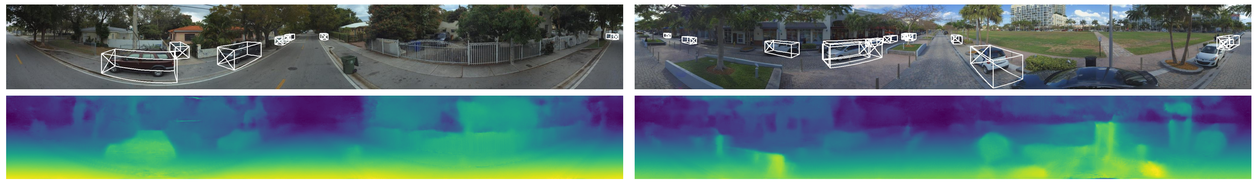
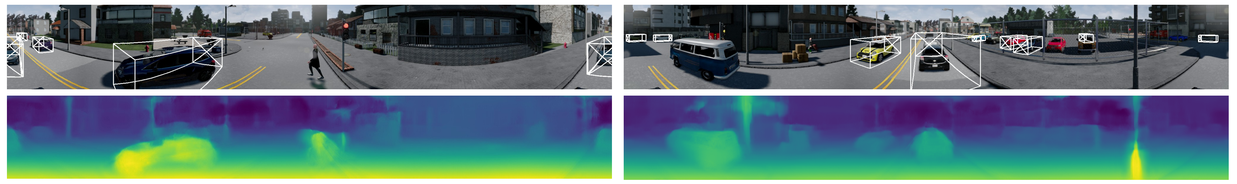
Eliminating the Blind Spot: Adapting 3D Object Detection and Monocular Depth Estimation to 360° Panoramic Imagery

Grégoire Payen de La Garanderie, Amir Atapour Abarghouei, and Toby Breckon
Department of Computer Science, Durham University, UK

https://gdlg.github.io/panoramic



Real-world imagery from Mapillary [4]



Synthetic imagery from CARLA simulator [5]

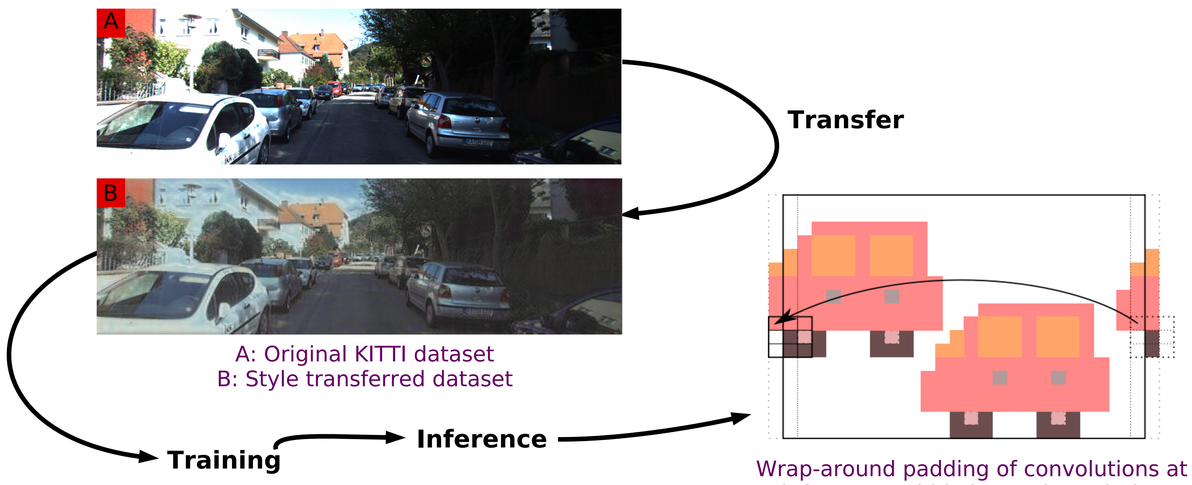
Issue: Provide 3D object detection and depth estimation within monocular 360° panoramic imagery for autonomous driving.

Motivation

(1) **Lack of 360° panoramic dataset for automotive applications.**
(2) **Can a monocular panoramic camera be used for 3D estimation tasks?**
(3) **Can existing models and training datasets be reused for panoramic imagery?**

Approach

(1) ***Reuse* existing neural network** architectures used for rectilinear imagery (MS-CNN [1], Monodepth [2]).
(2) ***Adapt existing datasets*** (KITTI) based on style (CycleGAN [3]) and projection transformation to be more similar to panoramic imagery used for testing.
(3) ***Extend*** network for monocular 3D detection in both rectilinear and panoramic imagery with additional outputs for object distance, orientation, size and position.
(4) ***Use wrap-around*** padding at inference time on convolutions to hide panoramic image boundaries, enabling a seamless processing.
(5) ***Test*** using real-world imagery from Mapillary [4] and synthetic imagery from the CARLA simulator [5].

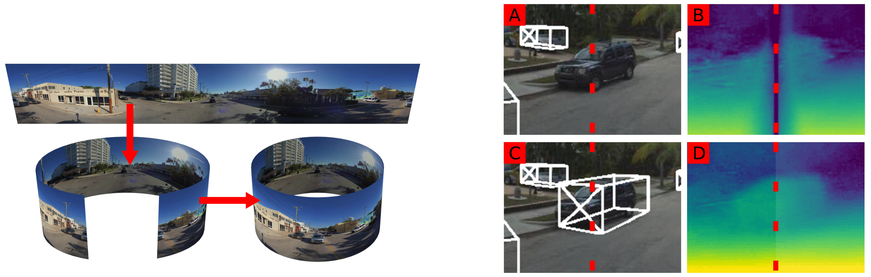


A: Original KITTI dataset
B: Style transferred dataset

Wrap-around padding of convolutions at inference to hide image boundaries

Results

(1) ***Qualitative analysis*** on real world imagery from Mapillary [4].
(2) ***Quantitative analysis*** on synthetic imagery from CARLA automotive simulator [5].



Effect of zero padding (A,B) and wrap-around padding (C,D)

		Detection	Depth Estimation				
Transform	Dataset	mAP	Abs. Rel.	Sq. rel.	RMSE	RMSE log	Depth acc. $\delta < 1.25$
none	K	0.34	0.25	7.7	3.5	0.46	0.70
proj.	K	0.24	0.25	7.4	3.5	0.44	0.73
style	C	0.35	0.26	7.7	3.6	0.48	0.69
style	M	0.36	0.26	7.9	3.6	0.47	0.68
style	M+C	0.38	0.23	6.3	3.6	0.47	0.68
style & proj.	C	0.26	0.30	9.6	3.7	0.47	0.72
style & proj.	M	0.31	0.30	10	3.8	0.47	0.72
style & proj.	M+C	0.36	0.23	6.4	3.6	0.46	0.72

Results on synthetic data from CARLA simulator

Conclusion

(1) Identify ***new set of challenging automotive problems*** about panoramic sensing.
(3) ***Exploration of dataset bias*** and transferrability from rectilinear to panoramic datasets.
(2) Provide ***new synthetic dataset*** as a performance benchmark.

[1] Cai, Z., Fan, Q., Feris, R.S., Vasconcelos, N.: A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection. In: Computer Vision – ECCV 2016. pp. 354–370. Springer International Publishing (2016)
[2] Godard, C., Mac Aodha, O., Brostow, G.J.: Unsupervised Monocular Depth Estimation with Left-Right Consistency. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6602–6611 (2017)
[3] Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2242–2251 (Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial)
[4] Mapillary: Mapillary Research. https://research.mapillary.com/
[5] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: An Open Urban Driving Simulator. CoRL (2017)