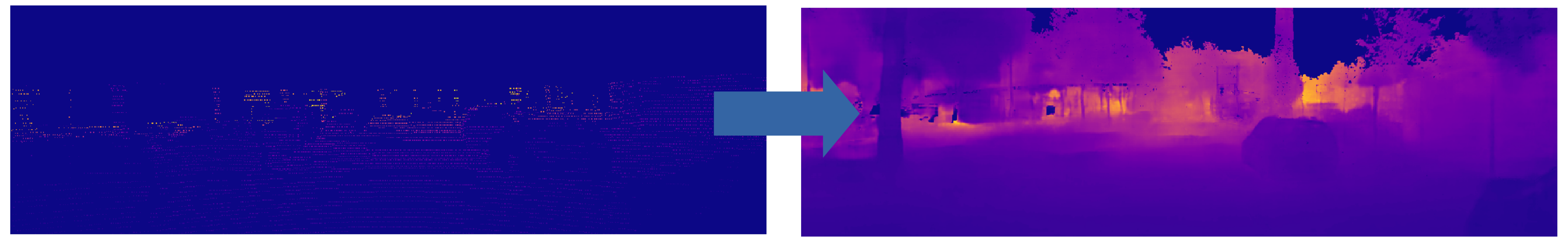
To complete or to estimate, that is the question:

A Multi-Task Approach to Depth Completion and Monocular Depth Estimation

Amir Atapour-Abarghouei and Toby P. Breckon, Durham University, UK

Monocular input → **Dense depth** output

Sparse depth (LiDAR) → **Dense depth** output

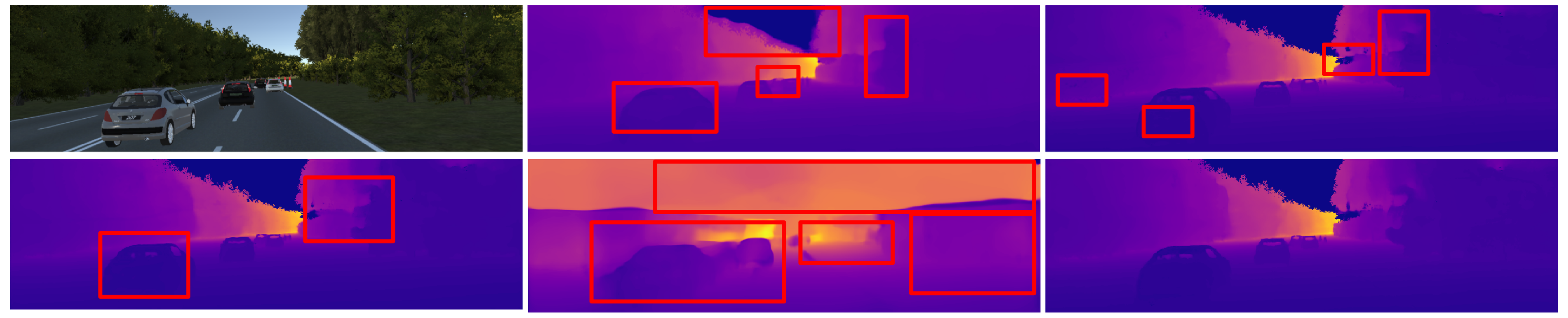


Contributions:

While the primary objective is **monocular depth estimation** via a mix of synthetic [1] and real-world [2] training images from urban driving scenarios, our method performs **sparse depth completion** in a **multi-task** dual-network approach without any retraining.

Our approach comprises **two networks**, is trained on **two data sets** and is based on a loss function with **three loss components**: L1 Loss, Adversarial Loss and a Smoothing Loss.

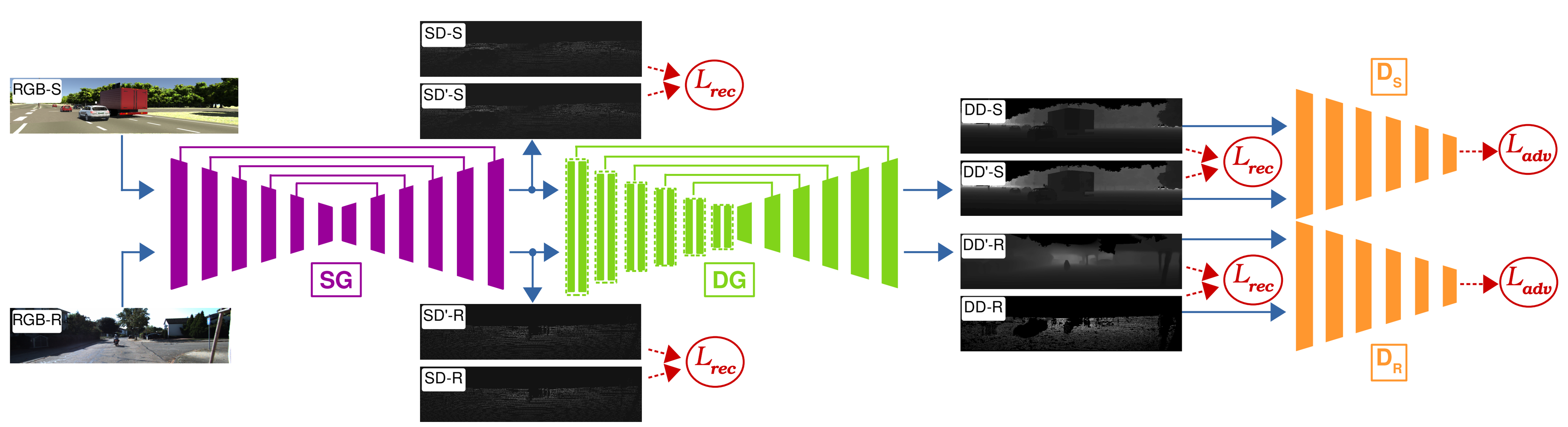
Extensive ablation studies and evaluations demonstrate the importance of every component of the approach.



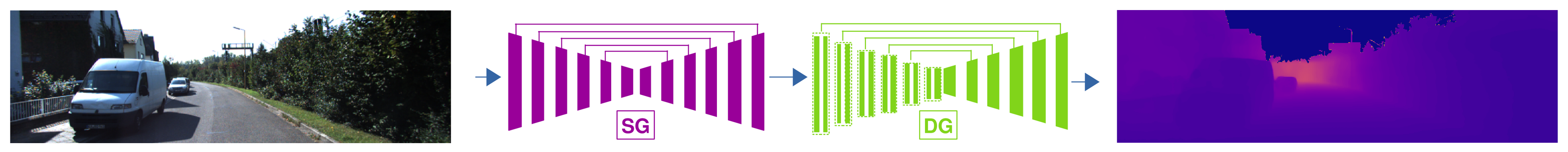
Ablation studies demonstrate the importance of every component of the approach.

Proposed Approach:

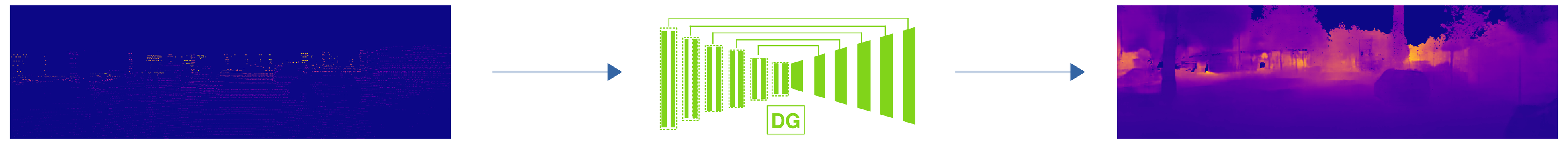
The first network, Sparse Generator, produces a sparse representation of the scene depth, which is then passed into the second network, Dense Generator, that produces the dense depth output. Two Discriminator networks ensure the overall quality of the final result.



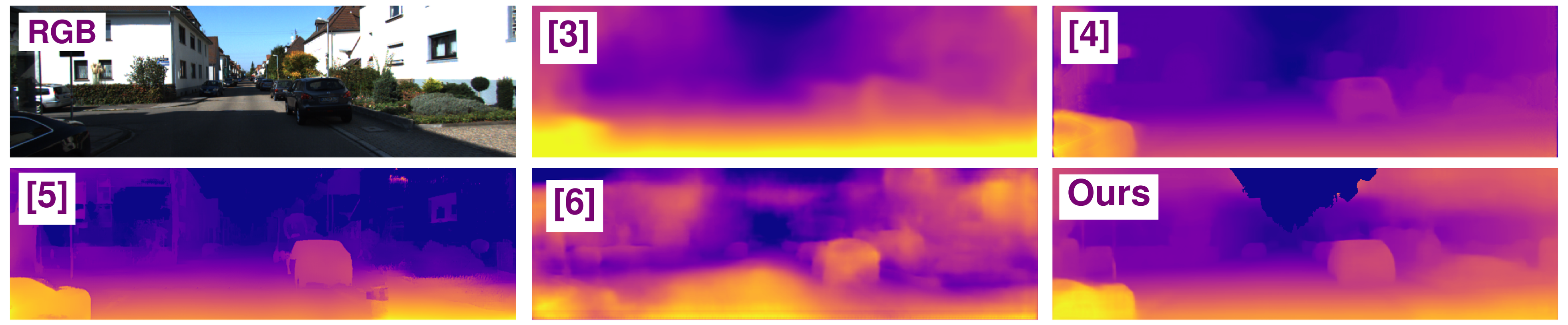
During inference, the entire model (Sparse Generator and Dense Generator) can be used to perform **monocular depth estimation**:



or the second network (Dense Generator) can be used alone to **complete a sparse depth image** captured using LiDAR without any need for retraining or fine tuning:



Results:



Superior qualitative results in **monocular depth estimation**.

Methods	Error Metrics				Accuracy Metrics		
	Abs. Rel.	Sq. Rel.	RMSE	RMSE log	$\sigma < 1.25$	$\sigma < 1.25^2$	$\sigma < 1.25^3$
Zhou et al. [3]	0.208	1.768	6.856	0.283	0.678	0.885	0.957
Godard et al. [4]	0.148	1.344	5.927	0.247	0.803	0.922	0.964
Atapour et al. [5]	0.110	0.929	4.726	0.194	0.923	0.967	0.984
Our Results	0.080	0.836	4.437	0.157	0.929	0.970	0.985

Superior quantitative results in **monocular depth estimation**.

Results of **sparse depth completion**. Note how the **upper regions** of the results are better synthesised in our approach due to the use of synthetic training data.

Generalisation capabilities are tested using data **locally** captured in Durham, UK.

[1] Francis et al., 'Exploring Spatial Context for 3D Semantic Segmentation of Point Clouds.' ICCV Workshops, 2017.
[2] Geiger, et al., 'Vision meets robotics: The KITTI dataset.' Robotics Research 2013.
[3] Zhou et al., 'Unsupervised learning of depth and ego-motion from video.' CVPR, 2017.
[4] Godard et al., 'Unsupervised monocular depth estimation with left-right consistency.' CVPR, 2017.
[5] Atapour et al., 'Real-time monocular depth estimation with domain adaptation via image style transfer.' CVPR, 2018.
[6] Kuznietsov, et al., 'Semi-supervised deep learning for monocular depth map prediction.' CVPR, 2017.
[7] Ma, et al., 'Sparse-to-Dense: Depth prediction from sparse depth samples and a single image.' ICRA, 2018.
[8] Shivakumar, et al., 'DFuseNet: Deep fusion of RGB and sparse depth information for image guided dense depth completion.' ArXiv, 2019.
[9] Eldesokey, et al., 'Confidence propagation through CNNs for guided sparse depth regression.' ArXiv, 2018.