FloodNet-to-FloodGAN : Generating Flood Scenes in Aerial Images

Shubham Goswami, Sagar Verma, Kavya Gupta, Siddharth Gupta

FloodNet-to-FloodGAN : Generating Flood Scenes in Aerial Images

Shubham Goswami
Granular AI
shubham@granular.ai

Sagar Verma
Granular AI
sagar@granular.ai

Kavya Gupta
Université Paris-Saclay, CentraleSupélec, Inria, Centre de Vision Numérique
kavya.gupta100@gmail.com

Siddharth Gupta
Granular AI
sid@granular.ai

Abstract

A global rise in the occurrences of natural disasters and human-borne conflicts has put a spotlight on the need for Earth Observation (EO) data in designing practical Humanitarian Assistance and Disaster Relief (HADR) interventions. Novel techniques that leverage remotely sensed data are leading to a paradigm shift in our understanding of such situations and improving the efficacy of our response. Aerial flood maps can provide localized insight into the extent of flood-related damage and the degree to which communities' access to shelter, clean water, and communication channels have been compromised. Unfortunately, such insights typically only emerge hours or days after a flooding event has occurred. Moreover, a dearth of available historical data restricts the development of practical machine learning based methods. This work examines the use of Generative Adversarial Networks (GANs) in simulating flooding in aerial images. We first introduce the *Houston UAV dataset*, an extension of the FloodNet dataset. Our dataset accommodates more well-defined semantic classes and significantly reduces the label noise in semantic masks. We propose a GAN-based pipeline to generate flood conditions in non-flooded regions, generating synthetic flooding scenes for predictive mapping. Code and dataset are available at https://github.com/granularai/flood-synthesis.

1 Application Context

In September of 2022 alone, the world saw 27 flood events, killing an estimated 1500 people and displacing millions from their homes. While much has been done to advance our understanding of when and where floods will occur, less progress has been made in understanding the nature and extent of the destructive impact of such events. Analysis of earth observation data can improve pre-flooding preparedness, situational readiness, and efficient post-flooding intervention. As such, evaluating post-flooding imagery is a crucial part of the HADR. Accurate evaluation helps organizations and governments to manage post-flooding humanitarian efforts like providing food, shelter, and financial aid. Relief-providing organizations must also know if impacted population segments have access to an alternate energy source, information, and mobility. Satellites and unmanned aerial vehicle (UAV) based imagery have emerged as the preferred sources for driving such evaluation and information extraction due to their relative availability, ease of tasking, and price.

Preprint. Under review.

In this paper, we demonstrate the generation of new training samples via a Generative Adversarial Neural network (GAN) to enhance inference performance for semantic segmentation tasks in post-flood scenarios. Scenes are generated to represent an aerial view of flooded residential properties. Within these scenes, features such as solar roofs, antennae, and vehicles are generated to provide additional context for the downstream segmentation tasks. Our main objective is to create flooded regions for the scenes comprising residential properties with objects like solar roofs, dish antennas, and vehicles. We have cleaned the label noise in the semantic masks and extended the FloodNet dataset [1] with more semantic class labels focused on HADR usability. We term this dataset *Houston UAV dataset* and use this dataset for training our GAN-based pipeline (FloodGAN), more precisely Pix2Pix GAN, for generating flooding zones in the residential areas.

2 Introduction

Data collection in case of disasters is tricky, and choosing the right data source is essential. The two most favorable data sources are satellites and UAVs. Satellite sources can suffer from bad weather, which is highly probable to happen during flooding. UAVs can be deployed in such cases but can not cover large areas. Data quality is often poor in post-disaster scenarios due to bad weather conditions, deformation of objects, submerged objects, and clutter. UAV has more mobility freedom and can capture data closely with high quality. Annotation quantity is constrained due to limited data collected and the fact that annotating post-disaster images in comparison to pre-disaster takes more time and effort to maintain quality. Satellites can capture data for a particular region for a more extended period than a UAV. However, this can not be achieved if weather conditions are bad. On the other hand, UAVs have limited flight time and the region they can capture. Using multiple UAV(s) can remove the cons associated with flight time and the area captured. Annotation quality suffers due to inadequate data quality, which can add more noise to the ground truth. Satellite-based data for both pre/post-disaster scenarios is more challenging to annotate than UAV-based data owing to resolution, identifying clutter, and marking object classes with precision. UAVs as a data source can serve better in evaluating flood-based disaster scenarios.

Recent advances in deep learning based algorithms, especially in computer vision, have shown potential in enabling rapid downstream analysis of such imagery over large areas. Deep learning based methodologies are well positioned to assist in humanitarian efforts due to their scalability, explainability, and robustness to variable conditions. In the short-term, such models can be employed to provide decision-makers with enhanced clarity on ground conditions. Moreover, these models can be extended to support autonomous assistive solutions. One of the critical requirements for generating such deep learning based solutions is the availability of high-quality training data, which poses a challenge due to the inherent dearth of post-flooding data. Attempting to train with insufficient data can lead to the model learning a poor representation due to the noise in the annotations. Generating new samples from existing data with negligible label noise can enable us to overcome such challenges.

3 Related work

There are various ways of segmenting flooded images into useful semantic maps. They range from texture-based feature learning [2] to neural network-based solutions. Recent literature on flood detection has witnessed the incremental use of GAN. GANs [3] are usually comprised of a generator and discriminator, where the generator learns to generate samples similar to actual data sample distribution, and the discriminator tries to classify them as real or fake data samples. GANs are widely used as a data generation technique that enables augmentation [4], super-resolution [5], etc. For the HADR, GANs are now being explored for weakly-supervised [6] and unsupervised domain adaptation [7]. Conditional GANs [8] is a variant of GAN where the generation of new samples is conditioned by the input of the generator [9, 10]. In generating new edge case post-flooding scenes, Conditional GAN can be used for the semantic mask to image translation, e.g., Pix2Pix [11] or domain adaptation CycleGAN[12]. DisasterGAN [13] utilizes CycleGAN and Pix2Pix to generate post-disaster images. Edge case disaster scenes generated by DisasterGAN sometimes completely differ from pre/post-disaster samples. This is because disentangled object classes have not been incorporated in the image translation for paired or unpaired images. Operating directly in RGB color space can lead to mode collapse, as shown in Appendix Fig 5.

4 Dataset

4.1 FloodNet Dataset

Rahnemoonfar et al. [1] introduced the FloodNet dataset comprising of high resolution (6cm) UAV images capturing post-hurricane Harvey in Houston, Texas, USA. The dataset contains 400 labeled images for semantic segmentation task with labels for eight object classes: 'Building-flooded', 'Building-non-flooded', 'Road-flooded', 'Road-non-flooded', 'Vehicle', 'Pool', 'Tree' and 'Water'. Floodnet dataset is available on GeoEngine [14, 15].

4.2 Houston UAV: Extending FloodNet

Our need for semantic classes like vehicles, solar roofs, swimming pools, trampolines, etc. is not satisfied in the FloodNet dataset. It only captures eight semantic classes making it unsuitable for generating post-disaster(flood) scenarios. Following are some other limitations in the FloodNet dataset:

1. **Label noise:** There is a significant presence of label noise in the semantic mask. Appendix Fig 6 shows the distribution of incorrect and missing polygons in each object class in the FloodNet dataset. This can negatively affect GAN training, mainly when working with insufficient data.
2. **Fewer semantic classes:** In order to generate new images similar to the provided images, more object classes should be captured. This is crucial to generate a dense mapping from a simulated semantic mask to an image. We must also have semantic labels informing HADR organizations about electricity, information, and mobility access.
3. **Absence of flood mask:** Flood mask is only available for 'Building' and 'Road,' which does not correctly capture the overall region affected by flood in a given data sample. For example, it would be good to know if electrical poles, swimming pools, trampolines, etc are in the flooded or non-flooded area.

We address the above issues by extending the FloodNet dataset into the Houston UAV dataset. We have corrected the labels for existing masks. For example, some non-flooded roads were assigned 'Road-flooded', and we changed them to 'Road.' We have also separated flooded polygons from semantic classes, and now we have 'Road', 'Building', 'Tree', etc. If flooding covers these parts of the image, then an overlapping polygon 'Flooded' is present. We have added several new semantic class labels. After these improvements, the final set of 20 semantic classes is: i) Property Roof, ii) Secondary Structure, iii) Swimming Pool, iv) Vehicle, v) Grass, vi) Trees / Shrubs, vii) Solar Panels, viii) Chimney, ix) Street Light, x) Window, xi) Satellite Antenna, xii) Garbage Bins, xiii) Trampoline, xiv) Road/Highway, xv) Under Construction / In Progress Status, xvi) Power Lines & Cables, xvii) Bridge, xviii) Water Tank / Oil Tank, xix) Parking Area - Commercial, xx) Sports Complex / Arena, xxi) Industrial Site, xxii) Dense Vegetation / Forest, xxiii) Water Body, xxiv) Flooded, xxv) Boat, and xxvi) Parking Area . Houston UAV dataset is available on GeoeEngine.

5 Proposed Pipeline

Houston UAV dataset enables us to exploit Pix2Pix GAN to simulate post-flooding scenarios. The architecture of FloodGAN is shown in Fig 1. Pix2Pix typically uses an RGB/gray-scale color mask. We have introduced RGB semantic mask to make distinct object classes independent from each other. Semantic masks keep the object feature space disentangled, leading to object labels being equally distant. Furthermore, it enables better control over the conditional generation by using a separate channel for the flood, which can overlap with other objects we want to be flooded. Using UNet [16] as the generator, has an added advantage of skip connections, which helps generate feature-rich images. To enhance the quality of house in generated image, the RGB mask of buildings is provided as a part of the input (semantic mask) to UNet. This particular modification enhances the quality of output and convergence of GAN training. During inference, a semantic mask of the pre-flooded image with a separate channel for the pixels we want to be flooded is given as input. Fig 2 shows the inference pipeline.

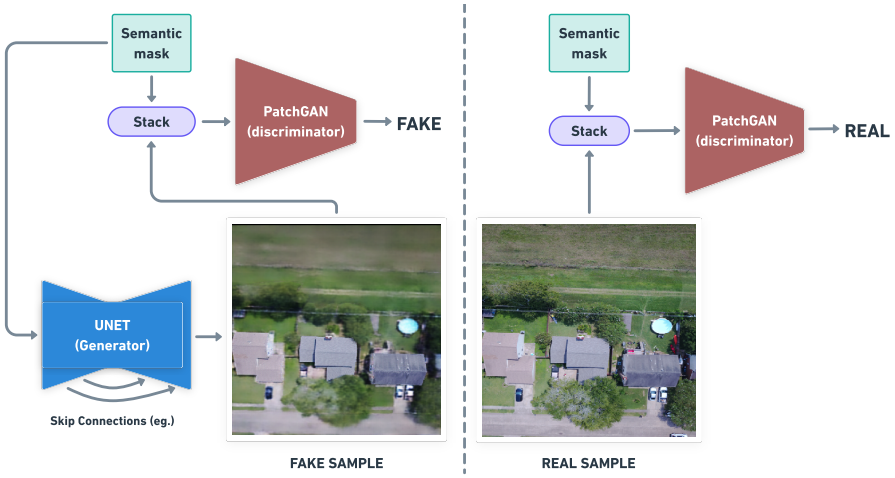

Figure 1: **FloodGAN** : Training Conditional GAN Pix2Pix to generate a new image from a given semantic mask. UNet and Patch-GAN are used as the generator and discriminator networks, respectively.

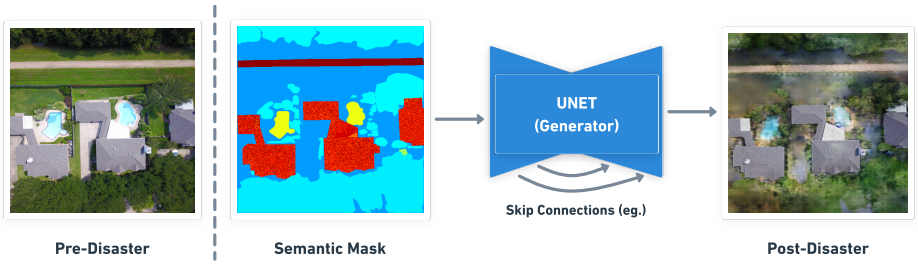

Figure 2: Generating post-flooded image using semantic mask of pre-flooded image.

In Pix2Pix, the objective for training is similar to a usual conditional GAN training:-

$$L_{\text{cGAN}}(G, D) = \mathop{\mathbb{E}}_{x,y}[\log D(x, y)] + \mathop{\mathbb{E}}_{x,z}[\log(1 - D(x, G(x, z)))] \tag{1}$$

$$G^* = \arg\min_G \max_D L_{\text{cGAN}}(G, D) + \lambda L_{\ell_1}(G) \tag{2}$$

Where Eq.2 is the training objective and x, y, and z are the input (semantic mask), ground truth image, and noise (dropout) for stochasticity, respectively. G is the conditional generator UNet, and D is the discriminator Patch-GAN [17]. λ is the regularization parameter controlling the effect of ℓ_1 loss.

6 Experiments

Houston UAV dataset contains 374 images [1] containing flood and non-flood scenarios. We split this into a training set comprising 262 samples and a hold-out set comprising 112 samples, selected in a stratified manner as determined by object count. For training, we use the train set with an image size of 512x512 with standard flip augmentation; Batch norm is used with no dropout. $UNet - 256$ architecture and discriminator $Patch - GAN$ (8 strided convolution-2D layers) are employed for the generator. We use 2 Amp RTX 3090 GPU and 2 Intel Xeon Silver 4208 CPU for training. Training ran for approximately 48 hours.

[1] Houston UAV has slightly fewer images because the original version of FloodNet has many redundant images in the form of successive frames captured within a few milliseconds of the exact location. To keep annotation inexpensive, we choose distinct scenes and annotate them.

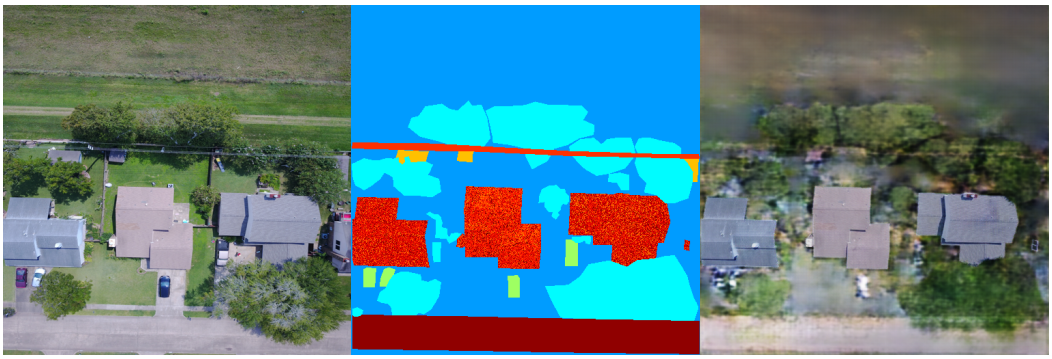

Figure 3: Flooded image generated from an image taken from higher altitude.

Fig 3 shows sample predictions on the test image of Houston UAV. Left column shows the non-flooded real image, the middle column shows the semantic mask of this real image, and on the right column, flooded images are generated predictions using our proposed pipeline. We condition grass, parking area, roads, and footpath mask channels in the input to the generator with random noise to simulate the flood in those areas. This scenario will simulate a low flooding situation in which only the ground is covered with water. Other examples can be found in Appendix Fig 7.

Figure 4: First image in the top column shows real image. Five other images show synthetic flooding at different locations with different coverage.

We can add a condition on mask channels like swimming pools, shrubs, vehicles, garbage bins, and power lines to synthesize an image where flooding is intense, and water has risen to meters. By controlling the location and size of the condition mask, we can even generate flooding images with different intensities. Fig 4 shows different synthetic flooded images generated from a single image where floods are of different sizes and occur in different locations.

7 Discussions

We observe that the generated post-flooded images are slightly blurred. Most of the blur is owed to ℓ_1 loss, which we try to counter using Patch-GAN. This leads to some degree of sharpness. A smaller patch size leads to rich fine-grained sharpness, and a large patch size leads to rich coarse features with blurred fine-grained details. Blurring in the generated post-flooded images can be minimized using multiple Patch-GANs operating with multi-scale patch size configuration. Such methods will help in training better models and hence can anticipate both pre- and post-disaster scenarios and increase preparedness in case of needed Humanitarian assistance.

References

[1] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. Murphy, "Floodnet: A high resolution aerial imagery dataset for post flood scene understanding," *IEEE Access*, 2020.

[2] D. Popescu, L. Ichim, and T. Caramihale, "Flood areas detection based on uav surveillance system," in *International Conference on System Theory, Control and Computing*, pp. 753–758, 2015.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *NeurIPS*, vol. 27, 2014.

[4] S. Mounsaveng, D. Vazquez, I. B. Ayed, and M. Pedersoli, "Adversarial learning of general transformations for data augmentation," *arXiv preprint arXiv:1909.09801*, 2019.

[5] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, pp. 4681–4690, 2017.

[6] J. Iqbal and M. Ali, "Weakly-supervised domain adaptation for built-up region segmentation in aerial and satellite imagery," *Photogrammetry and Remote Sensing*, vol. 167, pp. 263–275, 2020.

[7] B. Benjdira, Y. Bazi, A. Koubaa, and K. Ouni, "Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images," *Remote Sensing*, vol. 11, no. 11, p. 1369, 2019.

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

[9] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *CVPR*, pp. 8789–8797, 2018.

[10] G. Zhang, M. Kan, S. Shan, and X. Chen, "Generative adversarial network with spatial attention for face attribute editing," in *ECCV*, pp. 417–432, 2018.

[11] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *CVPR*, pp. 1125–1134, 2017.

[12] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, pp. 2223–2232, 2017.

[13] X. Rui, Y. Cao, X. Yuan, Y. Kang, and W. Song, "Disastergan: Generative adversarial networks for remote sensing disaster image generation," *Remote Sensing*, vol. 13, no. 21, p. 4284, 2021.

[14] H. Shin, N. Exe, U. Dutta, T. R. Joshi, S. Verma, and S. Gupta, "Europa: Increasing accessibility of geospatial datasets," in *IGARSS*, IEEE, 2022.

[15] S. Verma, S. Gupta, H. Shin, A. Panigrahi, S. Goswami, S. Pardeshi, N. Exe, U. Dutta, T. R. Joshi, and N. Bhojwani, "Geoengine: A platform for production-ready geospatial research," in *CVPRD*, pp. 21416–21424, June 2022.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, 2015.

[17] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *ECCV*, pp. 702–716, Springer, 2016.

A Appendix

Figure 5: Mode collapse when CycleGAN is used to generate flooded and non-flooded images. **Top-Left:** real non-flooded scene, **Bottom-Left:** fake flooded scene, **Top-Right:** real flooded scene, and **Bottom-Right:** fake non-flooded scene. In this case CycleGAN is used to flood a grassland and remove flood from a residential area, due to mode collapse, buildings appear in flooded grassland. Similarly building are removed along with flood and is converted to grassland.

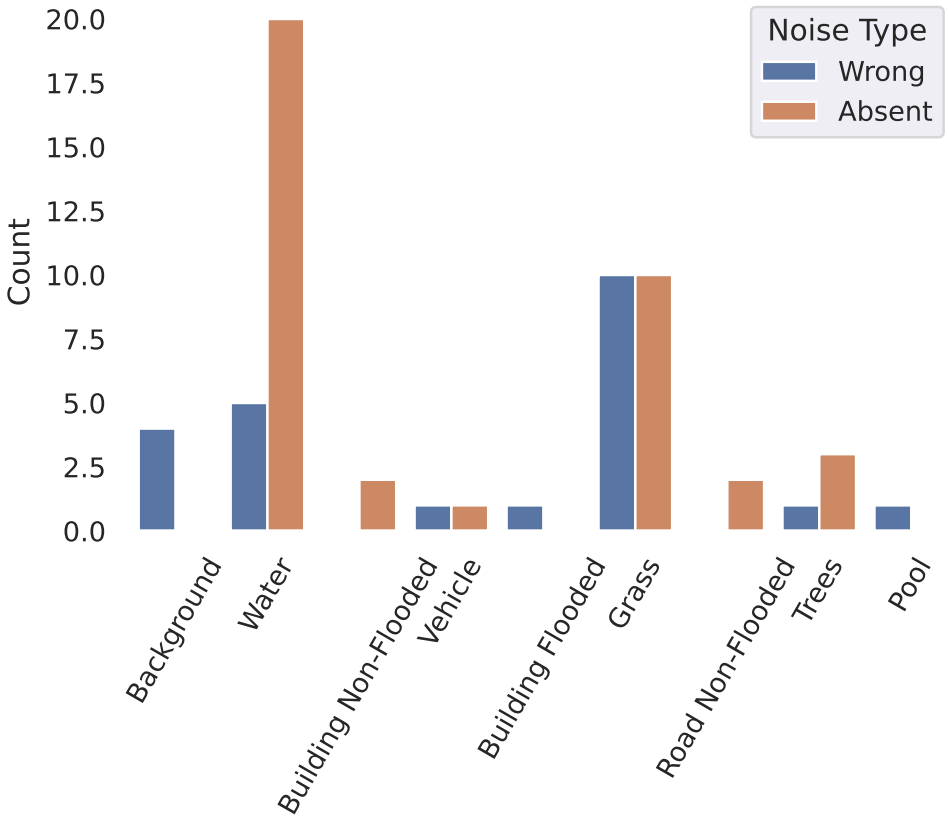

Figure 6: Label noise distribution in FloodNet Challenge out of 400 available labeled images

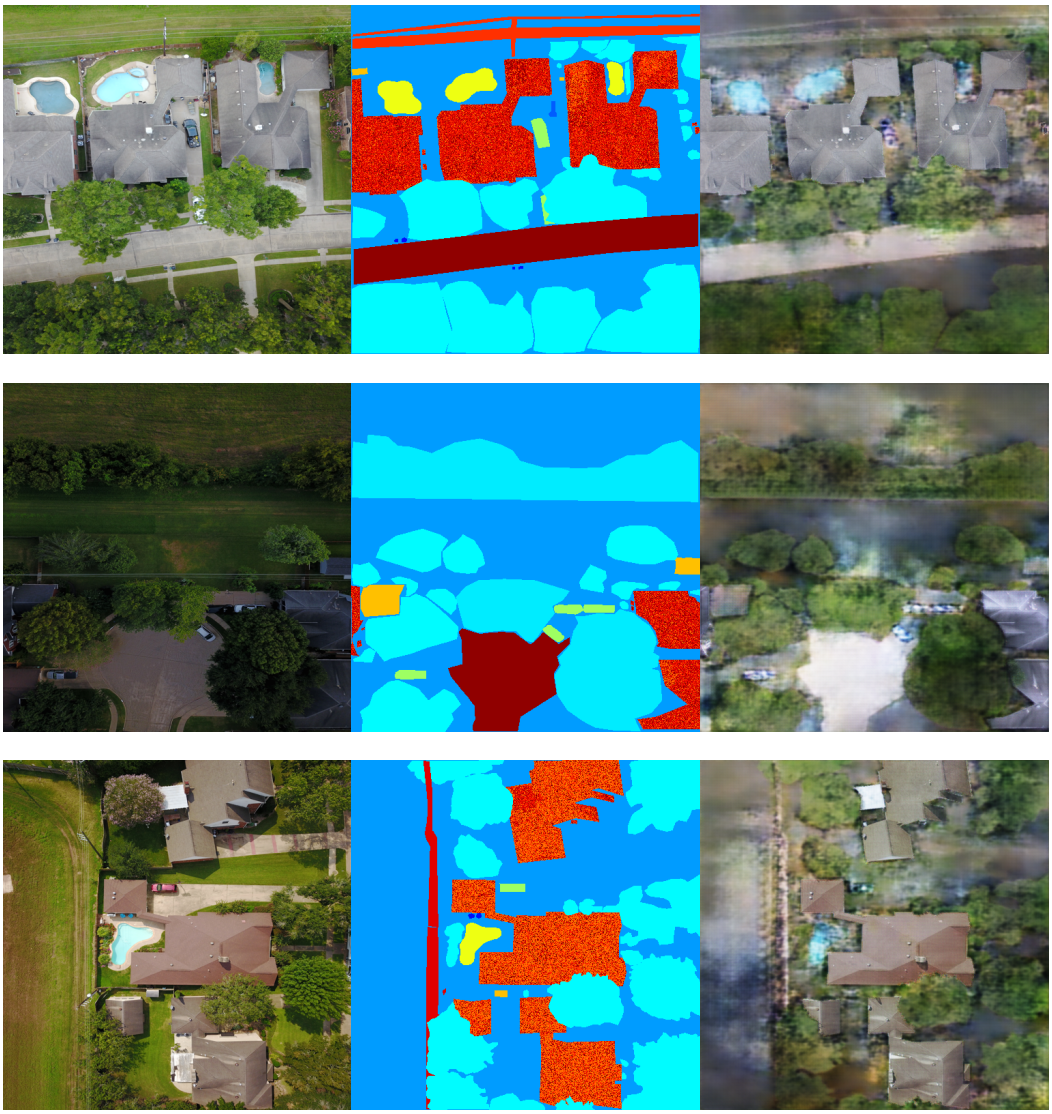

Figure 7: Inference results, left-to-right: pre-flooded, semantic mask with simulated flood channel, and generated post-flooded image.