Generating Synthetic Multispectral Images for Semantic Segmentation in Forestry Applications

Dominik Bittner[1], João Filipe Ferreira[2,3], Maria Eduarda Andrada[3], Jordan J. Bird[2] and David Portugal[3]

***Abstract*— In this paper, we introduce a GAN-based solution for generating synthetic multispectral images from fully-annotated RGB images for data augmentation purposes in forestry robotics applications at ground-level. Fully-annotated multispectral datasets are difficult to obtain with sufficient training samples when compared to RGB-based datasets, since annotation in this case is often very time-consuming and expensive due to the need for expert knowledge. In this text, a study comparing different GAN-based image translation models designed for data augmentation is presented. Synthetic images generated by the proposed solution are shown to be realistic enough to yield performance ratings comparable to what is obtained using real images.**

I. Introduction

Multispectral and hyperspectral imaging (MSI/HSI) have been extensively used in many recent applications, including precision agriculture [1], [2], crop [3], land resource [4] and water quality [5] monitoring, vegetation coverage [6], precision forestry [7], [8], and many others relating to promoting sustainability and protecting the environment. For precision forestry and forestry robotics in particular, multispectral imaging is expected to considerably improve perceptual capabilities compared to RGB imaging. This is due to the use of specialised filters to implement near-infrared (NIR) channels tuned to detecting chlorophyll, therefore enhancing plant detection and discrimination through machine learning and, more specifically, deep learning models [7], [8].

Unfortunately, fully-annotated multispectral datasets are difficult to obtain with sufficient training samples when compared to RGB-based datasets, since annotation is often time-consuming and expensive due to the need for expert knowledge (an issue compounded even further by the specific nature of multispectral images) [9], [10]. Two different approaches have been adopted to solve this gap between the significant number of training samples required by deep learning models and the annotated images that can be acquired in practice:

- deep learning models for MSI/HSI classification with few annotated samples – see [9] for an in-depth survey on this approach;
- data augmentation solutions to increase the amount of annotated images without the need for human expert intervention.

In this paper, we present a GAN-based solution for generating synthetic multispectral images from fully-annotated RGB images for data augmentation purposes, following the latter approach. In particular, we provide proof-of-concept showing that the synthetic images generated by this solution yield a level of performance in the semantic segmentation model that is being used in our application comparable to what is obtained in validation using real images. Moreover, we present a study comparing different GAN-based image translation models designed to choose the most appropriate one for this purpose. We close this text by drawing conclusions on our findings and setting out future work on MSI dataset augmentation.

II. Related Work

Synthetic images for data augmentation can be generated in various ways. Just to cite a few recent examples from different application areas, Oksuz et al. [11] use the k-space corruption scheme method to improve the accuracy in detection of motion artefacts in medical images, while style transfer methods, as presented by Mikołajczyk et al. [12], [13], have also been adopted to create synthetic images for data augmentation to handle the lack of sufficient amount of the image training data and improve classification and detection performance.

Generative Adversarial Networks (GANs) are a useful tool to perform data augmentation when datasets are small, fragmented, imbalanced or without labelling. Meister et al. demonstrated that a deep convolutional GAN is able to generate a large synthetic dataset from less than 50 input images to improve inspection processes [14]. In [15], Bird et al. found that the autonomous recognition ability of healthy and unhealthy fruits in smart agriculture could be improved when the training set was augmented with synthetic images of fruits generated by a Conditional GAN.

In [10], Kemker et al. generated synthetic MSI for semantic segmentation model training to overcome label scarcity and to make the segmentation process less prone to overfitting. Their results show that pretraining these models on synthetic multispectral imagery can improve their performance.

*This work has been mainly supported by an International Exchange program from the DAAD Germany's National Agency for EU Higher Education Cooperation (ref. 91820546). This work was also supported by the Safety, Exploration and Maintenance of Forests with Ecological Robotics (SEMFIRE, ref. CENTRO-01-0247-FEDER-03269) research projects co-funded by the Agência Nacional de Inovação within the Portugal2020 programme.

[1] Dominik Bittner is with the Faculty of Electrical Engineering and Information Technology, Ostbayerische Technische Hochschule Regensburg, Germany. E-mail: dominik.bittner@st.oth-regensburg.de

[2]João Filipe Ferreira and Jordan J. Bird are with the Department of Computer Science, School of Science and Technology, Nottingham Trent University, UK. E-mail: {joao.ferreira,jordan.bird}@ntu.ac.uk

[3]Maria Eduarda Andrada, David Portugal and João Filipe Ferreira are with the Institute of Systems and Robotics, University of Coimbra, Portugal. E-mail: {duda.andrada,davidbsp,jfilipe}@isr.uc.pt

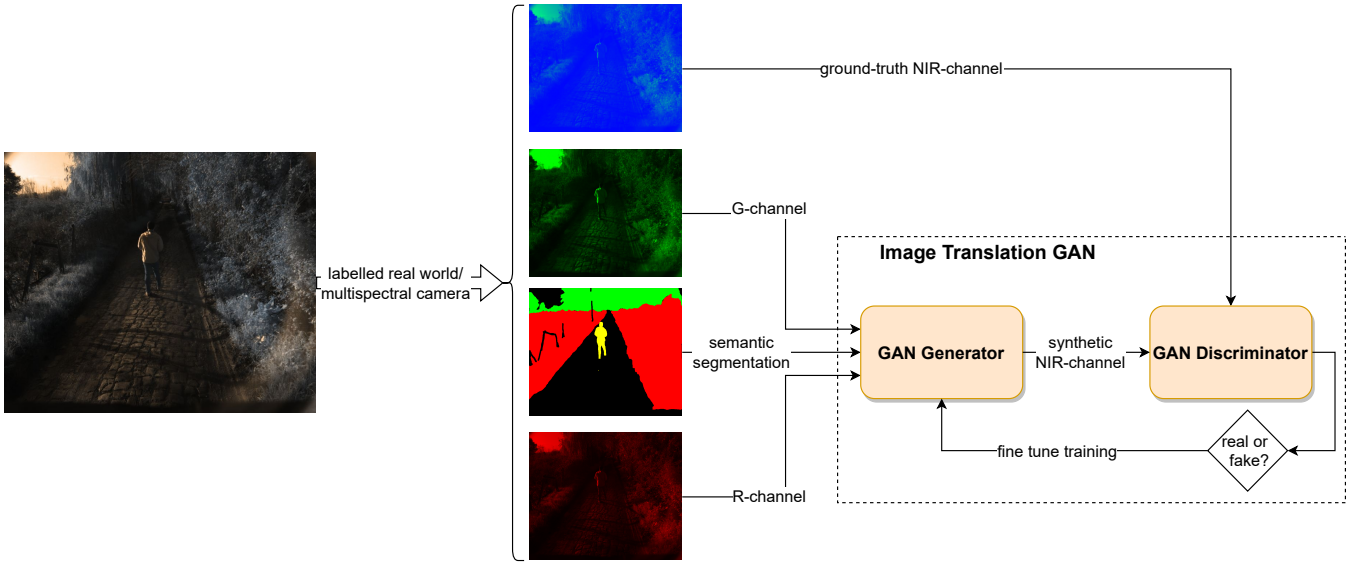


Fig. 1. GAN image translation training to generate corresponding NIR channels of multispectral images with an original multispectral image (left image); a model ground truth image, which the model attempts to predict (centre, top image); a green channel image, which is part of the model input image (centre, second image from the top); a semantic segmentation image, where its label values are part of the model input image (centre, third image from the top) and a red channel image, which is part of the model input image (centre, last image).

The same problem, namely the lack of a large number of samples with labels, was addressed in [16] using data imputation and matrix completion methods to deal with unacceptable model generalisations. According to their findings, augmenting data by imputation and matrix completion methods can increase the classification performance of multispectral images.

Abady et al. in [17] and Mohandoss et al. in [18] applied GANs to multispectral satellite imagery to also cope with the absence of a high number of samples but without labelling. They achieved satisfactory synthetic image quality for the data augmentation. Fawakherji et al. [19] created semi-artificial multispectral samples by employing a GAN to tackle the problem of label shortage for classification in precision farming. By replacing only the objects, e.g. plants and weeds, with synthetic objects, they enabled increased segmentation performance of state-of-the-art semantic segmentation convolutional networks with respect to common metrics.

Despite the success of recent studies data augmentation, and in particular synthetic image generation using GANs, we are unaware of any solution that specifically tackles multispectral image data for robotic forestry applications at ground-level – most of the solutions produced for forestry applications concern aerial and satellite MSI, as reported above.

III. Implementation

A. GANs for Image-to-Image Translation

Image-to-image translation is a method for converting an input image from a source domain to a target domain [20]. CNN-based image-to-image translation require fully comprehensive datasets containing images from the source domain as well as from the target domain in order to enable the model to understand the relationships between the two domains. Recent approaches for image translation include the use of autoencoders [21] and GANs.

Arguably the most commonly used approach, GAN-based solutions attempt to learn the recreation of specific ground truth images by optimising loss functions [22]. This approach is useful in terms of image quality, as GAN loss functions are designed to approximate human perception. Moreover, when corresponding labelling is available across domains, supervised GAN approaches yield accurate and highly precise samples of a dataset by taking the semantic segmentation mask and the specific channels of an input image as conditions.

TABLE I
Semantic classes and respective color coding

Class	Color
Background (BA)	Black
Live flammable material (aka Fuel) (FU)	Red
Trunks (TR)	Brown
Canopies (CA)	Green
Humans (HU)	Yellow
Animals (AN)	Purple

B. Synthetic Multispectral Image Generation Using GAN-based Solutions

To generate synthetic MSI data from RGB images, given what was explained in the previous section, we have decided to use a GAN-based approach for image-to-image translation from the RGB to the MSI domain.

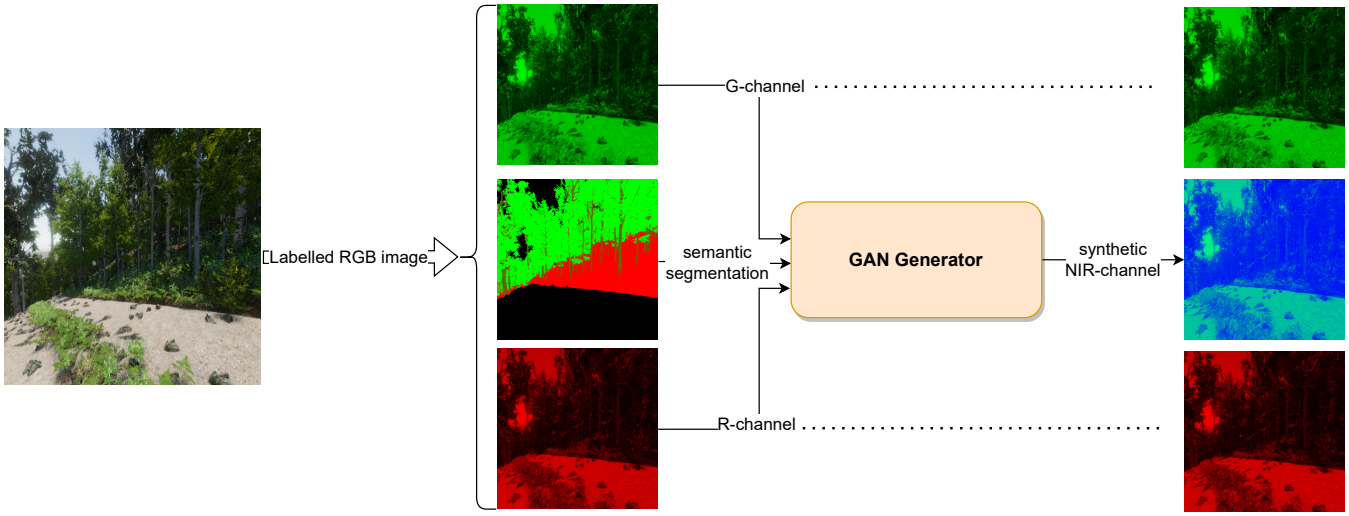


Fig. 2. GAN image translation generation of synthetic NIR channel and corresponding final "fake" multispectral image from a fully-annotated RGB input image (on the left); a green channel image, which is part of the model input image and is also feed forward to be merged after generation with the synthetic NIR channel (centre, top image and right, top image); a semantic segmentation image, where its label values are part of the model input image (centre, second image from the top); a red channel image, which is part of the model input image and is also feed forward to be merged after generation with the synthetic NIR channel (centre, last image and right, last image); a synthetic NIR channel image, which the model predicted and is merged afterwards with the real red and green channels as a synthetic multispectral image (right, second image from the top).

The multispectral images used in our application consist of red, green, and NIR (near-infrared) channels. Since our training dataset, described in section IV-A, includes ground-truth segmented labelling masks resulting from human expert annotation, these data are used for supervised training of a GAN framework to learn how to generate "fake" NIR-channel images, thereby implementing a 3-to-1 image-to-image translation model (Fig. 1). After the model is trained, it is expected to be able to generate synthetic MSI data from RGB images taken as input by only replacing the blue (B) channel with the "fake" NIR-channel, as depicted in Fig. 2.

Since we have access to labelling across domains, we have chosen to compare three state-of-the-art supervised learning-based GANs for image-to-image translation. The first solution, pix2pix GAN [23], is a conditional adversarial network, which offers a general-purpose solution to image-to-image translation problems. The network not only learns the mapping from input image to the output image, but also learns a loss function to train this mapping. This makes it possible to apply the same generic approach to problems that traditionally would require relatively different loss formulations. The second solution under scrutiny, pix2pixHD GAN [24], is an extension of pix2pix GAN. It is a method for synthesizing high-resolution photo-realistic images from semantic label maps using conditional GANs. Using pix2pixHD it is possible to generate 2048 $\times$ 1024 visually satisfying results, as it consists of adapted multi-scale generator and discriminator architectures. Finally, GauGAN [25], the third solution we tested which is in turn based on pix2pixHD, includes a spatially adaptive normalisation layer called "SPADE" to synthesise photorealistic images that specify a semantic layout.

(a) (b)

Fig. 3. Example of (a) an original multispectral image and (b) its ground truth labelling using the classes listed in Table I.

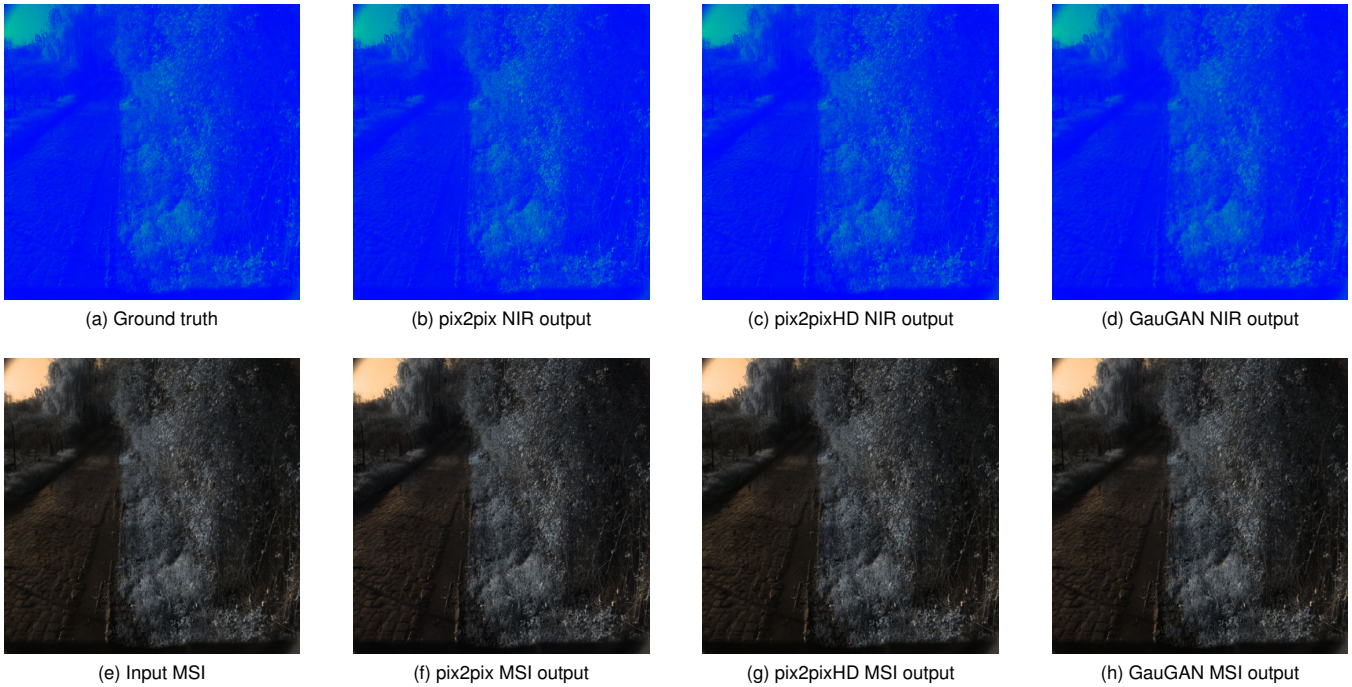
IV. Experimental Results

A. Experimental Setup and Metrics Definition

For training the different GANs on translating multispectral images into the corresponding NIR channels, we used the "2019_2020_quinta_do_bolao_coimbra" dataset from the SEMFIRE forestry robotics repository [26]. This dataset includes MSI data obtained using a Teledyne DALSA multispectral camera at a 720p resolution, totalling 920 images. The dataset also includes pixel-wise annotation masks resulting from human expert annotation according to the 6 classes listed in Table I (see also Fig. 3 for an illustrative example).

The evaluation of the appropriateness and quality of the synthetic images generated by the three GANs, which were trained for 500 epochs, was performed in three steps:

- First, we performed a qualitative evaluation by examining the quality of the resulting generated synthetic images by visual inspection.
- Then, the models were quantitatively tested by replacing the original validation image set with the corresponding



(a) Ground truth (b) pix2pix NIR output (c) pix2pixHD NIR output (d) GauGAN NIR output

(e) Input MSI (f) pix2pix MSI output (g) pix2pixHD MSI output (h) GauGAN MSI output

Fig. 4. Synthetic image generation per model for the same MSI input – first row ((a) - (d)): NIR-channels; second row ((e) - (h)): Multispectral images. First column: ground truth for NIR-channel and original multispectral image input. Second to last column: pix2pix, pix2pixHD and GauGAN "fake" image (i.e. prediction) generated for NIR-channel and the complete synthetic multispectral image, composed by original R- and G-channels and generated NIR-channel.

synthetic images that were generated. Subsequently, we compared the resulting performance of the semantic segmentation model proposed in [8] for all methods against the baseline represented by the original images.

- Finally, the models were benchmarked against each other with the most commonly used image-to-image translation metrics, which score each model according to how realistic the synthetic images would seem to human eyes.

The specific metrics used to evaluate semantic segmentation for the first step were the F1-Score, Recall, and IoU (Intersection over Union), in accordance with [8]. The mean value for these metrics was calculated across all classes and for each separate class as follows:

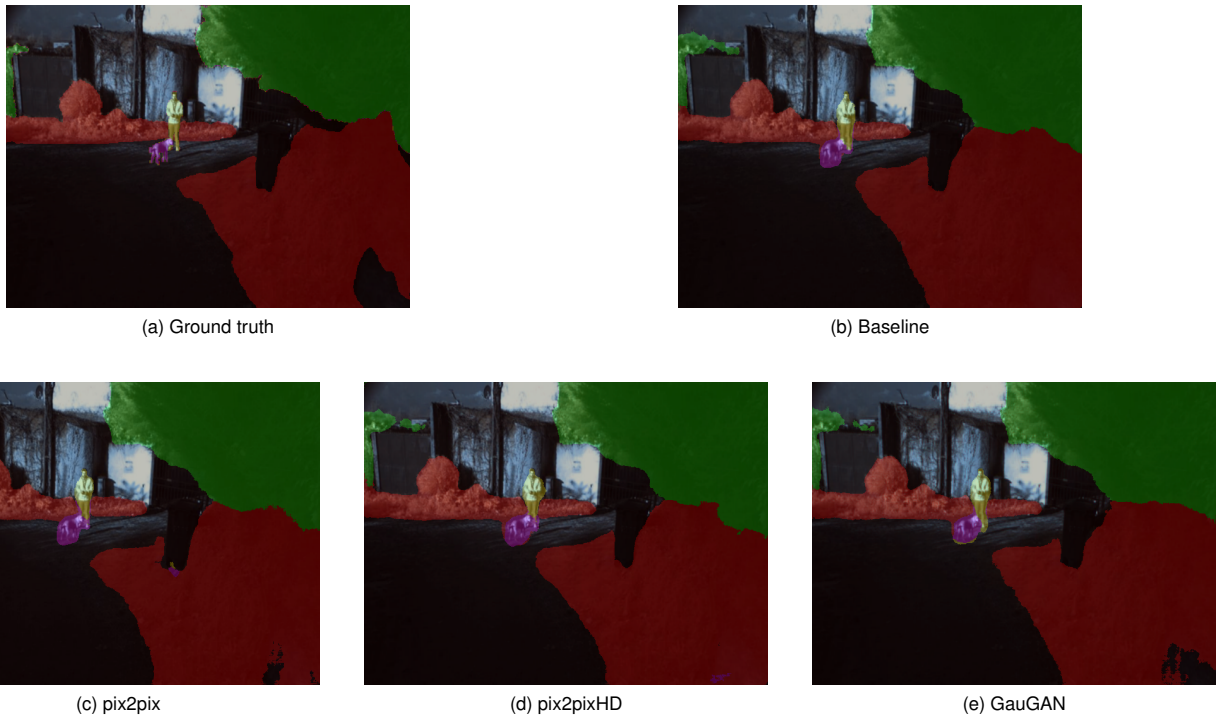
$$\text{F1-Score} = \frac{2TP}{2TP + FN + FP}; \tag{1}$$

$$\text{IoU} = \frac{TP}{TP + FN + FP}; \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3}$$

where TP stands for True Positive, FN for False Negative, FP for False Positive and TN stands for True Negative. To assess the appropriateness of each method we set a target of scoring **at least** 80% of the baseline performance for each metric across all classes.

As for quantitative comparative benchmarking, we used two of the most common metrics for evaluating how realistic generated images, the Inception Score (IS) and the Frechet Inception Distance (FID) – see [27] for an in-depth survey. The IS [28] uses a pretrained deep learning neural network model, the Inception v3 model [29] trained on the ImageNet dataset [30], for image classification of the generated images. It measures the average Kullback-Leibler divergence (KL divergence) between the conditional label distribution $p(y|x)$ of samples (expected to have low entropy for easily classifiable samples; better sample quality) and the marginal distribution $p(y)$ obtained from all the samples (expected to have high entropy if all classes are equally represented in the set of samples; high diversity). It favors low entropy of $p(y|\mathbf{x})$ but a large entropy of $p(y)$ [27]. The IS shows a reasonable correlation with the quality and diversity of generated images [28]. The FID has been proposed as an alternative or complement to the Inception Score [31]. FID embeds a set of generated samples into a feature space given by a specific layer (Inception v3 model or any other CNN). Viewing the embedding layers as a continuous multivariate Gaussian, the mean and covariance are estimated for both the generated images and the original images. The Frechet distance between these two Gaussians, which is also known as the *Wasserstein-2 distance* [32], is then used to quantify the quality of generated samples. FID performs reliably in terms of discriminability, robustness and computational efficiency. However, it assumes that features are of Gaussian distribution which is often not guaranteed. It has been shown that FID is consistent with human judgments and is more robust to noise than IS. In conclusion, a lower FID means smaller distances between synthetic and real data



(a) Ground truth (b) Baseline

(c) pix2pix (d) pix2pixHD (e) GauGAN

Fig. 5. Semantic segmentation using the model presented in [8] per GAN-based solution compared to (a) ground truth and (b) semantic segmentation Reference-Training as baseline. Figures (c) - (e) represent the Replaced-Validation Subset-Training results for pix2pix, pix2pixHD and GauGAN. In each case, the segmentation mask result is overlaid on the original multispectral images.

distributions, therefore it indicates more realistic generated images [27].

B. Qualitative Evaluation

Fig. 4 demonstrates the outcomes of the generation of the "fake" NIR-channel and respective synthetic MSI for each model as comparing to the corresponding ground-truth images. Fig. 5, shows an example of semantic segmentation results using the model proposed by [8], with the segmentation mask overlaid on the original multispectral image after training with the results of pix2pix, pix2pixHD and GauGAN as follows:

- **Reference-Training:** with the original multispectral image dataset as is.
- **Replaced-Validation Subset-Training:** retrained after replacing the near-infrared (NIR) channels of the multispectral images in the validation set with their synthetic versions generated by each GAN-based solution.

Both figures clearly demonstrate satisfactory qualitative results for synthetic MSI generation, with ground truth images, baseline segmentations, and model predictions nearly indistinguishable by visual inspection.

C. Quantitative Evaluation

Using the methodology for training described in the previous section, quantitative results were also obtained, more specifically by applying the F1-Score, IoU and Recall metrics defined in section IV-A – these are presented in Table II. The table indicates the absolute metrics per class, the mean value across all classes, and the scoring percentages against the baseline for each GAN-based solution, as defined in section IV-A.

These results show that all solutions yield a level of performance from the semantic segmentation model comparable to what is obtained in validation using real images, as the scoring percentages against the baseline are above the target value of 80% for all metrics established in section IV-A. Interestingly, the metrics scores for the "Humans" and "Animals" classes are furthest from the baseline scores among all classes, which is to be expected, given that these are the most under-represented in the dataset. In fact, it becomes clear that if only the other four classes would contribute for the mean values for the metrics, the performance of the GAN solutions would be even higher.

D. Comparative Benchmarking

When comparing the GAN-based methods with one another, the scoring percentages in Table II show that GauGAN approach is marginally better than the others for the F1-Score and IoU; however, this difference is nearly insignificant.

Table III, which compares all methods using the relative metrics for assessing image realism defined in section IV-A, also shows that the difference between the approaches is close to negligible, with pix2pix coming first for IS and pix2pixHD for FID, and GauGAN achieving in-between results for both metrics.

TABLE II
BENCHMARKING RESULTS OF ABSOLUTE METRICS (PLEASE REFER TO TABLE I FOR MEANING OF CLASS LABELS).

Training	Parameter	Mean	%Baseline Mean*	BA	FU	TR	CA	HU	AN
Reference-Training (baseline)	F1-Score	0.822	–	0.960	0.933	0.743	0.900	0.826	0.567
	IoU	0.718	–	0.923	0.875	0.591	0.819	0.704	0.395
	Recall	0.946	–	0.942	0.921	0.930	0.940	0.974	0.967
pix2pix	F1-Score	0.737	89.7%	0.951	0.931	0.661	0.880	0.669	0.328
	IoU	0.626	87.2%	0.907	0.871	0.494	0.786	0.503	0.196
	Recall	0.778	**82.2%**	0.930	0.916	0.875	0.934	0.709	0.303
pix2pixHD	F1-Score	0.739	90.0%	0.953	0.934	0.664	0.886	0.671	0.324
	IoU	0.629	87.6%	0.911	0.876	0.500	0.795	0.505	0.194
	Recall	0.778	**82.2%**	0.931	0.921	0.869	0.939	0.705	0.304
GauGAN	F1-Score	0.744	**90.5%**	0.952	0.934	0.675	0.884	0.671	0.346
	IoU	0.633	**88.2%**	0.909	0.877	0.509	0.792	0.504	0.209
	Recall	0.777	82.1%	0.933	0.922	0.876	0.932	0.707	0.292

* The **%Baseline Mean** column provides the proportion of the parameter means for each model as a percentage of the baseline parameter mean.

TABLE III
BENCHMARKING RESULTS OF RELATIVE METRICS

Solution	IS ↑	FID ↓
pix2pix	**3.77**	65.2
pix2pixHD	3.40	**60.1**
GauGAN	3.50	60.3

V. CONCLUSION

In this study, we investigated several solutions to generate synthetic multispectral images for semantic segmentation in forestry applications. These solutions were evaluated through different metrics and discussed for decisions on which model is the most suitable for our approach. In general, all tested methods provide satisfactory results for the pre-established targets. Taking into account both the absolute and comparative evaluation results, GauGAN marginally offers the most appropriate solution for future use in synthetic MSI sample generation for data augmentation.

Future work will focus on data augmentation for semantic segmentation by generating synthetic multispectral images from RGB images taken from a photo-realistic forestry simulator, which allows the capability of generating an unlimited amount of training images with ground-truth labelling for data augmentation. We expect, as a consequence, that the semantic segmentation performance reported in [8] is decisively improved.

ACKNOWLEDGMENTS

We would like to thank the SEMFIRE research team for their invaluable help with dataset labelling and for supporting this work.

REFERENCES

[1] M. Teke, H. Deveci, O. Haliloğlu, S. Gürbüz, and U. Sakarya, "A short survey of hyperspectral remote sensing applications in agriculture," in *6th International Conference on Recent Advances in Space Technologies (RAST)*, 2013, pp. 171–176.

[2] I. Sa, M. Popović, R. Khanna, Z. Chen, P. Lottes, F. Liebisch, J. Nieto, C. Stachniss, A. Walter, and R. Siegwart, "Weedmap: A large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming," *Remote Sensing*, vol. 10, no. 9, 2018.

[3] I. Strachan, E. Pattey, and J. Boisvert, "Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance," *Remote Sens. Environ.*, vol. 80, no. 2, pp. 213–224, 2002.

[4] A. Bannari, A. Pacheco, K. Staenz, H. McNairn, and K. Omari, "Estimating and mapping crop residues cover on agricultural lands using hyperspectral and ikonos data," *Remote Sens. Environ.*, vol. 104, no. 4, pp. 447–459, 2006.

[5] A. Mohamad, "Sea water chlorophyll-a estimation using hyperspectral images and supervised artificial neural network," *Ecol. Inf.*, vol. 24, 2014.

[6] C. Jänicke, A. Okujeni, S. Cooper, M. Clark, P. Hostert, and S. van der Linden, "Brightness gradient-corrected hyperspectral image mosaics for fractional vegetation cover mapping in Northern California," *Remote Sens. Lett.*, vol. 11, no. 1, pp. 1–10, 2020.

[7] M. S. Couceiro, D. Portugal, J. F. Ferreira, and R. P. Rocha, "SEMFIRE: Towards a new generation of forestry maintenance multi-robot systems," in *2019 IEEE/SICE International Symposium on System Integration (SII)*, 2019, pp. 270–276.

[8] M. Andrada, J. F. Ferreira, D. Portugal, and M. Couceiro, "Testing different cnn architectures for semantic segmentation for landscaping with forestry robotics," *WPPMFR 2020*, 2020.

[9] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, 2021.

[10] R. Kemker, C. Salvaggio, and C. Kanan, "Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, p. 60–77, Nov 2018.

[11] I. Oksuz, B. Ruijsink, E. Puyol-Antón, A. Bustin, G. Cruz, C. Prieto, D. Rueckert, J. A. Schnabel, and A. P. King, "Deep learning using k-space based data augmentation for automated cardiac mr motion artefact detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 250–258.

[12] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE, 2018, pp. 117–122.

[13] X. Zheng, T. Chalasani, K. Ghosal, S. Lutz, and A. Smolic, "Stada: Style transfer as data augmentation," *arXiv preprint arXiv:1909.01056*, 2019.

[14] S. Meister, N. Möller, J. Stüve, and R. M. Groves, "Synthetic image data augmentation for fibre layup inspection processes: Techniques to enhance the data set," *Journal of Intelligent Manufacturing*, vol. 32, no. 6, pp. 1767–1789, 2021.

[15] J. J. Bird, C. M. Barnes, L. J. Manso, A. Ekárt, and D. R. Faria, "Fruit quality and defect image classification with conditional gan data augmentation," *Scientia Horticulturae*, vol. 293, p. 110684, 2022.

[16] A. Acción, F. Argüello, and D. B. Heras, "A new multispectral data augmentation technique based on data imputation," *Remote Sensing*, vol. 13, no. 23, 2021.

[17] L. Abady, M. Barni, A. Garzelli, and B. Tondi, "GAN generation of synthetic multispectral satellite images," in *Image and Signal Processing for Remote Sensing XXVI*, L. Bruzzone, F. Bovolo, and E. Santi, Eds., vol. 11533, International Society for Optics and Photonics. SPIE, 2020, pp. 122 – 133.

[18] T. Mohandoss, A. Kulkarni, D. Northrup, E. Mwebaze, and H. Alemohammad, "Generating synthetic multispectral satellite imagery from sentinel-2," *arXiv preprint arXiv:2012.03108*, 2020.

[19] M. Fawakherji, C. Potena, A. Pretto, D. D. Bloisi, and D. Nardi, "Multi-spectral image synthesis for crop/weed segmentation in precision farming," *Robotics and Autonomous Systems*, vol. 146, p. 103861, 2021.

[20] P. Yingxue, L. Jianxin, Q. Tao, and C. Zhibo, "Image-to-image translation: Methods and applications," *arXiv preprint arXiv:2101.08629*, 2021.

[21] L. Cai, H. Gao, and S. Ji, "Multi-stage variational auto-encoders for coarse-to-fine image generation," pp. 630–638, 2019.

[22] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[23] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2018.

[24] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," 2018.

[25] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," 2019.

[26] D. Bittner, M. E. Andrada, D. Portugal, and J. F. Ferreira, "SEMFIRE forest dataset for semantic segmentation and data augmentation," Dec. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5751905

[27] A. Borji, "Pros and cons of gan evaluation measures," 2018.

[28] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," 2016.

[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015.

[30] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[31] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," 2018.

[32] L. N. Vaserstein, "Markov processes over denumerable products of spaces, describing large systems of automata," *Problemy Peredachi Informatsii*, vol. 5, no. 3, pp. 64–72, 1969.