.2 Super-Resolution

Current solutions to the super-resolution problem can be classified into two categories: *conventional* and *deep learning* approaches [11, 20, 26, 27, 30, 36, 44]. *Conventional approaches*: Lukin *et al.* explored a regularization-based image interpolation method for image enhancement by using filtering and convergence techniques on multiple degraded resolution images [30]. Shen *et al.* proposed a specialized super-resolution reconstruction algorithm for multiple images obtained from a satellite sensor called MODIS (Moderate Resolution Imaging Spectroradiometer) and applied a linear transformation approach to recover image features [36]. Yang *et al.* presented a morphing-based super-resolution method that leverages the complementary information contained in different sub-pixels (i.e., a denomination of a pixel) among multiple low-resolution frames to construct a single high-resolution image [44]. *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 and refine the generated images through three layers of convolution operations [11]. Ledig *et al.* developed a generative adversarial network framework to generate photo-realistic high-resolution images from corresponding low-resolution images through an optimization process regularized by adversarial and perceptual losses [26]. 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 [27].

However, we found the above approaches cannot solve our SISR problem well because they often failed to provide the assured perceptual quality of the reconstructed high-quality satellite images in remote sensing. More importantly, none of these solutions effectively quantified the uncertainty of the estimated RGB values in the reconstructed satellite images. In this paper, we develop an uncertainty-aware SISR scheme that integrates the uncertainty quantification model with the deep convolutional neural networks to provide high-resolution reconstructed satellite images with quality assurance.

2.3 Uncertainty-Aware Deep Learning

Our work is also related to the uncertainty-aware deep learning techniques, which have been studied in many areas such as reinforcement learning, computer vision, image generation, and Internet-of-Things (IoT) [18, 40, 46, 47]. For example, Houthooft *et al.* designed a curiosity-driven exploration strategy for high-dimensional deep reinforcement learning by incorporating variational inference in Bayesian neural networks [18]. Yasarla *et al.* proposed a multi-scale residual learning framework based on cycle spinning that gauges the uncertainty of prediction to learn optimized model weights for image de-raining tasks [47]. Tang *et al.* developed a multi-channel generative adversarial network that leverages cascaded semantic uncertainty to improve the performance of the cross-view image translation [40]. Yao *et al.* introduced a deep learning based uncertainty estimation approach to evaluate the reliability of sensory inference data using implicit Bayesian approximation [46]. However, a unique challenge in satellite-based remote imagery is the need for perceptual quality assurance, for which the existing solutions on uncertainty quantification are not designed to address. In contrast, the UA-CNN framework is the first work that aims to leverage the quantified uncertainty to reconstruct a high-resolution satellite image from a low-resolution image with high perceptual quality.

3 PROBLEM DESCRIPTION

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

Definition 3.1. **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 × 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 3.2. **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., 112×112 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 3.3. **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., 224×224 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 3.4. **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 3.5. **Uncertainty Matrix (Σ)**: 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 [24]:

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

where $H - \widehat{H}$ is the matrix to represent the error of estimated RGB values at all pixels in the image. Σ 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 3.6. **Perceptual Quality**: To evaluate the quality of $\widehat{H}$, we use the state-of-the-art perceptual metric [50] 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 [37]). $\Gamma(\cdot)$ 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 [6, 23].

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:

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

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 an UA-CNN scheme to address these challenges, which is elaborated in the next section.

4 SOLUTION

In this section, we present the UA-CNN framework to address the super-resolution problem formulated above. We first present an overview of the framework and then discuss its components in details.

4.1 Overview of UA-CNN Framework

UA-CNN is an uncertainty-aware convolutional neural network framework to address the SISR problem in remote sensing. The overview of the UA-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.

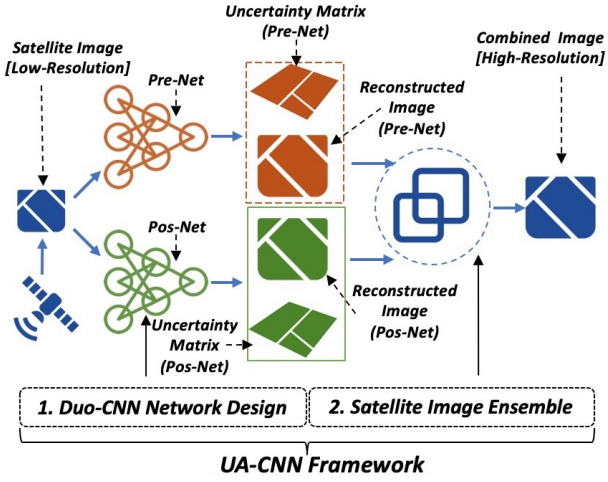


Figure 3: Overview of UA-CNN framework

4.2 Uncertainty-Aware Duo-CNN Architecture

In this subsection, we present the Duo-CNN architecture design in our framework. The Duo-CNN constructs 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 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 [14]. 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 [39]. 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 4.1. **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 Σ_{pre} as follows:

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

Table 1: Parameter Details of Pre-Net Architecture.

Module	Operation	Kernel Size	Padding	Input Channel	Output Channel
Bicubic Interpolation	N/A	N/A	N/A	3	3
Image Encoding	Reflection Pad	N/A	4	3	3
	Convolution	9	N/A	3	64
	Reflection Pad	N/A	0	64	64
	Convolution	1	N/A	64	32
Image Decoding	Reflection Pad	N/A	2	32	32
	Convolution	5	N/A	32	3
Uncertainty Matrix Generation	Reflection Pad	N/A	2	32	32
	Convolution	5	N/A	32	1

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 and Table 1, respectively. 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 [21] 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.

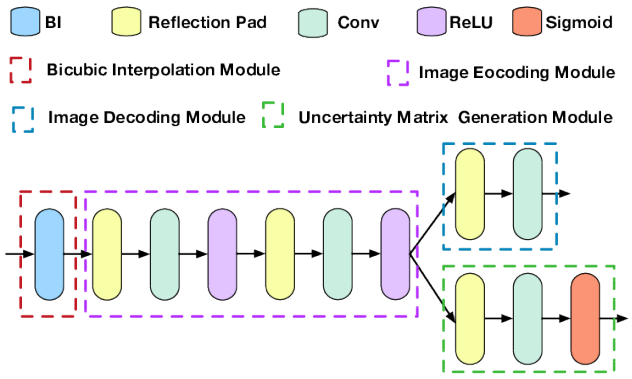


Figure 4: Illustration of Pre-upscaling Network (Pre-Net)

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

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

[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 [20].

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

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

where $\mathcal{L}_{\text{perceptual}}(H, \widehat{H}_{pre})$ is the perceptual loss [50] to quantify the perceptual difference between the actual and reconstructed images. $L_{\text{pixel}}(H, \widehat{H}_{pre})$ is the Mean Squared Error (MSE) loss [35] 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 Σ_{pre} given the difference between the *actual* and *reconstructed* satellite images (i.e., (H - $\widehat{H}_{pre}$) as defined in Definition 3.5). By observing such an estimation discrepancy often follows a normal distribution [24], we derive the likelihood function of our estimation as follows:

$$\mathbb{L}(\Sigma_{pre}|H-\widehat{H}_{pre}) = (2\pi||\Sigma_{pre}||_2)^{-\frac{1}{2}} exp(-\frac{1}{2||\Sigma_{pre}||_2}||H-\widehat{H}_{pre}||_2) \tag{7}$$

We can then derive the log-likelihood function accordingly:

$$log\mathbb{L}(\Sigma_{pre}|H-\widehat{H}_{pre}) = -\frac{1}{2}log2\pi - \frac{1}{2}log||\Sigma_{pre}||_2 - \frac{1}{2||\Sigma_{pre}||_2}||H-\widehat{H}_{pre}||_2 \tag{8}$$

Our goal is to maximize $log\mathbb{L}(\Sigma_{pre}|H-\widehat{H}_{pre})$ 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 $log\mathbb{L}(\Sigma_{pre}|H-\widehat{H}_{pre})$:

$$\mathcal{L}^{\text{pre}}_{\text{uncertain}} : \min\left(\frac{1}{2}log||\Sigma_{pre}||_2 + \frac{1}{2||\Sigma_{pre}||_2}||H-\widehat{H}_{pre}||_2 + \frac{1}{2}log2\pi\right) \tag{9}$$

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

Definition 4.2. **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

Table 2: Parameter Details of Pos-Net Architecture.

Module	Operation	Kernel Size	Padding	Input Channel	Output Channel
Image Encoding	ReflectionPad	N/A	4	3	3
	Convolution	9	N/A	3	64
Residual Block	5 × Residual blcoks	3	1	64	64
	Pixel Shuffle	3	1	64	64
Image Decoding	ReflectionPad	N/A	2	64	64
	Convolution	5	N/A	64	3
Uncertainty Matrix Generation	ReflectionPad	N/A	1	64	64
	Convolution	3	N/A	64	1

the corresponding uncertainty matrix Σ_{pos} as follows:

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

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 and Table 2, respectively. 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 [17] 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.

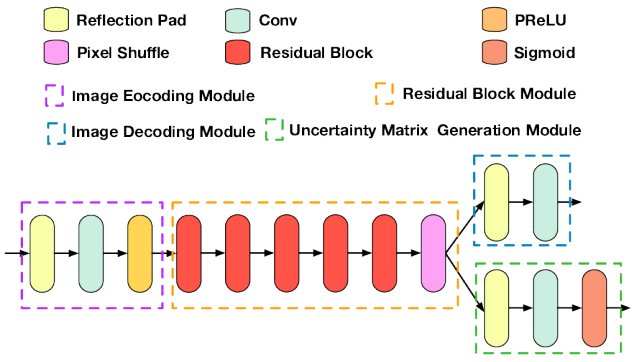


Figure 5: Illustration of Post-upscaling Network (Pos-Net)

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 Σ_{pos} as:

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

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

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

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:

$$\mathcal{L}^{\text{pos}}_{\text{uncertain}} : \min \left(\frac{1}{2} log||\Sigma_{pos}||_2 + \frac{1}{2||\Sigma_{pos}||_2} ||H - \widehat{H}_{pos}||_2 + \frac{1}{2} log 2\pi \right) \tag{13}$$

4.3 Uncertainty-driven Satellite Imagery Ensemble

In this subsection, we leverage the estimated uncertainty matrices (Σ_{pre} and Σ_{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 4.3. **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:

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

where Λ is a matrix to indicate the weights of each component at all pixels in the combined high-resolution image.

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., $\widehat{H}_{pre} \sim \mathcal{N}(H, \Sigma^2_{pre})$ and $\widehat{H}_{pos} \sim \mathcal{N}(H, \Sigma^2_{pos})$). Using these models, we can derive the distribution of $\widehat{H}_{combine}$ in Equation (14) as follows:

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

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 [1]. We enforce such an agreement by setting the variances of the two networks to be the same:

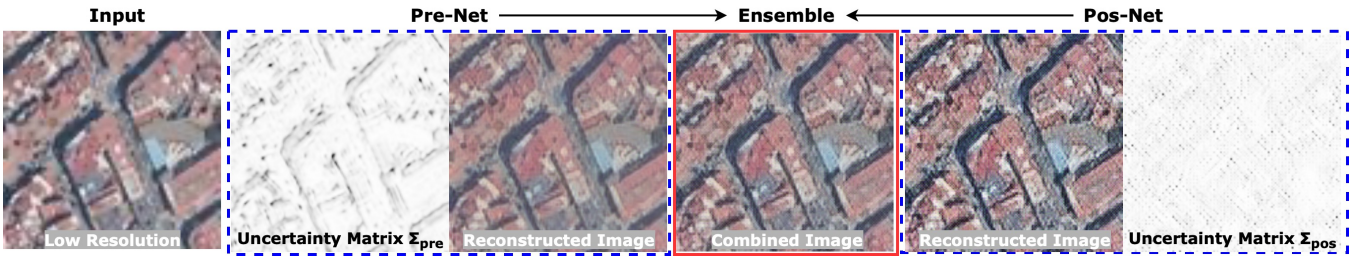


Figure 6: An Example of the Combined High-Resolution Image Generated by UA-CNN

$$((1-\Lambda)\cdot\Sigma_{pre})^2) = (\Lambda\cdot\Sigma_{pos})^2 \tag{16}$$

We can then derive the value of Λ as follows:

$$\Lambda = \frac{\Sigma_{pos}^{-1}}{\Sigma_{pre}^{-1}+\Sigma_{pos}^{-1}} \tag{17}$$

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

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

where $\widehat{H}_{combine}$ is the final output of our UA-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:

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

where $L_{\text{perceptual}}(H, \widehat{H}_{combine})$ 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 UA-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 Σ_{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 Σ_{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 UA-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 UA-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 UA-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 optimized using the Adaptive Moment Estimation (Adam) optimizer [25], which obtains the optimal parameters of both upscaling networks $PosNet^*$ and $PreNet^*$.

We summarize the UA-CNN framework in Algorithm 1. The input to the framework is the low-resolution satellite image L for each sensing cell. The output is the combined high-resolution satellite image $\widehat{H}$ for each sensing cell.

Algorithm 1 Summary of the UA-CNN Framework

```
input:L
output:Ĥ
initialize Pre-Net (Definition 4.1)
initialize Pos-Net(Definition 4.2)
epoch ← 0
while epoch < Δ do
    calculate L_pre (Equation (5))
    calculate L_pos (Equation (11))
    calculate L_combine (Equation (19))
    calculate L_overall (Equation (20))
    optimize L_overall using Adam optimizer
    update Pre-Net
    update Pos-Net
    epoch ← epoch + 1
end while
set current Pre-Net as Pre-Net*
set current Pos-Net as Pos-Net*
generate Ĥ_pre, Σ_pre from L (Equation (4))
generate Ĥ_pos, Σ_pos from L (Equation (10))
generate Ĥ_combine (Equation (18))
output Ĥ_combine as Ĥ
```

5 EVALUATION

In this section, we evaluate the performance of the UA-CNN scheme using the real-world satellite imagery data collected from two different cities in Spain, a region with diversified land features [38], through the publicly available *Google Maps Platform*. We compare the performance of UA-CNN with state-of-the-art conventional and deep learning schemes for single-image super-resolution task in

[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.

[2] Δ is usually set to be a large number (i.e., larger than 500) to ensure learned model quality. The training process also stops when $\mathcal{L}_{\text{overall}}$ is stable on validation set.

remote sensing. The results show that UA-CNN consistently outperforms the baselines by achieving the least perception errors in reconstructing high-resolution satellite images.

5.1 Dataset

In our experiment, we collect real-world satellite imagery datasets from two different cities in Spain (i.e., *Barcelona* and *Madrid*) that belong 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 the *high* resolution satellite image in our evaluation [4]. 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). In addition, we use the *Urban Atlas* dataset published by the European Environment Agency [5] to determine the land usage classes for all collected satellite images. 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

5.2 Baselines

We compare UA-CNN with the state-of-the-art *conventional* and *deep learning* baselines that are used to solve the SISR problem.

(1) Conventional Models
- **Nearest-neighbour (NN)** [19]: it is a conventional image upscaling scheme that fills each empty pixel with the same RGB value as the nearest available neighboring pixel.
- **Bi-linear/quadratic/cubic** [20]: it is a set of super resolution scheme that leverages the bi-linear/quadratic/cubic interpolation technique to generate an estimated RGB value for each empty pixel from its neighboring pixels.

(2) Deep Learning Models
- **SR-CNN** [11]: it utilizes the bi-cubic interpolation and conventional neural networks to generate the high-resolution image with a dedicated image refining process to improve the quality of reconstructed images.
- **SR-GAN** [26]: it imposes a generative adversarial network architecture that utilizes an image generator network and an image discriminator network to refine the reconstructed high-resolution images.
- **SR-ResNet** [27]: 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.

5.3 Evaluation Metrics and Settings

To evaluate the performance of all compared schemes, we use the perceptual metric (discussed in Definition 3.6), which has proven to be an accurate metric that is close to human perception in the recent computer vision studies [6, 23, 50]. In particular, the perceptual metric evaluates the image quality by comparing the difference between the deep features extracted from the *actual* and *reconstructed* satellite images using the ImageNet-trained deep convolutional neural networks (e.g., VGG [37]). Following [6, 50], we use three commonly used deep features extracted by 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 three error computing functions (i.e., $\Gamma(\cdot)$ in Definition 3.6): i) *Mean Absolute Error (MAE)*; ii) *Mean Squared Error (MSE)*; iii) *Log-Cosh Error (LCE)* [16] to calculate the difference between the deep features extracted from the *actual* and *reconstructed* satellite images. This is to ensure a comprehensive and robust evaluation of all compared schemes (e.g., MSE is robust in evaluating large errors, MAE is sensitive to small errors, and Log-Cosh achieves a balance between MSE and MAE). 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.

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 [25]. In particular, we set the learning rate to be 1e-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.

5.4 Evaluation Results on Perceptual Quality

Evaluation results on *urban fabric*: In the first set of experiments, we study the performance of all compared schemes in *Barcelona* and *Madrid*, where the land usage class of images is urban-fabric. The evaluation results are presented in Table 3 and Table 4. We observed that the UA-CNN scheme consistently outperforms all compared baselines across different deep features. For example,

[3] https://developers.google.com/maps/documentation/
[4] https://scikit-image.org/docs/dev/api/skimage.transform.html#skimage.transform.resize
[5] https://www.eea.europa.eu/data-and-maps/data/urban-atlas/

Table 3: Performance Comparisons (Class = *Urban Fabric*, City = *Barcelona*)

		Deep Feature = VGG_{4-1}			Deep Feature = VGG_{4-2}			Deep Feature = VGG_{4-3}		
Category	Algorithm	MAE	MSE	LCE	MAE	MSE	LCE	MAE	MSE	LCE
Conventional Model	Nearest-neighbor	1.1546	4.9464	0.8768	0.6487	2.5650	0.4842	0.4962	1.6637	0.3524
	Bi-linear	1.1396	5.0302	0.8689	0.6442	2.6017	0.4812	0.4891	1.6467	0.3458
	Bi-quadratic	1.1091	4.7499	0.8404	0.6222	2.4396	0.4619	0.4703	1.5286	0.3295
	Bi-cubic	1.1125	4.7782	0.8435	0.6253	2.4584	0.4645	0.4730	1.5450	0.3319
Deep Learning Model	SR-CNN	1.1138	4.5774	0.8375	0.6190	2.3757	0.4588	0.4634	1.4894	0.3253
	SR-GAN	1.0551	4.2401	0.7855	0.5780	2.1161	0.4210	0.4321	1.3126	0.2967
	SR-ResNet	1.0601	4.2572	0.7894	0.5800	2.1250	0.4226	0.4346	1.3255	0.2989
Our Model	**UA-CNN**	**1.0198**	**3.9821**	**0.7532**	**0.5563**	**1.9747**	**0.4016**	**0.4160**	**1.2291**	**0.2830**

Table 4: Performance Comparisons (Class = *Urban Fabric*, City = *Madrid*)

		Deep Feature = VGG_{4-1}			Deep Feature = VGG_{4-2}			Deep Feature = VGG_{4-3}		
Category	Algorithm	MAE	MSE	LCE	MAE	MSE	LCE	MAE	MSE	LCE
Conventional Model	Nearest-neighbor	1.2060	5.3422	0.9240	0.6962	2.8903	0.5274	0.5247	1.9278	0.3817
	Bi-linear	1.2458	5.9456	0.9678	0.7152	3.1382	0.5460	0.5293	1.9538	0.3838
	Bi-quadratic	1.2084	5.5693	0.9324	0.6893	2.9282	0.5229	0.5088	1.8113	0.3658
	Bi-cubic	1.2132	5.6149	0.9369	0.6934	2.9572	0.5265	0.5121	1.8341	0.3688
Deep Learning Model	SR-CNN	1.1604	4.9477	0.8813	0.6519	2.6021	0.4890	0.4814	1.6420	0.3441
	SR-GAN	1.1130	4.6791	0.8391	0.6157	2.3641	0.4549	0.4538	1.4706	0.3182
	SR-ResNet	1.1088	4.6512	0.8355	0.6144	2.3512	0.4538	0.4525	1.4598	0.3171
Our Model	**UA-CNN**	**1.0548**	**4.2422**	**0.7860**	**0.5798**	**2.1204**	**0.4227**	**0.4253**	**1.3131**	**0.2937**

the performance gains of UA-CNN over the best-performing baseline (i.e., SR-GAN) in Barcelona with the deep feature extracted by VGG_{4-1} on MAE, MSE, and LSE are 3.46%, 6.47%, and 4.28%, respectively. Such performance gains mainly come from the fact that UA-CNN judiciously learns the uncertainty of the estimated RGB values in the reconstructed high-resolution satellite image through an integrated Duo-CNN and MLE hybrid design. The obtained uncertainty matrix is explicitly used to guide the reconstruction of the combined satellite image from the ones generated by both *pre-upscaling* and *post-upscaling* networks.

Evaluation results on *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 UA-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 5 to Table 8. We observe that UA-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 UA-CNN compared to the best-performing baseline (i.e., SR-GAN) in the forest and green land class in Madrid with the deep feature extracted by VGG_{4-2} on MAE, MSE, and LSE are 3.87%, 7.16%, and 4.99%, respectively. Similar performance gains are also observed in the transportation class in both cities. Such consistent performance gains demonstrate the effectiveness and robustness of UA-CNN in learning the accurate uncertainty matrices to guide the convolutional neural networks to output high-quality super-resolution results across diversified classes of land usage in remote sensing applications. 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.

Cosine similarity for all compared schemes. In our experiment, we also adopt the cosine similarity [48] to measure the similarity between the deep features extracted from the *actual* and *reconstructed* satellite images for all compared schemes. The cosine similarity is known to be more robust than the distance-based evaluation metrics (e.g., MSE) to the *curse of dimensionality* problem (i.e., the high-dimensional deep feature vectors extracted from satellite images) [32]. Intuitively, a higher cosine similarity score represents a higher visual similarity between the *actual* and *reconstructed* satellite images, which provides an alternative and more intuitive perspective to evaluate the performance of all schemes.

Table 5: Performance Comparisons (Class = *Forrest and Green Land*, City = *Barcelona*)

		Deep Feature = VGG_{4-1}			Deep Feature = VGG_{4-2}			Deep Feature = VGG_{4-3}		
Category	Algorithm	MAE	MSE	LCE	MAE	MSE	LCE	MAE	MSE	LCE
Conventional Model	Nearest-neighbor	0.7527	2.4017	0.5255	0.4763	1.2522	0.3115	0.3819	0.9518	0.2422
	Bi-linear	0.6964	2.1695	0.4819	0.4325	1.0525	0.2741	0.3276	0.6925	0.1953
	Bi-quadratic	0.6865	2.0991	0.4726	0.4233	1.0136	0.2670	0.3195	0.6589	0.1888
	Bi-cubic	0.6890	2.1160	0.4750	0.4261	1.0246	0.2692	0.3219	0.6692	0.1909
Deep Learning Model	SR-CNN	0.6943	2.1025	0.4757	0.4280	1.0362	0.2716	0.3168	0.6529	0.1879
	SR-GAN	0.6193	1.7065	0.4098	0.3690	0.7846	0.2205	0.2712	0.4907	0.1508
	SR-ResNet	0.6180	1.7003	0.4088	0.3677	0.7788	0.2194	0.2701	0.4876	0.1501
Our Model	**UA-CNN**	**0.5991**	**1.6046**	**0.3923**	**0.3554**	**0.7351**	**0.2099**	**0.2617**	**0.4608**	**0.1436**

Table 6: Performance Comparisons (Class = *Forrest and Green Land*, City = *Madrid*)

		Deep Feature = VGG_{4-1}			Deep Feature = VGG_{4-2}			Deep Feature = VGG_{4-3}		
Category	Algorithm	MAE	MSE	LCE	MAE	MSE	LCE	MAE	MSE	LCE
Conventional Model	Nearest-neighbor	1.0164	4.0094	0.7545	0.6208	2.2225	0.4501	0.4795	1.5691	0.3359
	Bi-linear	0.9579	3.7372	0.7072	0.5729	1.9716	0.4086	0.4240	1.2430	0.2864
	Bi-quadratic	0.9363	3.5478	0.6868	0.5559	1.8635	0.3940	0.4108	1.1667	0.2750
	Bi-cubic	0.9394	3.5738	0.6897	0.5590	1.8807	0.3966	0.4133	1.1809	0.2772
Deep Learning Model	SR-CNN	0.9445	3.4894	0.6890	0.5549	1.8315	0.3930	0.4037	1.1477	0.2715
	SR-GAN	0.8396	2.8780	0.5964	0.4802	1.4159	0.3257	0.3458	0.8608	0.2206
	SR-ResNet	0.8483	2.9354	0.6042	0.4847	1.4428	0.3298	0.3487	0.8744	0.2231
Our Model	**UA-CNN**	**0.8095**	**2.6945**	**0.5695**	**0.4623**	**1.3212**	**0.3102**	**0.3328**	**0.8036**	**0.2098**

The results are shown in Figure 8. We observe that our UA-CNN scheme is able to achieve the highest cosine similarity score in all evaluation scenarios. Such performance gains further validate the effectiveness of our UA-CNN scheme in reconstructing satellite images with better quality than the state-of-the-art baselines.

6 CONCLUSION

In this paper, we develop an UA-CNN approach to address the single-image super-resolution problem in remote sensing applications. In particular, the UA-CNN scheme addresses two intrinsic challenges (i.e., perceptual quality assurance and uncertainty-aware super resolution). The UA-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 UA-CNN on a real-world remote sensing application involving land usage classification. The results demonstrate that our UA-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-resolution satellite images that are not always available to the applications (e.g., disaster assessment, poverty prediction, disease outbreak detection).

REFERENCES

[1] 2011. Uncertainty analysis: the variance of the variance. (2011).

[2] Salem Saleh Al-Amri, Namdeo V Kalyankar, and Santosh D Khamitkar. 2010. A comparative study of removal noise from remote sensing image. *International Journal of Computer Science Issues (IJCSI)* 7, 1 (2010).

[3] KA Al-Gaadi, MS Samdani, and VC Patil. 2011. Assessment of temporal land cover changes in Saudi Arabia using remotely sensed data. *Middle-East J. Sci. Res* 9 (2011), 711–717.

[4] Adrian Albert, Jasleen Kaur, and Marta C Gonzalez. 2017. 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, 1357–1366.

[5] Judit Bartholy, Rita Pongracz, Zoltan Barcza, and Zsuzsanna Dezso. 2004. Aspects of urban/rural population migration in the Carpathian basin using satellite imagery. In *Environmental Change and its Implications for Population Migration*. Springer, 289–313.

[6] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. 2018. The 2018 PIRM challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 0–0.

[7] Evan B Brooks, Valerie A Thomas, Randolph H Wynne, and John W Coulston. 2012. Fitting the multitemporal curve: A Fourier series approach to the missing data problem in remote sensing analysis. *IEEE Transactions on Geoscience and Remote Sensing* 50, 9 (2012), 3340–3353.

Table 7: Performance Comparisons (Class = *Transportation*, City = *Barcelona*)

Category	Algorithm	Deep Feature = VGG_{4-1}			Deep Feature = VGG_{4-2}			Deep Feature = VGG_{4-3}		
		MAE	MSE	LCE	MAE	MSE	LCE	MAE	MSE	LCE
Conventional Model	Nearest-neighbor	0.8189	3.0102	0.5917	0.5070	1.6158	0.3516	0.3964	1.1002	0.2619
	Bi-linear	0.7941	2.9851	0.5758	0.4973	1.5824	0.3430	0.3845	1.0659	0.2521
	Bi-quadratic	0.7707	2.7832	0.5535	0.4768	1.4610	0.3257	0.3673	0.9733	0.2376
	Bi-cubic	0.7736	2.8094	0.5563	0.4801	1.4788	0.3284	0.3701	0.9880	0.2399
Deep Learning Model	SR-CNN	0.7759	2.7210	0.5533	0.4761	1.4405	0.3249	0.3657	0.9611	0.2367
	SR-GAN	0.7690	2.6893	0.5460	0.4645	1.3871	0.3146	0.3552	0.9047	0.2269
	SR-ResNet	0.7683	2.6874	0.5455	0.4639	1.3850	0.3143	0.3543	0.8996	0.2261
Our Model	**UA-CNN**	**0.7442**	**2.5450**	**0.5242**	**0.4491**	**1.3130**	**0.3016**	**0.3429**	**0.8546**	**0.2168**

Table 8: Performance Comparisons (Class = *Transportation*, City = *Madrid*)

Category	Algorithm	Deep Feature = VGG_{4-1}			Deep Feature = VGG_{4-2}			Deep Feature = VGG_{4-3}		
		MAE	MSE	LCE	MAE	MSE	LCE	MAE	MSE	LCE
Conventional Model	Nearest-neighbor	1.0101	4.1899	0.7579	0.6157	2.2838	0.4503	0.4717	1.5432	0.3314
	Bi-linear	0.9788	4.1157	0.7348	0.5962	2.1983	0.4332	0.4492	1.4317	0.3114
	Bi-quadratic	0.9487	3.8388	0.7065	0.5727	2.0376	0.4128	0.4296	1.3143	0.2945
	Bi-cubic	0.9530	3.8773	0.7106	0.5769	2.0634	0.4163	0.4331	1.3349	0.2976
Deep Learning Model	SR-CNN	0.9601	3.7656	0.7106	0.5725	2.0186	0.4129	0.4286	1.3104	0.2949
	SR-GAN	0.9099	3.4976	0.6671	0.5330	1.7833	0.3768	0.3962	1.1355	0.2659
	SR-ResNet	0.9054	3.4727	0.6631	0.5304	1.7708	0.3745	0.3940	1.1249	0.2641
Our Model	**UA-CNN**	**0.8719**	**3.2451**	**0.6331**	**0.5103**	**1.6500**	**0.3567**	**0.3786**	**1.0481**	**0.2510**

[8] Guido Cervone, Emily Schnebele, Nigel Waters, Martina Moccaldi, and Rosa Sicignano. 2017. Using social media and satellite data for damage assessment in urban areas during emergencies. In *Seeing Cities Through Big Data*. Springer, 443–457.

[9] Gavin Q Collins, Matthew J Heaton, and Leiqiu Hu. 2019. Physically constrained spatiotemporal modeling: generating clear-sky constructions of land surface temperature from sparse, remotely sensed satellite data. *Journal of Applied Statistics* (2019), 1–21.

[10] Jadunandan Dash and Booker O Ogutu. 2016. Recent advances in space-borne optical remote sensing systems for monitoring global terrestrial ecosystems. *Progress in Physical Geography* 40, 2 (2016), 322–351.

[11] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2016. Image Super-resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 2 (2016), 295–307.

[12] Jigar Doshi, Saikat Basu, and Guan Pang. 2018. From satellite imagery to disaster insights. *arXiv preprint arXiv:1812.07033* (2018).

[13] Dronenerds. [n. d.]. Micasense Altum Sensor. ([n. d.]). https://www.dronenerds.com/products/cameras-sensors/multispectral/micasense/micasense-altum-multispectral-sensor-dji-skyport-kit-805-00040-micasense.html

[14] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. 2016. A learned representation for artistic style. *arXiv preprint arXiv:1610.07629* (2016).

[15] Michael J Falkowski, Michael A Wulder, Joanne C White, and Mark D Gillis. 2009. Supporting large-area, sample-based forest inventories with very high spatial resolution satellite imagery. *Progress in Physical Geography* 33, 3 (2009), 403–423.

[16] Anita S Gangal, PK Kalra, and DS Chauhan. 2007. Performance evaluation of complex valued neural networks using various error functions. *Enformatika* 23 (2007), 27–32.

[17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[18] Rein Houthooft, Xi Chen, Yan Duan, John Schulman, Filip De Turck, and Pieter Abbeel. 2016. Curiosity-driven exploration in deep reinforcement learning via bayesian neural networks. *arXiv preprint arXiv:1605.09674* (2016).

[19] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. 2015. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5197–5206.

[20] Zheng Hui, Xiumei Wang, and Xinbo Gao. 2018. Fast and accurate single image super-resolution via information distillation network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 723–731.

[21] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1125–1134.

[22] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. 2016. Combining satellite imagery and machine learning to predict poverty. *Science* 353, 6301 (2016), 790–794.

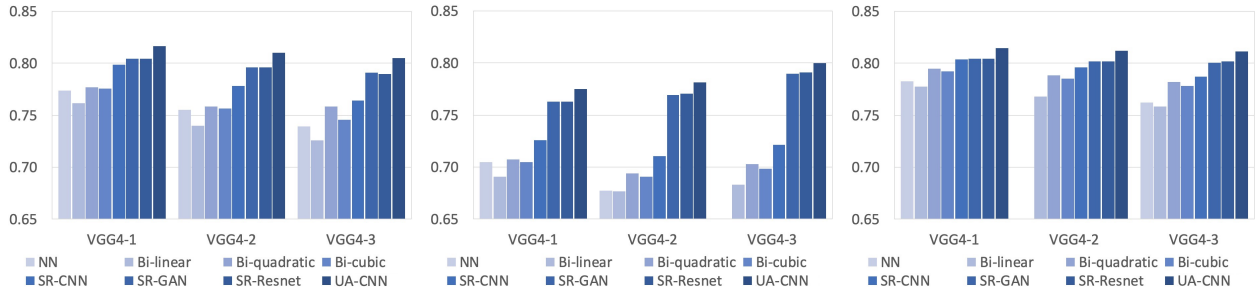
[23] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*. Springer, 694–711.

[24] Alex Kendall, Yarin Gal, and Roberto Cipolla. 2018. 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*. 7482–7491.

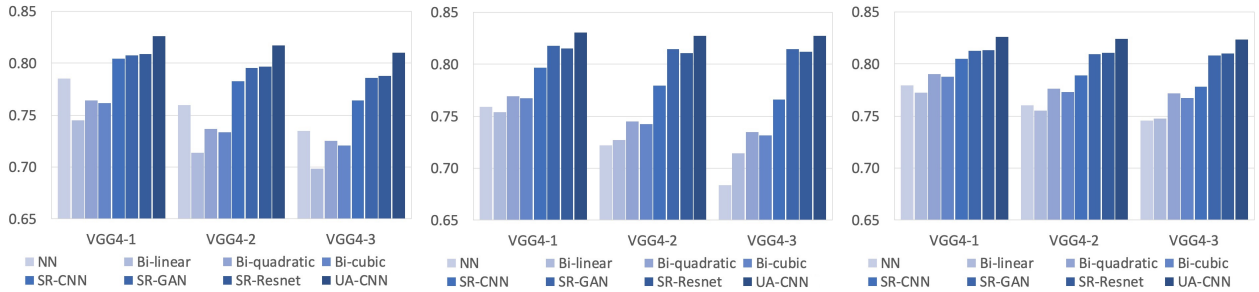
[25] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[26] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4681–4690.

[27] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*



(a) Class = *Urban Fabric*, City = *Barcelona* (b) Class = *Forrest and Green Land*, City = *Barcelona* (c) Class = *Transportation*, City = *Barcelona*



(d) Class = *Urban Fabric*, City = *Madrid* (e) Class = *Forrest and Green Land*, City = *Madrid* (f) Class = *Transportation*, City = *Madrid*

Figure 8: Average Cosine Similarity for All Compared Schemes

workshops. 136–144.

[28] Brad Lobitz, Louisa Beck, Anwar Huq, Byron Wood, George Fuchs, ASG Faruque, and Rita Colwell. 2000. Climate and infectious disease: use of remote sensing for detection of Vibrio cholerae by indirect measurement. *Proceedings of the National Academy of Sciences* 97, 4 (2000), 1438–1443.

[29] Dengsheng Lu and Qihao Weng. 2006. Use of impervious surface in urban land-use classification. *Remote Sensing of Environment* 102, 1-2 (2006), 146–160.

[30] Alexey Lukin, Andrey S Krylov, and Andrey Nasonov. 2006. Image interpolation by super-resolution. In *Proceedings of GraphiCon*, Vol. 2006. Citeseer, 239–242.

[31] Brian L Markham, James C Storey, Darrel L Williams, and James R Irons. 2004. Landsat sensor performance: history and current status. *IEEE Transactions on Geoscience and Remote Sensing* 42, 12 (2004), 2691–2694.

[32] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

[33] David J Mulla. 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems engineering* 114, 4 (2013), 358–371.

[34] Marc François Müller, Jim Yoon, Steven M Gorelick, Nicolas Avisse, and Amaury Tilmant. 2016. Impact of the Syrian refugee crisis on land use and transboundary freshwater resources. *Proceedings of the national academy of sciences* 113, 52 (2016), 14932–14937.

[35] Mehdi SM Sajjadi, Bernhard Scholkopf, and Michael Hirsch. 2017. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*. 4491–4500.

[36] Huanfeng Shen, Michael K Ng, Pingxiang Li, and Liangpei Zhang. 2007. Super-resolution reconstruction algorithm to MODIS remote sensing images. *Comput. J.* 52, 1 (2007), 90–100.

[37] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[38] M Stellmes, Achim Röder, T Udelhoven, and J Hill. 2013. Mapping syndromes of land change in Spain with remote sensing time series, demographic and climatic data. *Land Use Policy* 30, 1 (2013), 685–702.

[39] Ying Tai, Jian Yang, and Xiaoming Liu. 2017. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3147–3155.

[40] Hao Tang, Dan Xu, Nicu Sebe, Yanzhi Wang, Jason J Corso, and Yan Yan. 2019. 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*. 2417–2426.

[41] Caglayan Tuna, Gozde Unal, and Elif Sertel. 2018. Single-frame super resolution of remote-sensing images by convolutional neural networks. *International journal of remote sensing* 39, 8 (2018), 2463–2479.

[42] Sergey Venevsky. [n. d.]. Emergence of climate change ecology. ([n. d.]).

[43] Eric Ke Wang, Fan Wang, RP Sun, and Xi Liu. 2019. A new privacy attack network for remote sensing images classification with small training samples. *Mathematical biosciences and engineering: MBE* 16, 5 (2019), 4456–4476.

[44] Jianchao Yang and Thomas Huang. 2010. Image super-resolution: Historical overview and future challenges. *Super-resolution imaging* (2010), 20–34.

[45] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. 2019. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia* (2019).

[46] Shuochao Yao, Yiran Zhao, Huajie Shao, Chao Zhang, Aston Zhang, Dongxin Liu, Shengzhong Liu, Lu Su, and Tarek Abdelzaher. 2018. Apdeepsense: Deep learning uncertainty estimation without the pain for iot applications. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 334–343.

[47] Rajeev Yasarla and Vishal M Patel. 2019. Uncertainty Guided Multi-Scale Residual Learning-using a Cycle Spinning CNN for Single Image De-Raining. *arXiv preprint arXiv:1906.11129* (2019).

[48] Jun Ye. 2011. Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and computer modelling* 53, 1-2 (2011), 91–97.

[49] Ban Yifang, Peng Gong, and Chandra Gini. 2015. Global land cover mapping using Earth observation satellite data: Recent progresses and challenges. *ISPRS journal of photogrammetry and remote sensing (Print)* 103, 1 (2015), 1–6.

[50] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 586–595.

[51] Yang Zhang, Yiwen Lu, Daniel Zhang, Lanyu Shang, and Dong Wang. 2018. RiskSens: A Multi-view Learning Approach to Identifying Risky Traffic Locations in Intelligent Transportation Systems Using Social and Remote Sensing. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 1544–1553.

[52] Qin Zou, Lihao Ni, Tong Zhang, and Qian Wang. 2015. Deep learning based feature selection for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters* 12, 11 (2015), 2321–2325.