

Cross-View Regularization for Domain Adaptive Panoptic Segmentation (Supplemental Material)

1. More Illustrations in Domain Adaptive Panoptic Segmentation

We provide more qualitative illustrations that compare the proposed CVRN with state-of-the-art methods over three domain adaptive panoptic segmentation tasks as shown in Figures 1 to 3.

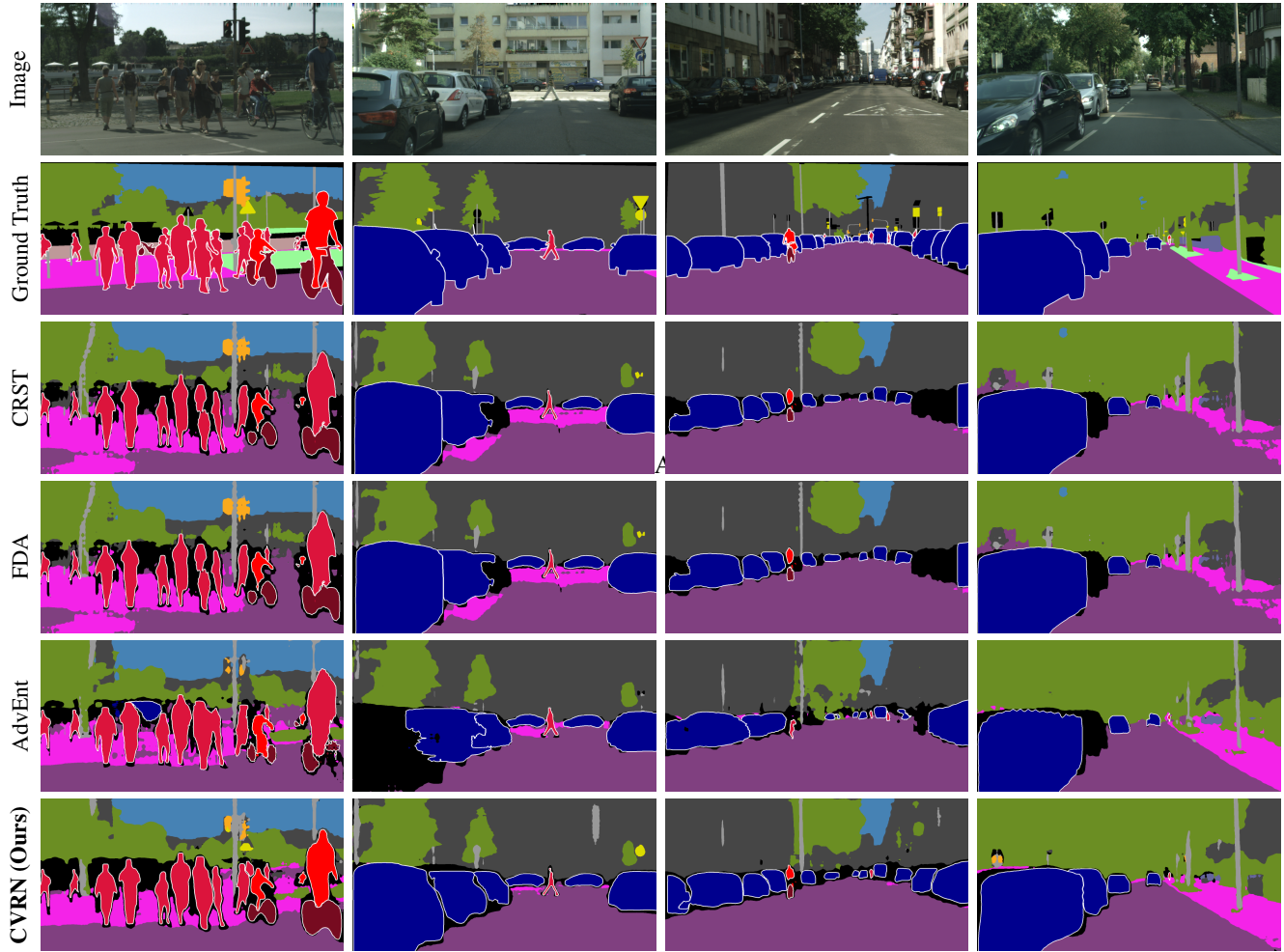


Figure 1. Qualitative comparison of CVRN with state-of-the-art methods *CRST* [6], *FDA* [5] and *AdvEnt* [4] for domain adaptive **panoptic segmentation** over the task “SYNTHIA \rightarrow Cityscapes” (synthetic-to-real). CVRN segments more accurate foreground “things” (*i.e.*, countable objects) and background “stuff” (*i.e.*, amorphous regions) in panoptic segmentation. The superior performance is largely attributed to the proposed cross-view regularization that encourages semantic consistency between different views. Best viewed in color and zoom in for details.

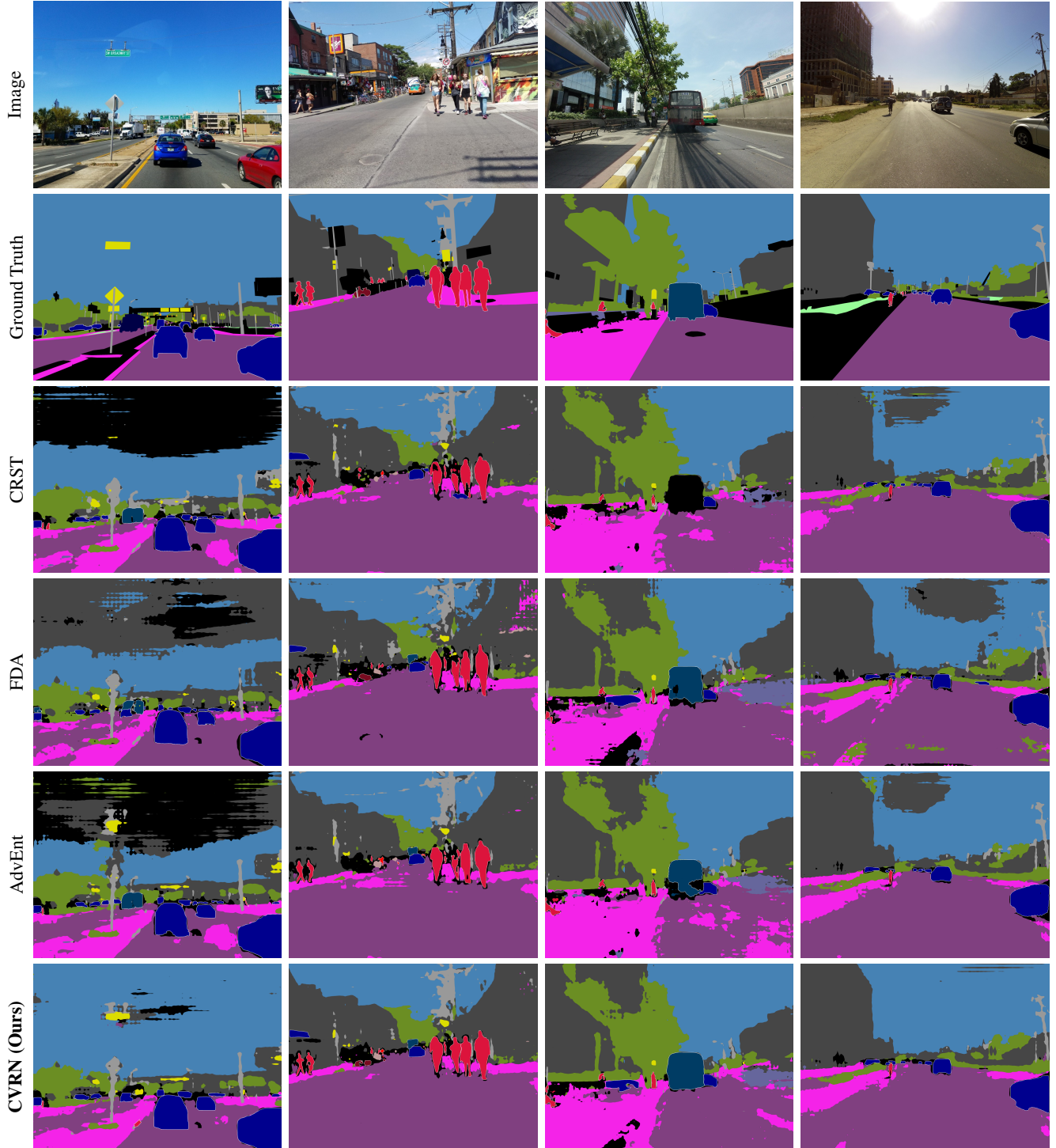


Figure 2. Qualitative comparison of CVRN with state-of-the-art methods *CRST* [6], *FDA* [5] and *AdvEnt* [4] for domain adaptive **panoptic segmentation** over the task “SYNTHIA \rightarrow Mapillary” (synthetic-to-real). CVRN segments more accurate foreground “things” (*i.e.*, countable objects) and background “stuff” (*i.e.*, amorphous regions) in panoptic segmentation. The superior performance is largely attributed to the proposed cross-view regularization that encourages semantic consistency between different views. Best viewed in color and zoom in for details.

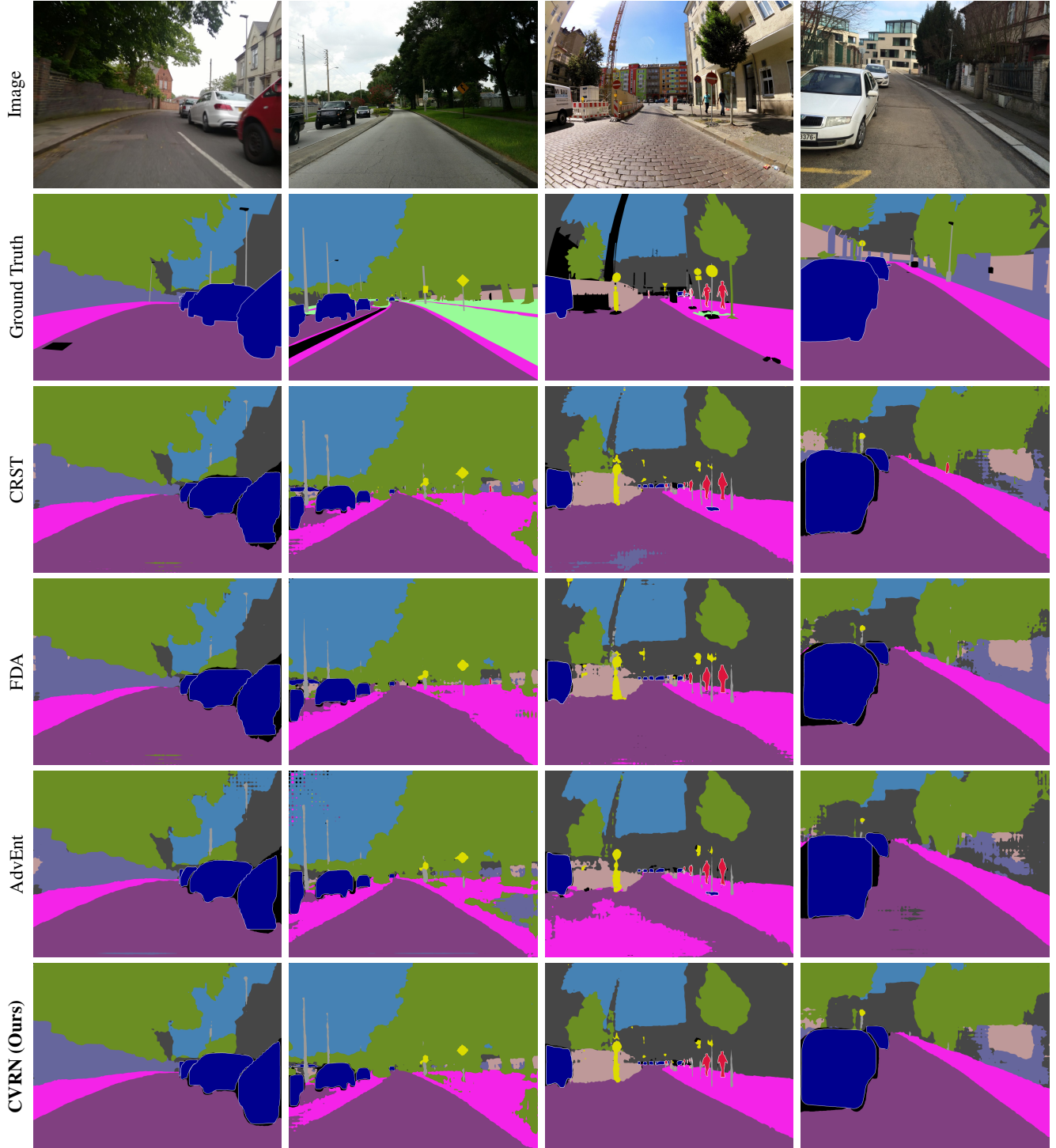


Figure 3. Qualitative comparison of CVRN with state-of-the-art methods *CRST* [6], *FDA* [5] and *AdvEnt* [4] for domain adaptive **panoptic segmentation** over the task “Cityscapes \rightarrow Mapillary” (real-to-real). CVRN segments more accurate foreground “things” (*i.e.*, countable objects) and background “stuff” (*i.e.*, amorphous regions) in panoptic segmentation. The superior performance is largely attributed to the proposed cross-view regularization that encourages semantic consistency between multiple views. Best viewed in color and zoom in for details.

2. More Illustrations in Domain Adaptive Semantic Segmentation

We provide more qualitative illustrations that compare the proposed CVRN with state-of-the-art methods over domain adaptive semantic segmentation task as shown in Figure 4.

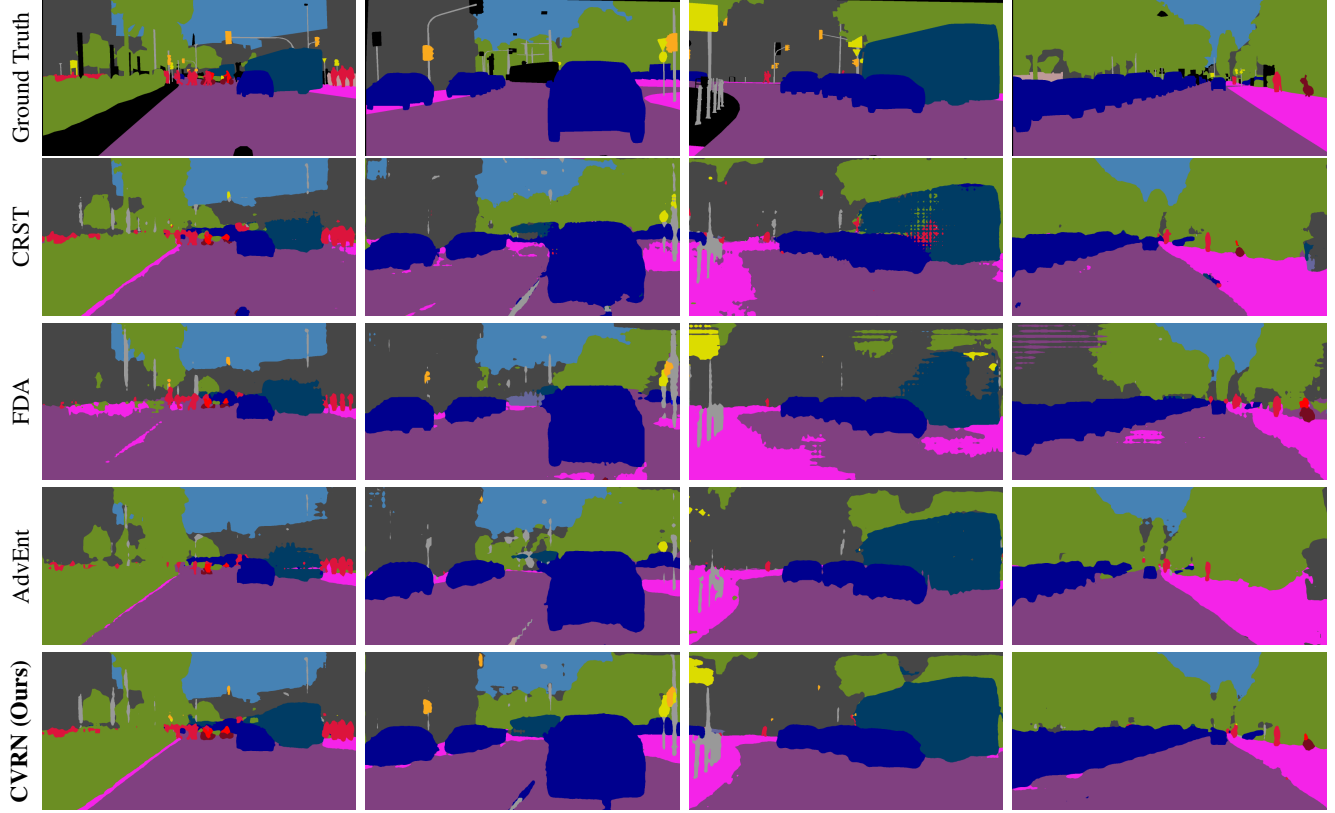


Figure 4. Qualitative comparison of CVRN with state-of-the-art methods *CRST* [6], *FDA* [5] and *AdvEnt* [4] for domain adaptive **semantic segmentation** over the task “**SYNTHIA** \rightarrow **Cityscapes**”. CVRN outperforms the state-of-the-art methods by providing more accurate pixel-level segmentation. The superior performance is largely attributed to the proposed cross-view regularization that exploits confident instance predictions to benefit semantic segmentation. Best viewed in color and zoom in for details.

3. Dataset Details

SYNTHIA [3] is a large-scale synthetic dataset with 9,400 images that are generated by random perturbation of virtual environments. This dataset provides pixel-level annotations for semantic segmentation as well as object-level labels for instance segmentation. Panoptic segmentation annotations can be obtained by fusing “stuff” regions as annotated for semantic segmentation with object labels as annotated for instance segmentation. All the images have the same resolution of 760×1280 .

Cityscapes [1] is a widely used autonomous driving dataset with images captured by an image acquisition system mounted in a driving vehicle. It consists of 2,975 training images and 500 validation images with dense manual annotations for panoptic segmentation. All the images have the same resolution of 1024×2048 .

Mapillary Vistas [2] is a large-scale autonomous driving dataset with images captured by different image acquisition sensors. It consists of 18,000 training images and 2,000 validation images with high-quality annotations for panoptic segmentation. The resolution of the dataset image varies from 768×1024 to 4000×6000 .

4. More Illustrations in Domain Adaptive Instance Segmentation

We provide more qualitative illustrations that compare the proposed CVRN with state-of-the-art methods over domain adaptive instance segmentation task as shown in Figure 5.

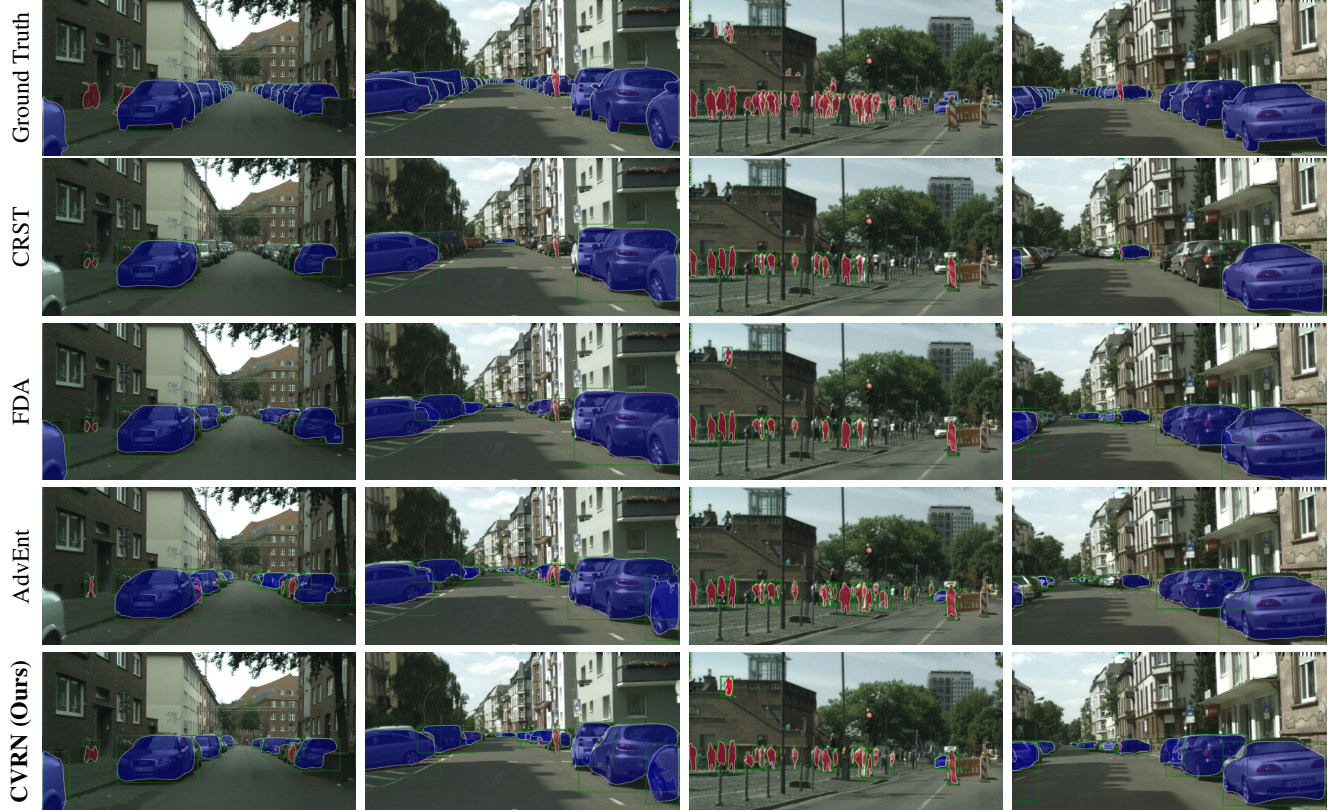


Figure 5. Qualitative comparison of CVRN with state-of-the-art methods *CRST* [6], *FDA* [5] and *AdvEnt* [4] for domain adaptive **instance segmentation** over the task “**SYNTHIA** \rightarrow **Cityscapes**”: CVRN outperforms the state-of-the-art by detecting more objects (*e.g.*, car and person) with more accurate boundaries and less false positives. The superior performance is largely attributed to the proposed cross-view regularization that exploits confident semantic predictions to benefit instance segmentation. Best viewed in color and zoom in for details.

References

- [1] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 5
- [2] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Buló, and Peter Kotschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4990–4999, 2017. 5
- [3] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3234–3243, 2016. 5
- [4] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2517–2526, 2019. 1, 2, 3, 4, 6
- [5] Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4085–4095, 2020. 1, 2, 3, 4, 6
- [6] Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. Confidence regularized self-training. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5982–5991, 2019. 1, 2, 3, 4, 6