Exploring the Potential of Reconstructed Multispectral Images for Urban Tree Segmentation in Street View Images

Tito Arevalo-Ramirez, Anali Alfaro, Jose M. Saavedra, Matías Recabarren, Mauricio Ponce-Donoso, and José Delpiano

***Abstract*—Deep learning has gained popularity in recent years for reconstructing hyperspectral and multispectral images, offering cost-effective solutions and promising results. Research on hyperspectral image reconstruction feeds deep learning models with images at specific wavelengths and outputs images in other spectral bands. Although encouraging results of previous works, it should be determined to what extent the reconstructed information can lead to an advantage over the captured images. In this context, the present work inspects whether or not reconstructed spectral images add relevant information to segmentation networks for improving urban tree identification. Specifically, we generate red-edge (ReD) and near-infrared (NIR) images from RGB images using a conditional Generative Adversarial Network (cGAN). The training and validation are carried out with 5770 multispectral images obtained after a custom data augmentation process using an urban hyperspectral dataset. The testing outcomes reveal that ReD and NIR can be generated with an average structural similarity index measure of 0.93 and 0.88, respectively. Next, the cGAN generates ReD and NIR information of two RGB-based urban tree datasets (i.e., Jekyll, 3949 samples, and Arbocensus, 317 samples). Subsequently, DeepLabV3 and SegFormer segmentation networks are trained, validated, and tested using RGB, RGB+ReD, and RGB+NIR images from Jekyll and Arbocensus datasets. The experiments show that reconstructed multispectral images might not add information to segmentation networks that enhance their performance. Specifically, the p-values from a T-test show no significant difference between the performance of segmentation networks.**

***Index Terms*—Image to image translation, multispectral features, neural networks, semantic segmentation, urban trees.**

I. Introduction

Hyperspectral and multispectral images are remotely sensed data that retrieve spatial and spectral knowledge in different electromagnetic spectrum bands. This information is commonly captured by passive sensors that capture the energy that is reflected or emitted by objects and can be used to improve the recognition and characterization of objects [1]. For instance, hyperspectral images, rich data cubes, have been widely used to detect and recognize objects (e.g., buildings and highways) and land covers [2], [3]. Furthermore, the reflectance captured by hyperspectral/multispectral cameras can be used to characterize vegetation, identify tree species, and infer water stress, among other tasks [4], [5], [6]. Therefore, spectral information is valuable for boosting the description of objects or regions of interest.

Although the advantages that hyperspectral/multispectral images can yield, they are not always accessible. Spectral information is retrieved by specialized sensors, which are high-cost due to their manufacturing technology. Further, different multispectral camera manufacturers have different preferences for central wavelengths and bandwidth, deriving different spectral reflectance values for the same vegetation [7]. Since the restrictions that hardware availability can impose on capturing spectral data, researchers have addressed the lack of it by software strategies. In particular, deep learning models have become an essential tool for inferring the object's spectral reflectance [8], [9], [10], [11], [12]. In recent years, the reconstruction of hyperspectral/multispectral images has been encouraged by the new trends in image restoration and enhancement competitions [13], [14]. Since the promising outcomes reported in these events, recent works have pushed forward deep learning strategies for multispectral reconstruction. For instance, Aslahishahri et al. [12] showed that a conditional generative adversarial network (cGAN) can be used for predicting near-infrared (NIR) images from RGB images with a structural similarity index measure (SSIM) of 92.28%. Furthermore, Deng et al. [8] has proposed a multispectral to hyperspectral network for determining hyperspectral images within a more comprehensive spectral range (380–2500 nm). The hyperspectral images can be estimated with a root mean squared error between 0.010 and 0.016 [8]. The works mentioned before support the idea that deep

Manuscript received 18 January 2024; revised 5 May 2024 and 7 June 2024; accepted 15 June 2024. Date of publication 26 June 2024; date of current version 12 July 2024. This work was supported by the Agencia Nacional de Investigación y Desarrollo (ANID) under Grant Fondecyt 11220510, Grant FONDEF ID21I10360, and Grant Tree Fund 21-JD-01. The work of Tito Arevalo-Ramirez was supported by the Pontificia Universidad Católica de Chile under Grant PIA-2023-VRA-PUC. The work of José Delpiano was supported by the Advanced Center of Electrical and Electronic Engineering, AC3E under Grant ANID/FB0008 and Grant Arbotag Chile. *(Corresponding author: Tito Arevalo-Ramirez.)*

Tito Arevalo-Ramirez is with the Department of Mechanical and Metallurgical Engineering, Department of Electrical Engineering, Pontificia Universidad Católica de Chile, Santiago 8331150, Chile (e-mail: tito.arevalo@uc.cl).

Anali Alfaro, Jose M. Saavedra, Matías Recabarren, and José Delpiano are with the Faculty of Engineering and Applied Sciences, Universidad de los Andes, Santiago 12455, Chile.

Mauricio Ponce-Donoso is with the Sociedad Chilena de Arboricultura, Santiago, Universidad de los Andes, Santiago 12455, Chile.

Digital Object Identifier 10.1109/JSTARS.2024.3419127

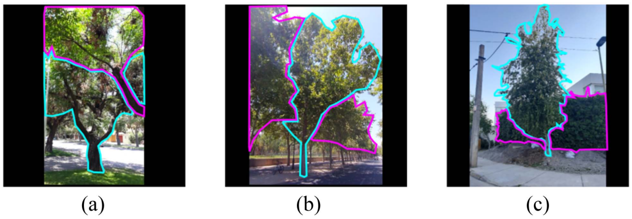


Fig. 1. Tree segmentation challenges. Tree of interest in enclosed by cyan lines. Unwanted objects that might affect tree segmentation are shown by magenta regions. (a) Occluded tree. (b) Combined crowns. (c) Complex background.

learning approaches can predict reliable spectral knowledge, which could be further exploited.

The generarion of hyperspectral images through GAN models have been crucial for classification task where there exist few training samples [15], [16], [17], [18], [19]. In particular, Alipour-Fard and Arefi [15] proposed a GAN for generating virtual training samples of hyperspectral images such that the identification of objects is improved. Specifically, a semisupervised framework extracts spectral features using a custom GAN for hyperspectral image classification. Further, GAN can also be used to address target and anomaly detection on hyperspectral images [18], [20], [21]. Regarding anomalies, GANs cannot effectively reconstruct the spectral information of them; thus, anomalies can be identified in areas where significant restoration errors exist [18]. Therefore, generative models can yield substantial spectral information to enhance the identification of objects and areas of interest.

Despite the works mentioned before supporting the application of generative networks, challenges remain and should be explored and addressed [15], [19], [21]. In particular, most of the previous works take advantage of generative models for boosting the detection of artificial objects from aerial perspectives, leaving aside the identification of individual trees or specific vegetation features from other perspectives rather than aerials. Based on the literature review, just one work explores the application of deep learning-generated spectral information for describing vegetation [7]. Specifically, the authors focused on characterizing the vegetation using the normalized difference vegetation index and normalized difference red-edge (ReD) index generated by a GAN network. Note that vegetation characterization is carried out using aerial images and detection tasks are not performed.

Since the works regarding tree and vegetation identification (e.g., semantic segmentation and classification) using deep learning-generated spectral images are still scarce, the current work explores whether GAN models can yield spectral information that improves the semantic segmentation of trees. The identification of trees is performed in urban scenes using street-view images. In particular, we are interested in urban trees because of their valuable ecosystem services (e.g., carbon capture) and the remaining segmentation challenges [22].

Pixel-wise identification of trees from street-view images is still arduous because of the inherent challenges to urban environments (e.g., tree occlusion and complex background); see Fig. 1. Specifically, most of the artificial intelligence-based strategies for the automatic identification of urban trees retrieve primary information by enclosing trees with bounding boxes [23], [24], [25], [26], [27], [28], [29]. Few works that address urban tree segmentation by neural networks perform fine segmentation of tree components, such as trunk and crown [22], [25], [27]. However, reliable segmentation of individual trees is achieved on ideal tree images (the tree is detectable without occlusion or overlapped by obstacles or other trees) [22].

In this context, the current work assesses whether or not reconstructed multispectral images at the NIR and ReD bands improve the segmentation of urban trees. In particular, we hypothesize that the knowledge retrieved by ReD and NIR regions might not be decoded by deep learning algorithms, which are trained solely with RGB images. Thus, feeding synthesized multispectral knowledge into segmentation networks might improve the identification of urban trees. The multispectral images at NIR and ReD regions are generated by a cGAN network known as Pix2Pix [30]. Two generative networks were trained for retrieving NIR and ReD images from RGB images, respectively. The cGAN models are trained, validated, and tested using the HSICityV2 dataset [31]. Next, cGAN-trained models are employed for generating multispectral images of urban tree datasets. The RGB and reconstructed multispectral images are concatenated to retrieve four-channel urban tree datasets, which are used for training, validating, and testing two neural network segmentation networks (i.e., DeepLabV3 [32] and SegFormer [33]). The strategy is validated by comparing the performance of segmentation networks trained with four-channel images and networks solely with RGB images. The experiments show no significant differences in segmentation performance can be achieved by reconstructing multispectral information.

The main contributions of the current work can be listed as follows.

1) We provide a comprehensive analysis of the impact of reconstructed multispectral information for pixel-wise urban tree identification in street-view images. Our findings challenge the notion that more spectral data should leads to better segmentation, offering critical insights into the actual value of reconstructed multispectral images in urban scenarios.
2) Assessment of the semantic segmentation networks (DeepLabV3 and SegFormer) when reconstructed multispectral images in ReD and NIR bands are used as additional sources of information alongside RGB images. We not only test the performance of these networks in a new context but also provide sharp perspectives regarding data configuration for tree identification in urban scenarios.
3) The evaluation of cGANs for reconstructing NIR and ReD channels from urban street-view RGB images. Through rigorous training and validation processes with an extensive dataset, we establish clear benchmarks for the expected reconstruction quality and its applicability in broader urban analysis.

The rest of this article is organized as follows. Description of publicly available and custom datasets and evaluation methodology are presented in Sections II and III, respectively. Section IV shows quantitative and quantitative results of the assessment.

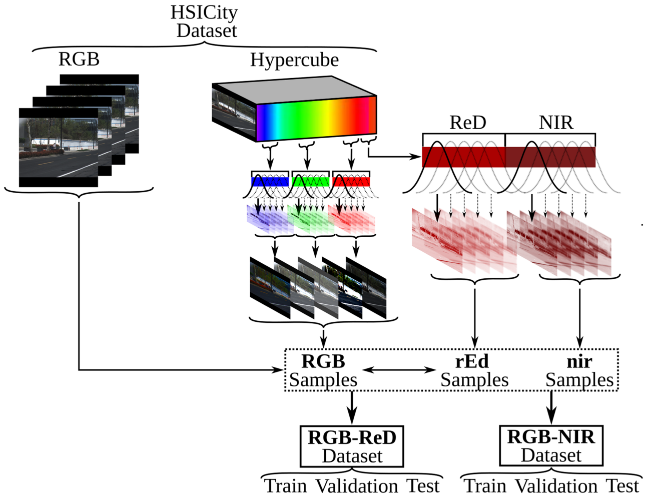


Fig. 2. Multispectral dataset generation based on the HSICityV2 dataset [31], [34].

Discussions of the results are detailed in Section V. Finally, Section VI concludes this article.

II. Data Acquisition

To answer our researh questions we employed one multispectral and two RGB-based urban tree datasets. The multispectral dataset are used to train, validate, and test the cGAN multispectral reconstruction models. On the other side, the RGB-based datasets allow to fine-tune the urban tree segmentation models. Further details about each datasets are as follows.

A. Hyperspectral City Dataset (HSICityV2)

The HSICityV2 published by the authors in [31] and [34] is an urban dataset collected in Shanghai at different light conditions using the LightGene Hypespectral sensor. This dataset is conformed by 1329 RGB images and hypercubes, where each one has 128 channels with in a spectrum range 450–950 nm and spectral resolution of 4 nm.

Based on the hypercubes, we determined a multispectral dataset. Specifically, the ReD and NIR multispectral images are determined by a weighted average of the hyperspectral images within the ReD (705–745 nm) and NIR (760–900 nm) regions [35], [36]. The weights are computed as a normal distribution with mean μ and standard deviation σ. It is essential to highlight that μ represents a central wavelength, which is computed randomly. This strategy is also used to generate new RGB images and simulate different sensors' spectral response. The blue, green, and red regions wavelengths are defined as follows 450–510, 530–590, and 640–670 nm, respectively. In this sense, we get a multispectral dataset that comprises more than five thousand RGB, ReD, and NIR images with different central wavelengths. The multispectral dataset is then divided into three subsets for ReD and NIR regions, training (5270), validation (500), and testing (880) datasets. These datasets were created by random selection of samples. Fig. 2 shows a general pipeline for generating these datasets.

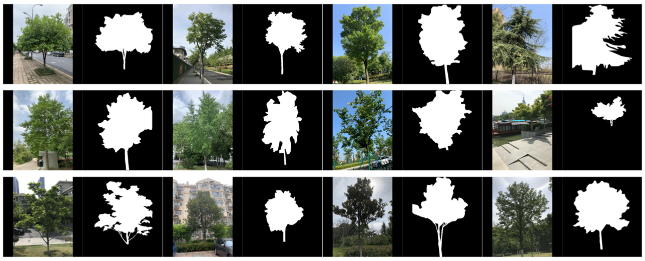


Fig. 3. Urban tree samples from the Jekyll dataset. Each RGB image's binary mask is displayed at its right.

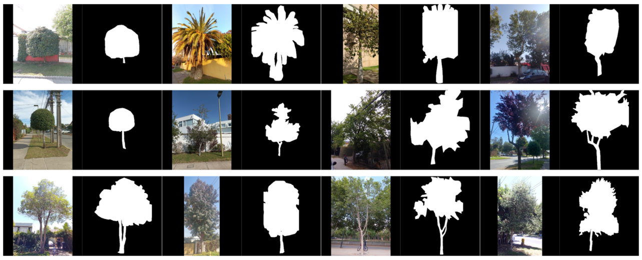


Fig. 4. Urban tree samples from the Arbocensus dataset. Each RGB image's binary mask is displayed at its right.

B. Urban Tree Dataset

1) Jekyll: Dataset is a comprehensive tree dataset of urban street view images covering over three thousand images of 22 tree species [37]. Specifically, this dataset is built by high-resolution images acquired with mobile devices in China's ten cities in spring, summer, fall, and winter; further information about data collection can be found in [37]. It is important to highlight that tree pixel-wise labeling was performed by a team of professional image annotators using LabelMe software [38]. The Jekyll dataset samples are grouped in three subsets of 3168, 395, and 386 samples for training, validation, and testing, respectively. Note that these sets are not divided randomly to guarantee tree species balance [37]. Jekyll dataset samples are shown in Fig. 3

2) Arbocensus: This is a custom dataset composed by RGB images that are captured by volunteers using smartphones. The volunteers (citizen scientists) capture about 3000 images of 100 species. The urban trees were mapped from Las Condes, and La Reina communes under the project Arbocensus in Santiago metropolitan region, Chile. Image labeling process was performed by Supervisely software [39] using polygons to outline individual objects. The arbocensus dataset generation is shown in Fig. 4. A total of 317 images were annotated and splited in 253, 32, and 32 subsets for training, evaluating, and testing segmentation algorithms, correspondingly. Samples for creating each set are selected randomly.

III. Methodology

Once the multispectral and urban tree datasets are generated, they are used to determine the cGAN multispectral reconstruction and urban tree segmentation models. Specifically, the cGAN model uses the multispectral dataset for learning a mapping from

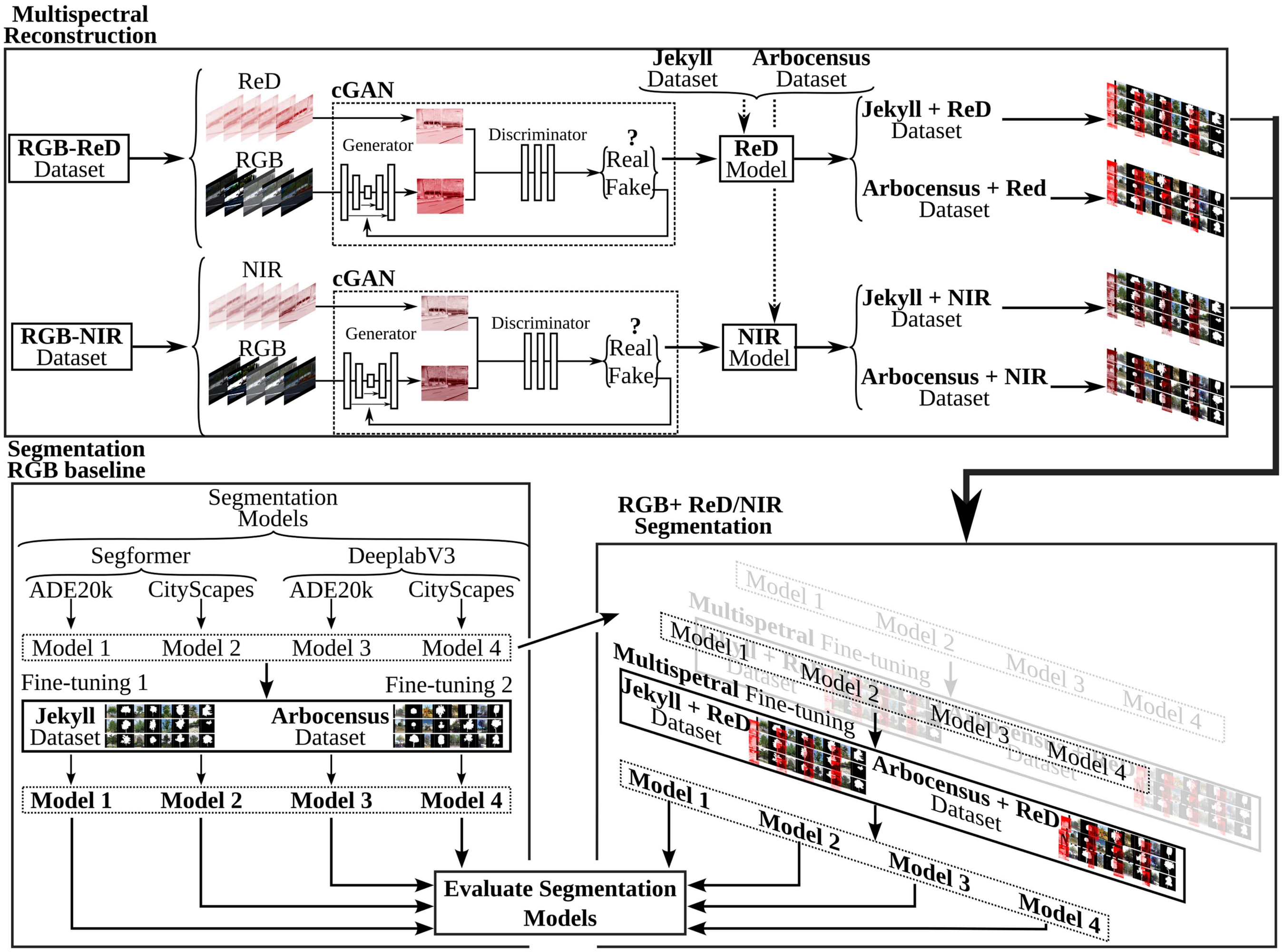


Fig. 5. General scheme of the proposed methodology for evaluating reconstructed ReD and NIR images in urban tree segmentation context.

RGB images to multispectral images in an urban environment. Next, the trained and validated cGAN model estimates the ReD and NIR spectral reflectance from the Jekyll and Arbocensus RGB images. In this sense, two new reconstructed multispectral datasets are generated for fine-tuning urban tree segmentation models. The outcomes achieved by the multispectral fine-tuned segmentation models are compared against RGB fine-tuned tree segmentation (baseline) models. The overall procedure is depicted in Fig. 5 It is essential to highlight that deep neural networks (segmentation models) are pre-trained with two urban datasets (ADE20k [40] and CityScapes [41]), which include urban vegetation instances.

A. ReD and NIR Reconstruction

Multispectral images are reconstructed by exploiting the potential of generative adversarial networks for mapping a pixel from a source image to a target image [30]. The outputs of a generator and discriminator achieve this mapping—the former attempt to produce an image similar to the target image. Next, the discriminator classifies the generated image as real or generated. The generation and discrimination process continues until the discriminator can not distinguish the real from the generated image [42]. Note that in cGAN, the generator and discriminator stages are conditioned on the source images. Therefore, in our case, the cGAN learns to reconstruct (generate) ReD and NIR images from the information retrieved by an RGB image. A general description of cGAN network is presented in Fig. 5.

1) Generator: Attempts to output a realistic multispectral image based on the RGB available information. Since the source and generated images represents the same objects, spatial details have to be kept; thus, an encoder–decoder network with skip connections is used as generator. This network is constructed by eight downsampling and upsampling blocks on the encoder and decoder sections. Specifically, each downsampling block performs a 2-D convolution, leaky linear rectification, and batch normalization. Conversely, each decoder block is composed of 2-D transposed convolution, batch normalization, and rectified linear unit layers. It is essential to highlight that skip connections are added between the i and $n - i$ layers.

2) Discriminator: Evaluates if the generator output image can be labeled as natural or generated. To perform the classification process, the discriminator classifies 70 × 70 pixels patches and averages all the individual patch outcomes to classify the image as real or generated. This procedure is performed by a Markovian discriminator, which can model the image as Markov random field. Note that pixels separated by more than a patch diameter are assumed to be independent.

In the current work, the implementation of generator and discriminator is achieved by following the guidelines exposed in [30] and [43]. Further details about generative adversarial networks can be found in [42].

B. RGB Segmentation Models

The pixel-wise identification is performed by two state-of-the-art semantic segmentation deep networks implemented in a PyTorch open-source toolbox MMSegmentation [44].

1) DeepLabV3: Is a deep convolutional neural network that exploits the potential of atrous convolution for improving its performance in semantic image segmentation tasks [32].

2) SegFormer: Is a semantic segmentation deep network that unifies transformers with lightweight multilayer perceptron decoders, which avoids complex decoders [33].

In particular, we select the DeepLabV3 with R-50-D8 backbone and SegFormer with MIT-B0 backbone from the MMSegmentation model zoo. These networks are chosen because of their performance in tree identification task ([25], [45]) and light computational requirements and acceptable mean intersection over union on ADE20k [40], and Cityscapes [41] datasets; see Benchmark and model zoo [44]. The former is a densely annotated dataset, which includes scenes, objects, parts of objects, and parts of parts pixel-wise annotations for scene understanding. We are interested in this dataset because it has more than 10 000 annotated instances of trees, buildings, persons, and walls [40]. The CityScapes dataset is a large-scale dataset of street scenes from 50 cities with fine and coarse pixel-label annotations. Specifically, it has more than 10^8 pixels finely annotated of nature (i.e., vegetation and terrain) [41]. Therefore, the DeepLabV3 and SegFormer networks pretrained on the datasets above should be capable of identifying trees. We used the segmentation models pretrained on the before-mentioned datasets in the present research. The pretrained models are publicly available in [44].

C. ReD/NIR Segmentation Models

Once the ReD and NIR reconstruction models are tested, they generate ReD and NIR new datasets using each urban tree dataset (see Section II-B). Using these new datasets DeepLabV3 and SegFormer models pretrained on ADE20 k and CityScapes dataset are fine-tuned with the new multispectral datasets. It is important to highlight that ReD and NIR channels are directly fed to segmentation networks; thus, the input layer of each network is modified to be capable of getting a four-channel image. This strategy, increasing the channels of the input layer, has been implemented by a previous work, which proposes a multispectral segmentation network and evaluates RGB-based segmentation models implemented on the MMSegmentation framework [46]. In contrast to [46], we use pretrained parameters as initial guesses. The pretrained parameters are the ones obtained using ADE20 k and Cityscapes datasets.

TABLE I
Quantitative Metrics for Evaluating the Performance of CRF

Confusion Matrix				**Metrics**
		Predicted		
		ToI	non–ToI	$IoU = a/(a+b+c)$
Ground-truth	ToI	a	b	$P = a/(a+c)$
	non–ToI	c	d	$R = a/(a+b)$

Where ToI is the tree of interest, IoU refers to the intersection over the union, P is the precision, and R the recall.

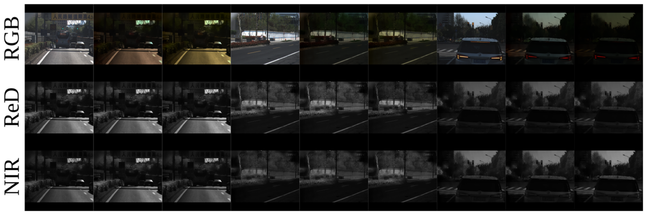


Fig. 6. Multispectral dataset generated from the HSICityV2 hyperspectral dataset. Dataset generation is described on Section II-A.

D. Training and Validation

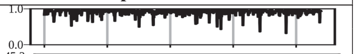
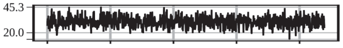
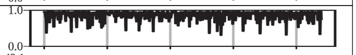
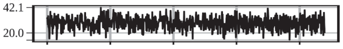
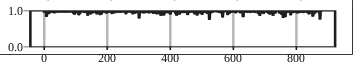
1) ReD and NIR Reconstructions: These models are determined by training, validating, and testing the cGAN using the multispectral dataset described in Section II-A. In particular, one model is generated for ReD reconstruction and the other for NIR reconstruction, as described in Fig. 5. Note that both reconstruction models are trained from scratch using a batch size of one and 3000 epochs. We set up batch size to one according to the optimization and inference configurations followed by [43]. Specifically, Zhu et al. [43] suggested setting the batch size to one and performing batch normalization because of its benefits for image generation tasks. The performance of these models is evaluated using the SSIM, Pearson product-moment correlation coefficients (R), and peak signal to noise ration (PSNR), between natural and generated images. These metrics are computed by the python image processing toolbox (Scikit-image) [47]. The multispectral assessment is performed using testing samples not seen in the training or validation stages.

2) Segmentation Models: Using urban tree datasets, we took advantage of the transfer learning procedure for training, validating, and fine-tuning RGB segmentation models. These models are used as a segmentation baseline to further evaluate the performance of ReD/NIR segmentation networks. In both RGB and multispectral segmentation cases, the networks are trained and validated with Jekyll and Arbocensus urban tree samples; see Section II-B. The fine-tuning description of segmentation models is illustrated in Fig. 5. Note that these models were trained and validated using MMSegmentation framework default parameters. Specifically, DeepLabV3 is trained with 80 thousand and SegFormer with 160 thousand iterations with a batch size of one. After training and validating segmentation models, we evaluate them using IoU, precision (P), and recall (R) quantitative metrics detailed in Table I and testing sets with samples that have not been seen in previous stages.

IV. Results

Fig. 6 shows samples of the generated multispectral dataset using the HSICityV2 hyperspectral dataset. Four RGB and

TABLE II
SSIM, PSNR, AND R OUTCOMES FOR ReD AND NIR RECONSTRUCTED IMAGES USING MULTISPECTRAL DATASET

Channel	Metrics			Sample Performance
ReD	SSIM	avg	0.93	
		std	0.08	
	PSNR	avg	30.8	
		std	5.39	
	R	avg	0.99	
		std	0.01	
NIR	SSIM	avg	0.88	
		std	0.11	
	PSNR	avg	28.5	
		std	5.34	
	R	avg	0.97	
		std	0.03	

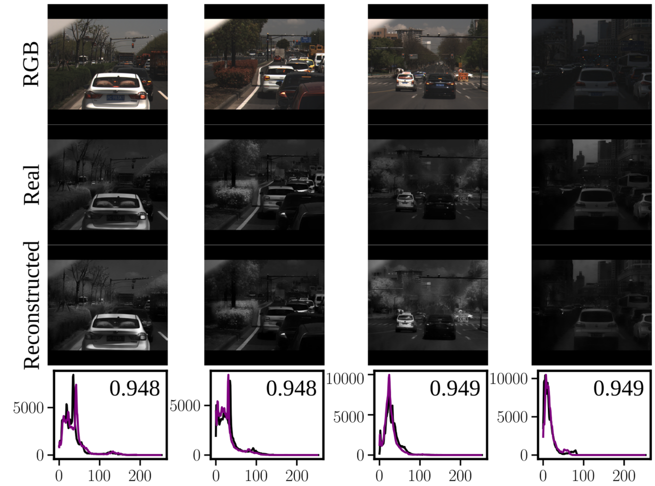


Fig. 7. Best ReD reconstructed images obtained by the cGAN model. The last row shows the current (black solid line) and reconstructed (magenta solid line) image histograms. The SSIM value is shown in the histogram figure.

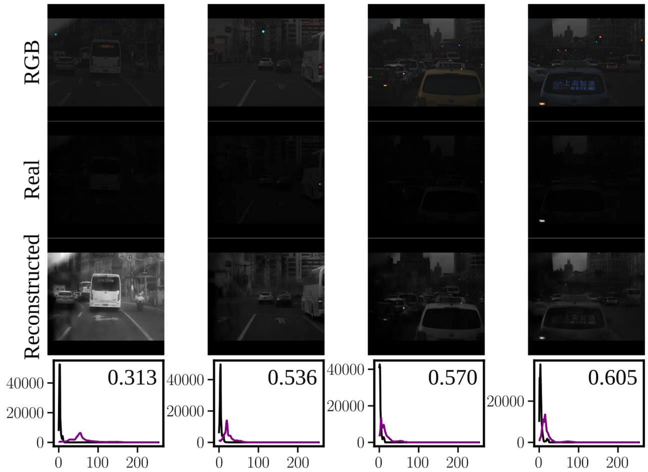


Fig. 8. Worst ReD reconstructed images obtained by the cGAN model. The last row shows the current (black solid line) and reconstructed (magenta solid line) image histograms. The SSIM value is shown in the histogram figure.

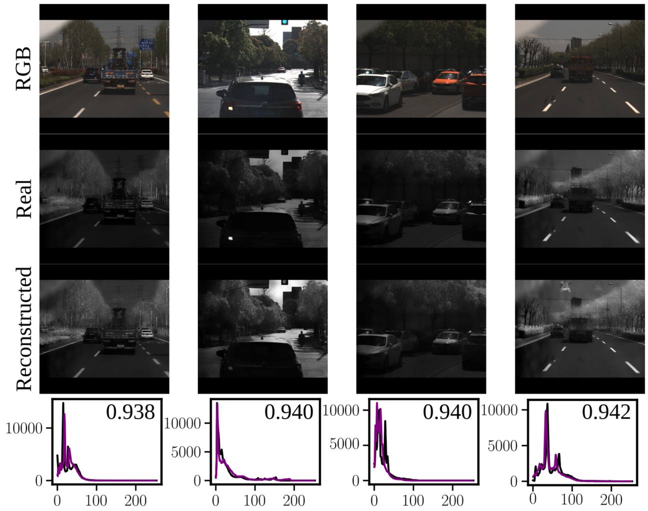


Fig. 9. Best NIR reconstructed images obtained by the cGAN model. The last row shows the current (black solid line) and reconstructed (magenta solid line) image histograms. The SSIM value is shown in the histogram figure.

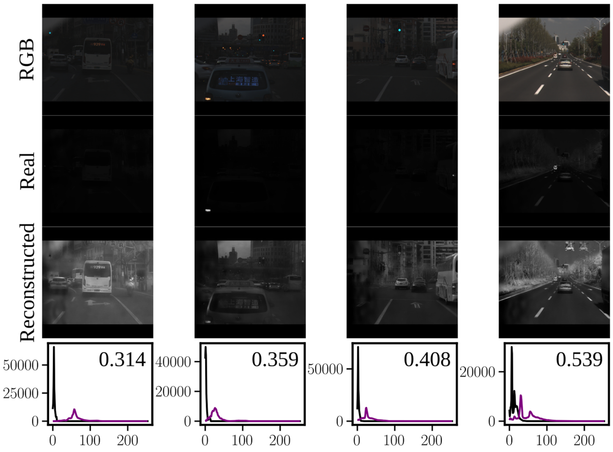


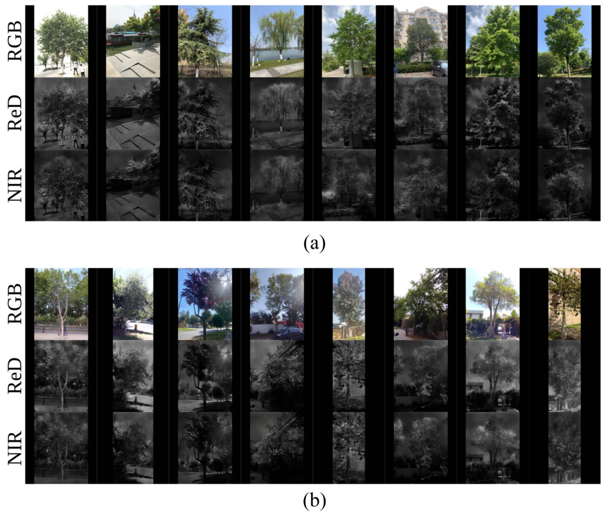
Fig. 10. Worst NIR reconstructed images obtained by the cGAN model. The last row shows the current (black solid line) and reconstructed (magenta solid line) image histograms. The SSIM value is shown in the histogram figure.

five multispectral images with random central wavelengths are generated to simulate different sensors' spectral responses and light conditions for the same environment. The remaining RGB images are available with HSICityV2 dataset. Fig. 6 shows three different urban locations with two RGB and multispectral augmented instances.

A. *ReD and NIR Reconstruction*

The cGAN was implemented, trained and validated using PyTorch framework following guidelines presented by Isola et al. [30]. The training and validation of the reconstruction models take four days for each model (ReD and NIR reconstruction models). However, the inference of multispectral images takes about one second per image. Quantitative outcomes about the performance of multispectral reconstruction models are shown in Table II. Figs. 7–10 show the best and worst multispectral reconstructed images using the HSICityV2 original RGB images.

The cGAN shows an acceptable reconstruction performance of Red and NIR images; see Table II. For instance, the average SSIM is 0.93 and 0.88 for ReD and NIR channels, respectively. Both SSIM values are within the range reported by previous works, 0.799–0.936, [7], [9], [12]. Moreover, the average PSNR



(a)

(b)

Fig. 11. Jekyll and Arbocensus Multispectral dataset generated using the cGAN trained with the multispectral dataset described in Section II-A. (a) Jekyll. (b) Arbocensus.

values for ReD and NIR images are 30.8 and 28.5 dB, correspondingly. These values also support that the trained and validated cGAN models are appropriate for reconstructing multispectral data. Specifically, PSNR values between 30 and 50 dB are considered satisfactory; higher values are better [48]. Note that PSNR below 20 dB might not be considered desirable. In our case, most PSNR values for ReD and NIR images are over 20 dB. Thus, reconstructed images can be viewed as acceptable representations of actual multispectral images in the context of the PSNR metric. Finally, the R values are above 0.9, revealing that the natural and generated multispectral images have a solid linear relationship. Therefore, the quantitative outcomes advocate applying cGAN models for generating multispectral images based on RGB data.

On the other side, qualitative results shown in Figs. 7–10 allow to realize that lousy lighting conditions affect the multispectral reconstruction of ReD and NIR bands. Specifically, Figs. 8 and 10 indicate that in low illumination conditions, the sensor LightGene Hyperspectral might not retrieve adequate ReD and NIR information, the images captured in these regions are very dark. Conversely, the cGAN model generates multispectral images in low illumination, as the illumination conditions might be sufficient. The histogram comparison between the actual and generated multispectral images shows the pixels' value offset of the generated image; see the last row on Figs. 8 and 10. We consider that cGAN models' response in different illumination conditions should be further investigated. However, urban tree images are expected to be captured in adequate light conditions. In addition, the deep analysis of cGAN outcomes is out of the scope of the current works.

Based on the quantitative and qualitative performance of the multispectral reconstruction models, both of them (ReD and NIR cGAN models) are considered suitable for generating multispectral information of Jekyll and Arbocensus datasets. The Jekyll and Arbocensus generated multispectral information is described in Fig. 11.

TABLE III
QUANTITATIVE METRIC VALUES FOR SEGMENTATION MODELS TRAINED WITH JEKYLL AND ARBOCENSUS DATASET

Pre-train	**Network**	**Metrics**		**Datasets**					
				Jekyll			**Arbocensus**		
				RGB	RGB+ReD	RGB+NIR	RGB	RGB+ReD	RGB+NIR
ADE20k	DeepLabV3	IoU	avg	0.90	0.90	0.90	0.79	0.78	0.79
			std	0.08	0.07	0.07	0.14	0.14	0.14
		P	avg	0.95	0.95	0.95	0.86	0.87	0.86
			std	0.06	0.06	0.06	0.14	0.14	0.14
		R	avg	0.95	0.95	0.95	0.90	0.88	0.91
			std	0.06	0.05	0.05	0.06	0.09	0.10
	SegFormer	IoU	avg	0.89	0.89	0.88	0.8	0.82	0.76
			std	0.08	0.09	0.09	0.11	0.09	0.15
		P	avg	0.95	0.94	0.94	0.90	0.92	0.91
			std	0.05	0.06	0.06	0.12	0.07	0.10
		R	avg	0.94	0.95	0.94	0.88	0.89	0.82
			std	0.08	0.07	0.09	0.07	0.09	0.14
Cityscapes	DeepLabV3	IoU	avg	0.90	0.90	0.90	0.80	0.80	0.79
			std	0.08	0.07	0.07	0.12	0.14	0.14
		P	avg	0.95	0.95	0.95	0.86	0.88	0.86
			std	0.06	0.06	0.06	0.13	0.13	0.16
		R	avg	0.94	0.95	0.95	0.91	0.89	0.90
			std	0.06	0.05	0.05	0.06	0.11	0.10
	SegFormer	IoU	avg	0.89	0.88	0.89	0.82	0.83	0.77
			std	0.08	0.09	0.08	0.11	0.10	0.14
		P	avg	0.95	0.94	0.94	0.89	0.91	0.90
			std	0.05	0.06	0.06	0.11	0.10	0.11
		R	avg	0.94	0.94	0.94	0.92	0.90	0.85
			std	0.07	0.08	0.07	0.06	0.07	0.13

Note that ReD and NIR images are predicted using the cGAN model described on Section III.

TABLE IV
P-VALUE FROM T-TEST FOR EVALUATING MEAN DIFFERENCE OF IOU METRIC VALUES

Pre-training set	**Network**	**Datasets**			
		Jekyll		Arbocensus	
		RGB+ReD	RGB+NIR	RGB+ReD	RGB+NIR
ADE20k	DeepLabV3	0.486	0.824	0.799	0.971
	SegFormer	0.495	0.048	0.319	0.221
Cityscapes	DeepLabV3	0.826	0.638	0.950	0.878
	SegFormer	0.077	0.191	0.841	0.138

B. Segmentation Models

The performance of segmentation models trained and validated with Jekyll and Arbocensus datasets are shown in Table III. On average, the DeepLabV3 and SegFormer models' segmentation network training and validation take 3 h and 30 min, respectively. Note that each model's inference takes less than a second. The RGB segmentation outcomes are used to evaluate whether reconstructed multispectral images add relevant information to improve urban tree segmentation. In general, the quantitative outcomes shown in Table III indicate that the segmentation networks (DeepLabV3, SegFormer) yield similar average metrics values (IoU, P, R) even if they are trained with predicted multispectral images. In particular, there is no significant difference between the outcomes retrieved by segmentation networks trained with RGB, RGB+ReD, or RGB+NIR; see p-values from the t-student test in Table IV. It is essential to highlight that for the SegFormer network pretrained with DeepLabV3 and fine-tuned using Jekyll RGB+NIR, the null hypothesis is rejected when compared with the same model fine-tuned solely by RGB images. However, in this case, the RGB+NIR segmentation model performs worse than the segmentation model solely trained with RGB images.

Since segmentation networks perform similar outcomes, we selected the SegFormer network for executing new experiments

TABLE V
QUANTITATIVE METRIC VALUES FOR SEGFORMER NETWORK TRAINED USING JEKYLL AND ARBOCENSUS DATASET

Metrics		**Datasets**					
		Jekyll			Arbocensus		
		RGB	RGB+ReD	RGB+NIR	RGB	RGB+ReD	RGB+NIR
IoU	avg	0.85	0.85	0.84	0.75	0.73	0.72
	std	0.12	0.10	0.13	0.18	0.13	0.11
P	avg	0.93	0.92	0.93	0.86	0.87	0.88
	std	0.08	0.08	0.07	0.15	0.12	0.12
R	avg	0.90	0.92	0.90	0.84	0.83	0.80
	std	0.12	0.09	0.13	0.18	0.13	0.20

The ADE20 k and cityscapes datasets are not used as pretraining sets. Note that ReD and NIR images are predicted using the cGAN model described on Section III.

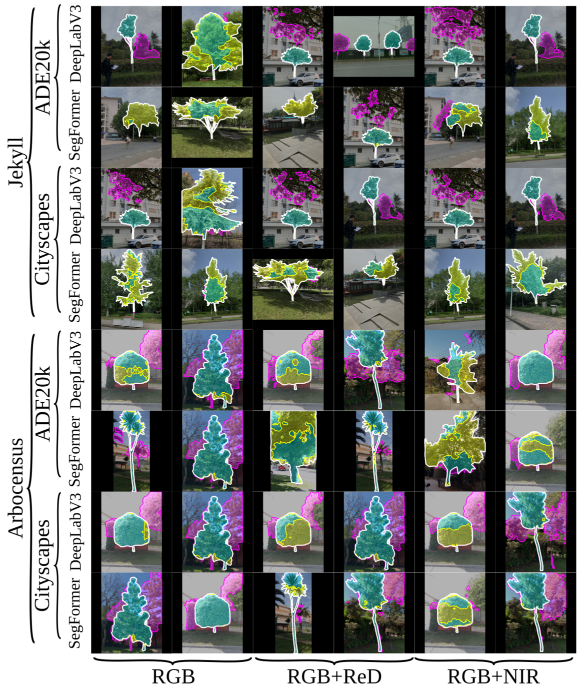


Fig. 12. Worst qualitative outcomes of segmentation networks. The cyan area represents the true positive pixels, magenta regions are false positive pixels, and yellow are false negative sections. Note that white solid line represent the ground truth boundaries.

because it is less expensive computationally. Specifically, we determined new segmentation models without a pretraining dataset (ADE20 k, Cityscapes). Table V shows quantitative metrics for the SegFormer network. Based on Table V, one can be aware that, using a pretraining dataset, one can expect better IoU metrics than not using a pretraining set. However, similar IoU, P, and R metric values are obtained with or without the information of reconstructed ReD and NIR images.

In general, the multispectral images generated by cGAN seem not to add information to segmentation networks to improve the identification of individual urban trees. Fig. 12 shows the worst qualitative outcomes for RGB and multispectral-based segmentation models. The worst outcomes are segmentation outputs with an IoU value within the first IoU quartil. Note that with or without multispectral information segmentation models do not overcome individual tree segmentation challenges (occluded trees, combined crowns, and complex background). Specifically, all networks fail to segment individual urban trees finely. In this sense, further experiments and studies are required to boost pixel-wise identification of urban trees.

V. Discussion

Similar to previous works, quantitative metrics, Table II, shows that cGAN models can be used for generating ReD and NIR images [8], [9], [10], [11], [12]. In particular, the SSIM for ReD and NIR channels is 0.93 and 0.88, respectively. Further, the average values of PSNR and R metrics advocate the potential of the implemented cGAN for retrieving multispectral information. It is essential to highlight that the best reconstruction outcomes could be achieved at adequate environment illumination as shown in Figs. 7 and 10. Note that the pixel distribution of actual and reconstructed images depict similar waveforms, supporting the cGAN reconstruction of multispectral images.

Although the encouraging multispectral reconstruction outcomes, it should be highlighted that inadequate illumination affects the generation of reliable images. Figs. 8 and 10 show urban images at low environmental light. In most cases, ReD and NIR images captured in bad lighting are very dark, which does not allow to discriminate the image's object visually. However, the guess about poor results on inadequate illumination might be sound because the last column in Figs. 7 and 10. The former shows an image with bad lighting, which ReD reconstructed channel yields an SSIM value of 0.949. The latter illustrates an image with good illumination, in which the NIR reconstructed channel outputs an SSIM value of 0.539. The lack of multispectral samples captured by different sensors hinders a deep analysis of these outliers. Nevertheless, further studies are encouraged to investigate multispectral reconstruction at low illumination conditions.

Previous works that address multispectral reconstruction might overcome bad lighting because of the environments mapped and data availability. For instance, the authors in [7], [9], and [12] train, evaluate, and test their reconstruction models using aerial photographs of vegetated areas. Specifically, they take advantage of diverse datasets captured by different sensors to generate reliable multispectral information on vegetated areas. Conversely, we computed RGB, ReD, and NIR images with different central wavelengths from a single sensor; see Section II-A. Despite the computed dataset having more than five thousand RGB, ReD, and NIR images, they might not reflect the actual response of RGB and multispectral commercial imagers. However, the data augmentation process helps to partly resolve the lack of street-view multispectral samples of urban environments.

Since the reconstruction metrics on the testing set, see Table II, support the application of cGAN for generating multispectral images; we computed two new multispectral datasets for urban tree segmentation using Jekyll and Arbocensus RGB datases, see Fig. 11. Despite no multispectral reference information for this new dataset, few guesses regarding the reconstructed multispectral channels can be made. First, the ReD and NIR images' vegetation has brighter pixels on top areas of branches and leaves. Higher reflectance on the vegetation's top would possibly be coherent since most reflectance is from the top of vegetation due to sunlight incidence. However, a contrast exists in reconstructed pixels for the same canopy tree, which might not be consistent with actual multispectral images. For instance, in actual ReD and NIR information from HSICityV2 retrieve similar pixels intensity for tree canopies disregarding the sunlight incidence; see Figs. 7 and 9.

The reconstructed pixel value differences within the same canopy can be explained because of the image's perspectives. Specifically, the Jekyll and Arbocensus datasets record urban trees from bottom to top, focussing on capturing one tree per image. In this context, vegetation and tree canopy could be populated by shadows in areas where there is no direct sunlight incidence. The salient tree's branches generate shadows within the tree canopy. Conversely, the HSICityV2 images capture vegetation from a broader perspective; they are not focused on capturing single trees. The samples that retrieve information about canopy shadows are scarce. Therefore, the cGAN models might need to be further trained and fine-tuned to map RGB to multispectral information when shadows exist within the same tree canopy. Actual ReD and NIR images of urban trees could corroborate the latter. However, the authors have not found publicly available data comprising RGB, ReD, and NIR of urban trees with image perspectives similar to those presented in Jekyll and Arbocensus datasets. This lack of multispectral information from street-view perspectives is encouraged to be addressed in future works.

Regarding pixel-wise tree identification and conversely to our expectations, Table III shows no significant difference in network segmentation outputs. Both models (DeepLabV3 and SegFormer) yield similar performance when fed with RGB, RGB+ReD, and RGB+NIR images, which is supported by the p-values shown in Table IV. The affinity on segmentation results reveals that the reconstructed ReD and NIR channels might not add valuable information to segmentation networks for improving the identification of urban trees. It is important to highlight that the training samples make no difference in DeepLabV3 and SegFormer performance. Both networks achieve similar outcomes, with IoU greater than 0.89 for the Jekyll dataset and greater than 0.75 for the Arbocensus dataset. Although the reconstructed multispectral images do not add relevant information for boosting tree segmentation, it should be noted that increasing the number of samples from about 300 (Arbocensus dataset) to 3000 (Jekyll datasets) improves the segmentation metric in about 0.1, 0.07, and 0.04 for IoU, P, and R, respectively. The boosting of segmentation metrics by increasing dataset samples can be considered as an expected behavior.

Furthermore, Table V shows segmentation metric values for the SegFormer network, which is not pretrained with ADE20 k or Cityscapes datasets. Those outcomes also depict that the reconstructed multispectral information might not be required for the network to improve its performance. In some cases (e.g., Arbocensus dataset), introducing a reconstructed multispectral channel could be ineffective and yield lower IoU values; see Table V.

The similar performance of segmentation networks (RGB and RGB+multispectral) might be a sign that, for segmentation purposes, the reconstructed multispectral channels are not useful. In other words, segmentation networks could probably decode the same information inferred by a cGAN model. The results of a previous work can support the latter. In particular, Deng et al. [8] proposed a deep-learning network for retrieving hyperspectral information based on several input bands (multispectral image). The consistency of reconstructed hyperspectral images is investigated by classifying their pixels using an unsupervised method (iterative self-organizing data analysis techniques algorithm, IsoData). The reported outcomes show no significant difference between the classification results of reconstructed hyperspectral pixels and actual multispectral pixels. Note that hyperspectral images are reconstructed using multispectral information. Moreover, the classification noise in multispectral images (recorded information) is advocated for extracting the multispectral bands from noisy bands in the original hyperspectral images. Note that our work and the proposed in [8] use different deep learning networks for inferring spectral information–furthermore, the study areas and images used in each work. For instance, our work uses street-view images, and Deng et al. [8] employed aerial-view images.

Since our work and previous research suggest that segmentation and classification outcomes might not be significantly improved by reconstructed reflectance, we advise performing new experiments regarding the application of reconstructed hyperspectral/multispectral images. For example, new research could focus on recording hyperspectral/multispectral images with a single sensor and investigate which electromagnetic spectrum bands retrieve reflectance information that leads to improvements in the semantic segmentation of trees. This could address one of our limitations; only reconstructed reflectance in the ReD and NIR bands was tested in the current work. For the data collection procedure, we suggest capturing images at different light conditions and perspectives of trees to get a heterogeneous dataset and tree representation. Generating a hyperspectral/multispectral dataset of urban trees from street-view perspectives will benefit the scientific community since such datasets are currently scarce. Moreover, future works are advised to investigate the behavior of generative models for inferring tree reflectance on different light conditions and irregular shadows. In the current work, we do not explore the effects of illumination or irregular shadows on the reconstruction performance of the cGAN network. Finally, it should be noted that the current work uses segmentation networks initially developed for using RGB images as input; thus, future works are encouraged to determine dedicated hyperspectral/multispectral semantic segmentation networks such that process reflectance information effectively for outperforming current semantic segmentation metrics.

It is essential to highlight that for urban tree segmentation, further studies are suggested to employ either DeepLabV3 or SegFormer pretrained with ADE20 k or CityScapes. Specifically, we recommend SegFormer because it is computationally less expensive. For instance, the pretraining of DeepLabV3 and SegFormer networks with ADE20 k requires 8.9 and 2.1 Gb of memory from graphical processing unit [44]. Further, it should mentioned that IoU reported in the current work using the Jekyll dataset is slightly similar to the ones reported by Yang et al. [37]. In particular, the best IoU value, obtained by Yang et al. [37], is 0.88 while our best IoU result is 0.90.

VI. Conclusion

The outcomes reported in this work support the generation of images at ReD and NIR regions from RGB street-view images by deep learning networks. In particular, we obtained average SSIM values of 0.93 and 0.88 for ReD and NIR images using cGAN models. These values are within the range of the ones presented in previous works that address hyperspectral or multispectral image reconstruction. A HSICityV2 was used to train, validate, and test cGAN models. Once the cGAN models are tested, we exploit them to generate ReD and NIR information from RGB-based urban tree datasets (Jekyll and Arbocensus). After visual inspection of generated ReD and NIR images, we became aware that the top face of the canopy and vegetation yield higher pixel values than lower or shadowed areas. This behavior might be expected since the top sections of tree canopy and vegetation reflect most solar light. Another factor influencing the pixel intensity values in a single canopy might be the perspective for taking tree pictures in Jekyll and Arbocensus datasets. It differs from the perspective used on the hyperspectral dataset. In this context, we suggest developing a multispectral urban tree dataset to analyze further and investigate multispectral generation's advantages in urban environments. Concerning the segmentation networks, we found that DeepLabV3 and SegFormer networks pre-trained with ADE20 k and Cityscapes datasets are appropriate for segmenting individual urban trees using RGB, RGB+ReD, and RGB+NIR images. Although the segmentation networks could retrieve acceptable IoU values (> 0.89 for the Jekyll dataset and > 0.75 for the Arbocensus dataset), incorporating multispectral information could have made more of a difference. Specifically, quantitative outcomes suggest that the reconstructed ReD and NIR images do not add practical knowledge that DeepLabV3 or SegFormer could manipulate to boost their performance. The results of segmentation networks are said to be similar based on p-values from a T-test. Besides the quantitative outcomes presented in this work, previous work has reported similar pixel classification metrics for actual and reconstructed spectral information. Therefore, reconstructed multispectral information might not be advised as an extra source of information for improving urban tree segmentation performance.

References

[1] A. F. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.

[2] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[3] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.

[4] E. B. Knipling, "Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation," *Remote Sens. Environ.*, vol. 1, no. 3, pp. 155–159, 1970.

[5] S. G. Yel and E. Tunc Gormus, "Exploiting hyperspectral and multispectral images in the detection of tree species: A review," *Front. Remote Sens.*, vol. 4, 2023, Art. no. 1136289.

[6] T. Arevalo-Ramirez et al., "Assessment of multispectral vegetation features for digital terrain modeling in forested regions," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2021, Art. no. 4405509.

[7] C. Davidson, V. Jaganathan, A. N. Sivakumar, J. M. P. Czarnecki, and G. Chowdhary, "NDVI/NDRE prediction from standard RGB aerial imagery using deep learning," *Comput. Electron. Agriculture*, vol. 203, 2022, Art. no. 107396.

[8] L. Deng et al., "M2H-Net: A reconstruction method for hyperspectral remotely sensed imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 323–348, 2021.

[9] X. Yuan, J. Tian, and P. Reinartz, "Generating artificial near infrared spectral band from RGB image using conditional generative adversarial network," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 279–285, 2020.

[10] L. An, J. Zhao, and H. Di, "Generating infrared image from visible image using generative adversarial networks," in *Proc. IEEE Int. Conf. Unmanned Syst.*, 2019, pp. 157–161.

[11] S. Illarionova, D. Shadrin, A. Trekin, V. Ignatiev, and I. Oseledets, "Generation of the NIR spectral band for satellite images with convolutional neural networks," *Sensors*, vol. 21, no. 16, 2021, Art. no. 5646.

[12] M. Aslahishahri et al., "From RGB to NIR: Predicting of near infrared reflectance from visible spectrum aerial images of crops," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1312–1322.

[13] B. Arad et al., "NTIRE 2018 challenge on spectral reconstruction from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 1042–1042.

[14] B. Arad, R. Timofte, O. Ben-Shahar, Y.-T. Lin, and G. D. Finlayson, "NTIRE 2020 challenge on spectral reconstruction from an RGB image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 446–447.

[15] T. Alipour-Fard and H. Arefi, "Structure aware generative adversarial networks for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5424–5438, 2020.

[16] Y. Zhan, D. Hu, Y. Wang, and X. Yu, "Semisupervised hyperspectral image classification based on generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 212–216, Feb. 2018.

[17] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018.

[18] D. Wang, L. Gao, Y. Qu, X. Sun, and W. Liao, "Frequency-to-spectrum mapping GAN for semisupervised hyperspectral anomaly detection," *CAAI Trans. Intell. Technol.*, vol. 8, no. 4, pp. 1258–1273, 2023.

[19] W. Xie, J. Zhang, J. Lei, Y. Li, and X. Jia, "Self-spectral learning with GAN based spectral–spatial target detection for hyperspectral image," *Neural Netw.*, vol. 142, pp. 375–387, 2021.

[20] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. Plaza, "BockNet: Blind-block reconstruction network with a guard window for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Nov. 2023, Art. no. 5531916.

[21] D. Wang, L. Zhuang, L. Gao, X. Sun, X. Zhao, and A. Plaza, "Sliding dual-window-inspired reconstruction network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Jan. 2024, Art. no. 5504115.

[22] T. Arevalo-Ramirez et al., "Challenges for computer vision as a tool for screening urban trees through street-view images," *Urban Forestry Urban Greening*, vol. 95, 2024, Art. no. 128316.

[23] S. Beery et al., "The auto arborist dataset: A large-scale benchmark for multiview urban forest monitoring under domain shift," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 21294–21307.

[24] S. Branson, J. D. Wegner, D. Hall, N. Lang, K. Schindler, and P. Perona, "From Google maps to a fine-grained catalog of street trees," *ISPRS J. Photogrammetry Remote Sens.*, vol. 135, pp. 13–30, 2018.

[25] K. Choi et al., "An automatic approach for tree species detection and profile estimation of urban street trees using deep learning and Google street view images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 190, pp. 165–180, 2022.

[26] D. S. Jodas et al., "A deep learning-based approach for tree trunk segmentation," in *Proc. 34th SIBGRAPI Conf. Graph., Patterns Images*, 2021, pp. 370–377.

[27] D. S. Jodas, T. Yojo, S. Brazolin, G. D. N. Velasco, and J. P. Papa, "Detection of trees on street-view images using a convolutional neural network," *Int. J. Neural Syst.*, vol. 32, no. 01, 2022, Art. no. 2150042.

[28] S. Lumnitz, T. Devisscher, J. R. Mayaud, V. Radic, N. C. Coops, and V. C. Griess, "Mapping trees along urban street networks with deep learning and street-level imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 175, pp. 144–157, 2021.

[29] Y. Wang et al., "Detecting occluded and dense trees in urban terrestrial views with a high-quality tree detection dataset," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4707312.

[30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134.

[31] Y. Huang, T. Ren, Q. Shen, Y. Fu, and S. You, "HSICityV2: Urban scene understanding via hyperspectral images," Jul. 2021. Accessed: Jul. 02, 2024. [Online]. Available: https://pbdl-ws.github.io/pbdl2021/challenge/index.html

[32] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[33] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," 2021, *arXiv:2105.15203*.

[34] S. You et al., "HyperSpectral city v1. 0 dataset and benchmark," 2019, *arXiv:1907.10270*.

[35] K. Smith, M. Steven, and J. Colls, "Use of hyperspectral derivative ratios in the red-edge region to identify plant stress responses to gas leaks," *Remote Sens. Environ.*, vol. 92, no. 2, pp. 207–217, 2004.

[36] G. J. Verhoeven, "Near-infrared aerial crop mark archaeology: From its historical use to current digital implementations," *J. Archaeological Method Theory*, vol. 19, pp. 132–160, 2012.

[37] T. Yang, S. Zhou, Z. Huang, A. Xu, J. Ye, and J. Yin, "Urban street tree dataset for image classification and instance segmentation," *Comput. Electron. Agriculture*, vol. 209, 2023, Art. no. 107852.

[38] K. Wada, "Labelme: Image polygonal annotation with Python," Accessed on: May 03, 2024. [Online]. Available: https://github.com/wkentaro/labelme

[39] D. Drozdov, M. Kolomeichenko, and Y. Borisov, "Supervisely: Annotation tool," Accessed on: May 03, 2024. [Online]. Available: https://supervisely.com/

[40] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ADE20 k dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 633–641.

[41] M. Cordts et al., "The CityScapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3213–3223.

[42] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[43] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2223–2232.

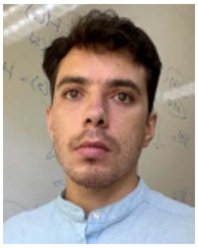
[44] M. Contributors, "MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark," 2020. [Online]. Available: https://github.com/open-mmlab/mmsegmentation

[45] M. E. Andrada, D. Russell, T. Arevalo-Ramirez, W. Kuang, G. Kantor, and F. Yandun, "Mapping of potential fuel regions using uncrewed aerial vehicles for wildfire prevention," *Forests*, vol. 14, no. 8, 2023, Art. no. 1601.

[46] Y. Gui, W. Li, X.-G. Xia, R. Tao, and A. Yue, "Infrared attention network for woodland segmentation using multispectral satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2022, Art. no. 5627214.

[47] S. van der Walt et al., "Scikit-image: Image processing in Python," *PeerJ*, vol. 2, 2014, Art. no. e453.

[48] D. R. Bull and F. Zhang, "Chapter 4 - digital picture formats and representations," in *Intelligent Image and Video Compression*, Eds., 2nd ed. Oxford, U.K.: Academic, 2021, pp. 107–142.



Tito Arevalo-Ramirez received the Ph.D. degree in electronics engineering from Universidad Técnica Federico Santa María, Valparaiso, Chile, in 2022.

Since 2023, he has been an Assistant Professor with the Department of Electric Engineering and Department of Mechanical and Metallurgic Engineering, Pontificia Universidad Catolica de Chile, Santiago, Chile. His research interests span remote sensing, simultaneous localization and mapping, agricultural robotics, environmental monitoring, autonomous navigation, and precision agriculture.

Anali Alfaro received the B.S. degree in informatic engineering from the Universidad Nacional de Trujillo, Trujillo, Perú, in 2005, and the master's degree in computer science from Pontificia Universidad Católica, Santiago, Chile, in 2017.

Currently, she is a Professor of computer science with the Universidad de Santiago de Chile, Estación Central, Chile. Her research interets include computer vision, video analytics, and deep learning applications.

Jose M. Saavedra received the Ph.D. degree in computer science from the Universidad de Chile, Santiago, Chile, in 2013.

Currently, he is a Professor with the Faculty of Engineering and Applied Science and Director of the computer vision lab CVLab there. He has also led diverse R&D projects on computer vision with applications in eCommerce and medical imaging. He has broad experience in computer science and deep learning. His research interets include self-supervision, multimodality and zero-shot detection.

Matías Recabarren received the Computer Engineering degree and Ph.D. degree in computer science engineering from Pontificia Universidad Católica de Chile, Santiago, Chile.

He is currently an Associate Professor with the Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Santiago, where he is involved in undergraduate and graduate teaching and research activities. His research interests include human–computer interaction, technology in education, engineering education, and STEM education in pre-K-12 levels including teacher education.

Mauricio Ponce-Donoso received the B.S. degree in forestry engineering from the Universidad de Talca, Talca, Chile, in 1991, and the Ph.D. degree in forestry engineering from the Politécnica Universidad de Madrid, Madrid, Spain, in 1998.

He was an Academic with the Universidad de Talca, where he was responsible for the course on Urban Forestry. He has been Visiting Professor with Universities in Argentina, Colombia, Ecuador, and Uruguay. He is currently CEO of Arbologia SpA, where he develops consultancies in arboriculture, highlighting the evaluation of visual and instrumental risk of urban trees.

José Delpiano received the B.S. degree in electrical engineering from the Pontificia Universidad Católica, Santiago, Chile, in 2003, and the Ph.D. degree in electrical engineering from the University of Chile, Santiago, in 2013.

He is currently an Associate Professor with the School of Engineering and Applied Sciences, Universidad de los Andes, Santiago, and a Researcher with the Advanced Center for Electrical and Electronic Engineering, Chile. His research interests include computer vision and bioimage analysis.