DeepSurfels: Learning Online Appearance Fusion – Supplementary Material –

A. Overview

In this supplementary document, we provide further details about our appearance learning pipeline (§ B), used baselines (§ C), and an extended ablation study (§ D).

B. Network architecture details

We provide further details on the Fusion Network and the Appearance Rendering module presented in Figure 3.

The Fusion Network is displayed in Figure B.1 and represents one part of the Appearance Fusion module (Figure 3). It takes as input three image maps – the upsampled input image that needs to be fused $I_t^k \in \mathbb{R}^{kH \times kW \times C}$, a feature map $\hat{F}_{t-1}$ that is rendered from existing scene content S_{t-1}, and optional meta features $\hat{M}_{t-1}$ – and produces a new blended feature map $\hat{F}_t$ that needs to be integrated into the representation.

This component consists of three modules *1)* the Feature Embedding learnable linear layer, implemented as a 1×1 convolutional layer, which compresses features of the concatenated input maps ($\hat{F}_{t-1} \otimes \hat{M}_{t-1} \otimes I_t^k$) into an intermediate feature map $\mathbb{R}^{kH \times kW \times 35}$, *2)* the Blending Network that comprises of four convolutional blocks interleaved with LeakyReLU and dropout layers, and *3)* the linear Feature Compression layer $W : \mathbb{R}^{kH \times kW \times 70} \mapsto \mathbb{R}^{kH \times kW \times c}$ that creates the new blended feature map $\hat{F}_t$. This new feature map is then integrated into the scene representation as described in the paper by updating the scene content ($S_{t-1} \mapsto S_t$).

The updated scene content is then rendered $\hat{F}'_t \in \mathbb{R}^{kH \times kW \times c}$ via the introduced differentiable projection module Π. The Appearance Rendering module (Figure B.2) takes this rendered feature map and decompresses its features into a higher resolution space with the linear Feature Decompression layer (transposed Feature Compression layer $W^T : \mathbb{R}^{kH \times kW \times c} \mapsto \mathbb{R}^{kH \times kW \times 70}$). The optional meta features are concatenated to the uncompressed feature channels and they are jointly propagated through the introduced masked average pooling operator to reduce the spatial dimension ($kH, kW \mapsto H, W$) and form an intermediate appearance feature map. This appearance feature map is then refined by the Rendering Network (5 convolutional blocks with a skip connection) and decoded as RGB values by the three-layer perceptron Feature Decoder.

C. Baseline experiments

Several baselines are used in the paper for results displayed in Figure 4, 5, and 7.

We used publicly released code with default parameters to run experiments for Fu *et al*. [23][3], SurfelMeshing [70][4], Waechter *et al*. [87][5], and NeRF [49][6]. The results for other baselines (Texture Fields [54], SRNs [73], DeepVoxels [72]) are released by the authors and we implemented the TSDF Coloring [17] baseline as a straightforward extension of TSDF Fusion that accumulates color information into voxel grids by the simple running mean algorithm.

Mesh files for Fu *et al*. [23] and Waechter *et al*. [87] for the experiment on the ShapeNet [12] cars (Table 1) are created by fusing depth frames into a grid with TSDF Fusion and then extracting the meshes with a standard marching cubes algorithm. These methods where provided by the ground truth meshes for the novel view synthesis experiment on the cat and the human dataset (Figure 4).

NeRF [49] was trained for each Replica room dataset (Figure 5) for two days on a 24GB NVidia Titan RTX GPU.

D. Ablation study

We provide an extended ablation study for 5 feature and 3 color channels (5+3 configuration) in comparison to the 3+3 configuration in Table D.1.

Quantitative and qualitative results (Table D.1, Figure D.3 and D.4) demonstrate that additional two feature channels are beneficial for the quality of rendered images.

[3] https://github.com/fdp0525/G2LTex
[4] https://github.com/puzzlepaint/surfelmeshing
[5] https://www.gcc.tu-darmstadt.de/home/proj/texrecon/
[6] https://github.com/bmild/nerf

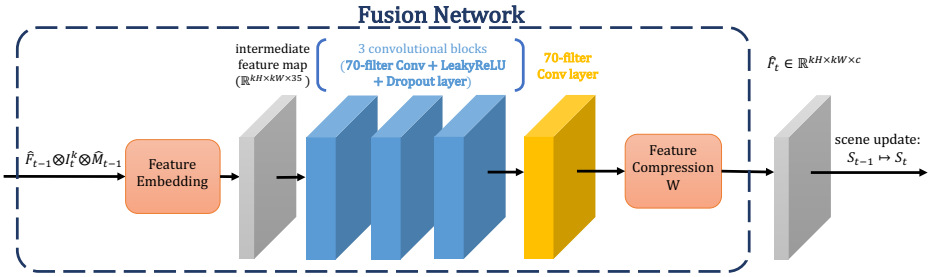

Figure B.1. **Fusion network architecture.** This module is a part of our learned appearance fusion pipeline (Figure 3). It creates a blended feature map $\hat{F}_t$ that needs to be integrated into DeepSurfel representation..

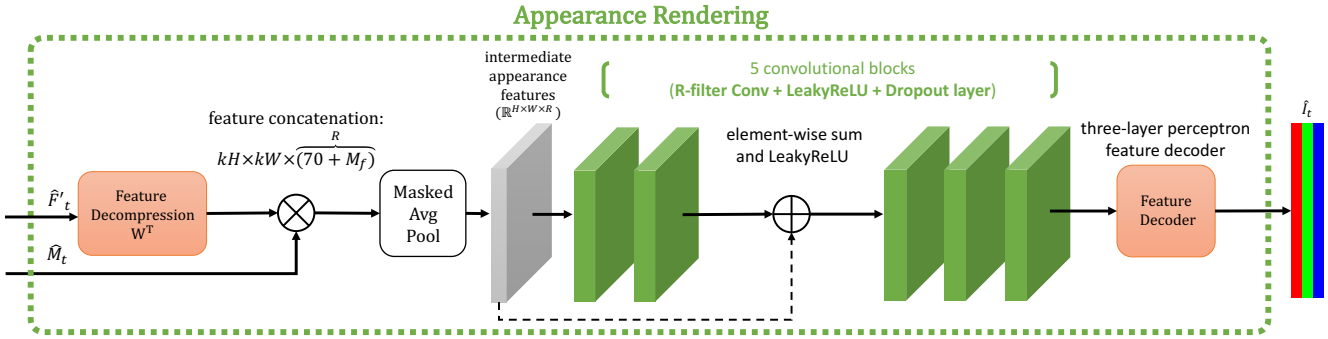

Figure B.2. **Appearance rendering module.** This module interprets rendered feature as RGB pixel values. M_f denotes the number of feature channels and R is the number of channels of the intermediate appearance features ($R = 70 + M_f$). The intermediate appearance features are refined by 5 convolutional blocks and decoded by a Feature Decoder. The Feature Decoder is implemented as a three-layer perceptron network with $\lfloor \frac{R}{2} \rfloor$, $\lfloor \frac{R}{4} \rfloor$, and 3 neurons respectively, each layer is followed by LeakyReLU activation function, except for the very last one that uses HardTanh to produce normalized RGB color values.

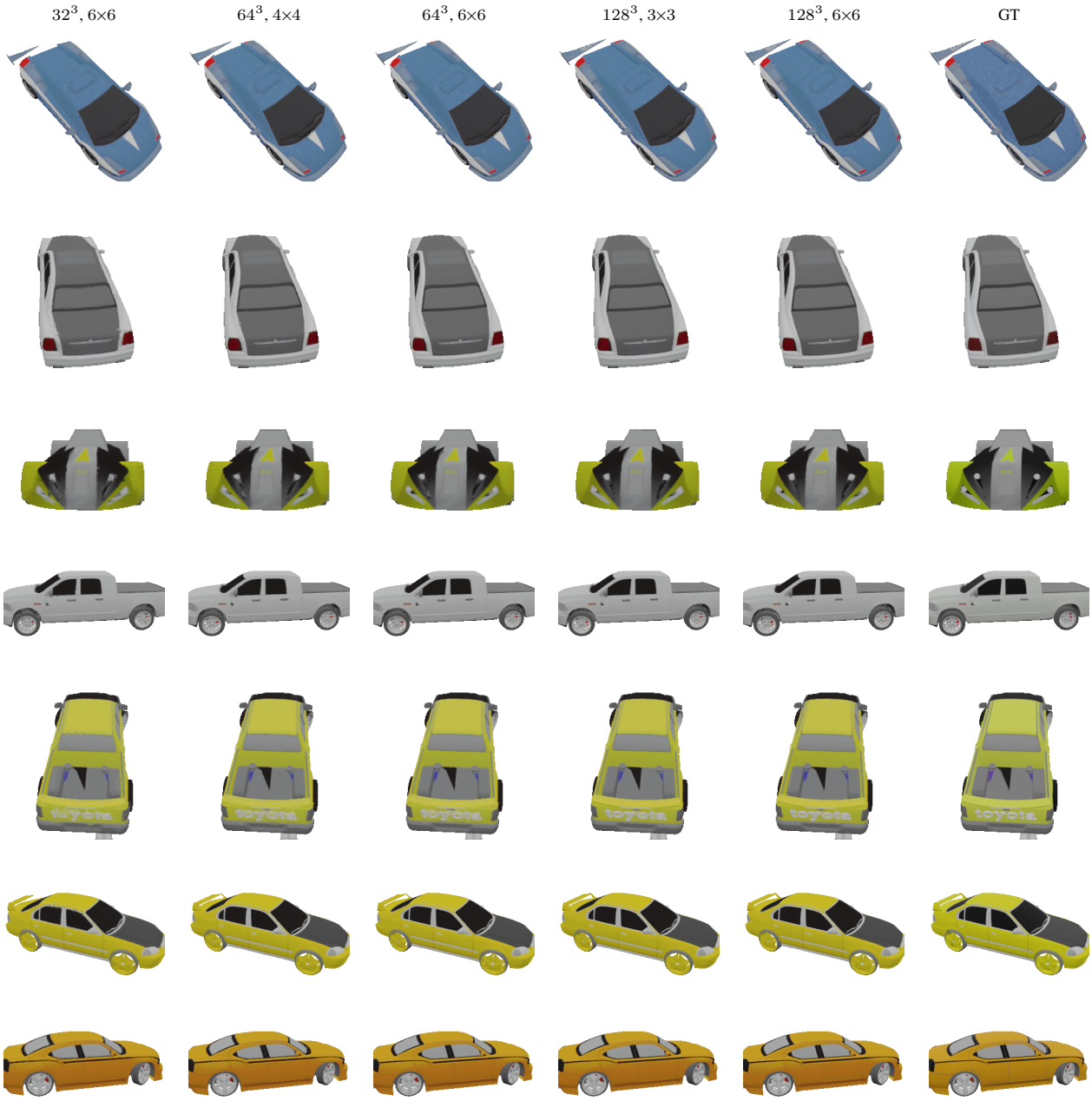

Figure D.3. **Qualitative results of our model on unseen ShapeNet [12] car scenes for different DeepSurfel parameters.** The column names denote DeepSurfel grid and patch resolution respectively. We used DeepSurfels with 3 feature and 3 color channels (3+3 configuration). A quantitative comparison is given in Table D.1.

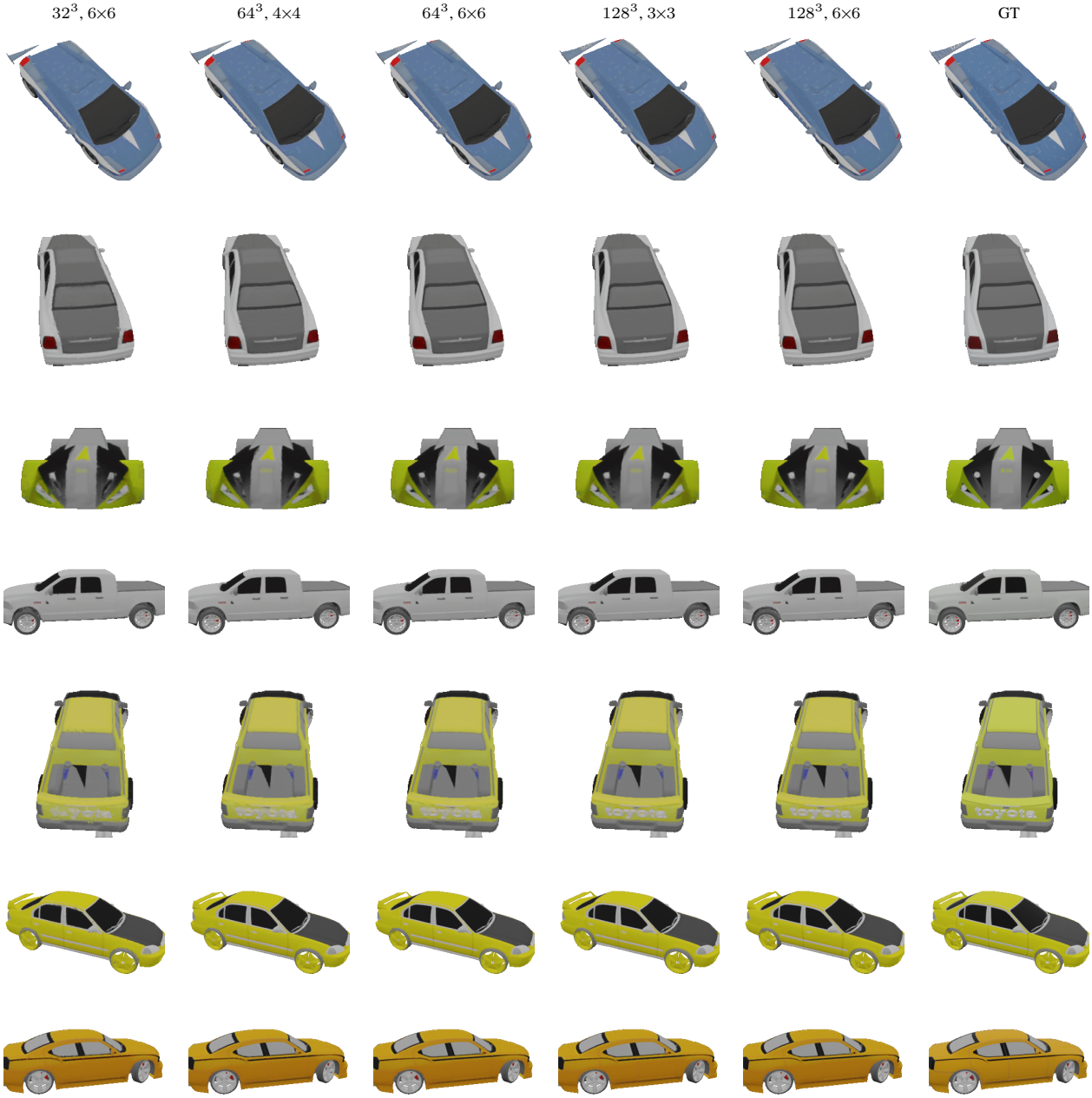

Figure D.4. **Qualitative results of our model on unseen ShapeNet [12] car scenes for different DeepSurfel parameters.** DeepSurfels with 5 feature and 3 color channels (5+3 configuration) demonstrate better results compared to our method with less channels (3+3) displayed in Figure D.3. Quantitative comparison is given in Table D.1. The column name denotes DeepSurfel gird and patch resolution respectively.

	Method	PSNR↑	SSIM↑
Baselines	SurfelMeshing [70]	13.92	0.2748
	Waechter *et al*. [87]	18.27	0.4753
	Fu *et al*. [23]	18.84	0.5196
	TSDF Coloring [17] (32^3)	21.57	0.6375
	TSDF Coloring [17] (64^3)	24.05	0.7552
	TSDF Coloring [17] (128^3)	26.68	0.8526
	Ours Det. (32^3, 6×6, 3)	27.20	0.8723
	Ours Det. (64^3, 4×4, 3)	28.73	0.9036
DeepSurfel Params (3+3)	32^3, 6×6, $3+3$	**28.89**	**0.8907**
	64^3, 4×4, $3+3$	29.92	0.9086
	64^3, 5×5, $3+3$	30.15	0.9126
	64^3, 6×6, $3+3$	**30.27**	**0.9147**
	128^3, 2×2, $3+3$	30.23	0.9133
	128^3, 3×3, $3+3$	30.51	0.9181
	128^3, 4×4, $3+3$	30.60	0.9196
	128^3, 5×5, $3+3$	30.63	0.9200
	128^3, 6×6, $3+3$	**30.64**	**0.9202**
DeepSurfel Params (5+3)	32^3, 6×6, $5+3$	**29.02**	**0.8955**
	64^3, 4×4, $5+3$	29.93	0.9118
	64^3, 5×5, $5+3$	30.12	0.9154
	64^3, 6×6, $5+3$	**30.22**	**0.9172**
	128^3, 2×2, $5+3$	30.21	0.9162
	128^3, 3×3, $5+3$	30.45	0.9206
	128^3, 4×4, $5+3$	30.54	0.9220
	128^3, 5×5, $5+3$	30.56	0.9224
	128^3, 6×6, $5+3$	**30.58**	**0.9226**

Table D.1. **Extended ablation study on ShapeNet [12] cars.** Comparison of baselines and our method on different grid and patch resolutions. The x+3 notation denotes disentangled x feature channels and 3 color channels. Results indicate that our method on a grid of 32^3 outperforms all baseline methods, including ones that require a much higher grid resolution (TSDF Coloring 128^3, Ours Deterministic 64^3). An increased number of channels and higher DeepSurfel resolution further benefits the quality of rendered images. Qualitative results for the 3+3 and 5+3 configuration are displayed in Figure D.3 and D.4 respectively.

References

[1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *Proc. International Conference on Machine Learning (ICML)*, 2018. 2

[2] Kara-Ali Aliev, Artem Sevastopolsky, Maria Kolos, Dmitry Ulyanov, and Victor Lempitsky. Neural point-based graphics. In *Proc. European Conference on Computer Vision (ECCV)*, 2020. 3

[3] Cédric Allène, Jean-Philippe Pons, and Renaud Keriven. Seamless image-based texture atlases using multi-band blending. In *Proc. International Conference on Computer Vision (ICCV)*, 2008. 3

[4] Matthieu Armando, Jean-Sébastien Franco, and Edmond Boyer. Adaptive mesh texture for multi-view appearance modeling. In *Proc. International Conference on 3D Vision (3DV)*, 2019. 3

[5] Fausto Bernardini, Ioana M. Martin, and Holly E. Rushmeier. High-quality texture reconstruction from multiple scans. *IEEE Transactions on Visualization and Computer Graphics*, 2001. 3

[6] Sai Bi, Nima Khademi Kalantari, and Ravi Ramamoorthi. Patch-based optimization for image-based texture mapping. *ACM Transactions on Graphics*, 2017. 3

[7] Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *Proc. European Conference on Computer Vision (ECCV)*, 2020. 3

[8] Andreas Bircher, Kostas Alexis, Michael Burri, Philipp Oettershagen, Sammy Omari, Thomas Mantel, and Roland Siegwart. Structural inspection path planning via iterative viewpoint resampling with application to aerial robotics. In *IEEE International Conference on Robotics and Automation*, 2015. 1

[9] Andreas Breitenmoser and Roland Siegwart. Surface reconstruction and path planning for industrial inspection with a climbing robot. In *Proc. International Conference on Applied Robotics for the Power Industry (CARPI)*, 2012. 1

[10] Andrew Brock, Theodore Lim, James M Ritchie, and Nick Weston. Generative and discriminative voxel modeling with convolutional neural networks. *arXiv preprint arXiv:1608.04236*, 2016. 3

[11] Rohan Chabra, Jan Eric Lenssen, Eddy Ilg, Tanner Schmidt, Julian Straub, Steven Lovegrove, and Richard Newcombe. Deep local shapes: Learning local sdf priors for detailed 3d reconstruction. In *Proc. European Conference on Computer Vision (ECCV)*, 2020. 3

[12] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 7, 8, 1, 3, 4, 5

[13] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[14] Hang Chu, Shugao Ma, Fernando De la Torre, Sanja Fidler, and Yaser Sheikh. Expressive telepresence via modular codec avatars. In *Proc. European Conference on Computer Vision (ECCV)*, Lecture Notes in Computer Science, 2020. 1

[15] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2020. 7

[16] Robert L. Cook, Thomas Porter, and Loren Carpenter. Distributed ray tracing. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1984. 4

[17] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1996. 3, 7, 8, 1, 5

[18] Paul E Debevec, Camillo J Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 1996. 1, 2, 3

[19] M. Eisemann, B. De Decker, M. Magnor, P. Bekaert, E. de Aguiar, N. Ahmed, C. Theobalt, and A. Sellent. Floating Textures. *Computer Graphics Forum*, 2008. 1, 2, 3

[20] Haoqiang Fan, Hao Su, and Leonidas J Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[21] John Flynn, Michael Broxton, Paul Debevec, Matthew DuVall, Graham Fyffe, Ryan Overbeck, Noah Snavely, and Richard Tucker. Deepview: View synthesis with learned gradient descent. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[22] Yanping Fu, Qingan Yan, Jie Liao, and Chunxia Xiao. Joint texture and geometry optimization for RGB-D reconstruction. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[23] Yanping Fu, Qingan Yan, Long Yang, Jie Liao, and Chunxia Xiao. Texture mapping for 3d reconstruction with RGB-D sensor. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 2, 3, 6, 7, 8, 5

[24] Matheus Gadelha, Subhransu Maji, and Rui Wang. 3d shape induction from 2d views of multiple objects. In *Proc. International Conference on 3D Vision (3DV)*, 2017. 3

[25] Ran Gal, Yonatan Wexler, Eyal Ofek, Hugues Hoppe, and Daniel Cohen-Or. Seamless montage for texturing models. *Computer Graphics Forum*, 2010. 3

[26] Santiago Garrido, María Malfaz, and Dolores Blanco. Application of the fast marching method for outdoor motion planning in robotics. *Robotics and Autonomous Systems*, 2013. 1

[27] Bastian Goldlücke, Mathieu Aubry, Kalin Kolev, and Daniel Cremers. A super-resolution framework for high-accuracy multiview reconstruction. *International Journal of Computer Vision*, 2014. 3

[28] Christian Häne, Lionel Heng, Gim Hee Lee, Friedrich Fraundorfer, Paul Furgale, Torsten Sattler, and Marc Polle-

feys. 3d visual perception for self-driving cars using a multi-camera system: Calibration, mapping, localization, and obstacle detection. *Image and Vision Computing*, 2017. 1

[29] Rana Hanocka, Gal Metzer, Raja Giryes, and Daniel Cohen-Or. Point2mesh: A self-prior for deformable meshes. *ACM Transactions on Graphics*, 2020. 2

[30] Jingwei Huang, Justus Thies, Angela Dai, Abhijit Kundu, Chiyu Jiang, Leonidas J Guibas, Matthias Nießner, and Thomas Funkhouser. Adversarial texture optimization from RGB-D scans. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[31] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, and Thomas Funkhouser. Local implicit grid representations for 3d scenes. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[32] Olaf Kähler, Victor Adrian Prisacariu, Carl Yuheng Ren, Xin Sun, Philip Torr, and David Murray. Very high frame rate volumetric integration of depth images on mobile devices. *IEEE Transactions on Visualization and Computer Graphics*, 2015. 3

[33] Angjoo Kanazawa, Shubham Tulsiani, Alexei A Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In *Proc. European Conference on Computer Vision (ECCV)*, 2018. 2

[34] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. International Conference on Learning Representations (ICLR)*, 2015. 6

[35] Kiriakos N Kutulakos and Steven M Seitz. A theory of shape by space carving. In *Proc. International Conference on Computer Vision (ICCV)*, 1999. 3

[36] Joo Ho Lee, Hyunho Ha, Yue Dong, Xin Tong, and Min H Kim. Texturefusion: High-quality texture acquisition for real-time rgb-d scanning. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3, 4

[37] Victor Lempitsky and Denis Ivanov. Seamless mosaicing of image-based texture maps. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007. 3

[38] Hendrik P. A. Lensch, Wolfgang Heidrich, and Hans-Peter Seidel. A silhouette-based algorithm for texture registration and stitching. *Graphical Models*, 2001. 3

[39] Yawei Li, Vagia Tsiminaki, Radu Timofte, Marc Pollefeys, and Luc Van Gool. 3d appearance super-resolution with deep learning. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[40] Yiyi Liao, Simon Donne, and Andreas Geiger. Deep marching cubes: Learning explicit surface representations. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[41] Shichen Liu, Tianye Li, Weikai Chen, and Hao Li. Soft rasterizer: A differentiable renderer for image-based 3d reasoning. In *Proc. International Conference on Computer Vision (ICCV)*, 2019. 2

[42] Z. Liu, Y. Cao, Z. Kuang, L. Kobbelt, and S. Hu. High-quality textured 3d shape reconstruction with cascaded fully convolutional networks. *IEEE Transactions on Visualization and Computer Graphics*, 2019. 3

[43] Stephen Lombardi, Jason M. Saragih, Tomas Simon, and Yaser Sheikh. Deep appearance models for face rendering. *ACM Transactions on Graphics*, 2018. 1

[44] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *ACM Transactions on Graphics*, 2019. 1, 3

[45] Robert Maier, Kihwan Kim, Daniel Cremers, Jan Kautz, and Matthias Nießner. Intrinsic3d: High-quality 3d reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In *Proc. International Conference on Computer Vision (ICCV)*, 2017. 3, 4, 8

[46] Lars M. Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[47] Marko Mihajlovic, Yan Zhang, Michael J. Black, and Siyu Tang. LEAP: Learning articulated occupancy of people. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 3

[48] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics*, 2019. 3

[49] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *Proc. European Conference on Computer Vision (ECCV)*, 2020. 1, 2, 3, 6, 8

[50] Richard A Newcombe, Shahram Izadi, Otmar Hilliges, David Molyneaux, David Kim, Andrew J Davison, Pushmeet Kohi, Jamie Shotton, Steve Hodges, and Andrew Fitzgibbon. KinectFusion: Real-time dense surface mapping and tracking. In *IEEE International Symposium on Mixed and Augmented Reality*, 2011. 1, 3, 4

[51] Richard A Newcombe, Steven J Lovegrove, and Andrew J Davison. Dtam: Dense tracking and mapping in real-time. In *Proc. International Conference on Computer Vision (ICCV)*, 2011. 1

[52] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[53] Matthias Nießner, Michael Zollhöfer, Shahram Izadi, and Marc Stamminger. Real-time 3d reconstruction at scale using voxel hashing. *ACM Transactions on Graphics*, 2013. 3

[54] Michael Oechsle, Lars Mescheder, Michael Niemeyer, Thilo Strauss, and Andreas Geiger. Texture fields: Learning texture representations in function space. In *Proc. International Conference on Computer Vision (ICCV)*, 2019. 1, 3, 6, 7

[55] Michael Oechsle, Michael Niemeyer, Christian Reiser, Lars Mescheder, Thilo Strauss, and Andreas Geiger. Learning implicit surface light fields. In *Proc. International Confer-*

ence on 3D Vision (3DV), 2020. 3

[56] Jeong Joon Park, Peter Florence, Julian Straub, Richard A. Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2, 3

[57] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *Proc. European Conference on Computer Vision (ECCV)*, 2020. 3

[58] Eric Penner and Li Zhang. Soft 3d reconstruction for view synthesis. *ACM Transactions on Graphics*, 2017. 3

[59] Hanspeter Pfister, Matthias Zwicker, Jeroen Van Baar, and Markus Gross. Surfels: Surface elements as rendering primitives. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 2000. 3

[60] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *Proc. International Conference on Machine Learning (ICML)*, 2019. 3

[61] Konstantinos Rematas and Vittorio Ferrari. Neural voxel renderer: Learning an accurate and controllable rendering tool. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[62] Danilo Jimenez Rezende, SM Ali Eslami, Shakir Mohamed, Peter Battaglia, Max Jaderberg, and Nicolas Heess. Unsupervised learning of 3d structure from images. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2016. 3

[63] Audrey Richard, Ian Cherabier, Martin R. Oswald, Vagia Tsiminaki, Marc Pollefeys, and Konrad Schindler. Learned multi-view texture super-resolution. In *Proc. International Conference on 3D Vision (3DV)*, 2019. 3

[64] Gernot Riegler and Vladlen Koltun. Free view synthesis. In *Proc. European Conference on Computer Vision (ECCV)*, 2020. 3

[65] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. PIFu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proc. International Conference on Computer Vision (ICCV)*, 2019. 3

[66] Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. PIFuHD: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[67] Manolis Savva, Abhishek Kadian, Oleksandr Maksymets, Yili Zhao, Erik Wijmans, Bhavana Jain, Julian Straub, Jia Liu, Vladlen Koltun, Jitendra Malik, Devi Parikh, and Dhruv Batra. Habitat: A Platform for Embodied AI Research. In *Proc. International Conference on Computer Vision (ICCV)*, 2019. 7

[68] Thomas Schöps, Martin R Oswald, Pablo Speciale, Shuoran Yang, and Marc Pollefeys. Real-time view correction for mobile devices. *IEEE Transactions on Visualization and Computer Graphics*, 2017. 1

[69] Thomas Schöps, Torsten Sattler, Christian Häne, and Marc Pollefeys. Large-scale outdoor 3d reconstruction on a mobile device. *Computer Vision and Image Understanding*, 2017. 1

[70] T. Schöps, T. Sattler, and M. Pollefeys. SurfelMeshing: Online surfel-based mesh reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 3, 4, 6, 7, 8, 1, 5

[71] Steven M Seitz and Charles R Dyer. Photorealistic scene reconstruction by voxel coloring. *International Journal of Computer Vision*, 1999. 3, 6

[72] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhofer. DeepVoxels: Learning persistent 3d feature embeddings. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2, 3, 7, 8

[73] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2019. 1, 2, 3, 7

[74] Pratul P Srinivasan, Richard Tucker, Jonathan T Barron, Ravi Ramamoorthi, Ren Ng, and Noah Snavely. Pushing the boundaries of view extrapolation with multiplane images. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3

[75] Frank Steinbrucker, Christian Kerl, and Daniel Cremers. Large-scale multi-resolution surface reconstruction from rgb-d sequences. In *Proc. International Conference on Computer Vision (ICCV)*, 2013. 3

[76] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J. Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, Anton Clarkson, Mingfei Yan, Brian Budge, Yajie Yan, Xiaqing Pan, June Yon, Yuyang Zou, Kimberly Leon, Nigel Carter, Jesus Briales, Tyler Gillingham, Elias Mueggler, Luis Pesqueira, Manolis Savva, Dhruv Batra, Hauke M. Strasdat, Renzo De Nardi, Michael Goesele, Steven Lovegrove, and Richard Newcombe. The Replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019. 6, 7

[77] David Stutz and Andreas Geiger. Learning 3d shape completion from laser scan data with weak supervision. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[78] Richard Szeliski and Polina Golland. Stereo matching with transparency and matting. In *Proc. International Conference on Computer Vision (ICCV)*, 1998. 3

[79] Takeshi Takai, Adrian Hilton, and Takashi Mastuyama. Harmonised texture mapping. In *Proc. International Conference on 3D Vision (3DV)*, 2010. 3

[80] Marco Tarini, Cem Yuksel, and Silvain Lefebvre. Rethinking texture mapping. In *ACM SIGGRAPH 2017 Courses*, 2017. 3

[81] A. Tewari, O. Fried, J. Thies, V. Sitzmann, S. Lombardi, K. Sunkavalli, R. Martin-Brualla, T. Simon, J. Saragih, M. Nießner, R. Pandey, S. Fanello, G. Wetzstein, J.-Y. Zhu, C. Theobalt, M. Agrawala, E. Shechtman, D. B Goldman, and M. Zollhöfer. State of the art on neural rendering. *Computer Graphics Forum*, 2020. 3

[82] Christian Theobalt, Naveed Ahmed, Hendrik P. A. Lensch, Marcus A. Magnor, and Hans-Peter Seidel. Seeing people

in different light-joint shape, motion, and reflectance capture. *IEEE Transactions on Visualization and Computer Graphics*, 2007. 3

[83] Justus Thies, Michael Zollhöfer, and Matthias Nießner. Deferred neural rendering: Image synthesis using neural textures. *ACM Transactions on Graphics*, 2019. 3

[84] Justus Thies, Michael Zollhöfer, Christian Theobalt, Marc Stamminger, and Matthias Nießner. Image-guided neural object rendering. In *Proc. International Conference on Learning Representations (ICLR)*, 2020. 3

[85] Vagia Tsiminaki, Wei Dong, Martin R. Oswald, and Marc Pollefeys. Joint multi-view texture super-resolution and intrinsic decomposition. In *Proc. of the British Machine and Vision Conference (BMVC)*, page 15. BMVA Press, 2019. 3

[86] Vagia Tsiminaki, Jean-Sébastien Franco, and Edmond Boyer. High resolution 3d shape texture from multiple videos. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 3

[87] Michael Waechter, Nils Moehrle, and Michael Goesele. Let there be color! large-scale texturing of 3d reconstructions. In *Proc. European Conference on Computer Vision (ECCV)*, 2014. 1, 2, 3, 6, 7, 8, 5

[88] Kaixuan Wang, Fei Gao, and Shaojie Shen. Real-time scalable dense surfel mapping. In *IEEE International Conference on Robotics and Automation*, 2019. 3, 4

[89] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *Proc. European Conference on Computer Vision (ECCV)*, 2018. 2

[90] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004. 7

[91] Silvan Weder, Johannes L Schönberger, Marc Pollefeys, and Martin R Oswald. RoutedFusion: Learning real-time depth map fusion. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[92] Silvan Weder, Johannes L Schönberger, Marc Pollefeys, and Martin R Oswald. NeuralFusion: Online depth fusion in latent space. In *Proc. International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 3

[93] Thomas Whelan, Michael Kaess, Hordur Johannsson, Maurice Fallon, John J Leonard, and John McDonald. Real-time large-scale dense rgb-d slam with volumetric fusion. *International Journal of Robotics Research*, 2015. 1

[94] Thomas Whelan, Renato F. Salas-Moreno, Ben Glocker, Andrew J. Davison, and Stefan Leutenegger. Elasticfusion: Real-time dense SLAM and light source estimation. *International Journal of Robotics Research*, 2016. 3, 4

[95] Daniel N Wood, Daniel I Azuma, Ken Aldinger, Brian Curless, Tom Duchamp, David H Salesin, and Werner Stuetzle. Surface light fields for 3d photography. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 2000. 3

[96] Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Interpretable transformations with encoder-decoder networks. In *Proc. International Conference on Computer Vision (ICCV)*, 2017. 3

[97] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2016. 3

[98] Qiangeng Xu, Weiyue Wang, Duygu Ceylan, Radomír Mech, and Ulrich Neumann. DISN: deep implicit surface network for high-quality single-view 3d reconstruction. In *Proc. Neural Information Processing Systems (NeurIPS)*, 2019. 3

[99] Cem Yuksel, John Keyser, and Donald H. House. Mesh colors. *ACM Transactions on Graphics*, 2010. 3

[100] Cem Yuksel, Sylvain Lefebvre, and Marco Tarini. Rethinking texture mapping. In *Computer Graphics Forum*, 2019. 3

[101] Ming Zeng, Fukai Zhao, Jiaxiang Zheng, and Xinguo Liu. A memory-efficient kinectfusion using octree. In *Proc. International Conference on Computational Visual Media*, 2012. 3

[102] Ming Zeng, Fukai Zhao, Jiaxiang Zheng, and Xinguo Liu. Octree-based fusion for realtime 3d reconstruction. *Graphical Models*, 2013. 3

[103] Xiuming Zhang, Sean Fanello, Yun-Ta Tsai, Tiancheng Sun, Tianfan Xue, Rohit Pandey, Sergio Orts-Escolano, Philip Davidson, Christoph Rhemann, Paul Debevec, et al. Neural light transport for relighting and view synthesis. *ACM Transactions on Graphics*, 2021. 3

[104] Michael Zollhöfer, Angela Dai, Matthias Innmann, Chenglei Wu, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Shading-based Refinement on Volumetric Signed Distance Functions. *ACM Transactions on Graphics*, 2015. 3, 4