From RGB to NIR: Predicting of near infrared reflectance from visible spectrum aerial images of crops

Masoomeh Aslahishahri
Computer Science, Univ. Saskatchewan
bia006@usask.ca

Kevin G. Stanley
Computer Science, Univ. Saskatchewan
kstanley@cs.usask.ca

Hema Duddu
Computer Science, Univ. Saskatchewan
hema.duddu@usask.ca

Steve Shirtliffe
Plant Sciences, Univ. Saskatchewan
steve.shirtliffe@usask.ca

Sally Vail
Agriculture and Agri-Food Canada
sally.vail@canada.ca

Kirstin Bett
Plant Sciences, Univ. Saskatchewan
k.bett@usask.ca

Curtis Pozniak
Plant Sciences, Univ. Saskatchewan
curtis.pozniak@usask.ca

Ian Stavness
Computer Science, Univ. Saskatchewan
ian.stavness@usask.ca

Abstract

Near infrared spectroscopy (NIR) provides rich information in agricultural operations and experiments to determine crop parameters which are not visible to the human eye. Collecting the NIR spectral band requires a multispectral camera which is typically more expensive and has lower resolution than a comparable RGB camera. We investigate image-to-image translation as a means to generate an NIR spectral band from an RGB image alone in aerial crop imagery. Aerial images were captured via a multispectral sensor mounted on an unmanned aerial vehicle (UAV) flown over canola, lentil, dry bean, and wheat breeding trials. A software workflow was created to preprocess raw aerial images creating a dataset suitable for training and evaluating deep learning based band inferencing algorithms. Two different experiments including in-domain and out-of-domain experiments over different crop types in our dataset were conducted to evaluate efficacy in an agricultural context.

1. Introduction

A growing world population and increasing climate instability are projected to strain global food production [44]. Crop breeding programs are key tool in increasing crop yield and stability under varying climate [12]. Crop breeding remains a labor intensive process, with experts manually surveying tens of thousands of breeding plots per trial. Several techniques have been developed to analyze plant phenotypes, but these traditional methods are slow, destructive and laborious [13]. Automated phenotyping could alleviate some of the labor burden of crop breeding and increase crop breeders' productivity [26]. In particular, image-based plant phenotyping has the potential to increase the speed and reliability of phenotyping by employing unmanned aerial vehicles (UAVs) to image large breeding fields quickly with relatively low overhead.

Aerial crop imaging has focused on multispectral imaging, due to the correlation between plant respiratory processes and emissions of particular wavelengths, with particular stress placed on the red edge (680 nm to 730 nm), near infrared (NIR) (800 nm to 2,500 nm), and infrared (700 nm to 1 mm) [53] portions of the spectra given their correlation with plant metabolic processes [24, 65]. While cameras exist to capture these spectral bands (for example the Micasense Rededge) they tend to be more expensive, lower resolution and more difficult to attach to UAVs than their RGB counterparts. A method to provide equivalent information on plant health and metabolism from an RGB camera would make UAV imaging of fields more lower cost, and allow access to the rapid development and economies of scale of consumer RGB cameras.

Cross modality of images can be achieved using machine learning, where a model is trained to map features in one band to another as long as sufficient correlation be-

tween the bands exist [65]. For example extensive work has been performed in inferring spectral bands from satellite data [19, 18], fixed cameras [51], and drone images of everyday objects [21, 15].

The existence of distribution shift is common in real-world image-to-image translation problem. In outdoor agricultural imagery, this issue is more obvious when training distribution is in a crop type dataset, or a modality differing from test distribution [27, 9]. A domain adversarial learning approach was proposed to detect plant organs in field images where domain shift exists between source and target datasets [3].

In this paper, we present a new dataset of three different crops imaged with a multispectral camera. Training existing machine learning models using this dataset, we were able to recreate the NIR channel from the RGB image with a high degree of fidelity, and demonstrate its utility in crop area measurement. Tests of this model, both within and across crop types indicate that RGB image data of crops can be used to approximate the NIR band in a number of circumstances. These findings indicate that, under many circumstances, flying less expensive RGB cameras and inferring the NIR band post-hoc is a viable approach for field phenomics in breeding programs.

The contributions of this study include:

- A new dataset containing color-calibrated aerial images of different agricultural crops, employing a multispectral sensor mounted on a UAV. The dataset consists of aerial images of canola, lentil, dry bean, and wheat crops ranging from early to late growth stages.
- An evaluation of image-to-image translation for generating NIR reflectance images from visible spectrum aerial images for different crop species.
- An out-of-domain experiment to evaluate the generalizability of the image-to-image translation models for generating NIR images when trained on one crop species and tested on a different crop species.
- Comparison of vegetation segmentation using visible spectrum excess-green index vs. vegetation segmentation using NDVI based on generated NIR images.

2. Background & Related work

Image-to-image translation (I2I) aims to transfer images from a source domain (A) to a target domain (B), preserving the source content representations. This is important in computer vision tasks when different style, restoration, modality or segmentation of the source domain are required [16]. In remote sensing, providing a large cover of different spectral bands, which has obvious benefits to employ techniques for image translation from one modality to another modality, especially in precision agriculture is required [35, 51]. Examples include the image translation from aerial images to maps [19], from satellite images to optical images [56], or from certain spectra to several different spectra [18]. The core idea of I2I was originally proposed as an image analogy task [16], where the transformation between two images, A and $A^{'}$, are learned. Once learned, the transformation can be used to generate the transformation between a different set of images, B and $B^{'}$. With the advent of deep generative models [60, 41], most recent I2I approaches use generative models to learn the mapping between different image domains. This approach creates models of the distribution of the target domain resulting in images drawn from the target domain distribution. In this paper, we focus on Generative Adversarial Network (GAN) based methods [14], rather than other I2I techniques [47, 48].

Current I2I tasks are divided into two-domain or multi-domain categories. In this study, we focus on the two-domain category. Two-domain I2I can be used in semantic segmentation [67], editor applications [30], and image super resolution [63, 64]. Two-domain I2I methods can be further classified into supervised I2I [22], unsupervised I2I [66], semi-supervised I2I [31], and few-shot I2I [32].

Supervised I2I translates source images to target images from a dataset of paired, corresponding images. In [22], a conditional GAN (cGAN) was employed to solve a wide range of supervised I2I problems. A version of cGAN was proposed to overcome the blurred output of high resolution images [57]. In [54], a GAN model was proposed to solve the cross-view translation problem in which source and target images have little or no overlap. Disentangled representation is another approach to produce diverse translated images, where each feature is encoded as a separate dimension [10, 23]. A supervised deep learning approach was employed to segment apple trees in occluded images [8].

Unsupervised I2I employs two unpaired training sets to translate images from one domain to another. In [66], a GAN model constrained using cycle-consistency was proposed to translate two-cross domain representations. In [52], NIR spectral band was obtained from its grayscale counterpart in which minimal spatial misalignment exists between input and target images. A shared latent space assumption was introduced to map a pair of corresponding images from different domains to the same latent code in the shared latent space [33]. In [1], a Siamese network was used to output a transformation vector over the two images from each domain and minimize the distance between the two image vectors. Semi-supervised I2I is used in specialized applications such as artistic reconstruction. In this approach, fewer labeled data are required to guide the algorithm. A semi-supervised I2I model was introduced employing transformation consistency regularization to learn a mapping between two-domain images [39]. In [4], a semi-

supervised deep learning technique was used to generate synthetic agricultural images including bell pepper images for further image-based agricultural analysis. Few-shot I2I is still in its infancy. In [32], a few-shot I2I model was introduced to incorporate several domain translation tasks where each task consists of limited examples.

Image processing techniques cannot translate RGB images to other spectrum when there is no overlap between wavelengths. In image-based plant phenotyping, a multispectral camera, which is usually expensive with low resolution, or applying a filter over an RGB camera capturing low quality images with a different wavelength, are employed. The lack of a paired dataset from different modalities, which has a sufficient coverage of plant species and growth stages, limits the ability to apply new I2I models to plant phenotyping. There are a few datasets that include different modalities captured with satellite imagery, thermal cameras, or stationery cameras [7, 49, 50], but so far high resolution aerial crop images have not been investigated for I2I across spectra.

Date	Canola	Lentil	Dry bean-1	Dry bean-2	Wheat
28 Jun 2018	293	-	-	-	-
27 Jul 2018	-	113	-	-	115
30 Jul 2018	938	-	-	-	-
29 Aug 2018	-	114	-	-	114
31 Aug 2018	370	-	-	-	-
26 Jul 2019	646	-	-	-	-
27 Jul 2019	-	-	-	-	942
06 Aug 2019	832	-	-	-	-
13 Aug 2019	-	-	-	-	972
30 Aug 2019	-	-	-	-	1009
14 Jun 2020	-	-	371	278	-
25 Jun 2020	-	-	-	416	-
04 Jul 2020	-	-	433	-	-
30 Jul 2020	-	-	412	-	-
06 Aug 2020	-	-	556	327	-
20 Aug 2020	-	-	384	362	-
Total	3079	227	2156	1383	3152

Table 1: The number of images in each dataset captured for different acquisition dates.

3. Materials & Methods

Our code and dataset is available at `https://github.com/p2irc/rgb2nir`.

3.1. Dataset Collection

We collected aerial datasets consisting of raw multispectral images, captured by the Micasense Rededge camera mounted on a UAV, as depicted in Figure 1. Similar to many multispectral cameras, the Micasense camera includes five separate sensors that are not synchronized, which causes spatial misalignment among images from different bands.

Figure 1: The Micasense Rededge sensor with five separated spectral bands [37] mounted on a UAV.

Therefore band-to-band spatial registration was required. The resulting aligned dataset enables training a model in an end-to-end fashion to translate aerial RGB images to NIR.

We obtained aerial images from breeding trial fields for four crops: canola, wheat, lentils and dry beans. Two different dry bean trials were imaged (dry bean-1 and dry bean-2). The lentil and dry beans trials were aggregated into a "pulse" dataset, representing edible plant seeds in the legume family. We combined all crops to create the "all-crop" dataset.

The images were acquired over multiple growing seasons, between 2018 and 2020, and include images for each crop from early, mid to late growth stages, as reported in Table 1. Each dataset covers a diverse range of genotypes where substantial variations exist in appearance in the dataset.

The Micasense camera covered blue, green, red, NIR, and red edge wavelengths centering at 477 nm, 560 nm, 668 nm, 840 nm, and 717 nm, respectively [53]. The camera was mounted on an Ronin-MX gimbal on a DJI Matrice 600 UAV. The flight altitude was not consistent over all trial fields in different years. The differences in flight altitudes lead to various ground sampling distance (GSD) among datasets, meaning a larger GSD represents a smaller spatial resolution. The flight altitudes varied from 20 to $30m$ above the ground for different crops. As the field of view for the Micasense Rededge is $47.2°$, the GSD varies between approximately 12 mm/pixel and 18 mm/pixel across images in the dataset.

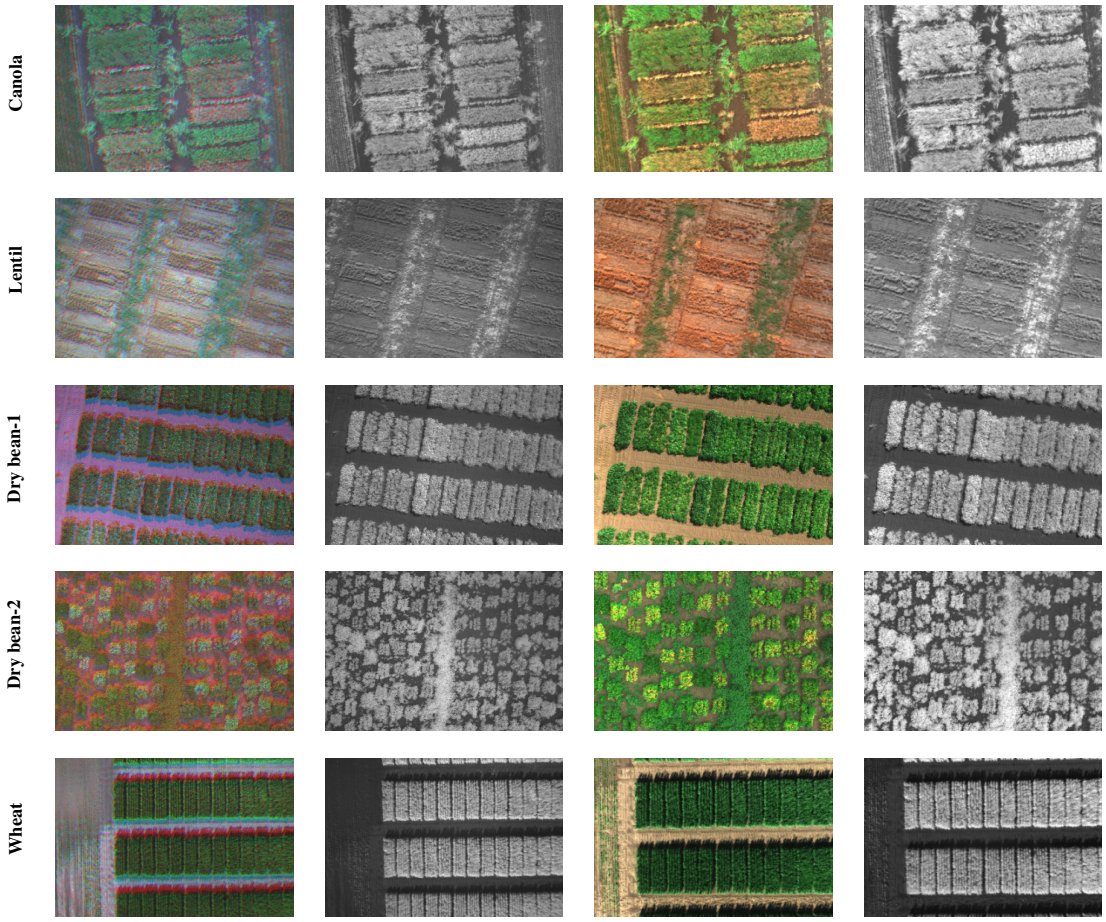

Figure 2: Example images from our canola, lentil, dry bean-1, dry bean-2, and wheat datasets showing: (a) unaligned raw RGB image, (b) the corresponding NIR band, (c) radiometrically calibrated and aligned RGB composite, (d) the corresponding radiometrically calibrated and aligned NIR band.

3.2. Image Pre-processing

In passive imaging, reflectance measurement is influenced by natural disturbances such as scattering light in the atmosphere, the spectral and directional reflectance of an object, and the topography of an object [17]. We employed the proscribed process for the Micasense Rededge to convert raw multispectral pixels to radiance and then to reflectance as unitless images [2, 36]. To obtain a unitless band, an image of a calibrated reflectance panel (CRP) is captured before or after a flight to measure the amount of light in the surrounding environment, which is used to cancel out the illumination and energy from multispectral images. To determine the transfer function from raw pixels to reflectance for the *i-th* band of the multispectral image:

$$F_i = \frac{\rho_i}{avg(L_i)} \tag{1}$$

where F_i is the calibrated reflectance factor for band i, ρ_i denotes the average reflectance of the CRP for the *i-th* band, and $avg(L_i)$ represents the average value of the radiance for pixels inside the panel for the *i-th* band.

Because the multispectral bands are spatially misaligned and supervised learning requires pixel-to-pixel alignment between the input and target images, we register the RGB and NIR bands pixel-wise, where every pixel of the RGB image corresponds to a pixel of the NIR counterpart. We employ the scale invariant feature transform (SIFT) algorithm to perform image registration [34]. We followed a traditional approach for multispectral image registration, where a band is selected as a reference image and the other bands are aligned accordingly [45]. We chose the NIR image as the reference.

Our dataset contains 9997 paired images with the size of $(800, 1100)$ pixels including canola, pulse and wheat crops where 80% and 20% of each dataset build training set and

test set, respectively. Table 1 represents the number of crop images in each captured date and the total number of images per dataset. The pulse dataset containing lentil, dry bean-1 and dry bean-2 crops has 3766 images in total.

3.3. Experiments

Having created a paired image dataset with corresponding images in the source (RGB) and target (NIR) domains, we chose to employ supervised I2I translation. Several different models have been introduced for supervised I2I translation [22, 55, 42, 57]. These models differ in their network structure and vary substantially in the required computational cost for training and inference. Because our dataset consists of self-similar and repeating plant structures, we chose to use the cGAN architecture [22], which is a relatively small network. We expect that this simpler I2I network will have sufficient capacity for our task.

The cGAN [22] is a conditional GAN model that learns a mapping from an observed image x and random noise vector z to a target image y. In our case, the observed image is the aligned RGB image and the target is the corresponding NIR band. For the generator, we employed UNet, a standard architecture for image segmentation [46], and we chose PatchGAN for the discriminator [22, 38]. In the original cGAN [22], L_1 distance along with the GAN objective was used to generate more realistic images, but we replaced the L_1 objective with Charbonnier penalty function which is a smoothed form of L_1 to improve loss minimization during model training [29].

We train the cGAN [22] on our dataset from scratch without employing pretrained model. We crop one patch per image during the training and inference. We train the model with stochastic cropped images to enhance the model's performance. The input data is normalized to $[0, 1]$. Data augmentation is performed by randomly rotating and horizontally flipping the input. The Adam solver [25] with default parameters ($\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$) is employed to optimize the network parameters. The learning rate is set at 10^{-4} and is decayed linearly. The model is trained for 1000 epochs.

To test the generalizability of the model on aerial crop images, we conduct two in-domain and out-of-domain experiments to evaluate whether training a model on each individual crop species is required, or if training a model on a single or combined dataset is sufficient.

Experiment1: In-domain. For the in-domain experiment, we train the model with RGB and NIR images from a single crop and test the model with images from the same crop. We also include a all-crop model trained on images from all crops combined. A random uniform 256×256 patch of the RGB image is used as input for the model and it is translated to NIR image with the same size. At inference, we compare the performance of the trained model with a test set from the same dataset, but with larger patches of size 512×512 as input. This experiment assesses the performance of the algorithm within the same data distribution to provide an approximate upper bound of the performance, because it solves the simpler in-domain problem.

Experiment2: Out-of-domain. For the out-of-domain experiment, we use the models trained models from the in-domain experiment on single crops and test them with images from different crops. At inference, we crop a larger patch of size 512×512 as input from a different type of crop. This experiment is meant to evaluate how well the models trained on a single or combined aerial crop dataset generalize to other unseen aerial crop datasets. We include tests with the all-crop trained model as well, as an upper bound for out-of-domain performance.

3.3.1 Evaluation Measures

To evaluate the performance of the I2I model, we employ four standard measures, including: the Structural Similarity Index (SSIM) [58], the Peak Signal-to-Noise Ratio (PSNR) [20] used in many computer vision tasks, the Dice Similarity Coefficient (DSC) [11] to assess vegetation segmentation accuracy, and Spectral Angle Mapping (SAM) [28] to quantify spectral similarity between the generated and reference spectral bands for each pixel. Higher SSIM, PSNR and DSC values represent a better performance of the model, while lower SAM shows the generated spectrum is closer to its target.

Vegetation segmentation. To evaluate the image quality in terms of vegetation segmentation, which is important for characterizing a crop's vigor and canopy coverage[61], we computed Normalized Difference Vegetation Index (NDVI) [6] for the real NIR images and the generated NIR images. We also computed excess green index (ExG) [59] exclusively from RGB images to compare vegetation segmentation obtained from NIR-based index with an RGB-based index.

The NDVI quantifies the presence of living vegetation using reflected visible light and NIR bands. NDVI is an indicator of the density and plant health of each pixel [43]. Generated NDVI means the NDVI computed using the real red band and the generated NIR image, whereas real NDVI refers to using real red and NIR images. Excess green index provides a grey scale image, outlining a plant region of interest and is computed as follows:

$$ExG = 2g - r - b \tag{2}$$

where g, r, b represent green, red and blue bands respectively. We obtained vegetation segmentations by applying Otsu's thresholding [40] to both the generated and real NDVI images along with the ExG images. We compared the real and generated NDVI segmentations to the ExG based

Train \ Test	Experiments 1 & 2															
	Canola				Pulse				Wheat				All-crop			
	PSNR	SSIM	SAM	DSC	PSNR	SSIM	SAM	DSC	PSNR	SSIM	SAM	DSC	PSNR	SSIM	SAM	DSC
Canola	30.91	0.92	0.03	0.96	29.17	0.85	0.05	0.92	29.35	0.89	0.06	0.95	29.66	0.89	0.05	0.94
Pulse	29.12	0.88	0.05	0.92	31.71	0.93	0.03	0.96	29.20	0.89	0.05	0.9	29.95	0.9	0.05	0.93
Wheat	29.15	0.87	0.05	0.94	28.85	0.89	0.05	0.94	31.11	0.93	0.03	0.95	29.7	0.9	0.05	0.94
All-crop	30.85	0.92	0.03	0.96	31.73	0.93	0.03	0.96	31.06	0.92	0.03	0.94	31.21	0.93	0.03	0.95

Table 2: Average PSNR, SSIM, SAM and DSC values for the in-domain and out-of-domain experiments. DSC was applied over the generated and real NDVI segmentations. The shaded experimental evaluations correspond to the in-domain tests. The all-crop model tested on single crop datasets represents an upper bound because its distribution covers the other crop's distribution.

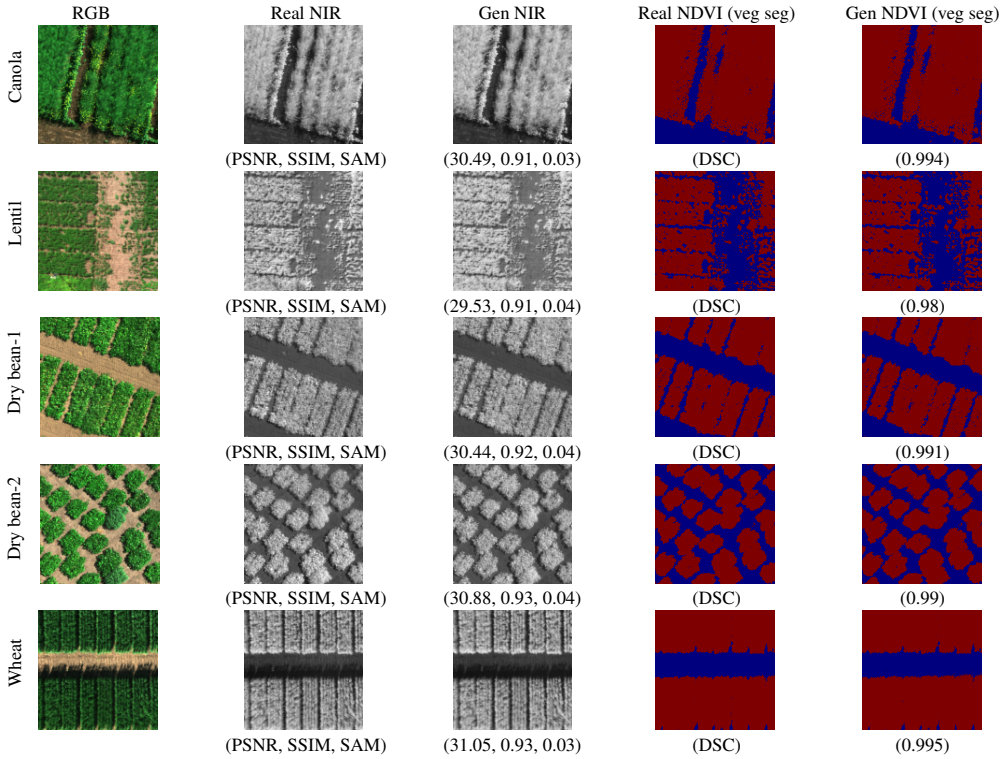

Figure 3: Example test images from the in-domain experiment for the canola, pulse, and wheat datasets in the mid growth stage. The real and generated NIR images are shown in grayscale, as well as the vegetation segmentation produced with NDVI for each ('Veg Seg'). The type of crop is shown beside each row. PSNR, SSIM, SAM, and DSC values are listed below each image. In the segmentation images, red and blue pixels denote vegetation and ground, respectively.

Train \ Test	Experiments 1 & 2 (segmented ExG vs. segmented NDVI)							
	Canola		Pulse		Wheat		All-crop	
	Real NDVI	Gen NDVI	Real NDVI	Gen NDVI	Real NDVI	Gen NDVI	Real NDVI	Gen NDVI
Canola	0.8	0.79	0.85	0.85	0.69	0.7	0.78	0.78
Pulse	0.82	0.79	0.85	0.85	0.71	0.69	0.8	0.78
Wheat	0.81	0.79	0.85	0.84	0.7	0.7	0.79	0.78
All-crop	0.8	0.79	0.86	0.85	0.7	0.69	0.78	0.78

Table 3: Average DSC values for comparison between segmented ExG and NDVI values in the in-domain and out-of-domain experiments. The DSC values representing the overlap between the ExG and real NDVI segmentations and the overlap between the ExG and generated NDVI segmentations are reported below 'Real NDVI' and 'Gen NDVI', respectively. The shaded experimental evaluations correspond to the in-domain tests. The all-crop model tested on single crop datasets represents an upper bound because its distribution covers the other crop's distribution.

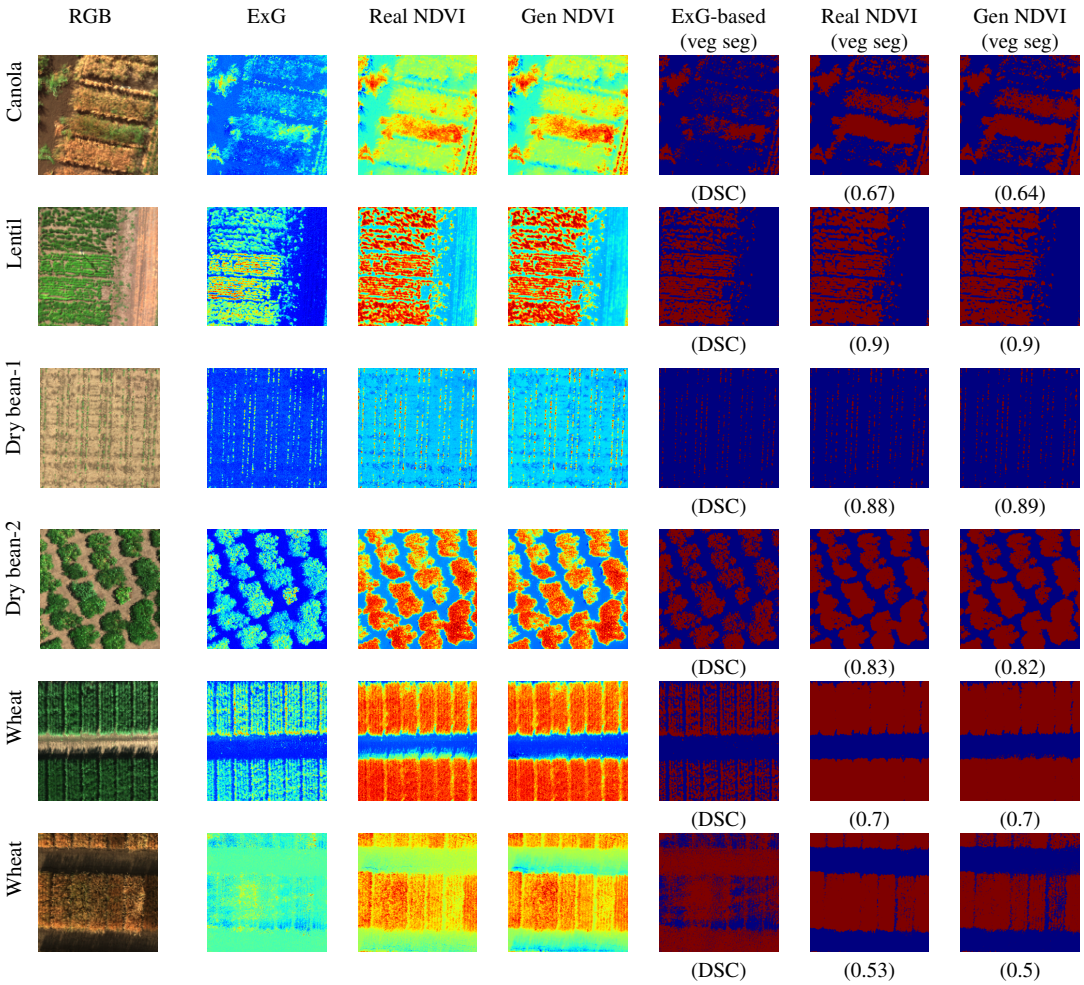

Figure 4: Example test images from the in-domain experiment for the canola, pulse, and wheat datasets for different growth stages. False colored (*'jet'* colormap) ExG, real NDVI and generated NDVI images are shown from columns 2 to 4 respectively, as well as the vegetation segmentation produced with each ('Veg Seg'), shown in the last three columns. The type of the crop is shown beside each row. The DSC value representing the overlap between the real NDVI and ExG segmentations and the overlap between the generated NDVI and ExG segmentations are listed below 'Real NDVI (veg seg)' and 'Gen NDVI (veg seg)', respectively. In the segmentation images, red and blue pixels denote vegetation and ground, respectively.

segmentation using DSC to demonstrate the difference between using generated NIR vs. visible spectrum alone for vegetation segmentation.

4. Results

Table 2 represent the PSNR, SSIM, SAM, and DSC values in both the in-domain and out-of-domain experiments. In the in-domain experiment where the model tested on the crop of interest, all datasets achieved above 92% accuracy in SSIM values meaning the distribution of the generated images is closely matched with the target distribution. SAM values have shown low numbers which is another proof for having substantial overlap between the two generated and target image distributions. PSNR values show that noise was removed from the generated NIR image. DSC values have an almost complete overlap between the generated and real NDVI segmentations. The results for the in-domain experiment indicate that the model can produce high fidelity results required for further agricultural analysis when a model is trained on the crop of interest. The out-of-domain experiment where the model tested on different crop types, have reported above 90% overlap between the generated and real NDVI segmentations via DSC values, meaning the model trained on a single crop type can generalize well on unseen data from another crop. SSIM, PSNR, and SAM also achieved relatively high accuracy for a shifted domain problem. The all-crop dataset is not considered as out-of-

domain evaluation because the distribution of the dataset covers a larger data distribution consisting of every single crop type, and represents a generalization upper bound.

Figure 3 shows the quality of the generated images both quantitatively and qualitatively for the in-domain experiment. The pulse dataset including lentil, dry bean-1, dry bean-2 crops, contains imbalanced number of images of the lentil grain. However, the model has successfully learned the target domain distribution despite of the lowest number of lentil images in the dataset, visualized in the second row of Figure 3.

Table 3 report the overlap (DSC) between the NDVI and ExG based segmentations in the in-domain and out-of-domain experiments across the entire growing season. In general, segmentations produced by generated NDVI are a better match to actual NDVI than ExG based segmentations. We observe a plant type and plant growth stage effect in this result. Figure 4 demonstrates that these two indices are closer in the mid growth stage for crops with relatively low density in vegetation. ExG predicts less plant pixels and consequently it shows a closer match with NDVI in the canola and pulse datasets. The overlap between the ExG and NDVI segmentations is lower in the wheat dataset where the wheat plants are narrow and dense. In the late season examples, both segmented ExG and NDVI are inconsistent because of less prominent reflection of NIR and green spectral bands.

5. Discussion

The in-domain experiment achieved the most promising performance because there is no concern of domain shift, as shown in Table 2. The all-crop dataset in the in-domain experiment has shown a comparable result with each single crop dataset, where the model was trained on a larger sample size covering more variations and tested on a smaller population. In the out-of-domain experiment, the all-crop dataset has shown a close performance to each single crop dataset in the in-domain experiment, because of its cover over the other datasets. The generalizability of the model is encouraging for applying this model to new crops or differing field conditions with little or no retraining.

Many remote sensing research studies have investigated the cross modality inference problem using satellite imagery. In [62], a cGAN with the same architecture as our model was employed to produce NIR spectral band from RGB satellite images. They assessed their results with Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and SSIM measures. They achieved 92.28% SSIM using the L_1 robust loss function, which is similar to our results for the in the in-domain experiment. Despite the substantially increased ground resolution of our UAV images as compared to satellite images, RGB to NIR translation performance remained consistent.

Vegetation segmentations produced from real versus generated NDVI were nearly identical across all crops and growth stages. This is a promising result for using generated NIR in agricultural applications. The generated NIR images can be used in several different agricultural indices combining with visible spectrum to enhance the contribution of vegetation properties. For example NDRE [5], which is more sensitive than NDVI for a certain period of crop maturation.

The segmentations produced by generated NDVI were also mostly consistent with the standard ExG based segmentations. In general, the ExG segmentations appeared tighter to the vegetation than the NDVI based segmentations. This may be due to NDVI becoming saturated and providing slightly less vegetation discrimination than ExG. However, in later growth stages, as seen in the mature canola and wheat in Figure 4, NDVI provided a better segmentation of plant pixels. This is expected due to the lack of green reflectance after senescence in wheat and canola.

While our work has made substantial contributions to the analysis of crops from aerial imagery, it is not without shortcomings. First, we only investigated four crops. Other staple crops such as corn, cassava, or soybean might have significantly different responses to the model. Further model verification with additional crops would be beneficial. Second, we only evaluated a single multispectral camera, and only inferred a single band. Additional hardware and frequency band characterization would extend this work. Finally, we examined well-maintained breeding crops, and did not consider crops under disease or environmental stress. Extending the model to include these data would enhance utility and generalizability.

6. Conclusion

In this work, we have demonstrated the effectiveness of generating an NIR reflectance band from RGB bands for aerial crop images using image-to-image translation. We collected raw sensor data via a multispectral sensor mounted on a UAV over four trial fields. Raw images were processed to be radiometrically calibrated and pixel-wise aligned with their NIR counterpart. We conducted two deep-learning based experiments trained and tested on different combinations of our dataset. We performed extensive quantitative and qualitative assessments considering standard and agricultural domain-specific measures to evaluate the quality of the generated NIR images. Our evaluation demonstrates that the generated NIR images are comparable to real NIR images across different crop species and growth stages. These results show that combining inexpensive RGB aerial imaging with image translation methods to synthetically generate NIR images is a promising approach for image-based plant phenotyping.

References

[1] Matthew Amodio and Smita Krishnaswamy. Travelgan: Image-to-image translation by transformation vector learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8983–8992, 2019.

[2] Masoomeh Aslahishahri, Kevin G Stanley, Hema Duddu, Steve Shirtliffe, Sally Vail, and Ian Stavness. Spatial super resolution of real-world aerial images for image-based plant phenotyping. *Remote Sensing*, 13(12):2308, 2021.

[3] Tewodros W Ayalew, Jordan R Ubbens, and Ian Stavness. Unsupervised domain adaptation for plant organ counting. In *European conference on computer vision*, pages 330–346. Springer, 2020.

[4] Ruud Barth, JMM IJsselmuiden, Jochen Hemming, and Eldert J van Henten. Optimising realism of synthetic agricultural images using cycle generative adversarial networks. In *Proceedings of the IEEE IROS workshop on Agricultural Robotics*, pages 18–22, 2017.

[5] Boris Boiarskii and Hideo Hasegawa. Comparison of ndvi and ndre indices to detect differences in vegetation and chlorophyll content. *Journal of Mechanics of Continua and Mathematical Sciences*, 4:20–29, 2019.

[6] Tomasz Borowik, Nathalie Pettorelli, Leif Sönnichsen, and Bogumiła Jedrzejewska. Normalized difference vegetation index (ndvi) as a predictor of forage availability for ungulates in forest and field habitats. *European journal of wildlife research*, 59(5):675–682, 2013.

[7] M. Brown and S. Süsstrunk. Multispectral SIFT for scene category recognition. In *Computer Vision and Pattern Recognition (CVPR11)*, pages 177–184, Colorado Springs, June 2011.

[8] Zijue Chen, David Ting, Rhys Newbury, and Chao Chen. Semantic segmentation for partially occluded apple trees based on deep learning. *Computers and Electronics in Agriculture*, 181:105952, 2021.

[9] Etienne David, Simon Madec, Pouria Sadeghi-Tehran, Helge Aasen, Bangyou Zheng, Shouyang Liu, Norbert Kirchgessner, Goro Ishikawa, Koichi Nagasawa, Minhajul A Badhon, et al. Global wheat head detection (gwhd) dataset: a large and diverse dataset of high-resolution rgb-labelled images to develop and benchmark wheat head detection methods. *Plant Phenomics*, 2020, 2020.

[10] Emily Denton and Vighnesh Birodkar. Unsupervised learning of disentangled representations from video. *arXiv preprint arXiv:1705.10915*, 2017.

[11] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[12] Dániel Fróna, János Szenderák, and Mónika Harangi-Rákos. The challenge of feeding the world. *Sustainability*, 11(20):5816, 2019.

[13] Robert T Furbank and Mark Tester. Phenomics–technologies to relieve the phenotyping bottleneck. *Trends in plant science*, 16(12):635–644, 2011.

[14] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

[15] YanFeng Gu, XuDong Jin, RunZi Xiang, QingWang Wang, Chen Wang, and ShengXiong Yang. Uav-based integrated multispectral-lidar imaging system and data processing. *Science China Technological Sciences*, 63(7):1293–1301, 2020.

[16] Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 327–340, 2001.

[17] Eija Honkavaara and Ehsan Khoramshahi. Radiometric correction of close-range spectral image blocks captured using an unmanned aerial vehicle with a radiometric block adjustment. *Remote Sensing*, 10(2):256, 2018.

[18] Svetlana Illarionova, Dmitrii Shadrin, Alexey Trekin, Vladimir Ignatiev, and Ivan Oseledets. Generation of the nir spectral band for satellite images with convolutional neural networks. *arXiv preprint arXiv:2106.07020*, 2021.

[19] Vaishali Ingale, Rishabh Singh, and Pragati Patwal. Image to image translation: Generating maps from satellite images. *arXiv preprint arXiv:2105.09253*, 2021.

[20] Michal Irani and Shmuel Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of visual communication and image representation*, 4(4):324–335, 1993.

[21] Brian KS Isaac-Medina, Matt Poyser, Daniel Organisciak, Chris G Willcocks, Toby P Breckon, and Hubert PH Shum. Unmanned aerial vehicle visual detection and tracking using deep neural networks: A performance benchmark. *arXiv preprint arXiv:2103.13933*, 2021.

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

[23] Liming Jiang, Changxu Zhang, Mingyang Huang, Chunxiao Liu, Jianping Shi, and Chen Change Loy. Tsit: A simple and versatile framework for image-to-image translation. In *European Conference on Computer Vision*, pages 206–222. Springer, 2020.

[24] Sami Khanal, John Fulton, and Scott Shearer. An overview of current and potential applications of thermal remote sensing in precision agriculture. *Computers and Electronics in Agriculture*, 139:22–32, 2017.

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

[26] Joshua CO Koh, German Spangenberg, and Surya Kant. Automated machine learning for high-throughput image-based plant phenotyping. *Remote Sensing*, 13(5):858, 2021.

[27] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton Earnshaw, Imran Haque, Sara M Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. Wilds: A benchmark

of in-the-wild distribution shifts. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 5637–5664. PMLR, 18–24 Jul 2021.

[28] Fred A Kruse, AB Lefkoff, JW Boardman, KB Heidebrecht, AT Shapiro, PJ Barloon, and AFH Goetz. The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163, 1993.

[29] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017.

[30] Hsin-Ying Lee, Hung-Yu Tseng, Qi Mao, Jia-Bin Huang, Yu-Ding Lu, Maneesh Singh, and Ming-Hsuan Yang. Drit++: Diverse image-to-image translation via disentangled representations. *International Journal of Computer Vision*, 128(10):2402–2417, 2020.

[31] Xinyang Li, Jie Hu, Shengchuan Zhang, Xiaopeng Hong, Qixiang Ye, Chenglin Wu, and Rongrong Ji. Attribute guided unpaired image-to-image translation with semi-supervised learning. *arXiv preprint arXiv:1904.12428*, 2019.

[32] Jianxin Lin, Yijun Wang, Tianyu He, and Zhibo Chen. Learning to transfer: Unsupervised meta domain translation. *arXiv preprint arXiv:1906.00181*, 2019.

[33] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. *arXiv preprint arXiv:1703.00848*, 2017.

[34] G Lowe. Sift-the scale invariant feature transform. *Int. J*, 2(91-110):2, 2004.

[35] Zhuoran Ma, Feifei Wang, Weizhi Wang, Yeteng Zhong, and Hongjie Dai. Deep learning for in vivo near-infrared imaging. *Proceedings of the National Academy of Sciences*, 118(1), 2021.

[36] Baabak Mamaghani and Carl Salvaggio. Multispectral sensor calibration and characterization for suas remote sensing. *Sensors*, 19(20):4453, 2019.

[37] MicaSense Rededge Support. Getting started with rededge, micasense knowledge base, 2018.

[38] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[39] Aamir Mustafa and Rafal K Mantiuk. Transformation consistency regularization-a semi-supervised paradigm for image-to-image translation. *arXiv preprint arXiv:2007.07867*, 2020.

[40] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[41] Achraf Oussidi and Azeddine Elhassouny. Deep generative models: Survey. In *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pages 1–8. IEEE, 2018.

[42] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2337–2346, 2019.

[43] Nathalie Pettorelli. *The normalized difference vegetation index*. Oxford University Press, 2013.

[44] Jean B Ristaino, Pamela K Anderson, Daniel P Bebber, Kate A Brauman, Nik J Cunniffe, Nina V Fedoroff, Cambria Finegold, Karen A Garrett, Christopher A Gilligan, Christopher M Jones, et al. The persistent threat of emerging plant disease pandemics to global food security. *Proceedings of the National Academy of Sciences*, 118(23), 2021.

[45] Gustavo K Rohde, Sinisa Pajevic, Carlo Pierpaoli, and Peter J Basser. A comprehensive approach for multi-channel image registration. In *International Workshop on Biomedical Image Registration*, pages 214–223. Springer, 2003.

[46] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[47] Ruslan Salakhutdinov and Geoffrey Hinton. Deep boltzmann machines. In *Artificial intelligence and statistics*, pages 448–455. PMLR, 2009.

[48] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798, 2007.

[49] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1):7–42, 2002.

[50] Michael Schmitt, Lloyd Haydn Hughes, Chunping Qiu, and Xiao Xiang Zhu. Sen12ms–a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. *arXiv preprint arXiv:1906.07789*, 2019.

[51] Guy Shacht, Dov Danon, Sharon Fogel, and Daniel Cohen-Or. Single pair cross-modality super resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6378–6387, 2021.

[52] Patricia L Suárez, Angel D Sappa, Boris X Vintimilla, and Riad I Hammoud. Image vegetation index through a cycle generative adversarial network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[53] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[54] Hao Tang, Dan Xu, Nicu Sebe, Yanzhi Wang, Jason J Corso, and Yan Yan. Multi-channel attention selection gan with cascaded semantic guidance for cross-view image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2417–2426, 2019.

[55] Chao Wang, Haiyong Zheng, Zhibin Yu, Ziqiang Zheng, Zhaorui Gu, and Bing Zheng. Discriminative region proposal adversarial networks for high-quality image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 770–785, 2018.

[56] Lei Wang, Xin Xu, Yue Yu, Rui Yang, Rong Gui, Zhaozhuo Xu, and Fangling Pu. Sar-to-optical image translation us-

ing supervised cycle-consistent adversarial networks. *IEEE Access*, 7:129136–129149, 2019.

[57] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018.

[58] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[59] David M Woebbecke, George E Meyer, Kenneth Von Bargen, and David A Mortensen. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, 38(1):259–269, 1995.

[60] Jungang Xu, Hui Li, and Shilong Zhou. An overview of deep generative models. *IETE Technical Review*, 32(2):131–139, 2015.

[61] Jinru Xue and Baofeng Su. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors*, 2017, 2017.

[62] Xiangtian Yuan, Jiaojiao Tian, and Peter Reinartz. Generating artificial near infrared spectral band from rgb image using conditional generative adversarial network. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:279–285, 2020.

[63] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 701–710, 2018.

[64] Yongbing Zhang, Siyuan Liu, Chao Dong, Xinfeng Zhang, and Yuan Yuan. Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution. *IEEE transactions on Image Processing*, 29:1101–1112, 2019.

[65] Fan Zhao, Wenda Zhao, Libo Yao, and Yu Liu. Self-supervised feature adaption for infrared and visible image fusion. *Information Fusion*, 2021.

[66] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[67] Peihao Zhu, Rameen Abdal, Yipeng Qin, and Peter Wonka. Sean: Image synthesis with semantic region-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5104–5113, 2020.