Scene Construction from Depth Map Using Image-to-Image Translation Model
Görüntüden Görüntüye Dönüşüm Modeli Kullanılarak Derinlik Haritası Girdisiyle Sahne Görseli Üretimi

Hasan Avşar[1], Mehmet Sarıgül[1], Levent Karacan[1]
[1]Department of Computer Engineering, Iskenderun Technical University, Hatay, Turkey
ORCIDs: 0000-0001-5365-1040, 0000-0001-7323-6864, 0000-0003-2764-5258
E-mails: hasanavsar.mfbe20@iste.edu.tr, mehmet.sarigul@iste.edu.tr, levent.karacan@iste.edu.tr

***Abstract*—In recent years, deep learning approach to solve the image and video processing problems have become very popular. Generative Adversarial Networks (GANs) are one of the most popular deep learning-based models. GANs form a generative model utilizing two sub-models, namely, generator and discriminator. The generator tries to generate indistinguishably realistic outputs where the discriminator tires to classify the outputs of the generator as real or fake. These two models work together to achieve a successful generation of realistic outputs. This study aims to reconstruct daytime image of a given depth map data recorded with a camera or a sensor which can capture the depth map data during night time or in a lightless environment. Our model was used for reconstructing the 2D images for a given depth map representation of a known scene. The model was trained with the chess scene from 7-scenes dataset and realistic 2D images were successfully generated for the given input maps.**

Keywords*—deep learning; generative adversarial networks; Pix2pixHD

***Özetçe*—Son yıllarda, görüntü ve video işleme problemlerinin çözümlerinde derin öğrenme tabanlı yaklaşımların kullanımı büyük bir popülerlik kazanmıştır. Üretici Çekişmeli Ağlar (Generative Adversarial Networks veya kısaca GAN), en çok tercih edilen derin öğrenme modelleri arasında bulunmaktadır. Üretici Çekişmeli Ağlar, üretici ve ayırt edici olmak üzere iki farklı alt modelin bir araya gelmesiyle oluşan üretici yapıya sahip derin modellerdir. Üretici alt modelin amacı gerçeğe en yakın çıktıları üretebilmek iken ayırt edici alt modelin amacı ise üretici alt model tarafından üretilen çıktıları gerçek veya yapay olarak etiketlemektir. Bu çalışmamızın amacı, ışıksız ortamda özel kamera veya sensörler ile elde edilen derinlik haritası verisini kullanarak ortamın renkli görsellerini üretmektir. Modelimiz, bilinen bir sahneye ait derinlik haritası verisinden renkli görsellerin üretimi yapılarak denenmiştir. 7-scenes veri seti içinde bulunan satranç sahnesi ile eğitilen modelimiz derinlik haritası girdilerinden renkli görsel üretimini başarıyla gerçekleştirmiştir.**

Anahtar Kelimeler*—derin öğrenme; üretici çekişmeli ağlar; Pix2pixHD

I. INTRODUCTION

Photo-realistic image rendering for a specific scene or an object problem and there are many studies focusing on it. Representing a scene in a model learned from a data set makes it possible to generate successful outputs. The scene can be modelled with different kind of data inputs. Camera location and orientation, point clouds, scene coordinate maps and depth maps can be used along with the corresponding real RGB images to represent a scene. After a successful training, photo-realistic images of a scene can be constructed for desired inputs [1]. Devices such as a LIDAR camera or a Kinect sensor can capture depth map of a scene even in the lightless environment. Using these kinds of devices, captured depth map data in a dark place or during the night time can be converted to daytime images. This conversion problem can be solved with the use of deep learning algorithms [2], [3]. Deep neural networks are solid structures and also are popularly used in different problem domains. This popularity enabled the development of many different variations of deep learning algorithms. Generative Adversarial Networks (GANs) [4] are a deep learning based generative algorithm popularly used for applications such as image-to-image conversion, style transfer, scene construction. GANs are basically algorithms that utilize two sub-networks, namely generator and discriminator. These networks challenge each other to update and upgrade the generated outputs of the main algorithm. The generator generates outputs and the discriminator classifies whether the output is real or fake. This challenge continues until the generated results are indistinguishable from the real-life data.

Pix2pixHD [5] is a type of GAN which performs image-to-image transformation. It aims to solve the generated image quality problem at higher resolution images with the use of multi-scale generator and discriminator architectures. Pix2pixHD is inspired from a previous study named Pix2pix [6] and with its multi-scale approach it can generate better results. In this study, Pix2pixHD was used for modelling a scene with depth maps and corresponding real images. Then de photo-realistic images were successfully generated for unseen depth map data inputs gathered from the same scene.

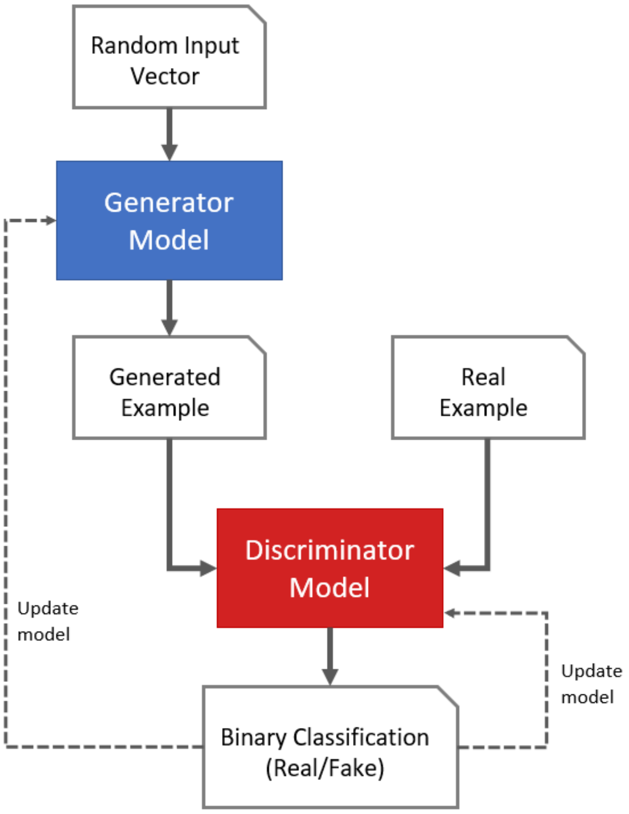

Figure 1: Generative Adversarial Network structure

II. MATERIALS AND METHODS

A. Generative Adversarial Networks

Generative Adversarial Networks, abbreviated as GANs, are basically a pair of models competing with each other in order to be able to generate outputs which are indistinguishable from the real ones. The term "adversarial", comes from this nature of models in competition. Starting with randomized input data, the generator model generates fake images which become more and more realistic as the training continues. The discriminator model is fed with the fake images along with the real images from training dataset and makes a binary prediction whether the generated image is fake or real. At the end of each competition round, both generator and discriminator models are updated based on the success of generated (fake) images. The basic flowchart of a GAN is given in Fig. 1.

B. Pix2PixHD

Pix2pixHD is an extended GAN model which bring new ideas to the basic structure in order to overcome the problems arising as the resolution of generated images increase. It includes multi-scale generator and discriminator architectures which makes it possible to generate realistic images at 2048 x 1024 resolution. A generator network G_1 is trained on lower resolution (coarse) and then it is appended to another network G_2 to be jointly trained for higher resolution (fine) scale. The 2-scale structure of generator model $G = G_1, G_2$ is given in Fig. 2. As the target resolution increases, the generator can be modified to have more scale levels (G = $G_1, \cdots, G_n$).

Various number of discriminator networks are utilized to handle the discrimination task on different scales of given input images. The images are down sampled and are fed to into discriminator networks to be processed in a coarse-to-fine manner. The result is then used to update individual discriminator networks and the generator network (Fig. 3).

C. 7-Scenes Dataset

Chess scene of 7-Scenes dataset was used for this study. The training dataset includes 4000 consecutive frames of a video stream where the test data set includes 2000 frames.

D. Experiment

The Pix2pixHD was structured with 4 levels of generator models and 2 layers of discriminator models. Training was performed with 44 epochs.

III. RESULTS AND DISCUSSION

Using the Pix2pixHD model formed after training, realistic images are generated successfully. Fig. 4 shows the comparison of generated images with the real images. Fig. 5 shows the loss trends achieved during the training.

GANs are one of the most preferred deep learning-based algorithms used for image-to-image conversion. Pix2PixHD is a type of GAN which targets to increase the generated output efficiency even for the images with high resolution. Pix2PixHD uses multiscale network structures enabling coarse-to-fine approach to the realistic image generation problem. In this study, Pix2pixHD was used to generate realistic daytime images of a given depth map from a known scene. Although there are imperfections in the generated images, the realistic images were successfully generated by the model.

The success of this study shows that the depth map data recorded by a camera or a sensor can be converted to daytime images. LIDAR cameras, Kinect sensors or other similar devices can capture depth map data in lightless environment. The model in this study can be used to generate daytime images of a scene during the night time or when the scene is in the complete darkness. On the other hand, the model can also be further developed for generating daytime video of a scene just by converting the depth map data captured in the darkness.

APPENDICES

Author Contributions

All authors equally contributed on writing the paper.

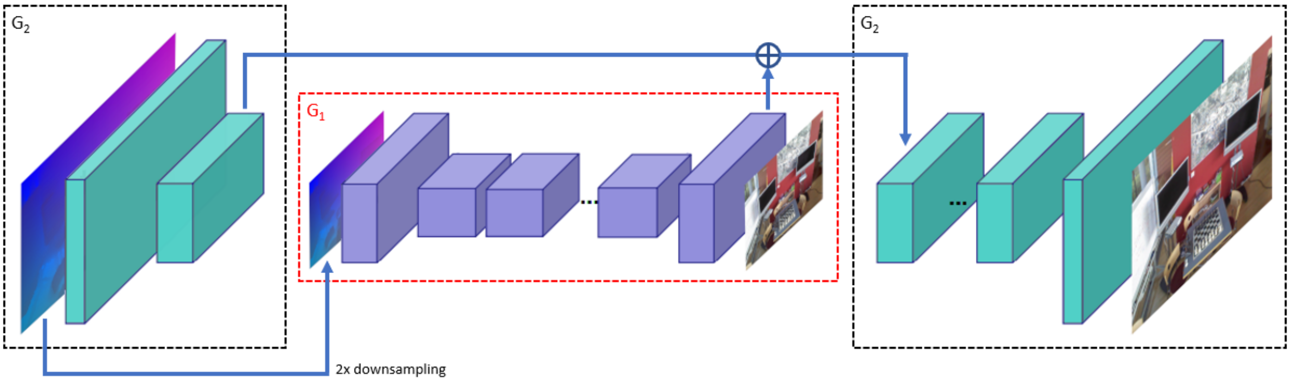

Figure 2: Multi-scale generator model structure of Pix2pixHD

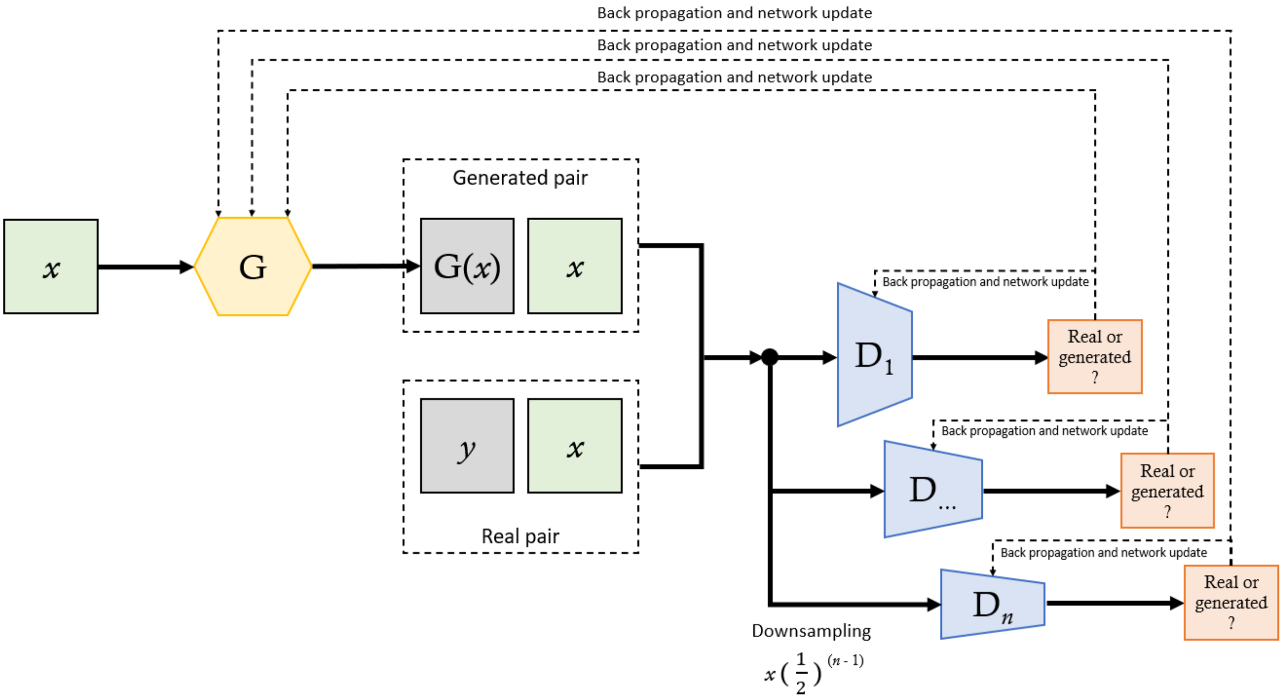

Figure 3: Multi-scale discriminator model structure of Pix2pixHD

Acknowledgments

The study was supported by grant 120E447 from the TUBITAK.

Conflicts

None declared.

Ethical Declaration

This article does not contain any studies involving human participants and/or animals performed by any of the authors.

REFERENCES

[1] Yildirim O, Ucar A, Baloglu UB. Recognition of real-world texture images under challenging conditions With deep learning. Journal of Intelligent Systems with Applications 2018; 1(2): 122-126.

[2] Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks 2015; 61: 85-117.

[3] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436–444.

[4] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Bengio Y. Generative ddversarial nets. Advances in Neural Information Processing Systems 2014; 521: 436–444.

[5] Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B. High-resolution image synthesis and semantic manipulation with conditional GANs. In

Images generated by Pix2pixHD Real Images

Figure 4: Comparison of generated images with real images

arXiv Preprint Archive on Computer Vision and Pattern Recognition 2018.

[6] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In arXiv Preprint Archive on Computer Vision and Pattern Recognition 2017.

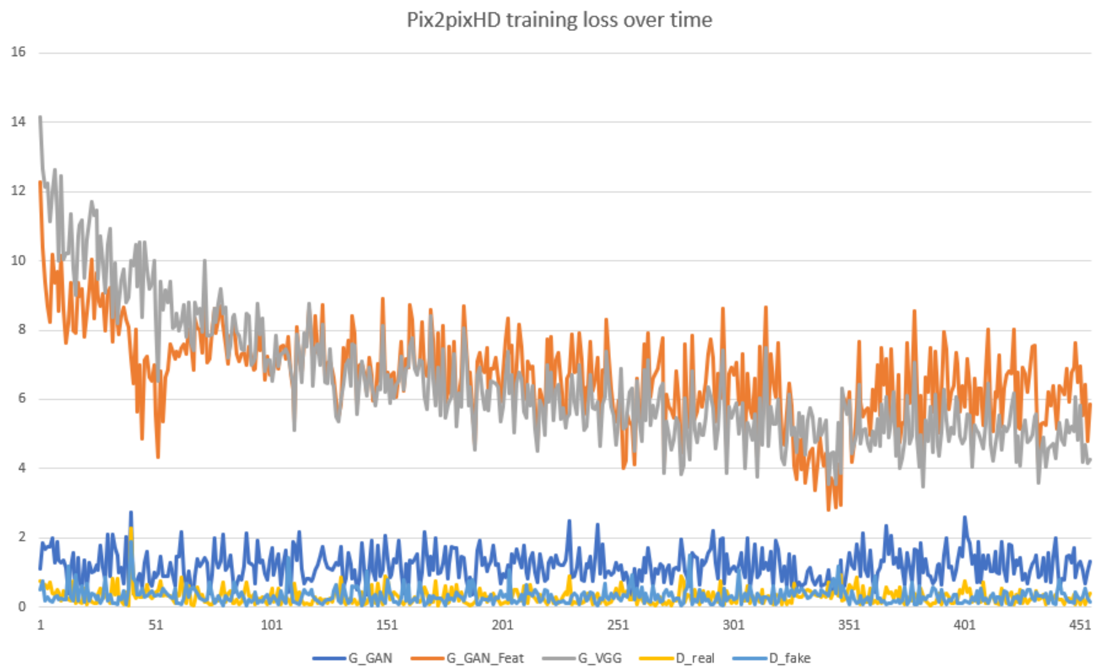

Figure 5: Training loss trends