Pal-GAN: Palette-conditioned Generative Adversarial Networks

Adam Balint

School of Engineering, University of Guelph
Vector Institute

Graham W. Taylor

School of Engineering, University of Guelph
Vector Institute

Email: {balint, gwtaylor}@uoguelph.ca

Abstract

Recent advances in Generative Adversarial Networks (GANs) have shown great progress on a large variety of tasks. A common technique used to yield greater diversity of samples is conditioning on class labels. Conditioning on high-dimensional structured or unstructured information has also been shown to improve generation results, e.g. Image-to-Image translation. The conditioning information is provided in the form of human annotations, which can be expensive and difficult to obtain in cases where domain knowledge experts are needed. In this paper, we present an alternative: conditioning on low-dimensional structured information that can be automatically extracted from the input without the need for human annotators. Specifically, we propose a Palette-conditioned Generative Adversarial Network (Pal-GAN), an architecture-agnostic model that conditions on both a colour palette and a segmentation mask for high quality image synthesis. We show improvements on conditional consistency, intersection-over-union, and Fréchet inception distance scores. Additionally, we show that sampling colour palettes significantly changes the style of the generated images.

1 Introduction

The past several years have seen an explosion of deep learning brought on by an increase in freely available datasets with many examples and high quality labels. The availability of such datasets enables deep neural networks to solve many problems, especially those involving visual reasoning. Deep Neural Networks have been able to solve classification and detection tasks to super-human levels[1], and have shown potential in image generation and image translation tasks [2–5].

The performance of deep learning models is generally contingent on having access to a large amount of high quality labelled data for training. As a result, some industries have not been able to take full advantage of state of the art research. For example, the medical and remote sensing industries require expert knowledge in the field and a large workforce to generate labels for existing data. Along with other concerns, such as privacy, it is difficult to obtain the quantity of labelled data needed to train deep learning models.

A promising approach to solving data sparsity is dataset augmentation. Classical dataset augmentation applies common geometric and photometric distortions that can be modelled by a set of simple operations [6–9]. More advanced dataset augmentation techniques involve modelling the data generating distribution and sampling new data from the model [10, 11]. Generative adversarial networks have shown some effectiveness in data-scarce domains, such as the medical field [12, 13] and remote sensing images [14] in conditional settings. Conditioning on information-dense labels improves the quality of image generation [15]. Even though high specificity information-dense labels allow the network to focus on modelling relationships not contained in the labels, generating these labels comes at the cost of expensive human-labelled annotations.

In this paper, we explore other sources of information for conditioning GANs that can be automatically extracted from the input without any human annotation. Implicit information sources exist in the form of low-dimensional, structured information. Motivated by the desire to increase colour richness of generated images, as well as the ability to decrease the complexity of spatial conditions, we investigate conditioning on a colour palette.

Here, we only consider the setting where the annotation-free conditioning system complements the existing annotated information, specifically binary building masks. We believe this approach is general enough to assist in the unconditional setting. We apply our image generation technique to the dataset augmentation task and explore the effects of introducing varying amounts of generated data into a dataset to train a variety of segmentation networks.

In experiments conducted on the Amazon SpaceNet dataset, the Palette-conditioned GAN (Pal-GAN) architecture produces higher quality results when compared to both classical encoder-decoder and U-Net based GAN architectures. We assess the generated images quantitatively through three key metrics: Fréchet inception distance, conditional consistency IOU, and Learned Perceptual Image Patch Similarity. We then briefly explore Pal-GAN for dataset augmentation on a segmentation task through the IOU metric.

2 Related Work

2.1 Generative Adversarial Networks

Generative adversarial networks (GANs) form a family of generative models trained using two models that compete against each other to implicitly learn data distribution [16]. GANs have been applied to in-painting [5, 17, 18], image translation [2, 4], and image generation [16, 18]. The GAN network architecture consists of a generator $G(z)$ and a discriminator $D(x)$. The generator is typically a feed-forward neural network that aims to convert a random noise vector z into a target image x. The discriminator is a neural network that distinguishes between samples produced by the generator and those that are obtained from the dataset. The generator network is trained exclusively through feedback from the classification made by the discriminator. The discriminator learns to differentiate between real and generated examples through sampling both the generator and the dataset. The two components compete in a minimax game where the generator attempts to maximize $D(G(z))$ and the discriminator attempts to maximize $D(x)$ while minimizing $D(G(z))$. GANs implicitly learn data distributions without specifying any prior. As a result, GANs are susceptible to mode collapse; a state in which the generator learns to model a small subset of modes contained in the data distribution well enough to fool the generator. Subsequently, the generated data does not capture the entirety of the data distribution. To help counteract mode collapse, conditional variants of GANs are often used [2, 4, 15].

2.2 Conditional Generative Adversarial Networks

Conditional GANs are a modification of the basic GAN architecture that allow for more control over the generated images based on some additional context. We use the notation $G(z, c)$ for a generator conditioned on context c. It is common practice to condition on class labels [15]. However, it is also possible to incorporate more structured conditions such as segmentation masks or pose co-ordinates. Structured conditions allow GANs to generate entities based on a strict spatial mapping [4, 19], while text based conditioning guides generation in a more subtle way [20]. Another consideration for conditioning information is the expense at which it is obtained. In this paper, we focus on structured information obtained at no additional cost.

2.3 Remote Sensing

The field of remote sensing has increasingly adopted machine learning techniques to analyze data for industrial and governmental purposes. Some works from the machine learning community aim to detect roads and buildings [4], or classify crops [21]. Most machine learning techniques in the remote sensing field focus on classification[22, 23], detection[24, 25], and segmentation[21, 26] tasks. Some work has been conducted on dataset augmentation using GANs [27] in multi-sensor settings, converting between RGB, semantic segmentation, LIDAR, and 2d multi-class box labels. Using a CycleGAN architecture for dataset augmentation, a modest improvement in accuracy was achieved. Although the network improved performance on large objects, such as soccer fields and swimming pools, it had difficulty generating small objects and complex objects such as boats and planes. Our work directly addresses this limitation of the CycleGAN approach.

	Training	Validation	Test	Total
Rio	4,441	1,111	1,388	6,940
Las Vegas	2,464	616	771	3,851
Paris	733	184	230	1,147
Shanghai	2,932	733	917	4,582
Khartoum	647	162	203	1,012
Total	11,217	2,806	3,509	17,532

Table 1: Areas of interest and data splits of the SpaceNet dataset.

2.4 Colour Palette Based Image Generation

The use of colour palettes for an artistic image-to-image translation task has been shown to work in a well controlled environment with high-information conditions. For instance, PaletteNet [28] uses a set of 1611 curated image-palette pairs. The set is augmented in a three-stage augmentation process to produce images containing the same content, but displaying different colour palettes. The augmentation technique is applied manually to each image to generate a set of visually appealing results along with their palettes. The architecture is trained in two stages: first using a pixel-wise loss between the generated and the expected image, followed by an adversarially trained refinement step. The PaletteNet approach has the ability to produce visually appealing results, with the drawback of a slow manual dataset generation process. Pal-GAN improves the process by automatically extracting colour palettes and modifying the training method to better suit a one-to-many image generation task.

3 Methodology

3.1 Dataset

We train and evaluate our model on the Amazon SpaceNet dataset [29]. SpaceNet consists of over 17,000 remote sensing images, along with road and building information covering five geographical areas of interest: Rio, Las Vegas, Paris, Shanghai, and Khartoum. For our purposes, we only utilize the building footprint labels to produce binary masks of building locations. In addition to the RGB images and building footprint masks, we create a binary mask to indicate the region of interest by setting all non-black pixels to white. The region of interest mask is created because the SpaceNet dataset contains images where portions of the image extend beyond the city limits and are set to black pixels. Generating a region of interest binary mask allows for the penalization of only the areas that are within city limits. This allows the generator to produce realistic images of the city space without needing to also model the artificially black regions from the SpaceNet dataset.

For image generation, we divided the dataset into three components. The test set is 20% of the entire dataset proportionally divided among the different areas of interest. The remaining 80% of the data is then split into a training and validation set at a ratio of 4:1 (Table 1).

3.2 Colour Palette Extraction

There are many ways to extract colour palettes from a given image. For instance, an averaging approach is a relatively simple calculation but it may lead to loss of colour information. On the other hand, the k-means approach may better capture outlier colours, but at a cost of decreased performance and non-determinism. To avoid both problems, we propose a deterministic colour palette extraction technique that does not average values, rather it selects representative colours based on the distribution of colours within the image. Our palette extraction method first converts the image to HSV. Next, the hue and saturation are converted into an 8-bit binary number. Finally, the hue and saturation are concatenated and sorted. The array is then split into k buckets where k is the size of the colour palette, and the middle value of each bucket is selected. The selected colour is converted back into RGB and stored. This colour palette extraction technique allows for variable-sized palettes with limited impact on performance.
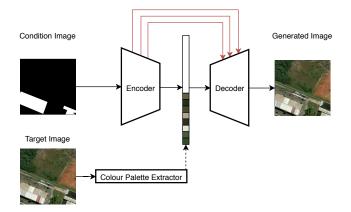

Fig. 1: Configuration style of the Pal-GAN model along with the flow of data during the pre-processing and training steps. The red lines correspond to connections that only exist in the U-Net style architecture. The discriminator is not shown.

3.3 Network Configuration

To start, we establish three baseline GAN results: unconditional, conditional encoder-decoder, and conditional U-Net. For both conditional models, building footprint binary masks were provided. Next, we perform experiments with a colour palette conditioned variant of the encoder-decoder and U-Net architectures (Figure 1). Each experiment was run for 500 epochs with the FID score recorded every fifth epoch to reduce computation time. All images were resized to 128×128 pixels. All networks were trained using the Adam optimizer with a learning rate of 1e-3. Each experiment was run 5 times recording the observed means and standard deviations throughout the 5 runs (Table 2).

We explored the use of the generated baselines for dataset expansion. However, due to space restrictions, only a brief summary is presented at the end of the paper.

3.4 Model Evaluation

The Pal-GAN architecture was evaluated using three different evaluation metrics: Fréchet inception distance (FID), intersection-over-union (IOU), and Learned Perceptual Image Patch Similarity (LPIPS).

The Fréchet inception distance (FID) [30] was calculated over the testing set to quantitatively evaluate the quality of the generated images. The FID metric compares the statistics of generated samples to the statistics of real samples using the Fréchet distance between two multivariate Gaussians. It is defined as:

$$\text{FID} = ||\mu_r - \mu_g||^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}) \quad (1)$$

where r and g represent the Inception-V3 [31] activations obtained from an intermediate layer for the real and generated images respectively. The mean and the covariance of the activations are represented by μ and Σ respectively. A lower FID score suggests that the activations between the real and the generated images are similar. This corresponds to an improvement in generated images both in terms of quality and variety.

Conditional consistency IOU is an explicit way of measuring how closely the generated image follows the given context. To evaluate the IOU score, we segment our generated image using our best segmentation baseline network and calculate the IOU between the results of the segmentation and the provided condition. Our best segmentation network trained on the full dataset provided an IOU of 0.5561 and will be used as the target IOU score.

LPIPS measures the diversity of generative models. LPIPS is calculated by generating multiple images using the same condition and calculating the average embedding distance between each generated image. Higher average distances correspond to more diversity in the generated images [32].

3.5 Experiments

Beginning with image generation experiments, we first establish baseline results by testing simple unconditional and mask conditioned GAN models. Our goal is to improve the textures and rep-

Model	FID (↓)	LPIPS (↑)	Conditional Consistency (↑)
Unconditional	202.3188 ± 53.4466	-	-
Encoder-Decoder	198.9274 ± 9.6079	0.0943 ± 0.0101	0.0479 ± 0.0014
U-Net	202.1443 ± 11.7878	0.0866 ± 0.0123	0.2065 ± 0.0296
Pal-GAN Encoder-Decoder	175.4146 ± 9.4739	0.1199 ± 0.0205	0.0497 ± 0.0058
Pal-GAN U-Net	111.3036 ± 5.6897	0.1165 ± 0.0057	0.2996 ± 0.0639

Table 2: Baseline FID, LPIPS, and conditional consistency IOU results for the unconditional model and both conditional models.
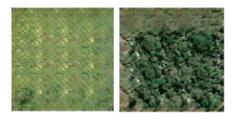

Fig. 2: Generated examples from the unconditional GAN

Expansion(%) \ Model	U-Net	SegNet
Baseline	0.4507 ± 0.0042	0.4468 ± 0.0052
10	0.4565 ± 0.0040	0.4502 ± 0.0075
30	0.4607 ± 0.0045	0.4455 ± 0.0031
50	0.4669 ± 0.0019	0.4511 ± 0.0072
100	0.4771 ± 0.0034	0.4401 ± 0.0031

Table 3: IOU performance on the small dataset with different expansion factors. Higher IOU corresponds to better results.

resentational capability of these architectures by injecting colour information into the latent space to guide the image generation. To test if injecting colour information improves the quality of generated images, we add a colour palette of size 8 to both conditional architectures.

We conduct additional visual experiments to test edge cases and explore other properties that the Pal-GAN architecture may have. These tests include transferring palettes between areas of interest, brightening and darkening colour palettes, and shifting the hues of the colour palette.
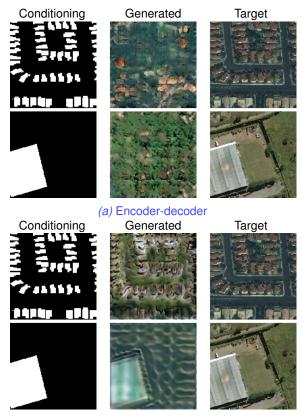
Finally, we test the effectiveness of our image generation in a downstream building segmentation task. In this set of experiments, we first begin by re-establishing a Pal-GAN baseline, and baselines on the U-Net and SegNet architectures using half of the available training data to simulate a low data environment. We continue by generating data using Pal-GAN and creating datasets 10%, 30%, 50% and 100% larger than the small dataset, where 100% expansion corresponds to a dataset of the same size as the original SpaceNet dataset. We expect the 100% expanded small SpaceNet dataset to produce similar results to the baseline segmentation models trained on the full dataset. The generation was only done once before training to increase training speed. However, a more robust generation during training method may also be used.

4 Results

We evaluate the images generated by our model and compare them to baseline GAN architectures. Next, we conduct exploratory experiments to assess the behaviour of different palette configurations imposed during image generation. Finally, we discuss dataset expansion results.

4.1 Image Generation

For traditional GAN architectures, image generation using low-information conditions can be a challenging problem to solve. In the case of the SpaceNet dataset, binary building masks only provide information on building shape, size, and location, along with spatial relationships between different buildings. Compared to other possible conditions, e.g. a map, there are many other missing but relevant features. For instance, maps provide a layout of roads, location of parks and other large objects. Maps may also contain fine-grained information such as the location of sidewalks, type of vegetation, and bodies of water. The outlined fine-grained features greatly improve the simplicity of the problem as the network

(a) Encoder-decoder

(b) U-Net

Fig. 3: Generated test examples from encoder-decoder and U-Net conditional baseline models.

no longer needs to model complex relationships between objects implicitly. To establish a comparable baseline, we first train traditional models conditioned on the building footprint mask exclusively. Our baselines are an unconditional GAN, a simple encoder-decoder based GAN, and a U-Net based GAN. The unconditional GAN model produces repeating patterns rather than realistic satellite images (Figure 2). It is also more likely to exhibit mode collapse.

The conditional baseline models show an improvement in perceived image quality observed through more building-shaped and building-coloured generations. The encoder-decoder style architecture maintains a similar texture generation property as the unconditional GAN (Figure 3a). Note, however, that there are portions of the image that demonstrate more structured outputs in the form of shrubbery and square reddish boxes. The U-Net baseline shows a strong improvement in the representation of the buildings, though much of the texture detail is lost (Figure 3b). The ability of the U-Net architecture to represent buildings is apparent in the 5× improvement in the conditional consistency IOU score when compared to the encoder-decoder architecture (Table 2, top).

The addition of colour information in the generation step greatly improves the FID results (Table 2, bottom). For the U-Net architecture, a large conditional consistency increase is also seen. Finally, a small diversity increase can also be observed. The encoder-decoder model maintains a similar quality to the mask conditioned model. However, the colours contained in the image more closely resemble the colours expected (Figure 5a). The Pal-GAN U-Net model shows a large improvement compared to the mask conditioned U-Net model, as evidenced by i) the more realistic colours of the buildings, ii) the reduction of artifacts, and iii) the presence of elements such as roads (Figure 5b).

Fig. 4: Image generation as the brightness is decreased or increased. The middle column, in red, contains the generation with the true colour palette.
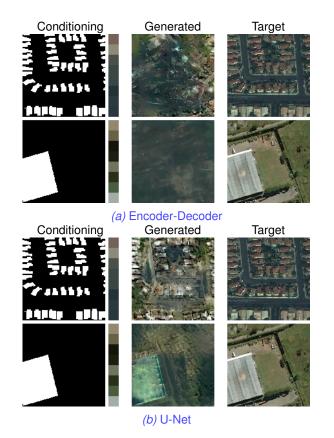

(a) Encoder-Decoder

(b) U-Net

Fig. 5: Generated test examples from both the 8 colour Pal-GAN models.

4.2 Colour Experiments

Now that it has been shown that colour palettes improve the quality of image generation, we experiment with how much control the palette provides.

4.2.1 Palette Transfer

In the palette transfer experiment, we tested the effects of using a mask from one area of interest and a colour palette from a different area of interest. We found that several areas of interest have similar colour palettes. When a mask is paired with a similar palette from a different area of interest, the quality of the generated image is significantly better than those from an area of interest with a different palette distribution. Furthermore, the fewer buildings the mask tends to contain, the closer the palette colours are followed.

4.2.2 Brightening and Darkening

To test how the network behaves when out-of-distribution palettes are provided, we modulate the brightness. Brightening and darkening greatly influences the generated image (Figure 4). The generator tends to produce more foliage as the palette is lightened. The

palette is more precisely followed when there are fewer buildings in the mask.

4.2.3 Hue Shifting

Next, we tested out-of-distribution hues. Hue shifting to the left corresponds to a counter-clockwise rotation of the hue in HSL space while shifting to the right corresponds to a clockwise rotation. Shifting the hue left towards more within-distribution palettes increased the quality of the generated images. Alternatively, a shift right provided more out-of-distribution palettes resulting in degraded image generation. However, when the mask contains many buildings, out-of-distribution detection style behaviours can be seen. Once the colour palette is out-of-distribution, the generator defaults to the most common palette and produces a more realistic image without following the palette. When the network is provided with random colour palettes, this same out-of-distribution behaviour occurs.

4.3 Dataset Expansion

We found that U-Net had a small improvement over the small baseline results, reducing the gap between the small and original network by 25%. The SegNet architecture did not see an improvement (Table 3). We interpret these results as showing potential for use in dataset expansion. Commissioning cheaper annotation labels and converting them into usable data is a power factor. Taken further, learning to generate the simpler mask conditions and generating completely new data from scratch in a two-stage data generation process is also possible. However, more research is needed to explore these ideas.

5 Conclusion and Future Work

This paper introduces a new variant of GANs that explicitly uses factored out colour information and provides it to the generator. Pal-GAN allows for the production of realistic remote sensing images, lowering the cost of creating large scale datasets. The Pal-GAN approach is architecture agnostic and does not require any additional information to be explicitly labelled or gathered. We test our approach on both a classical encoder-decoder model and a U-Net style architecture. In both cases, additionally conditioning on a colour palette significantly improved FID scores and the visual quality of the images. Colour palette conditioning also improves the diversity of generated samples. Furthermore, the U-Net style architecture showed significant improvement in conditional consistency. Pal-GAN contains out-of-distribution detection behaviours and may default to more appropriate colour palettes if enough information is contained in the binary mask.

Further exploration of various colour palette organization techniques may provide more structured information to the generator. Currently, only one palette may be specified per image. Thus, exploring more versatile colour conditioning to allow for more nuanced palette control is also needed. More experiments are required to explore the limits of the out-of-distribution detection. The current architecture does not explicitly penalize deviations from the provided palette. Stricter ways of enforcing colour palettes within the generated image also warrant further investigation.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, vol. abs/1502.01852, 2015. [Online]. Available: http://arxiv.org/abs/1502.01852

[2] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242–2251, 2017.

[3] A. Antoniou, A. J. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," *ArXiv*, vol. abs/1711.04340, 2017.

[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, 2016.

[5] U. Demir and G. B. Ünal, "Patch-based image inpainting with generative adversarial networks," *ArXiv*, vol. abs/1803.07422, 2018.

[6] Y. Bengio, F. Bastien, A. Bergeron, N. Boulanger-Lewandowski, T. Breuel, Y. Chherawala, M. Cissé, M. Côté, D. Erhan, J. Eustache, X. Glorot, X. Muller, S. P. Lebeuf, R. Pascanu, S. Rifai, F. Savard, and G. Sicard, "Deep learners benefit more from out-of-distribution examples," in *AISTATS*, 2011.

[7] G. E. Hinton, A. Krizhevsky, and I. Sutskever, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1106–1114, 2012.

[8] "Deep image: Scaling up image recognition," *CoRR*, vol. abs/1501.02876, 2015, withdrawn. [Online]. Available: http://arxiv.org/abs/1501.02876

[9] T. Devries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *CoRR*, vol. abs/1708.04552, 2017. [Online]. Available: http://arxiv.org/abs/1708.04552

[10] V. Sandfort, K. Yan, P. Pickhardt, and R. Summers, "Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks," *Scientific Reports*, vol. 9, 11 2019.

[11] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," *CoRR*, vol. abs/1612.07828, 2016. [Online]. Available: http://arxiv.org/abs/1612.07828

[12] O. Bailo, D. Ham, and Y. M. Shin, "Red blood cell image generation for data augmentation using conditional generative adversarial networks," *CoRR*, vol. abs/1901.06219, 2019. [Online]. Available: http://arxiv.org/abs/1901.06219

[13] H. Uzunova, J. Ehrhardt, F. Jacob, A. Frydrychowicz, and H. Handels, "Multi-scale gans for memory-efficient generation of high resolution medical images," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, Eds. Cham: Springer International Publishing, 2019, pp. 112–120.

[14] Y. Shi, Q. Li, and X. X. Zhu, "Building footprint generation using improved generative adversarial networks," *CoRR*, vol. abs/1810.11224, 2018. [Online]. Available: http://arxiv.org/abs/1810.11224

[15] M. Mirza and S. Osindero, "Conditional generative adversarial nets. arxiv 2014," *arXiv preprint arXiv:1411.1784*, 2014.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[17] R. A. Yeh, C. Chen, T.-Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6882–6890, 2016.

[18] C. Burlin, Y. Le Calonnec, and L. Duperier, "Deep image inpainting," 2017.

[19] S. E. Reed, Z. Akata, S. Mohan, S. Tenka, B. Schiele, and H. Lee, "Learning what and where to draw," *ArXiv*, vol. abs/1610.02454, 2016.

[20] S. E. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," *ArXiv*, vol. abs/1605.05396, 2016.

[21] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, pp. 778–782, 2017.

[22] M. Pritt and G. Chern, "Satellite image classification with deep learning," in *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2017, pp. 1–7.

[23] C. Zhang, I. Sargent, X. Pan, H. Li, A. Gardiner, J. Hare, and P. M. Atkinson, "Joint deep learning for land cover and land use classification," *Remote Sensing of Environment*, vol. 221, pp. 173 – 187, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0034425718305236

[24] Q. Jiang, L. Cao, M. Cheng, C. Wang, and J. Li, "Deep neural networks-based vehicle detection in satellite images," in *2015 International Symposium on Bioelectronics and Bioinformatics (ISBB)*, 2015, pp. 184–187.

[25] M. Kartal and O. Duman, "Ship detection from optical satellite images with deep learning," in *2019 9th International Conference on Recent Advances in Space Technologies (RAST)*, 2019, pp. 479–484.

[26] T. G. J. Rudner, M. Rußwurm, J. Fil, R. Pelich, B. Bischke, V. Kopačková, and P. Biliński, "Multi3net: Segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 702–709, Jul. 2019. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/3848

[27] S. Milz, T. Rüdiger, and S. Süss, "Aerial ganeration: Towards realistic data augmentation using conditional gans," in *ECCV Workshops*, 2018.

[28] J. Cho, S. Yun, K. Lee, and J. Y. Choi, "Palettenet: Image recolorization with given color palette," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1058–1066, 2017.

[29] A. V. Etten, D. Lindenbaum, and T. M. Bacastow, "Spacenet: A remote sensing dataset and challenge series," *ArXiv*, vol. abs/1807.01232, 2018.

[30] 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," in *NIPS*, 2017.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, 2015.

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