DeepDR: Deep Structure-Aware RGB-D Inpainting for Diminished Reality

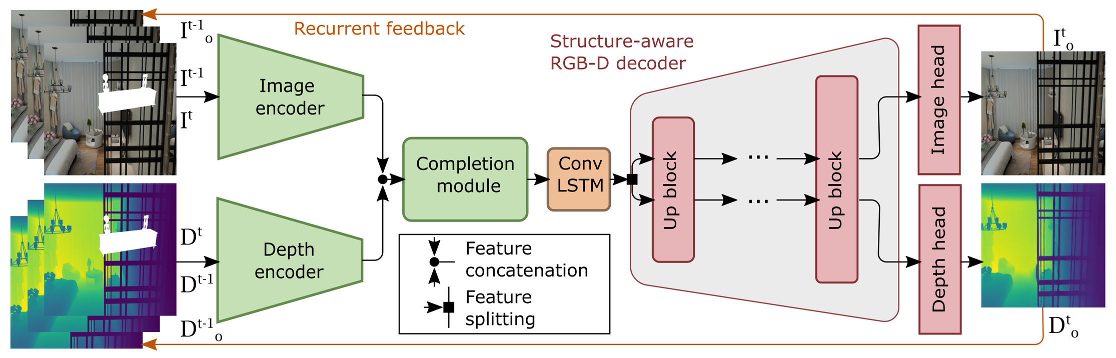
Christina Gsaxner[1], Shohei Mori[1], Dieter Schmalstieg[1,2], Jan Egger[1,3], Gerhard Paar[4], Werner Bailer[4] and Denis Kalkofen[1,5].

[1]Graz University of Technology, [2]University of Stuttgart, [3]University of Duisburg-Essen, [4]Joanneum Research, [5]Flinders University

Introduction

Diminished reality (DR) involves virtually removing real objects from the environment using inpainting techniques. However, existing methods struggle with maintaining coherent structure and 3D geometry, particularly for advanced tasks like 3D scene editing. In response, we introduce DeepDR, a real-time RGB-D inpainting framework tailored for DR, ensuring both realistic image and depth inpainting with minimal artifacts, achieved through a structure-aware generative network explicitly conditioned on scene semantics.

Network architecture



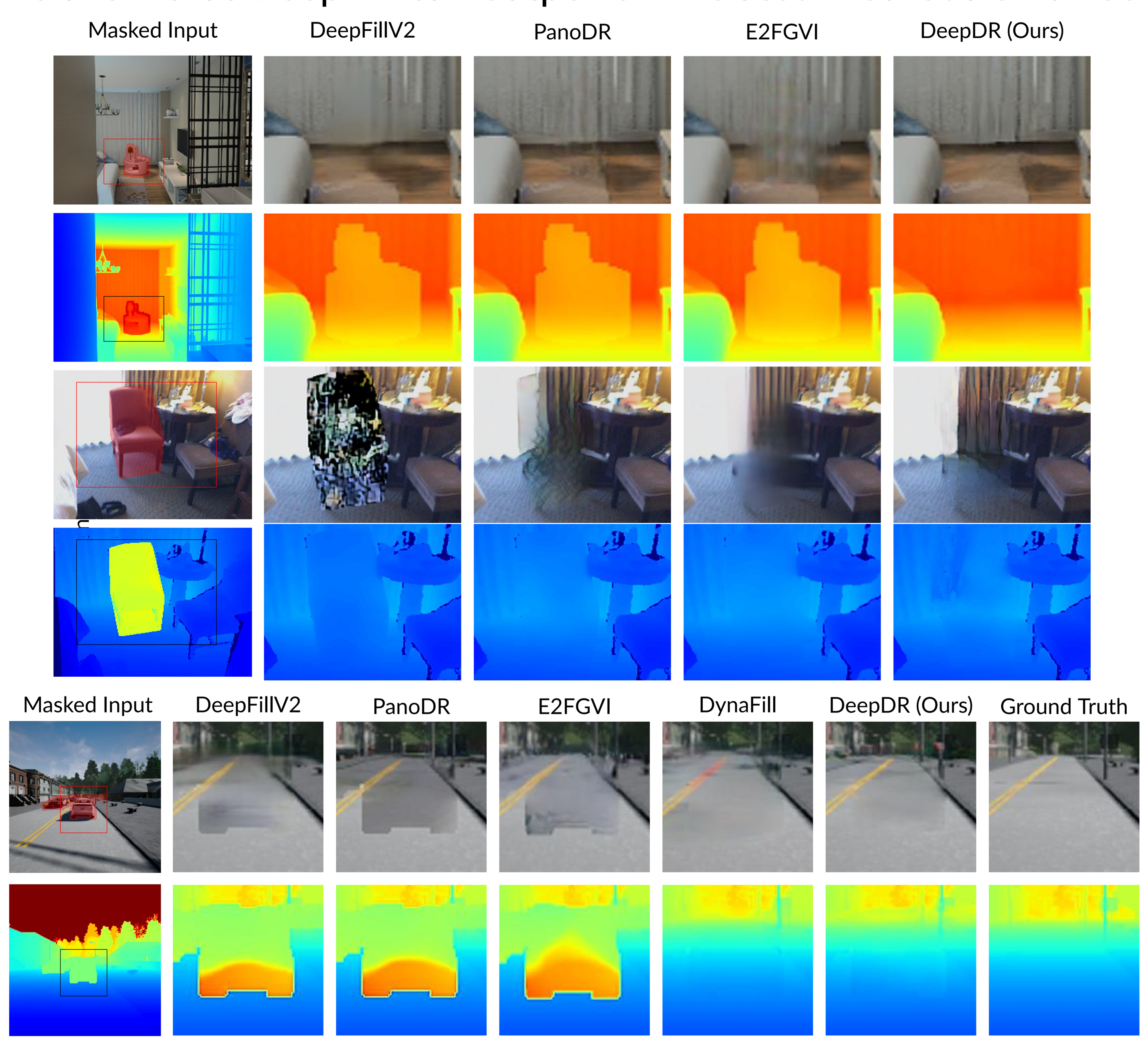
Input images I and depth maps D are encoded separately, then fused on a higher dimension for joint completion. Our semantics-aware decoder comprises up blocks conditioning features on semantic information. Outputs I_o^t and D_o^t serve as auxiliary inputs in a recurrent feedback loop for subsequent time steps.

DeepDR is a framework for **Diminished Reality** that removes objects from the environment via real-time, structure-aware **image** and **depth inpainting.**

Take a picture to get the paper!

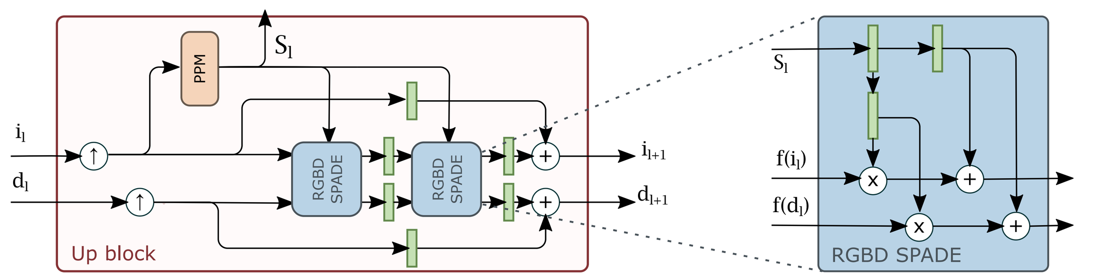
Results

We show that DeepDR can outperform related methods on three datasets qualitatively and quantitatively.



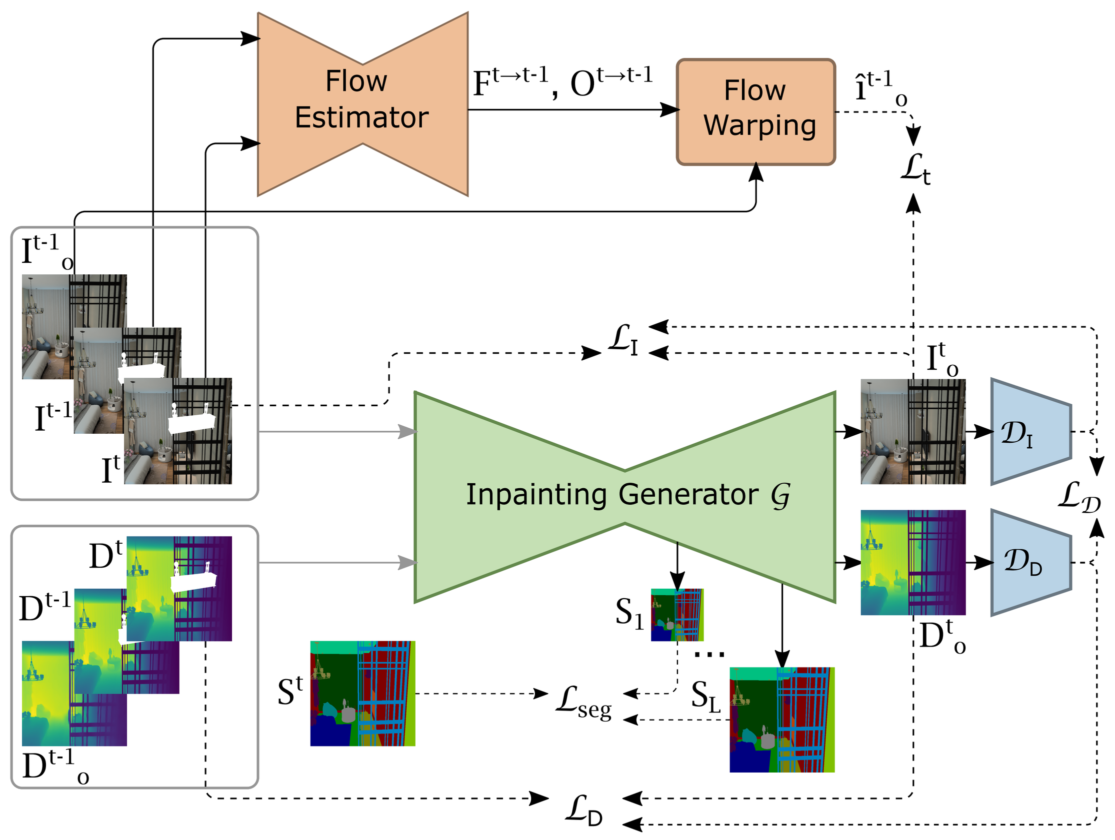
	Method	LPIPS ↓	FID ↓	Depth RMSE ↓
InteriorNet [5]	DeepFillV2 [1]	0.0150	0.448	0.572
	PanoDR [2]	0.0128	0.606	0.564
	E2FGVI [3]	0.0131	0.363	0.563
	DeepDR (Ours)	**0.0104**	**0.218**	**0.278**
DynaFill [4]	DeepFillV2 [1]	0.0238	4.122	7.92
	PanoDR [2]	0.0250	5.579	8.12
	E2FGVI [3]	0.0169	2.826	7.83
	DynaFill [4]	0.0197	2.665	7.78
	DeepDR (Ours)	**0.0168**	**2.415**	**4.51**
ScanNet [6]	DeepFillV2 [1]	0.0208	0.693	0.508
	PanoDR [2]	0.0119	0.348	0.536
	E2FGVI [3]	0.0110	0.295	0.512
	DeepDR (Ours)	**0.0108**	**0.292**	**0.484**

Structure-aware RGB-D decoder



Our up blocks predict a semantic segmentation map S_l from the up-sampled image feature i_l via a pyramid pooling module. This segmentation guides the image and depth feature generation consistently.

Training



We employ loss terms and adversarial learning to enhance the accuracy and realism of inpainted RGB and depth data ($\mathcal{L}I$, $\mathcal{L}D$). Temporal consistency is ensured through the computation of a temporal loss $\mathcal{L}t$ using optical flow. We enforce structural learning by incorporating semantic supervision via $\mathcal{L}seg$.

Acknowledgement: The work was funded by the FFG project "TRIP" (BRIDGE 883658) and the FWF project "enFaced 2.0" (KLI 1044). We thank xCAD Solutions GmbH for their continuous support.

[1] Yu et al., "Free-form image inpainting with gated convolution". In ICCV, 2019.

[2] Giktas et al., "PanoDR: Spherical panorama diminished reality for indoor scenes". In CVPR Workshops, 2021.

[3] Li et al., "Towards an end-to-end framework for flow-guided video inpainting". In CVPR, 2022.

[4] Bešić et al., "Dynamic object removal and spatio-temporal RGB-D inpainting via geometry-aware adversarial learning". IEEE trans. intell. veh., 2022.

[5] Li et al., "InteriorNet: Mega-scale Multi-sensor Photo-realistic Indoor Scenes Dataset". In BMVC, 2018.

[6] Dai et al., "Scannet: Richly-annotated 3d reconstructions of indoor scenes". In CVPR, 2017.

3DV 2024 International Conference on 3D Vision Davos, Switzerland March 18-21, 2024

TU Graz

JOANNEUM RESEARCH

xcad interior software