Multi-Domain Conditional Image Translation: Translating Driving Datasets from Clear-Weather to Adverse Conditions

Vishal Vinod, K. Ram Prabhakar, R. Venkatesh Babu, Anirban Chakraborty

Department of Computational and Data Sciences, Indian Institute of Science, Bangalore, India

2021 ICCV OCTOBER 11-17 VIRTUAL

Introduction

Vision systems for autonomous navigation must perform well in unstructured and degraded scenarios. In most driving datasets, there is a bias toward clear-weather as compared with extreme-weather owing to the difficulty in capturing and annotating large-scale image datasets degraded by weather.

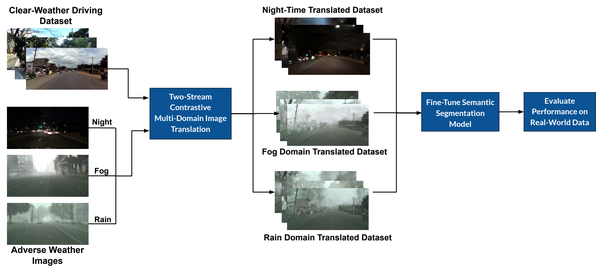

Figure 1:Overview of the multi-domain translation method.

In this work, we present a conditional multi-domain image translation method to translate clear-weather datasets to adverse-weather. We evaluate our method on the semantic scene understanding task and demonstrate superior translation results from clear-weather conditions to adverse-weather domains such as Rain, Night and Fog and show improved domain-invariant content disentanglement, and segmentation methods trained with datasets translated using the proposed method have improved performance over baselines.

Experimental Methodology

An image $\{S_n\}$ from the clear-weather source domain $\mathbf{S}$ is the content stream input and an image $\{T_m\}$ from the adverse-weather (Fog, Rain and Night-time) target domain ($\{T_{fog}, T_{rain}, T_{night}\} \in \mathbf{T}$), is the style stream conditioning input. The generator $\mathbf{G}$ draws a latent vector z_0 from a random Gaussian distribution and utilizes, (a) the style feature map m^i_{style} and style features f^i_{style} to fuse the target domain style with the content, and (b) the content feature map $m^i_{content}$ and style features $f^i_{content}$ to fuse the source domain content with the style at each of the i-th blocks ($i = 7$) to perform multi-domain image-to-image translation. The output from the generator is the translated image in the conditioned domain. We compute the adversarial loss ($\mathcal{L}_{GAN}$) with a multi-scale discriminator to improve the realism of the translation, and compute a modified perceptual loss ($\mathcal{L}_{Perceptual}$) for improved content consistency. A multi-scale patch-wise contrastive loss ($\mathcal{L}_{CTR}$) [3] is used to preserve source domain content and also translate the image with a less restrictive assumption than methods that impose cyclic consistency.

We show that using the above architecture, we are able to successfully perform the task of multi-domain conditional translation with improved results for the semantic scene understanding task. The proposed method can then be used to translate clear-weather datasets such as Cityscapes and IDD to each of the three adverse-weather domains as validated by the results.

Setup and Result Images

Figure 2:(Top) Two-Stream Contrastive Translation model. (Bottom) Qualitative results for the Cityscapes Dataset.

Quantitative Evaluation

Table 1. Comparison of performance for semantic segmentation models pre-trained on Cityscapes and fine-tuned on the respective domain-shifted datasets. The fine-tuned models are evaluated on adverse-weather datasets (in Mean IoU).

Fine-Tune (↓)	Foggy Zürich	ACDC-Rain	Dark Zürich
Cityscapes	32.55	41.37	14.61
CycleGAN	38.45	43.67	31.50
MUNIT	37.12	42.71	29.94
DRIT++	38.63	44.05	30.36
CUT	37.48	44.29	32.02
TSIT	39.81	44.41	33.46
Proposed	**41.69**	**44.56**	**35.36**

We see that our method outperforms other single-domain and multi-domain image translation methods including TSIT [1], DRIT++ [2] and MUNIT. The improvement over TSIT and DRIT++ especially for the Fog and Rain datasets is in part due to the better disentanglement of rain and fog styles when trained with a composite of Foggy-Cityscapes and Rainy-Cityscapes (and BDD100K-Night). This is because the synthetic rain images in Rainy-Cityscapes dataset has been created based on the visual effects of rain in real-world images using scene depth information to synthesize rain streaks and fog as a function of distance from the camera. Since both domains have synthetic fog at different intensities, it is thus necessary to effectively disentangle the rain and fog styles.

Qualitative Results

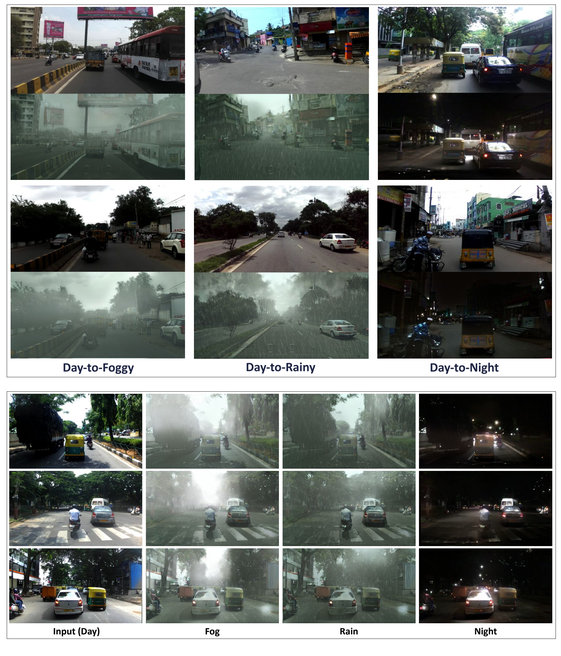

Figure 3:Qualitative results for the Indian Driving Dataset.

Discussion and Conclusion

In this work, we propose a two-stream multi-domain image translation method to effectively translate a driving dataset from clear-weather to adverse-weather domains: fog, rain and night-time. Our method is able to effectively capture domain-invariant content from the content stream using an adaptive denormalization method and fuses style from style stream with adaptive instance normalization at multiple feature scales. The proposed method uses a modified perceptual loss and a multi-layer patch-wise contrastive loss to disentangle and preserve content and structure from the source domain, and also utilizes multi-scale discriminators to learn the translational mapping from one domain to many domains when conditioned with the target domain image.

Table 2. Semantic segmentation generalization performance (in mIoU) when translation models are trained with 501 total images (167 images from each domain) strategy as Table 1.)

Fine-Tune (↓)	Foggy Zürich	Rainy-Cityscapes	Dark Zürich
TSIT [1]	34.11	54.89	29.64
Proposed	**34.98**	**56.12**	**30.02**

References

[1] Liming Jiang, Changxu Zhang, Mingyang Huang, Chunxiao Liu, Jianping Shi, and Chen Change Loy. Tsit: A simple and versatile framework for image-to-image translation. In *European Conference on Computer Vision*, pages 206–222. Springer, 2020.

[2] Hsin-Ying Lee, Hung-Yu Tseng, Qi Mao, Jia-Bin Huang, Yu-Ding Lu, Maneesh Singh, and Ming-Hsuan Yang. Drit++: Diverse image-to-image translation via disentangled representations. *International Journal of Computer Vision*, 128(10):2402–2417, 2020.

[3] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European Conference on Computer Vision*, pages 319–345. Springer, 2020.

Acknowledgements

This work was supported by a project grant from WIRIN (Wipro IISc Research Innovation Network). We would like to thank the anonymous reviewers for their valuable suggestions.

Contact Information

- Email: vishalvinod@acm.org
- Website: https://val.cds.iisc.ac.in/