2023 NDIA MICHIGAN CHAPTER
GROUND VEHICLE SYSTEMS ENGINEERING
AND TECHNOLOGY SYMPOSIUM
AUTONOMY, ARTIFICIAL INTELLIGENCE & ROBOTICS TECHNICAL SESSION
AUGUST 15-17, 2023 - NOVI, MICHIGAN

CROSS-VIEW IMAGE TRANSLATION USING CONDITIONAL GAN FOR AUTONOMOUS DRIVING

Dan Zhang[1,*], Bradley Sanders[1,*], Grayson Byrd[1], , Feng Luo[1,#], Venkat Krovi[2], David Gorsich[3], Jonathon M. Smereka[3], Mark Brudnak[3]

[1]School of Computing, Clemson University, Clemson, SC
[2]School of Mechanical and Automotive Engineering, Clemson University, Clemson, SC
[3]US Army DEVCOM Ground Vehicle Systems Center, Warren, MI

ABSTRACT

A major benefit of intelligent and autonomous vehicles is their ability to navigate through hazardous environments that pose a significant danger to humans. In such environments, eventual damage to vehicle sensors is often inevitable. To address this threat to vehicle function, we propose a more robust system in which information from alternative sensors is leveraged to restore navigation capabilities in the case of primary sensor failure. This system employs image translation methods that enable the vehicle to use images generated from an auxiliary camera to synthesize the display of the primary camera. In this work, we present a conditional Generative Adversarial Network (cGAN) based method for view translation coupled with a Residual Neural Network for imitation learning. We evaluate our approach in the CARLA simulator and demonstrate its ability to restore navigation capabilities to a real-world vehicle by generating a front-view image from a left-camera view.

Citation: Dan Zhang, Bradley Sanders, Grayson Byrd, Feng Luo, Venkat Krovi, David Gorsich, Jonathon M. Smereka, Mark Brudnak, "Cross-view Image Translation using Conditional GAN for Autonomous Driving". In *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium* (GVSETS), NDIA, Novi, MI, Aug. 15-17, 2023.

* These authors contributed equally to the manuscript as first authors
# To whom correspondence should be addressed.

1. INTRODUCTION

In the context of real-world autonomous vehicles, sensor perception plays a crucial role in the functioning of these systems. Damage to sensors can impact the vehicle's ability to gather information about its environment, thus hindering its mobility. Utilizing information from auxiliary sensors to reinstate navigation capabilities in the event of main sensor failure could address this problem. In this work, we leverage an alternate left camera view feed to synthesize a front camera view which is necessary for vehicle autonomy. This simulates a scenario where the main sensor was necessary for vehicle autonomy, but was Our task can be framed as an image-to-image translation problem.

Isola et al.[19] first introduced the use of conditional GANs [18] to solve various image-to-image translation tasks, such as transforming semantic maps to real views and generating cats from user sketches, where the input and target images share similar outlines and scene structures. Differing from these applications, our task is particularly challenging due to various viewing angles, which leads to the presence of objects in one view that is absent from the other view. Additionally, details such as shadows and lights can differ significantly between the two views. Several works [11, 30] have proposed methods to address cross-view translation problems by using semantic maps of the target domain to guide the generation process. However, we argue that the success of these existing methods is limited to their benchmark data sets since prior information on the environment is not available in many practical applications. Additionally, the synthesized images in our application are required to produce accurate driving controls for an autonomous vehicle, which is unprecedented in current studies.

To tackle the above challenges, we propose an image translation framework adopted from the pix2pix method [19] and incorporate a perceptual loss which has been shown to be useful for image super-resolution [24] and style transfer [23]. We fine-tune this model in conjunction with a fixed, pre-trained imitation model that provides driving autonomy to a vehicle. The baseline imitation model is adopted from [3] and is trained using behavioral cloning in an end-to-end driving format. The network is comprised of a perception backbone and a control head and outputs driving controls based on an input image. We show that through a modified objective as well as a two-step training strategy, we can synthesize images that are not only visually appealing but also faithful to the ground truths in terms of driving functionality.

Our proposed image translation model is evaluated by comparing the driving performance when using the generated fake image to the performance of real front view image in the Carla simulator environment [4]. We conduct experiments following the NoCrash benchmark [3] and the Carla benchmark [4], to compare driving performance of the original imitation learning model with that of our proposed image translation model.

2. RELATED WORK

Here we discuss related works in the fields of GANs and behavioral cloning, focusing on the context of autonomy.

2.1. GANs and conditional GANs

Generative Adversarial Networks(GANs) were introduced by [5], which explored the viability of adversarial networks. These networks displayed successful results in generating images by utilizing two models, a generator and a discriminator, which are trained in conjunction with one another. The generator is trained to fool the discriminator and the discriminator is trained to detect fake images created by the generator. This competition produces consistent and concurrent optimization of each model until the generator produces images that cannot be distinguished from the ground truth.

Conditional GANs [9] built on this concept by providing images with their actual labels during training to help the model learn the difference between different classes of images. Regmi and Borji [11] propose a solution to the cross-view image synthesis problem using a cGAN approach. They enforce the network to generate a semantic map of the target view, which facilitates the generation of a real image. Tang et al. [30] address this challenge with their proposed SelectionGAN, which leverages semantic information through a two-stage process. In contrast, our model does not require additional information from the target domain, demonstrating a clear advantage for real-world applications.

2.2. Imitation Learning

Imitation learning in autonomy was first introduced with ALVINN [10], where a supervised learning approach was taken to produce the necessary curvature the vehicle would need to follow. Simulated road images and lidar data were used as inputs and expert outputs were provided based on the actual road curvature. This was extended with inverse reinforcement learning [12], where the agent is allowed to infer a reward function from expert demonstrations. Inverse reinforcement learning tends to have poor sample efficiencies, which is improved by generative adversarial imitation learning [6]. Here they connect imitation learning with generative adversarial networks by setting up a discriminator which distinguishes between expert policy and learned policy and pitting it against a generator that produces samples using the learned policy. The GAN is a more efficient way to match the expert data distribution with the learned data distribution than behavioral cloning. In the interest of autonomous driving, Carla Simulator [4] is a popular platform for research and provides a high-fidelity simulated environment where we can set up the autonomous agent with multiple camera sensors. [3] introduced an imitation learning model where the learning is conditioned on a high-level direction signal. The vehicle's velocity is used as an input, along with an RGB camera image and the directional command to produce controls for the throttle, brake, and steering. Generative Adversarial Imitation Learning was extended to the Carla simulator and tested in an end-to-end setting by [2, 7]. They introduce multi-stage training where a privileged agent is used to train a standard unprivileged agent.

3. IMITATION LEARNING BASELINE

Given sensor data observation $\mathbf{o}$ and high-level command $\mathbf{c}$, it is trained to output vehicle controls that are close to ground truth $\mathbf{a}$. In particular, observation $\mathbf{o}$ contains a single camera view and ego car speed. Ground truth $\mathbf{a}$ indicates controls for the throttle, brake, and steering of the vehicle, which is obtained from an expert driving AI agent that leverages privileged information about the scene to drive naturally and perform well in a simulator environment. The loss function of the imitation model $I(\cdot)$ can be defined as:

$$\mathcal{L}_{imi} = \|I(\mathbf{o}, \mathbf{c}) - \mathbf{a}\|_1 \quad (1)$$

To provide a baseline for testing our view translation network in the end-to-end driving scenario, we utilize a pre-trained conditional imitation learning model from [3]. This model adopts ResNet [20] architecture as the perception backbone, learning reactions to dynamic objects and traffic lights in complex urban environments. To reduce the dependency on input speed as the sole indicator of the scene's dynamics, they jointly train a network to predict the vehicle's speed. The decision to re-utilize a checkpoint is made as the model has been trained on 100 hours of realistic simulated data, which is significantly more than the size of our current data set.

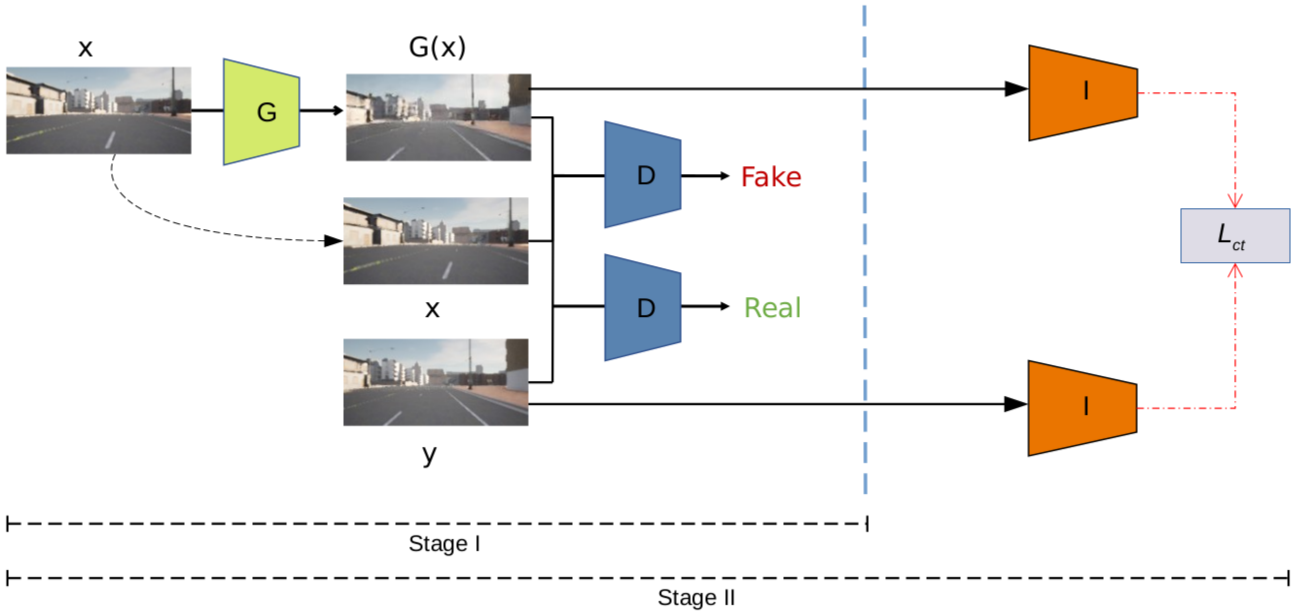

Figure 1: Overview of the proposed method. In stage I, we employ the improved adversarial loss to train a conditional GAN for generating front view image G(x) from left view image x. We further fine-tuned it with imitation learning model ***I*** in the second stage by comparing the vehicle controls generated by ground truth y with those produced by the synthesized view G(x).

4. IMAGE TRANSLATION MODEL

Given a pair of images (**x**,**y**), our objective is to learn a mapping from an auxiliary left view image **x** to front view image **y** in the simulated driving environment. Specifically, the generated image is intended to produce driving controls that match those derived from the front-view image when passed through the imitation model. We propose a conditional adversarial framework for the task and explain the model in two stages. We first review the baseline model pix2pix[19] and then describe modifications we made for image quality improvements(Sec.4.1). Next, we explain how we fine-tune our image translation model in conjunction with the imitation learning model, which ensures the generated images produce the correct vehicle controls when passed through the imitation model. (Sec.4.2). An illustration of our method is depicted in Fig.1.

4.1 The pix2pix baseline and improvements

We start from the baseline pix2pix[19] model, which adopts the U-Net[25] architecture as the generator and a patch-based, fully convolutional network[26] as the discriminator. The input to the generator is the left-view image without Gaussian noise, which leads to deterministic outputs. The input to the discriminator is a channel-wise concatenation of the source image and the corresponding real front-view image. We adopt a non-saturating adversarial loss in [5] to train the network,

$$\begin{aligned} \mathcal{L}_{cGAN} = & \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim p_{\text{data}}(\mathbf{x},\mathbf{y})}[\log D(\mathbf{x},\mathbf{y})] + \\ & \mathbb{E}_{\mathbf{x}\sim p_{(\mathbf{x})}}[1 - \log D(\mathbf{x}, G(\mathbf{x}))], \end{aligned} \tag{2}$$

where $G(\cdot)$ and $D(\cdot)$ denote the generator and discriminator, respectively.

Following [19], we also utilize L1 loss to ensure the generator captures the low frequencies beyond modeling the distribution of the target domain and synthesizes images close to desired target,

$$\mathcal{L}_{L1} = \|\mathbf{y} - G(\mathbf{x})\|_1. \quad (3)$$

Improved adversarial loss. We improve the GAN loss in the pix2pix model by incorporating a perceptual loss term LPIPS[21]. This loss stabilizes the training as the generator has to produce similar feature maps at multiple scales. In particular, the generator is forced to synthesize an image with matching intermediate representations of the real image by extracting from multiple layers of the pre-trained network. The feature matching loss is defined as:

$$\mathcal{L}_{LPIPS} = \|F(\mathbf{y}) - F(G(\mathbf{x}))\|_2^2, \quad (4)$$

where $F(\cdot)$ denotes a pre-trained VGG[21] network used for perceptual feature extraction.

Our final loss for the first stage of learning is a weighted sum of the above losses:

$$\mathcal{L}_{GAN} = \mathcal{L}_{cGAN} + \lambda_1 \mathcal{L}_{L1} + \lambda_2 \mathcal{L}_{LPIPS}, \quad (5)$$

where λ_1 and λ_2 are parameters to control the relative importance of different components.

4.2 Fine-tune with Imitation Learning model

Ideally, the synthesized image should be able to generate exact or similar controls to those generated by the actual front-view image. To achieve this, we further fine-tune our image translation model with pre-trained imitation learning model mentioned in Sec.3. As the high-level command **c** and current car speed are shared between the left and front view image pair, we simplify the control loss here as:

$$\mathcal{L}_{ct} = \|I(\mathbf{y}) - I(G(\mathbf{x}))\|_1, \quad (6)$$

where $I(\cdot)$ refers to imitation learning model. Intuitively, Eq.(6) acts as a regularization that forces the generator to output an image with control related features that match those from the corresponding ground truth image. Combined with the loss in the first stage, our objective function for stage two is

$$\mathcal{L}_{fine} = \mathcal{L}_{GAN} + \lambda_3 \mathcal{L}_{ct}. \quad (7)$$

During training, we keep $I(\cdot)$ untouched and only back-propagate the gradients from $\mathcal{L}_{ct}$ to the generator.

Note that GANs are often difficult to train and collapse without carefully selected hyperparameters and regularization [27, 28]. We observed that incorporating $\mathcal{L}_{ct}$ at the initial stage of training can impede the convergence of the GAN model. To mitigate this issue, we adopt such two-step training approach, where the control regularization term is only employed once the generation process has stabilized.

5. EXPERIMENTS

5.1. Data Collection

In accordance with the methodology described in [3], data were collected in the Carla Simulator using an expert driving agent. The data collection process involved recording images, velocity, and directional inputs during a series of episodes. Each episode consisted of a designated start location and target location, with data only being recorded if the expert agent successfully completed the episode. To add diversity to the data, the simulator was populated with randomly generated actors, including vehicles and pedestrians, and the weather was randomly selected from four pre-designated training weather conditions. The left camera was positioned with an outward angle of 15 degrees. To ensure the correct synchronization of data, the data from each sensor was recorded at the same timestamp. In summary, we collected a dataset of 207,000 realistic simulated driving samples, each comprising a pair of left-view images, front-view images, and the corresponding measurements.

5.2 Training details

We adapt the model architecture from the pix2pix method and modify it to be compatible with our image size of 200x88. During the initial training stage, we set $\lambda_1 = \lambda_2 = 10$ in Eq.(5). We employ minibatch SGD using the Adam solver [29], with a learning rate of 0.0002, and momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$. In the subsequent fine-tuning stage, we adjust the learning rate to 0.00001 and use $\lambda_3 = 200$ in Eq.(7) to compensate for the relatively small signal from the imitation model. Our model is trained for 500 epochs and fine-tuned for 200 epochs.

5.3. Experimental Setup

We evaluate the performance of our image translation model on two benchmarks in the Carla Simulator. The first benchmark is the NoCrash benchmark, which is more complex and was the primary focus of our fine-tuning and improvement efforts. The second benchmark is the standard Carla Benchmark, which includes simpler tasks.

NoCrash benchmark, introduced in [3], measures the driving performance on a set of complex episodes under various conditions. Each episode consists of a designated start and end point and typically requires the vehicle to navigate through several turns. The benchmark includes multiple testing conditions, including the training conditions, new weather, new town, and new town and weather. The training conditions feature the same four weather conditions and town used in the data collection process. The new weather scenario adds two previously unseen weather conditions, WetSunset and SoftRainSunset. The new town scenario changes the town to Town02, which is similar to the training town, Town01, but is smaller and has different buildings and backgrounds. The new town and weather scenario combines the previous two additions to create a more challenging scenario. The simulation is also run with three different tasks: empty, regular, and dense, which describe the number of other actors in the simulation, with more actors increasing the difficulty of navigation. A given episode is considered a failure if a collision occurs or the vehicle does not reach the target destination within the allotted time. As shown in Table 1, the results are reported as the percentage of completed episodes.

Carla Benchmark evaluates the driving performance on four tasks that increase in difficulty. The tasks include straight, one turn, navigation, and navigation dynamic, with the only task that includes other actors being the navigation dynamic. The other three tasks are all tested in an empty town and mainly assess the vehicle's ability to follow the lane and complete a turn. The conditions refer to the weather and town scenarios, with the new weather and town scenario being the most challenging for generalization.

5.4. Results

Qualitative Results. The offline image translation results are displayed in Fig.2, where the input images are loaded as mini-batches from the hard drive. Our model is observed to generate visually appealing images that closely match the ground truth front view. In particularly, the generated images accurately capture crucial driving details, such as solid lines, road shoulders, and oncoming vehicles. Furthermore, our model is capable of real-time image translation, where the input left views are directly obtained from the left camera on an ego car in the Carla environment. As demonstrated in Fig.3, the generated front views are faithful and are produced without latency (better viewed in video), ensuring the self-driving functionality is not compromised when the signal from main sensor is unavailable. These results indicate that our method can be effectively applied in empirical end-to-end driving scenarios.

Quantitative Results. As demonstrated by Table 1 and Table 2, our translation model exhibits a favorable comparison with the imitation learning baseline when evaluated under the training

Figure 2: Input left view, ground truth front view, and our synthesized front view are shown respectively

conditions. More importantly, our generated images even outperform the baseline in three tasks from the Carla benchmark and one task from the NoCrash benchmark. These results are noteworthy as the performance of the model does not significantly decrease as more agents are introduced into the simulation, which highlight the capability of our model in translating images from different angles without losing important features crucial for driving controls.

However, our method exhibits a decrease in driving performance compared to the baseline as new weather conditions and a new town are introduced. In particular, when evaluated in the Empty task scenario from the NoCrash benchmark, a 55% decrease in performance is observed with the introduction of unseen weather. Similar performance drops are also observed in the New Town scenario and the more challenging New Town and Weather scenario. This degradation in performance can be attributed to the inherent limitations of GANs, which tend to struggle in generalizing to unseen scenarios. Fine-grained information is required by the generator to synthesize ground truth images, making it challenging to perform well on unseen weather or town conditions without prior training data.

NoCrash	**Empty**	**Regular**	**Dense**
Training Conditions	**0.97**/0.96	**0.81**/0.88	**0.31**/0.49
New Weather	**0.45**/1.0	**0.44**/0.82	**0.20**/0.46
New Town	**0.29**/0.67	**0.22**/0.53	**0.08**/0.25
New WeatherTown	**0.10**/0.6	**0.02**/0.54	**0.06**/0.22

Table 1: Results from NoCrash benchmark. Driving performances are displayed based on the percentage of episodes completed in each experiment. Results from our best translation model are displayed in bold to the left of the baseline performance.

Carla	**Straight**	**One turn**	**Dyn**	**Nav. Dyn**
Training Conditions	**1.0**/1.0	**0.99**/0.97	**0.95**/0.99	**0.95**/0.90
New Weather	**0.56**/1.0	**0.38**/0.82	**0.16**/0.46	**0.04**/0.98
New Town	**0.29**/0.67	**0.22**/0.53	**0.08**/0.25	**0.35**/0.62
New WeatherTown	**0.78**/1.0	**0.56**/0.96	**0.28**/0.98	**0.30**/0.96

Table 2: Results from Carla benchmark. Driving performances are displayed based on the percentage of episodes completed in each experiment. Results from our best translation model are displayed in bold to the left of the baseline performance.

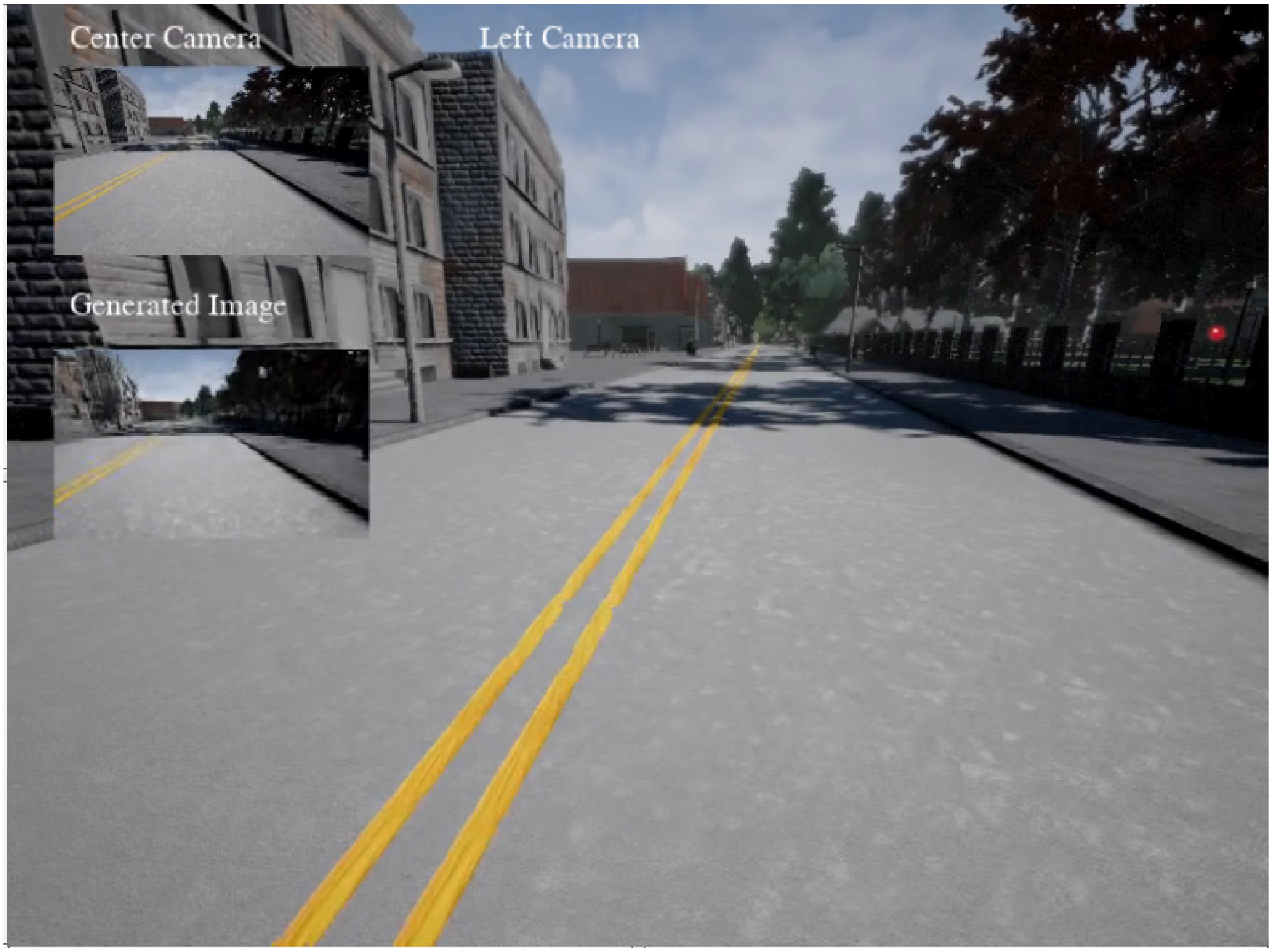

Figure 3: Screenshot of real-time image translation in Carla. Left view captured by the sensor is shown in the middle while synthesized front view is below ground truth at the top left corner. Note that we do not input front view to the imitation model and vehicle is controlled by the outputs from synthesized view.

6. CONCLUSION

The results in this paper suggest that conditional GANs can handle cross-view image translation tasks without any auxiliary information from the target domain as guidance. By incorporating a perceptual loss and fine-tuning with an imitation learning model, our approach exhibits strong performance in driving scenarios that share a similar statistical distribution with training data. Despite the inherent lack of generalization for unseen data in GANs, our method highlights the potential for using an alternate camera to synthesize missing information necessary for autonomous vehicles. We anticipate that this work can be extended to other scenarios, such as addressing wider camera angle discrepancy or generating ground views from aerial camera streams, to create a robust autonomous system.

Acknowledgment: This work was supported by the Automotive Research Center (ARC), a US Army Center of Excellence for modeling and simulation of ground vehicles, under Cooperative Agreement W56HZV-19-2-0001 with the US Army DEVCOM Ground Vehicle Systems Center (GVSC).

7. REFERENCES

[1] Abraham, T., Todd, A., Orringer, D. A., & Levenson, R. (2021). Applications of artificial intelligence for image enhancement in pathology. Artificial Intelligence and Deep Learning in Pathology, 119–148.

[2] Chen, D., Zhou, B., Koltun, V., & Krähenbühl, P. (2019). Learning by Cheating. CoRL. http://arxiv.org/abs/1912.12294

[3] Codevilla, F., Santana, E., Lopez, A., & Gaidon, A. (2019). Exploring the limitations of behavior cloning for autonomous driving. Proceedings of the IEEE International Conference on Computer Vision, 2019-Octob(Cvc), 9328–9337. https://doi.org/10.1109/ICCV.2019.00942

[4] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017). CARLA: An Open Urban Driving Simulator. CoRL, 1–16.

[5] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139–144.

[6] Ho, J., & Ermon, S. (2016). Generative adversarial imitation learning. Advances in Neural Information Processing Systems, 4572–4580.

[7] Karl Couto, G. C., & Antonelo, E. A. (2022). Generative Adversarial Imitation Learning for End-to-End Autonomous Driving on Urban Environments. 1–7. https://doi.org/10.1109/ssci50451.2021.9660156

[8] Karras, T., Laine, S., & Aila, T. (2021). A Style-Based Generator Architecture for Generative Adversarial Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(12), 4217–4228. https://doi.org/10.1109/TPAMI.2020.2970919

[9] Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets. 1–7. http://arxiv.org/abs/1411.1784

[10] Pomerleau, D. A. (1989). ALVINN: an autonomous land vehicle in a neural network (Technical Report CMU-CS-89-107). Advances in Neural Information Processing Systems, 305–313.

[11] Regmi, K., & Borji, A. (2018). Cross-View Image Synthesis Using Conditional GANs. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 3501–3510. https://doi.org/10.1109/CVPR.2018.00369

[12] Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., & Abbeel, P. (2017). Domain randomization for transferring deep neural networks from simulation to the real world. IEEE International Conference on Intelligent Robots and Systems, 2017-Septe, 23–30. https://doi.org/10.1109/IROS.2017.8202133

[13] Zhu, J., Park, T., Isola, P., & Efros, A.A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 IEEE International Conference on Computer Vision (ICCV), 2242-2251.

[14] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 8798–8807, 2018.

[15] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In Proceedings of the IEEE/CVF Conference on

Computer Vision and Pattern Recognition, pages 8188–8197, 2020.

[16] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision (ECCV), pages 172–189, 2018.

[17] Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A Efros, Oliver Wang, and Eli Shechtman. Toward multimodal image-to-image translation. In Advances in neural information processing systems, pages 465–476, 2017.

[18] M. Mirza and S. Osindero. Conditional generative adversarial nets. CoRR, abs/1411.1784, 2014. 1, 2, 3

[19] Isola, Phillip et al. Image-to-Image Translation with Conditional Adversarial Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 5967-5976.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun.Deep residual learning for image recognition. In CVPR,2016.

[21] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition.arXiv preprint arXiv:1409.1556, 2014.

[22] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 586–595, 2018.

[23] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In European Conference on Computer Vision (ECCV),2016

[24] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image superresolution using a generative adversarial network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017

[25] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015.

[26] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015

[27] Andrew Brock, Theodore Lim, J. M. Ritchie, and Nick Weston. Neural photo editing with introspective adversarial networks. arXiv:1609.07093, 2016

[28] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. arXiv:1809.11096, 2018.

[29] D. Kingma and J. Ba. Adam: A method for stochastic optimization. ICLR, 2015.

[30] Tang, Hao et al. “Multi-Channel Attention Selection GAN With Cascaded Semantic Guidance for Cross-View Image Translation.” 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019): 2412-2421.