Supplementary Material

In this supplementary material we discuss limitations and ethical implications of our method. Additionally, we provide more samples including their segmentation masks for our pseudo-target domain.

A. Discussion of Limitations

If data of a real domain is available, then the performance of current unsupervised domain adaptation (UDA) methods is still better than that of domain generalization (DG) method. However, given that UDA methods employ real-world data, and we only employ images that are generated by a diffusion model without accessing any real data, a comparison of these methods would not be adequate. As shown in Figure 4 and Figures 6, 7, 8, and 9 the text prompt has a crucial impact on the content of the generated image. We utilize a novel, systematic and modular text prompt strategy to obtain a diverse pseudo-target domain and validate it in Table 3. However, we believe that there is a strong potential to improve the prompt design and thereby the quality and diversity of the generated content. One idea might be automated prompt generation. Further details are left for future work.

Although our method does not require any data from any target domain, it should be noted that the employed diffusion models were naturally trained with real data.

B. Discussion of Ethical Implications

The application of DIDEX includes the generation of data with diffusion models. Biases in this generated data have to be carefully considered since they are dependent on how the employed diffusion models were trained. Additional biases can be caused by the prompt generation, particularly, by the location string. When different locations shall be represented this may lead to ethnic biases, which was not within the scope of this paper. Biases in the employed real datasets also have to be considered when assessing the generalization performance. Of our four real datasets two (Cityscapes [6] and ACDC [40]) were collected in Central Europe. BDD100k [56] contains images from the USA, and only Mapillary Vistas [28] represents multiple regions from all over the world. For real-world applications, especially when safety-critical, these biases have to be considered.
Although we have developed our method for applications such as automated driving or robotics, there is of course also the possibility of unintended use, e.g., surveillance or military applications. However, this is a general problem of methods that aim to make computer vision more robust.

C. Further Example Images

In the following we will provide more example images from both employed diffusion models with their respective prompt $\mathbf{\Phi}_n$ in Figures 6, 7, 8, and 9. We display the text prompt at the top of each respective row and the input image $\mathbf{x}_n^{\mathrm{S}}$ at the left of the row. The image $\mathbf{x}_n^{\mathrm{PT}}$ is generated based on the prompt displayed in the middle. Additionally it is constrained by a depth estimation which is not shown. The segmentation result of the generated image is displayed at the right. Our best domain-generalizing segmentation model $\mathbf{M}^{\mathrm{DG}}$ was used to create the prediction $\mathbf{M}^{\mathrm{DG}}(\mathbf{x}_n^{\mathrm{PT}})$. Similar to Figure 4 we can observe that generated images show semantic and structural inconsistencies. As expected, `ControlNet` [60] shows a higher degree of consistency than `SD2.0` [38]. However, even with strong inconsistencies the model adapts well to our pseudo-target domain as the segmentation predictions reveal. Please note that since we do not have labels for the pseudo-target domain we cannot calculate the corresponding mIoU.

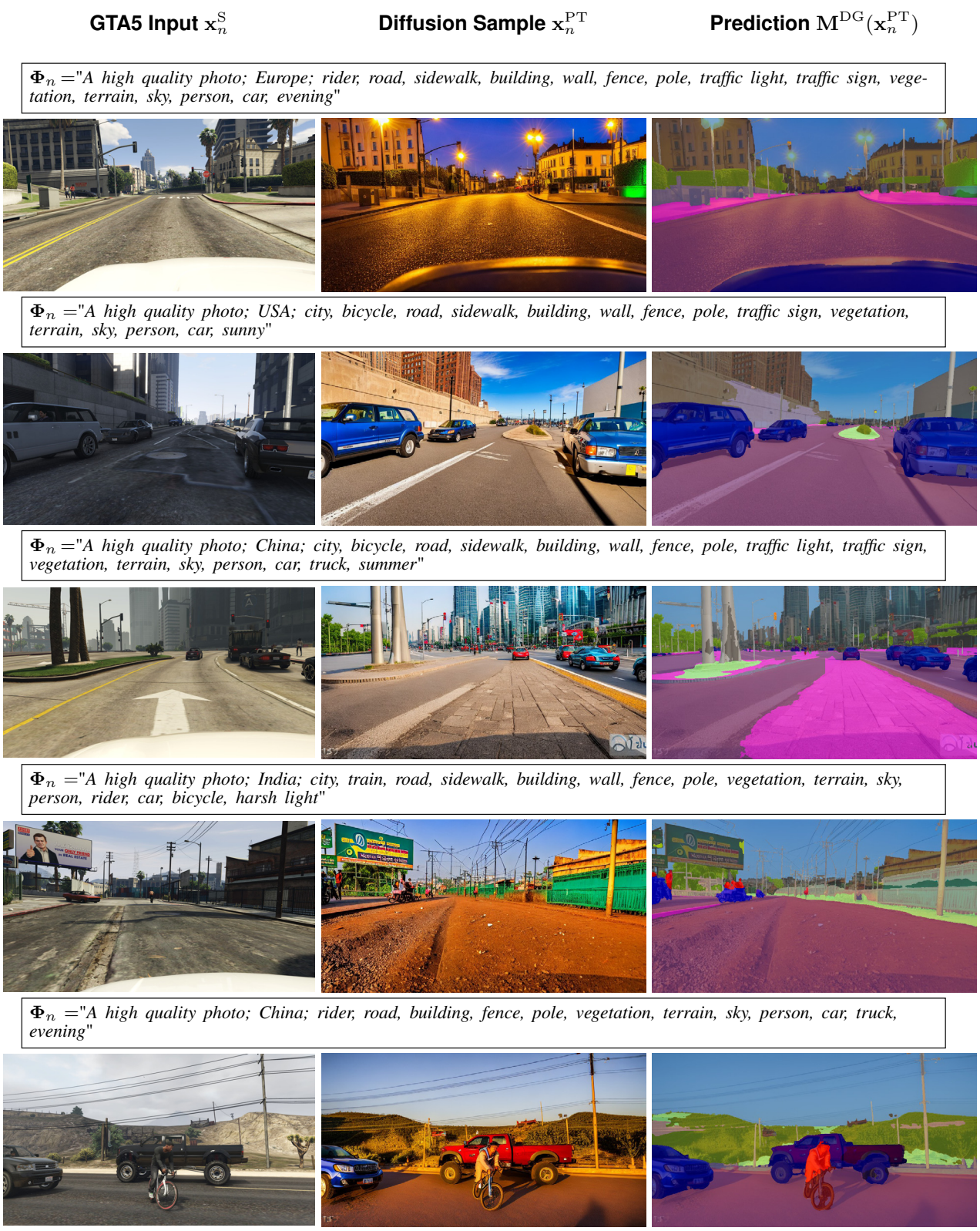


Figure 6. **Samples from the pseudo-target domain** $\mathcal{D}^{\mathrm{PT}}$ generated with `ControlNet` [60] with depth constraints and based on $\mathcal{D}^{\mathrm{S}} = \mathcal{D}^{\mathrm{GTA5}}_{\mathrm{train}}$ images. The predictions were obtained with our best domain-generalizing model $\mathbf{M}^{\mathrm{DG}}$.

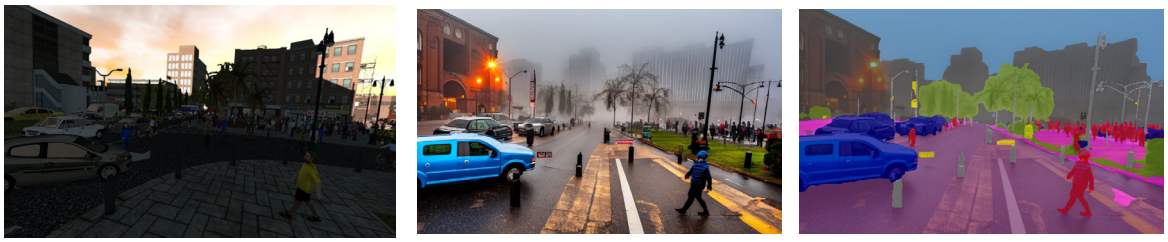
GTA5 Input $\mathbf{x}_n^{\mathrm{S}}$ **Diffusion Sample $\mathbf{x}_n^{\mathrm{PT}}$** **Prediction $\mathbf{M}^{\mathrm{DG}}(\mathbf{x}_n^{\mathrm{PT}})$**

$\boldsymbol{\Phi}_n$ ="*A high quality photo; Europe; terrain, road, sidewalk, building, wall, fence, pole, vegetation, sky, person, rider, car, bus, motorcycle, bicycle, snowy*"

$\boldsymbol{\Phi}_n$ ="*A high quality photo; India; Highway, truck, road, sidewalk, building, fence, pole, traffic light, traffic sign, vegetation, sky, person, rider, car, motorcycle, bicycle, spring*"

$\boldsymbol{\Phi}_n$ ="*A high quality photo; India; truck, road, sidewalk, building, fence, pole, traffic light, traffic sign, vegetation, sky, person, rider, car, bus, motorcycle, bicycle, sunny*"

$\boldsymbol{\Phi}_n$ ="*A high quality photo; USA; City, truck, road, sidewalk, building, wall, fence, pole, traffic light, traffic sign, vegetation, sky, person, rider, car, bus, motorcycle, bicycle, fog/mist*"



$\boldsymbol{\Phi}_n$ ="*A high quality photo; China; city, terrain, road, sidewalk, building, fence, pole, traffic light, traffic sign, vegetation, sky, person, rider, car, bus, motorcycle, bicycle, morning*"

Figure 7. **Samples from the pseudo-target domain $\mathcal{D}^{\mathrm{PT}}$** generated with `ControlNet` [60] with depth constraints and based on $\mathcal{D}^{\mathrm{S}} = \mathcal{D}^{\mathrm{SYN}}_{\mathrm{train}}$ images. The predictions were obtained with our best domain-generalizing model $\mathbf{M}^{\mathrm{DG}}$.

Figure 8. **Samples from the pseudo-target domain** $\mathcal{D}^{\mathrm{PT}}$ generated with `SD2.0` [38] with depth constraints and based on $\mathcal{D}^{\mathrm{S}} = \mathcal{D}^{\mathrm{GTA5}}_{\mathrm{train}}$ images. The predictions were obtained with our best domain-generalizing model $\mathbf{M}^{\mathrm{DG}}$.

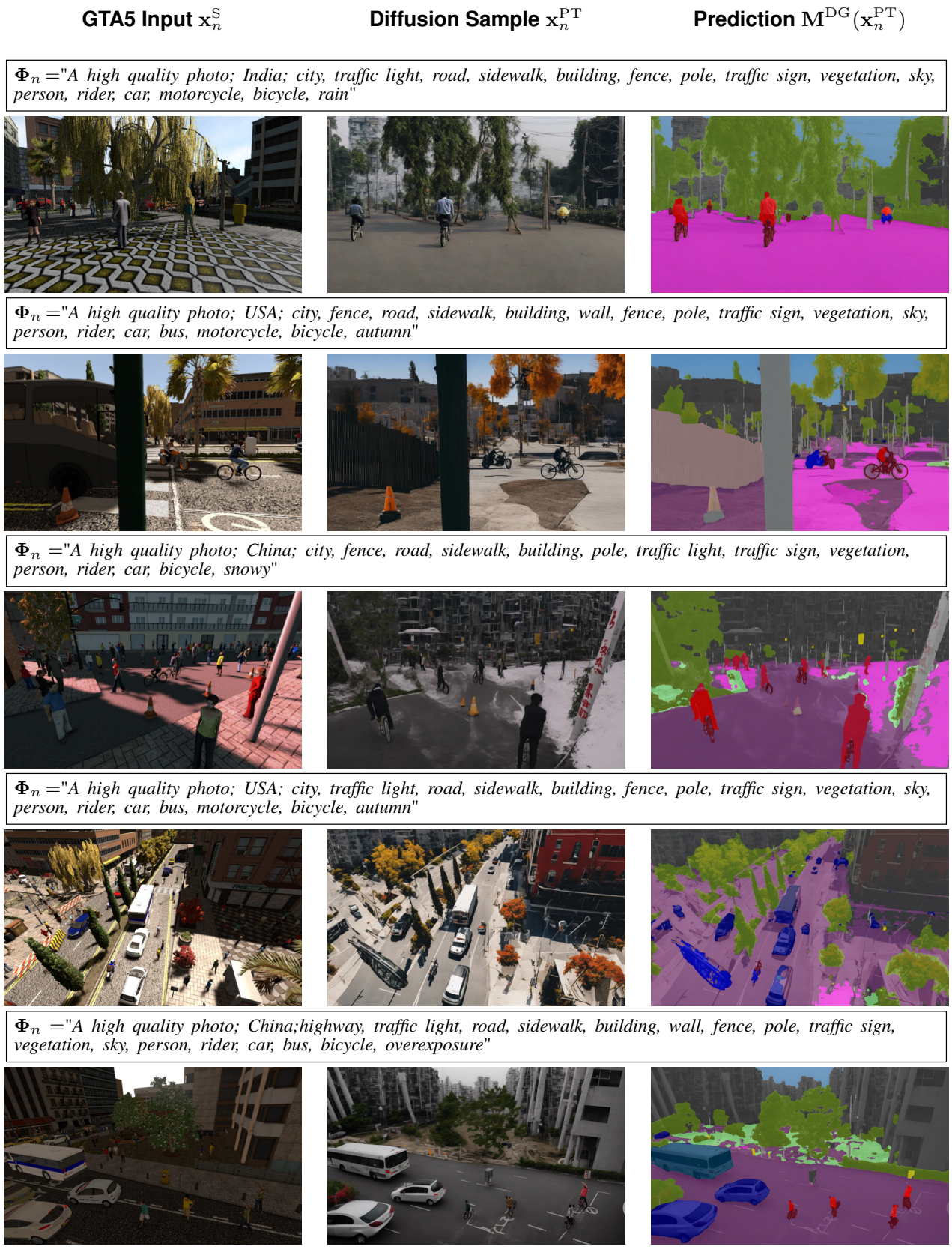


Figure 9. **Samples from the pseudo-target domain** $\mathcal{D}^{\mathrm{PT}}$ generated with `SD2.0` [38] with depth constraints and based on $\mathcal{D}^{\mathrm{S}} = \mathcal{D}^{\mathrm{SYN}}_{\mathrm{train}}$ images. The predictions were obtained with our best domain-generalizing model M^{DG}.