**TCENet: Texture-Color Enhancement Network in Synthetic Video Generation**

**Walia Farzana**

**Abstract:**

Flood water and car segmentation is required in floodwater depth estimation which is critical component of flood monitoring system. A large-scale annotated real flood dataset required to train segmentation models. To handle data scarcity challenge, synthetic generated videos of flood can be utilized to train deep learning segmentation model and tested on real flood images. However, because of the domain gap between synthetic and real the model trained on synthetic flood image underperformed when tested on real flood data. To address the gap between real and synthetic flood image, we proposed GAN based texture color enhancement network TCENet to enhance the texture and color of synthetic flood image similar to real flood image. The TCENet comprised of enhancement network and segmentation network. The enhancement network is trained on unpaired synthetic and real flood images to enhance the texture and color of the synthetic images. Finally, the enhanced synthetic images are provided as input to segmentation model and the segmentation model is tested on real flood image to segment flood water and car. To best of our knowledge, for the first time in the literature, this work trained a semantic model on enhanced synthetic flood images and tested on real flood image data. The proposed method can be applied to semantic segmentation of flood water and car with unpaired synthetic and real flood data in absence of large real flood image dataset.

**Introduction**

One of the most catastrophic natural disasters worldwide is flooding. Flooding risk is the outcome of intricate interactions between hydrological risks (such as riverine flooding due to heavy rain), exposure, vulnerability (such as the possibility of structure damage or human casualties) [1]. According to one research, the rising sea level will make repeated flooding typical for many US coastal communities by 2050, and it will probably happen thirty or more days a year [2]. Weather-related disasters are becoming more frequent and managing flood risk has become a major issue for many countries. For assessing flood danger and damage, image data from many sources, for instance satellites, smartphones, drones, and surveillance cameras, has become an invaluable resource [3]. The majority of earlier flood detection literature concentrated on employing distant sensing technologies like satellites and airplanes [6]. Camera-based approaches employing computer vision are far less costly than typical sensor-based systems [7]. Studies has addressed the issue satellite imagery and digital elevation models (DEM) [8] but there are few studies [9] that employed local street view for flood images.

Flooding causes inconvenience for motorists, especially if there is little information accessible about their route. The measurement of floodwater depth is a necessary component of a flood monitoring system. In pre-existing flood dataset [5] do not translate well in real world scenarios as there are distinct images and no temporal relation between the images [9] compared to generated video frames. In real scenario, a live video broadcast from traffic cameras would be the input for a real-time flood detection system.

Researchers are developing computer vision techniques using machine learning and deep learning to detect flood water on roadways [4] and detecting flood water depth [5]. Recent years have seen a significant growth in the application of ML for flood water detection. However, most of these techniques still require extensive preprocessing of the data, which varies depending on the region [10]. The absence of large-scale annotated datasets reflecting complex sceneries with fluctuations in lighting, reflections, floodwater ripples, etc. is a major obstacle for image-based algorithms for flood monitoring [9]. To address the challenge, one of our research works [9] propose a GAN-based Synthetic Urban Recurrent Flooding (SURF) to produce synthetic video pictures of flooding on streets. The suggested SURFGenerator is unique because it can create flood sceneries for videos instead of still photographs, captures complicated fluid dynamics (such ripples and waves due to driving cars on partially submerged highways), and creates floodwater of different depths. One of the promising approaches to facilitate research on a real-time flood management system that can alert motorists about floods and its extent is to train ML models with generated synthetic data. However, because to the domain gap between the synthetic and real datasets, models trained on the synthetic data underperformed when applied to the real datasets [11]. Using style transfer or cross domain mapping networks to stylize the synthetic domain images to real domain and train the segmentation models in this stylized space is one approach that might be used to address the domain gap challenge. To address the domain gap in synthetic and real flood data, we proposed texture color enhancement network (TCENet) in synthetic video for flood water and car segmentation. The proposed TCENet comprised of enhancement network and segmentation network. For enhancement network we have applied a Generative adversarial Network (GAN) based model with adversarial, identity and perceptual reconstruction loss that focuses on texture and color enhancements in synthetic images because the gap between synthetic and real image emerge from low-level difference such as color and texture. In the segmentation network, we experiment with different segmentation model and LinkNet with ResNet50 based encoder performed better compared to other segmentation models. The main contribution of this work is to align synthetic and real flood images in common feature space to enhance the color and texture of synthetic images and deploying a segmentation model that is trained on enhanced synthetic images and tested on real flood images.

The remainder of the paper is structured as follows: Section II discuss on background on understanding the enhancement network followed by segmentation network and relevant research work on flood water detection and segmentation. Section III present the overall pipeline utilized for enhancing synthetic image texture and color, training, and test of the models. Section IV present the results of segmentation network trained on enhanced synthetic data and tested on real flood data. Section V concludes and discuss our future work.

**2. Background:**

**2.1. Related Work**

Previous related works, namely [4], [5], [12], [13], [14] focused mainly on flood water detection and segmentation, level estimation from real images using extensive image pre-processing, feature extraction, applying deep learning models. One of the research works [13] focuses on introducing image dataset that is collected from different sources and evaluate the performance of the deep learning models with inclusion and exclusion of dataset. In the same line, the study [12] a sample dataset from nuisance flooding events in Norfolk, VA is created which contain location-based matched image. The proposed pipeline includes water edge detection, image inpainting and contrast correction for handling dynamic scenario in crowdsourced images. For flood level estimation work [14] a object detection Convolutional Neural Network (CNN) has been applied to detect a particular object in the image which is utilized as an input to Neural Net and merge the water level information from different object to achieve a single level from image. However, in the previous work the models are trained and tested using real crowdsourced flood image data. Therefore, the methods applied in the literature in not ideal case for our scenario. Since these models are trained and tested in real flood image domain and in some cases [12] location-based paired images are required which involve difference in image resolution, lighting, presence of dynamic objects.

In addition, it is a trivial task to obtain images from same location with and without presence of flood water considering the dynamic environment. Training the model with generated synthetic data might underperform when tested on real flood image data. In our proposed method, we tackle the domain gap between synthetic and real flood data utilizing enhancement network that does not require reference paired data and the segmentation network in trained enhanced synthetic flood data and tested on real flood data.

**2.2. Enhancement Network for Synthetic Image**

Flood water monitoring involves flood depth estimation which is dependent on flood water detection and segmentation. Previous research work focused on applying deep learning or machine learning to learn mapping between image features and segmentation mask comprising of real image and segmentation mask pair. An apparent limitation is requirement of large dataset of real images with corresponding segmentation mask. Due to the requirement for large dataset to train deep learning model, creating such expansive datasets for certain scene types is effort and cost intensive. To alleviate the requirement of large real image-segmentation mask pairs, generating synthetic 3D flooding scenario [9] to train deep learning with synthetic image for utilization on real flood images, particularly for scenario where synthetic data can easily be generated. Inspired by these methods, we have researched flood water and car segmentation method that utilized synthetic image-segmentation mask pairs instead of real paired data, but which also explores the the wide availability of unpaired real and synthetic flood image data. In summary our scenario is the following: we have limited availability of real flood image and corresponding segmentation mask. We also have access to large set of synthetic 3D flood scene [9] which can render multiple synthetic flood image from different viewpoint in urban street and their corresponding segmentation mask. The main goal is to segment flood water and car from real flood image.

Diagram

Description automatically generated

Figure 1: Possible approach to segmentation of flood water and car using synthetic image-segmentation mask pairs (**IS, YS**) pairs and real flood image **(IR). G** indicates generatorand **D** indicates discriminator**.** Please refer to text for details.

Considering the first two scenarios:

1. Training a segmentation model (SEG) using only synthetic image and segmentation mask, and hope that the model will generalize to real flood imagery. (Scenario 1 in Fig.1)
2. Using a two-stage pipeline in which synthetic flood images is mapped into real-image domain using a GAN, and then train the segmentation model. (Scenario 2 in Fig.1)

The limitation with scenario 1 it is improbable the segmentation model is indifferent to the difference between synthetic and real flood imagery. In scenario 2, while a GAN might stimulate synthetic images to map the distribution of real flood images, it does not specifically require the enhanced (synthetic to real) image (**ȊS**) to have physically correct correlation to its corresponding segmentation map, meaning the trained segmentation model will not perform effectively with real input. This somehow be mitigated by instigating a regularization loss to try and keep the enhanced synthetic (**ȈS** indicates converting the image **ȊS** back to synthetic domain using **GR>S**) image “similar in content” to its original synthetic images. By content we indicate geometric and semantic differences.

In this work we propose a pipeline where the trained segmentation model with enhanced synthetic image should work well when tested on real flood image. To attain the goal, we want to the synthetic-to-real enhancement network to behave as a close loop which is addressed by cycle consistent loss that allow images from either domain to be enhanced and then reconstructed. Basically, the cycle consistent loss comprised of adversarial loss inherited from GAN-based model that focus on match the distribution of enhanced images to real image domain and the perceptual reconstruction loss ensure the dual mapping knowledge is consistent with the images.

1. **Method:**

Our main goal is to train a segmentation (**SEG**) network on Fig.1 such that when presented with real flood image, it predicts corresponding flood water and car from the image accurately.

Regarding data accessibility, we assume that we have limited access to collection of real-world flood image **IR** with its corresponding segmentation mask. Instead, we have access to collection of 3D flood synthetic videos, from where a range of synthetic flood images and related segmentation masks can be rendered denoted as (**IS, YS**).

In contrast to directly training the **SEG** on the synthetic (**IS, YS**) data, we anticipate that the synthetic flood images will not be sufficiently comparable to the real flood images, necessitating the use of the earlier enhancement network to make synthetic image features comparable to real image feature domain. However, a regularized loss function is required to preserve the geometric and semantic content of the enhanced synthetic image. The motivation is **GS>R** implicitly learn to apply the minimum change to synthetic image to make it realistic and consider this the most reasonable technique to regularize a network while maintaining the segmentation relevant shape semantics. To achieve our goal, we propose a pipeline as shown in Fig.2, which we call texture color enhancement network (TCENet) to highlight the texture color enhancement of synthetic flood image that is utilized as input to the segmentation network.

Graphical user interface

Description automatically generated with medium confidence

Figure 2. The overall pipeline of TCENet.

The enhancement network indicates the enhancing the texture and color of synthetic flood images using unpaired synthetic (**IS**) and real flood images (**IR**) as input to the enhancement network. The segmentation network depicts the enhanced synthetic image **ȊS** with corresponding segmentation mask **YS** are input to the segmentation network which is tested on real flood images it will give segmentation mask **ŶS** .In addition to predict the segmentation mask on real flood images, we also evaluate the activation mapping of segmentation model using the layer just before the output layer to observe model prediction for flood car and water.

* 1. **Adversarial loss with Perceptual Reconstruction Loss**

In contrast to high-level geometric and semantic differences, low-level variations in color and texture (such as those seen in roads and trees) account for the apparent gap between synthetic and realistic images [15]. An ideal enhancement network, for usage with a segmentation architecture, requires producing images that cannot be discriminated from real images while retaining the original scene geometry included in the synthetic input images to bridge this gap between the two domains.

For cross domain image mapping, these traditional GAN-based models typically use annotated image pairings from the source domain and the target domain. Nevertheless, gathering such side-by-side matching pairs requires a lot of time and effort. In real world scenario, it is very simple to gather various images in the source domain and target domain separately [16,17]. Therefore, a variant of GAN termed as CycleGAN [18] is proposed to translate an image from one domain to target domain without any paired image. The unpaired GAN model shares similar idea to GAN model. In vanilla GAN version there is only one generator (G) and one discriminator (D) which is trained on paired images. However, CycleGAN utilize domain defined supervision. For cross domain image mapping from a domain S to domain R, there are two GANs with one primal GAN with Generator GS>R learns to translate an image from domain S to R and a corresponding GAN with generator GR>S that learns to invert the translated image in R to the original domain S with an additional reconstruction loss. In these unpaired GAN for image cross-domain mapping task, the performance significantly depends on designed optimization loss function [19]. All GAN-based model optimization function includes adversarial loss that identify real or fake images. In addition, some (unpaired) GAN variations have included pixel level loss in the modeling process [20,21,22], for example, the error between them with L1 or L2 distance loss [22], to guarantee the resemblance between an original image and its generated version. As the human visual system typically concentrates on the higher-level abstractions of images for perceptual quality evaluation, the generated images are not satisfactory for human perception even though pixel level matching consistency is good [23,24]. Therefore, to generate high quality images researchers proposed to include a content-based loss in the optimization loss function where the content-based loss is formulated from pre-trained deep networks between source image and the corresponding generated image [25,26,27]. Therefore, we focused on incorporating adversarial loss with perceptual reconstruction loss and emphasis on constructing loss function that can be applied to synthetic flood image enhancement without paired real world flood image.

The pixel-wise loss methods have the drawback of inadequately simulating the human visual experience, which frequently results in blurry areas [19]. Convolutional Neural Networks (CNN) have demonstrated remarkable performance in automatically extracting high level content structure information from images. Numerous empirical studies [28,29] have confirmed that the CNN network's higher layers capture perceptual abstractions of image. Thus, it is suggested to use feature maps in CNNs as perceptual quality measures for a variety of image processing-related tasks, including image resolution [24, 30], style transfer [25, 31], and unsupervised depth and motion estimation [32] and incorporate the perceptual quality into the optimization function to produce quality-aware images [24, 25].Typically, the visual characteristics from a pretrained VGG network [33] were represented by the perceptual loss function [9]. These models might be used for the unpaired synthetic image enhancement because they don't require any labeled data for quality assessment.

The enhancement network GAN structure contains three losses: the adversarial loss **LADV (S, R)**inherited from vanilla GAN model that attempts to match the generated images to real image domain distribution, the perceptual reconstruction loss **LPR (S, R)** to ensure learned mapping complement the images, the identity loss, inspired from Cycle-GAN, **LI (S, R)**to preserve color composition between output and input image of the enhancement network.

**(1)**

In the above equation, for adversarial loss, the objective function as:

**(2)**

where the first row represents the adversarial loss for mapping from S (synthetic) to R (real) domain and the second row defines the adversarial loss for mapping from R to S. GS>R map the synthetic image domain S to real image domain R and DR focus on classify the mapped image GS>R(S) and real image R. The adversarial process GS>R focus on minimize the objective against DR that attempt to maximize it. Similarly, the second row present a similar adversarial loss for mapping from R to S.

For the cycle consistent reconstruction, GS>R and GR>S met the backward reconstruction consistency as: **S GS>R (S) GR>S (GS>R (S)) ~ S** and R GR>S (R) GS>R(GR>S (R)) ~R. Generally, the reconstruction loss is defined as pixel wise mean square error (MSE) loss. The optimization of MSE lack high-frequency content which results in blurriness and have poor perceptual quality [24]. Therefore, instead of pixel-wise loss, we utilized VGG loss [24] based on ReLU activation layers of the pre-trained 19 layer VGG network [33]. With ϴ(i,j) indicates feature map obtained by j-th convolutional (after activation) before the i-th maxpooling layer within the VGG19 network. The perceptual reconstruction loss is defined as the euclidean distance between the feature representations of a reconstructed image **GR>S (GS>R (S))** and the reference image S.

**2 (3)**

Where, and indicates the dimension of the respective feature map and x,y indicates image dimension.

**=(] + ] (4)**

In general, including an identity loss aids in maintaining color and intensity in cross domain mapping so that the generator does not map a daytime image to nighttime image.

Given the adversarial loss in Eq.(2), perceptual reconstruction loss in Eq.(3) and identity loss in Eq.(4), the GAN based enhancement network focus on optimizing the following objective function:

**= (5)**

* 1. **Network Architecture**

The enhancement network, the generator (GS>R, GR>S) is a ResNet [34] based UNet architecture similar to Unet based CycleGAN [35]. The ResNet-Unet architecture utilized the traditional Unet structure which is comprised of encoder (down-sampling) and decoder (up-sampling) structure. The down-sampling encoder is replaced by ResNet-34 blocks. In the discriminator (DS>R, DR>S) we use PatchGANs [18] to distinguish between real or fake image patches. For the perceptual reconstruction loss, the perceptual model [25] was based on VGG19 pre-trained network.

For segmentation of flood water and car using the enhanced synthetic image, we first utilized the encoder-decoder based segmentation network with atrous separable convolution [36]. To expand the receptive field, atrous convolution adds dilations to the conventional convolution. It can successfully address the issue of decreased spatial resolution effected by down-sampling. For comparison of different segmentation model, we have utilized pytorch based pre-trained backbone in segmentation models [37]. We experiment with six (DeepLabV3 [36], Unet[38], Unet++[39], PSPNet[40], LinkNet[41], FPNet[42]) different segmentation model architecture with ResNet-50 based encoder with pre-trained (ImageNet) weight initialization. The comparison among segmentation models allows us to understand which segmentation network performed better while tested on real flood images.

1. **Experimental Results**

We evaluated the flood water and car segmentation model on the Deep Flood [5], Cem Sazara [4] and Water Segmentation Dataset [6]. The evaluation dataset comprised of 667 real flood images which have flood water and car within the image and corresponding segmentation mask. For training of enhancement network, we utilized unpaired real flood images from aforementioned dataset in conjunction with images rendered from 3D scenario which is created by Blender software. Total 7,200 or 24 synthetic flooded videos are generated by SURFGenerator [9]. Considering the images with flooded water, excluding the dry conditions, we utilized 5864 synthetic flood images in this work. Therefore, the trained enhancement network is applied to enhance color and texture of synthetic image and the enhanced images are utilized as training data for the segmentation models.

**4.1. Implementation Details**

**Training Details:**

In the enhancement network, for the adversarial loss , the generator minimizing the least square loss instead of log-likelihood loss [15] during the adversarial training. The loss for in equation 2, by minimizing

**]**

and trained by minimizing

**] + ())].**

The enhancement model is trained using Tensorflow. During the optimization, different loss component in the total is set to weights, the least square loss is given a weight of 1.0, cycle-loss with 10.0 and perceptual cycle loss is 0.1\* cycle loss. The enhancement network is trained from scratch, with initial learning rate of 10-4 for the first 10 epoch and linearly decaying for the next 10 epochs and the model is trained for total 1000 epoch. For the segmentation model, we utilized weight initialization in the encoder network and the model is trained for 100 epochs with an initial learning rate of 0.0001. For both enhancement and segmentation model Adam solver is utilized.

**Benchmark Models for Segmentation:**

Besides implementing the TCENet model, we tested the segmentation model for the following scenarios: (1) an “all-real” scenario, in which we utilized real flood images and segmentation map for training where we expect an upper bound compare to full TCENet performance, and (2) an “synthetic-real” scenario, in which we train the segmentation model with synthetic flood image-segmentation mask pair and tested the model on real flood image, for which we presume a lower bound compare to our model performance.

**Evaluation Metrics:**

We evaluated the performance of our methods using the semantic segmentation metrics for each class as reported in [43] for multi-class segmentation. Dice Similarity Coefficient (DSC) is calculated to measure the similarity of predicted and ground truth mask for each class and accuracy is measured for all classes. Dice loss which is adapted as loss function for the segmentation model is also calculated for each class.

where, TP, FP, TN, FN indicates true positive, false positive, true negative and false negative respectively.

A picture containing text, different, shop

Description automatically generated

Figure 3. Example Enhanced Image for our enhancement network in TCENet for blender generated synthetic images. First row: Synthetic flood image render from SURFGenerator [9], Second row: corresponding output of CycleGAN, Third row: corresponding output of our enhancement network, Fourth row: real flood images (no correspondence to above rows).

**4.2 Qualitative Results of Enhanced Image**

Fig.3 depicts the output of our enhancement network using with perceptual reconstruction loss and it is visible that compare to CycleGAN results, our results are perceptually better in terms of color and texture. We observe that in terms of CycleGAN, the color of the enhanced images (second row in Fig.3) got distorted and the texture representation of buildings, signboards does not correspond to the original synthetic images. The visual distinctions between real flood and synthetic flood images are apparent: real flood sceneries have more sophisticated colors, textures, lighting, and shadows than synthetic ones. In terms of low-level appearance, the enhanced synthetic versions of enhancement network (third row in Fig.3) resemble real flood images more than synthetic ones. In addition, water features such as ripples and waves from cars in the SURFGenerator generated synthetic images complement the real flood images water features.

**4.3 Segmentation Results**

In Table 1, we report the performance of different segmentation model that have been utilized to segment flood water and car from the real flood image dataset. The grey color rows present the results when the models are trained with only synthetic images and tested on the real flood images. The blue rows indicate the models are trained with enhanced synthetic images from the enhancement network and then the trained model tested on real flood images. The white row

**Table 1.** Semantic Segmentation results for real flood image data. Grey rows indicate the segmentation models are trained with only synthetic images. Best semantic segmentation results are marked with \* when model is trained with enhanced synthetic images via enhancement network.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Lower is better** | | **Higher is better** | | |
| **Models** | **Dice loss(water)** | **Dice loss(car)** | **Dice Score(water)** | **Dice Score (Car)** | **Accuracy (%)** |
| DeepLabV3+ (syn) | 0.378 | 0.756 | 0.621 | 0.243 | 65.66% |
| Unet (syn) | 0.373 | 0.749 | 0.626 | 0.250 | 62.38% |
| Unet+ (syn) | 0.363 | 0.733 | 0.636 | 0.266 | 62.74% |
| PSPNet (syn) | 0.425 | 0.783 | 0.574 | 0.216 | 63.28% |
| LinkNet (syn) | 0.398 | 0.614 | 0.601 | 0.385 | 68.74% |
| FPNet (syn) | 0.345 | 0.720 | 0.654 | 0.279 | 64.51% |
| **LinkNet (all-real)** | **0.055** | **0.069** | **0.944** | **0.930** | **96.22%** |
| DeepLabV3+ | 0.254 | 0.342 | 0.745 | 0.658 | 74.27% |
| Unet | 0.295 | 0.308 | 0.704 | 0.691 | 79.5% |
| Unet+ | 0.267 | 0.273 | 0.732 | 0.726 | 80.72% |
| PSPNet | 0.308 | 0.426 | 0.691 | 0.573 | 73.80% |
| LinkNet | 0.218\* | 0.229\* | 0.781\* | 0.770\* | 84.24%\* |
| FPNet | 0.224 | 0.256 | 0.775 | 0.743 | 83.48% |

* syn indicates the segmentation is trained on synthetic flood data
* all-real indicates the segmentation is trained and test on real flood data

indicates the segmentation model is trained and tested on real flood dataset and will be considered as upper band baseline for comparison between segmentation models. We observe LinkNet performed better compared to other segmentation models, therefore, in case of only real flood images we considered the LinkNet model for training and testing on the real flood dataset. As we expected “all-real” and “synthetic-real” (refer to section 4.1) scenarios provides the upper bound and lower bound respectively for segmentation compared to our proposed method.

We observe for each class (flood water, car) dice loss is smaller when the model is trained with enhanced synthetic images and tested on real flood images. In terms of, dice score and accuracy the performance of the segmentation models (trained on enhanced synthetic images) is better than the segmentation models when trained with only synthetic flood image data. Our proposed TCENet create a clear gap in the performance of the segmentation of flood water and car compared to synthetic-only benchmark.

**4.4 Activation mapping of segmentation model**

The interpretability of the segmentation model on the real flood image can be understood utilizing class activation mapping. We use Grad-CAM [44], or gradient-weighted class activation mapping, to present and analyze the results of different segmentation model that is trained on enhanced synthetic image and tested on real flood dataset. The discriminative pixel areas utilized by the segmentation model to identify the class (flood water and car) are shown in the activation heatmap. Heatmaps that closely resemble the logits of the specified class and the output segmentation mask are produced by feature maps that are situated between the bottleneck and final layer [45]. Therefore, we considered the layer before the final layer for each of the segmentation models to visualize the activation heatmap and interpret results of the segmentation models. In Fig.4, we present the activation mapping for target class (flood water, car) for different segmentation models. The first column represents a real flood image from the test data. The second column shows the pixels that are focused by the respective segmentation model to decide on the segmentation of car from the image and third column indicates the pixels for flood water segmentation decision making by the model. The green and red pixels in the activation mapping indicates the important pixel region for segmentation of a particular target class. For instance, in case of LinkNet segmentation model, in Fig.3 (2nd column 2nd row), for flood car segmentation the important pixels for flood car segmentation is presented with green pixels and we can see that model only focused on the car region to segment the flood car and similarly for segmenting the flood water it focused on the flood water presented by the green pixels. From the Table.1, we observe that compare to other segmentation model the performance of LinkNet is better. Therefore, if we see the activation map--ing of the other segmentation model compared to LinkNet model we can see that in case of LinkNet the pixel of interest by the segmentation model is more concentrated on the target class. However, from activation mapping we observe in case of all segmentation model, the trained model could not correctly distinguish between trees and water and most of the time considers trees or green scenario as flood water. The underlying reason could be the absence of trees or green color objects in the generation of synthetic image from SURFGenerator[9].

1. **Conclusion and Future Work**

We present our TCENet deep neural network for flood water and car segmentation, that require unpaired synthetic flood image and real flood image for model training. The overall pipeline comprises an enhancement network and a segmentation network. It can enhance the synthetic flood images color and texture via a learning framework that combines adversarial loss, perceptual reconstruction loss and identity loss. The enhanced synthetic images are input to the segmentation model which is optimized with dice loss. The trained segmentation network is tested on real flood images to segment flood water and car. The TCENet does not require paired synthetic and real flood images for training and can produce good results when segmentation model trained with enhanced synthetic image. In future, we will extend the framework for flood depth estimation from real flood images when the model will be trained on enhanced synthetic images.

|  |  |  |
| --- | --- | --- |
|  | A picture containing water, outdoor  Description automatically generatedA picture containing text, water  Description automatically generatedA picture containing water, outdoor  Description automatically generated  FPNet  PSPnet  Unet  Unet+  DeepLabV3+  LinkNet | A picture containing colorful  Description automatically generated |

Figure 4. Class activation mapping of segmentation models. First column presents a real flood image. Second column indicates activation mapping for “flood car” for different segmentation models. Third column depicts activation mapping for “flood water” for different segmentation models.

**References**

[1][1] J. Knighton, K. Hondula, C. Sharkus, C. Guzman, and R. Elliott, “Flood risk behaviors of United States riverine metropolitan areas are driven by local hydrology and shaped by race,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, no. 13, 2021, doi: 10.1073/pnas.2016839118.

[2] W. V. Sweet and J. Park, “From the extreme to the mean: Acceleration and tipping points of coastal inundation from sea level rise,” *Earth’s Futur.*, vol. 2, no. 12, 2014, doi: 10.1002/2014ef000272.

[3] N. S. Romali, Z. Yusop, M. Sulaiman, and Z. Ismail, “Flood risk assessment: A review of flood damage estimation model for Malaysia,” *J. Teknol.*, vol. 80, no. 3, 2018, doi: 10.11113/jt.v80.11189.

[4] C. Sazara, M. Cetin, and K. M. Iftekharuddin, “Detecting floodwater on roadways from image data with handcrafted features and deep transfer learning,” 2019, doi: 10.1109/ITSC.2019.8917368.

[5] P. Chaudhary, S. D’Aronco, J. P. Leitão, K. Schindler, and J. D. Wegner, “Water level prediction from social media images with a multi-task ranking approach,” *ISPRS J. Photogramm. Remote Sens.*, vol. 167, 2020, doi: 10.1016/j.isprsjprs.2020.07.003.

[6] V. Klemas, “Remote sensing of floods and flood-prone areas: An overview,” *Journal of Coastal Research*, vol. 31, no. 4. 2015, doi: 10.2112/JCOASTRES-D-14-00160.1.

[7] B. Arshad, R. Ogie, J. Barthelemy, B. Pradhan, N. Verstaevel, and P. Perez, “Computer vision and iot-based sensors in flood monitoring and mapping: A systematic review,” *Sensors (Switzerland)*, vol. 19, no. 22, 2019, doi: 10.3390/s19225012.

[8] L. Hawker, P. Bates, J. Neal, and J. Rougier, “Perspectives on Digital Elevation Model (DEM) Simulation for Flood Modeling in the Absence of a High-Accuracy Open Access Global DEM,” *Front. Earth Sci.*, vol. 6, 2018, doi: 10.3389/feart.2018.00233.

[9] S. Lamczyk, “SURFGenerator : Generative Adversarial Network Modeling for Synthetic Flooding Video Generation.”

[10] A. Mosavi, P. Ozturk, and K. W. Chau, “Flood prediction using machine learning models: Literature review,” *Water (Switzerland)*, vol. 10, no. 11. 2018, doi: 10.3390/w10111536.

[11] S. Sankaranarayanan, Y. Balaji, A. Jain, S. N. Lim, and R. Chellappa, “Learning from Synthetic Data: Addressing Domain Shift for Semantic Segmentation,” 2018, doi: 10.1109/CVPR.2018.00395.

[12] M. A. Witherow, C. Sazara, I. M. Winter-Arboleda, M. I. Elbakary, M. Cetin, and K. M. Iftekharuddin, “Floodwater detection on roadways from crowdsourced images,” *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, vol. 7, no. 5–6, 2019, doi: 10.1080/21681163.2018.1488223.

[13] M. Zaffaroni and C. Rossi, “Water segmentation with deep learning models for flood detection and monitoring,” in *Proceedings of the International ISCRAM Conference*, 2020, vol. 2020-May.

[14] P. Chaudhary, S. D’Aronco, M. Moy De Vitry, J. P. Leitão, and J. D. Wegner, “FLOOD-WATER LEVEL ESTIMATION from SOCIAL MEDIA IMAGES,” in *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019, vol. 4, no. 2/W5, doi: 10.5194/isprs-annals-IV-2-W5-5-2019.

[15] C. Zheng, T. J. Cham, and J. Cai, “T2 Net: Synthetic-to-realistic translation for solving single-image depth estimation tasks,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11211 LNCS, pp. 798–814, doi: 10.1007/978-3-030-01234-2\_47.

[16] Z. Yi, H. Zhang, P. Tan, and M. Gong, “DualGAN: Unsupervised Dual Learning for Image-to-Image Translation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-October, doi: 10.1109/ICCV.2017.310.

[17] M. Wang, R. Hong, X. T. Yuan, S. Yan, and T. S. Chua, “Movie2Comics: Towards a lively video content presentation,” *IEEE Trans. Multimed.*, vol. 14, no. 3 PART 2, 2012, doi: 10.1109/TMM.2012.2187181.

[18] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-October, doi: 10.1109/ICCV.2017.244.

[19] L. Chen, L. Wu, Z. Hu, and M. Wang, “Quality-Aware Unpaired Image-to-Image Translation,” *IEEE Trans. Multimed.*, vol. 21, no. 10, 2019, doi: 10.1109/TMM.2019.2907052.

[20] D. He *et al.*, “Dual learning for machine translation,” 2016.

[21] X. Mao and Q. Li, “Unpaired multi-domain image generation via regularized conditional GANs,” *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2018-July, pp. 2553–2559, 2018, doi: 10.24963/ijcai.2018/354.

[22] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January, doi: 10.1109/CVPR.2017.632.

[23] Z. Chen, T. Jiang, and Y. Tian, “Quality assessment for comparing image enhancement algorithms,” 2014, doi: 10.1109/CVPR.2014.384.

[24] C. Ledig *et al.*, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January, doi: 10.1109/CVPR.2017.19.

[25] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9906 LNCS, doi: 10.1007/978-3-319-46475-6\_43.

[26] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” 2018, doi: 10.1109/CVPR.2018.00068.

[27] G. Guo, H. Wang, C. Shen, Y. Yan, and H. Y. M. Liao, “Automatic Image Cropping for Visual Aesthetic Enhancement Using Deep Neural Networks and Cascaded Regression,” *IEEE Trans. Multimed.*, vol. 20, no. 8, 2018, doi: 10.1109/TMM.2018.2794262.

[28] J. Schmidhuber, “Deep Learning in neural networks: An overview,” *Neural Networks*, vol. 61. 2015, doi: 10.1016/j.neunet.2014.09.003.

[29] C. Luo, B. Ni, S. Yan, and M. Wang, “Image Classification by Selective Regularized Subspace Learning,” *IEEE Trans. Multimed.*, vol. 18, no. 1, 2016, doi: 10.1109/TMM.2015.2495248.

[30] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S. Huang, “Robust Single Image Super-Resolution via Deep Networks with Sparse Prior,” *IEEE Trans. Image Process.*, vol. 25, no. 7, 2016, doi: 10.1109/TIP.2016.2564643.

[31] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, “CamStyle: A Novel Data Augmentation Method for Person Re-Identification,” *IEEE Trans. Image Process.*, vol. 28, no. 3, 2019, doi: 10.1109/TIP.2018.2874313.

[32] Z. Yin and J. Shi, “GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose,” 2018, doi: 10.1109/CVPR.2018.00212.

[33] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.

[34] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, doi: 10.1109/CVPR.2016.90.

[35] D. Torbunov *et al.*, “UVCGAN: UNet Vision Transformer cycle-consistent GAN for unpaired image-to-image translation,” pp. 1–5, 2022, [Online]. Available: http://arxiv.org/abs/2203.02557.

[36] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11211 LNCS, doi: 10.1007/978-3-030-01234-2\_49.

[37] “GitHub - qubvel/segmentation\_models.pytorch: Segmentation models with pretrained backbones. PyTorch.” https://github.com/qubvel/segmentation\_models.pytorch (accessed Sep. 19, 2022).

[38] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351, doi: 10.1007/978-3-319-24574-4\_28.

[39] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018, vol. 11045 LNCS, doi: 10.1007/978-3-030-00889-5\_1.

[40] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January, doi: 10.1109/CVPR.2017.660.

[41] A. Chaurasia and E. Culurciello, “LinkNet: Exploiting encoder representations for efficient semantic segmentation,” in *2017 IEEE Visual Communications and Image Processing, VCIP 2017*, 2018, vol. 2018-January, doi: 10.1109/VCIP.2017.8305148.

[42] L. Pi and J. Wu, “FPNet: Fusion Attention Instance Segmentation Network Based on Pose Estimation,” 2021, doi: 10.1109/CCDC52312.2021.9602451.

[43] D. Müller, I. Soto-Rey, and F. Kramer, “Towards a guideline for evaluation metrics in medical image segmentation,” *BMC Res. Notes*, vol. 15, no. 1, pp. 1–7, 2022, doi: 10.1186/s13104-022-06096-y.

[44] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, 2020, doi: 10.1007/s11263-019-01228-7.

[45] K. Vinogradova, A. Dibrov, and G. Myers, “Towards interpretable semantic segmentation via gradient-weighted class activation mapping (student abstract),” *AAAI 2020 - 34th AAAI Conf. Artif. Intell.*, pp. 13943–13944, 2020, doi: 10.1609/aaai.v34i10.7244.

[2]

[3][4][5][6][7][8] [9][10][11][12][13][14][15][16][17][18][19][20]

[21][22][23][24][25][26][27][28][29][30][31][32][33]

[34][35][36][37][38][39][40][41][42][43][44][45]