


FACULTY OF MATHEMATICS,
PHYSICS AND INFORMATICS
Comenius University
Bratislava

3D Vision

Lecture 10: 3D Object Detection, 3D Pose Estimation, Camera Pose Estimation

Ing. Viktor Kocur, PhD.

9.5.2023

Contents

Frustum PointNet

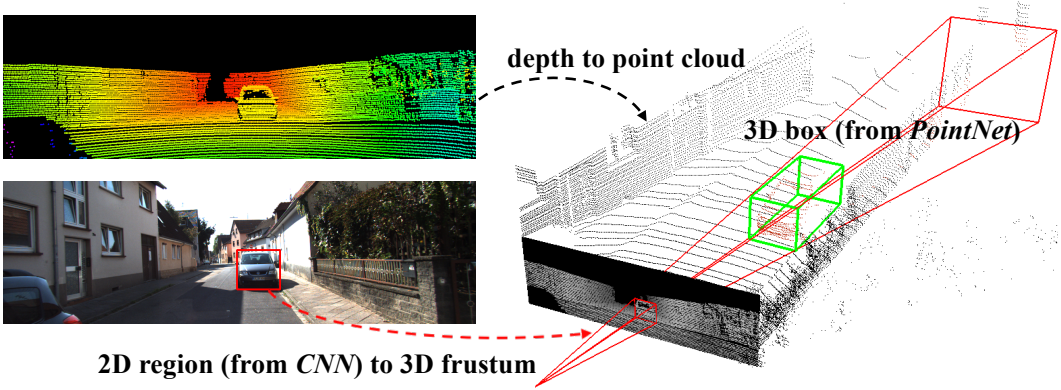


Image adopted from: Charles R Qi et al. “Frustum pointnets for 3d object detection from rgb-d data.” In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 918–927

Frustum PointNet

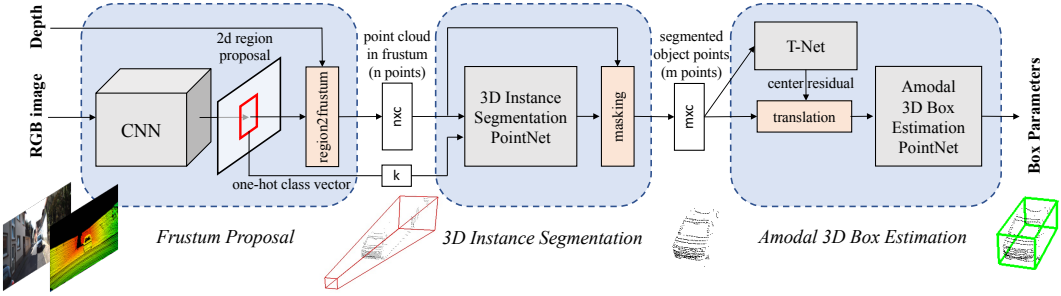


Image adopted from: Charles R Qi et al. “Frustum pointnets for 3d object detection from rgb-d data.” In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 918–927

Point Pillars

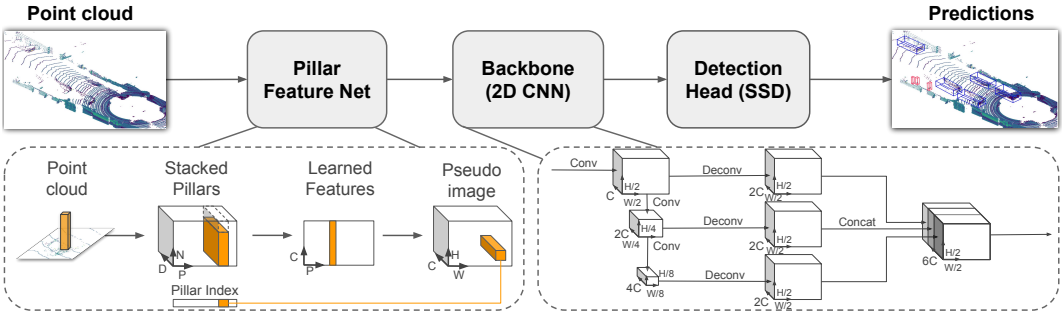


Image adopted from: Alex H Lang et al. “Pointpillars: Fast encoders for object detection from point clouds.” In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 12697–12705

Pseudolidar

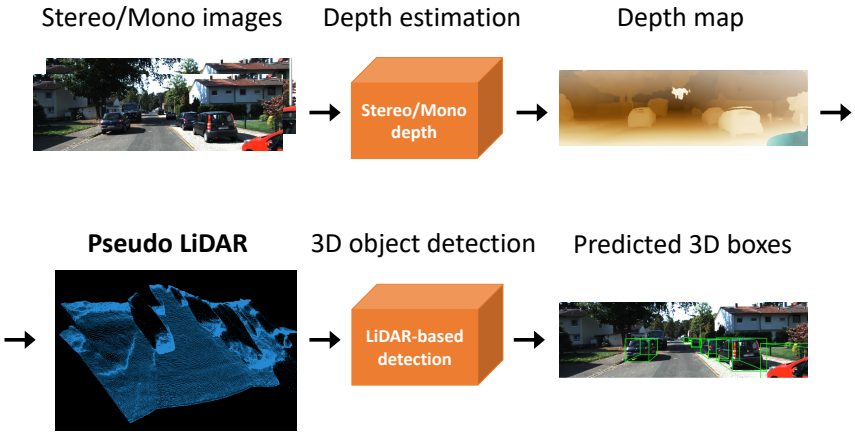


Image adopted from: Yan Wang et al. “Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 8445–8453

Pseudolidar - differentiable

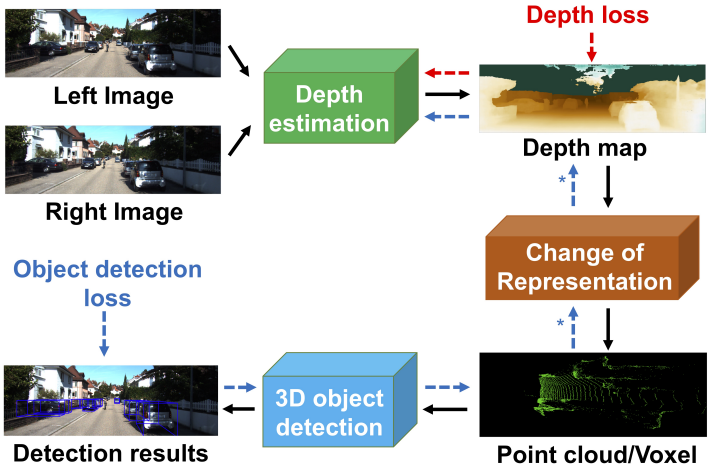


Image adopted from: Rui Qian et al. “End-to-end pseudo-lidar for image-based 3d object detection.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 5881–5890

Depth-aware Transformers

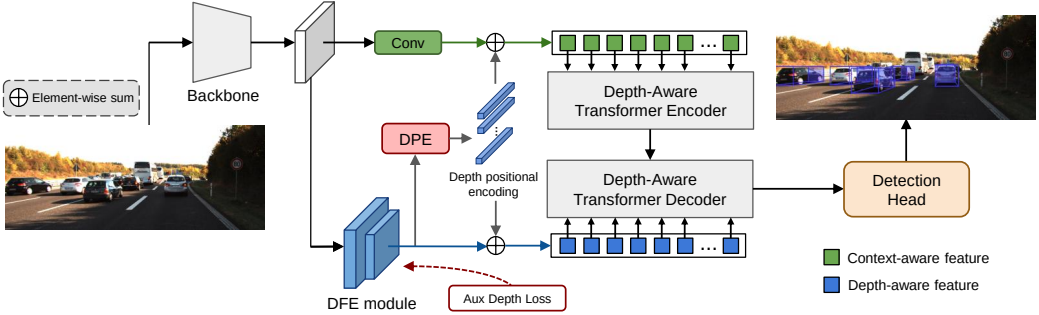


Image adopted from: Kuan-Chih Huang et al. “Monodtr: Monocular 3d object detection with depth-aware transformer.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 4012–4021

Sensor Fusion

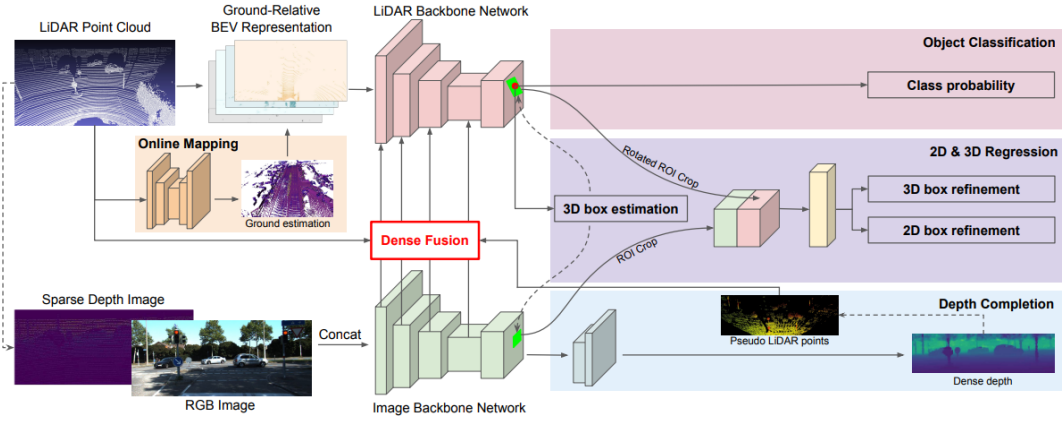


Image adopted from: Ming Liang et al. “Multi-task multi-sensor fusion for 3d object detection.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 7345–7353

Human Pose Estimation 2D to 3D

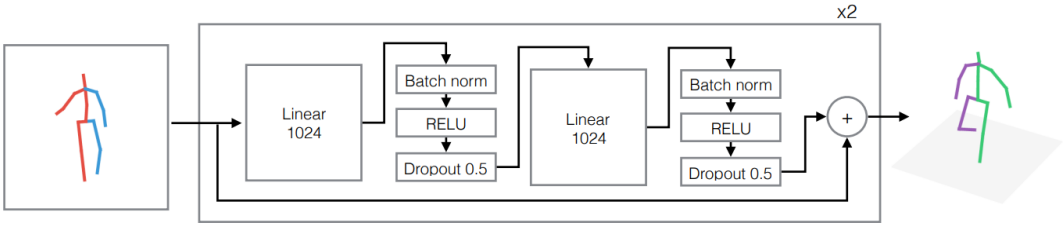


Image adopted from: Julieta Martinez et al. “A simple yet effective baseline for 3d human pose estimation.” In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2640–2649

Weakly-supervised Human Pose Estimation

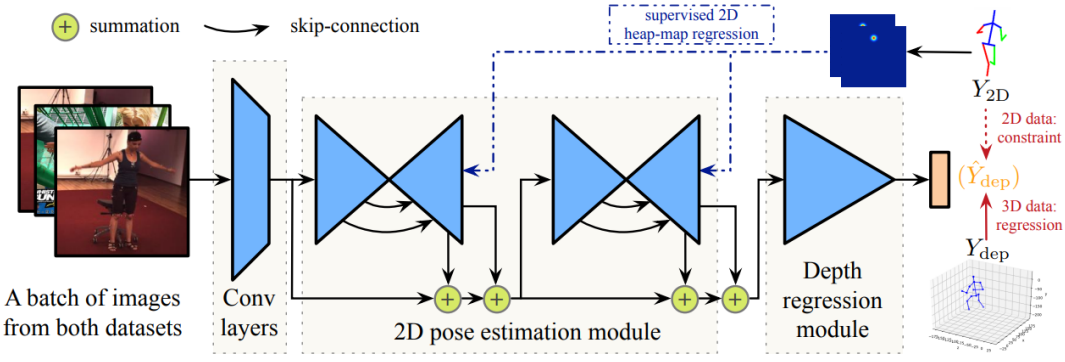


Image adopted from: Xingyi Zhou et al. “Towards 3d human pose estimation in the wild: a weakly-supervised approach.” In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 398–407

3D Hand Pose Estimation

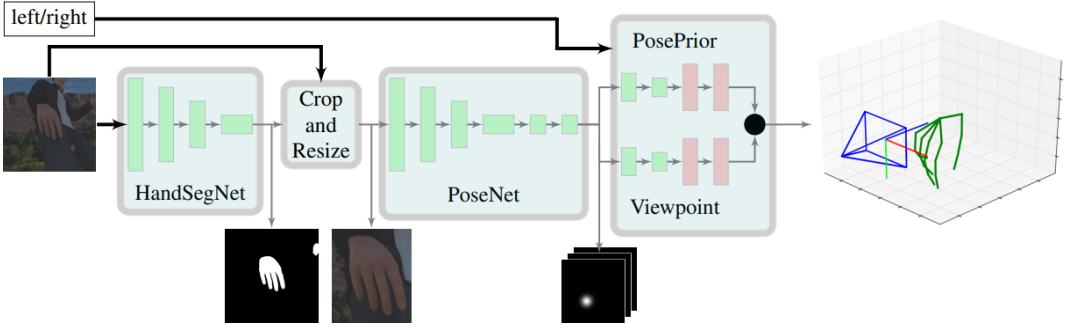


Image adopted from: Christian Zimmermann and Thomas Brox. “Learning to estimate 3d hand pose from single rgb images.” In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 4903–4911

3D Hand PosePrior

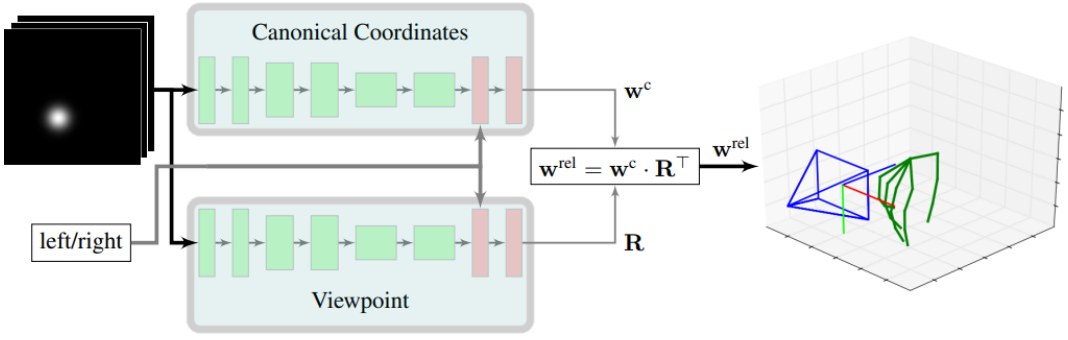


Image adopted from: Christian Zimmermann and Thomas Brox. “Learning to estimate 3d hand pose from single rgb images.” In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 4903–4911

Generating Synthetic Annotations Using GANs

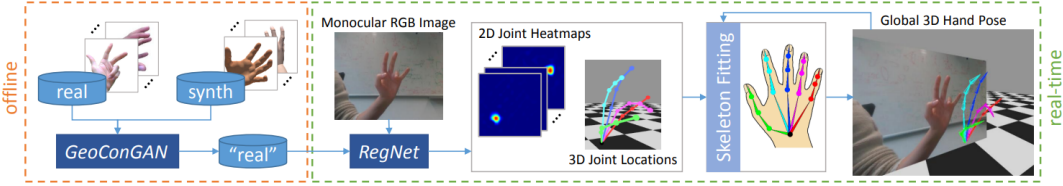


Image adopted from: Franziska Mueller et al. "GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB." In: *Proceedings of Computer Vision and Pattern Recognition (CVPR)*. 2018. URL: https://handtracker.mpi-inf.mpg.de/projects/GANeratedHands/

Weakly Supervised 3D Hand Pose Estimation

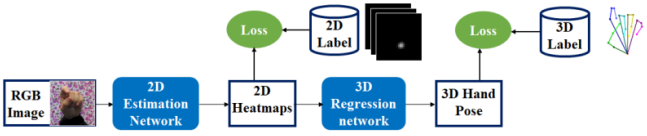


Image adopted from: Yujun Cai et al. “Weakly-supervised 3d hand pose estimation from monocular rgb images.” In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 666–682

Detecting Hands Grasping Objects

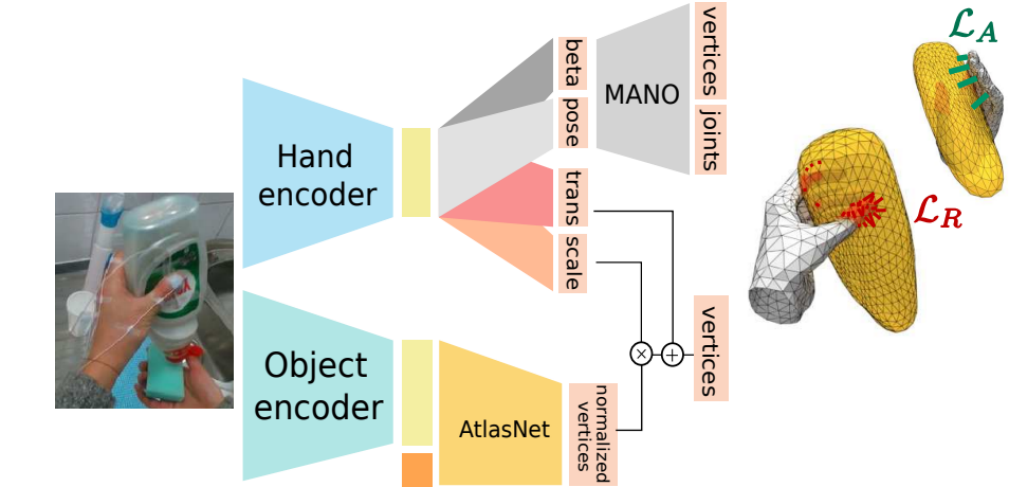
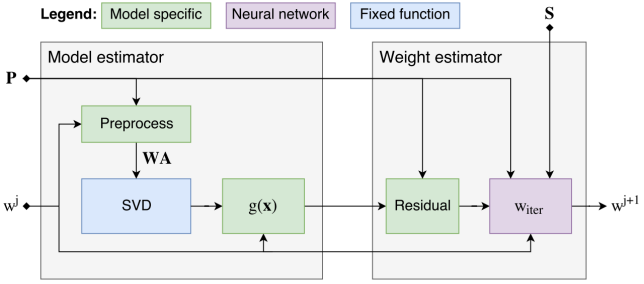


Image adopted from: Yana Hasson et al. “Learning joint reconstruction of hands and manipulated objects.” In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pp. 11807–11816

Deep F Estimation



The network uses as inputs point correspondences and learns to obtain weights for iterative least squares loop.

Image adopted from: René Ranftl and Vladlen Koltun. "Deep fundamental matrix estimation." In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 284–299

Deep F without Correspondences

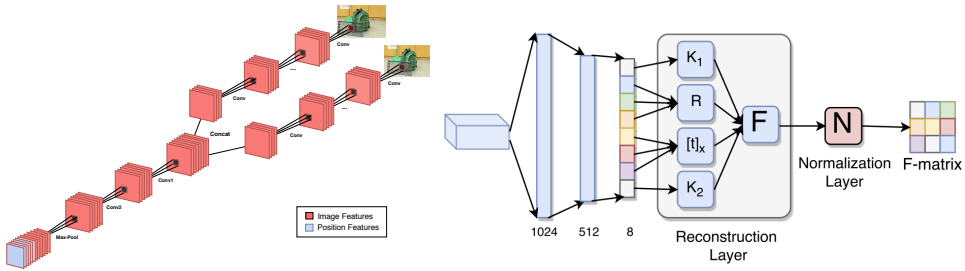


Image adopted from: Omid Poursaeed et al. “Deep fundamental matrix estimation without correspondences.” In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 2018, pp. 0–0

Learning to find good models in RANSAC

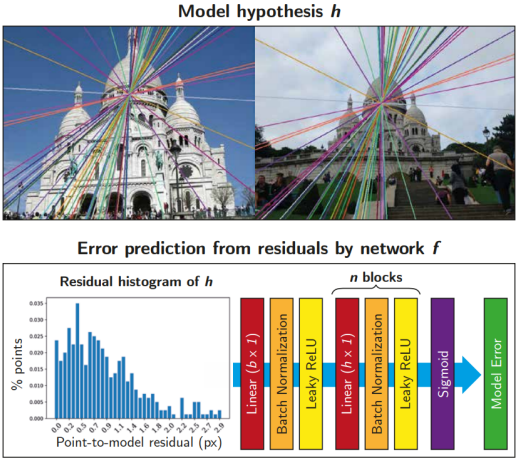


Image adopted from: Daniel Barath, Luca Cavalli, and Marc Pollefeys. “Learning to find good models in RANSAC.” In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 15744–15753

Posenet

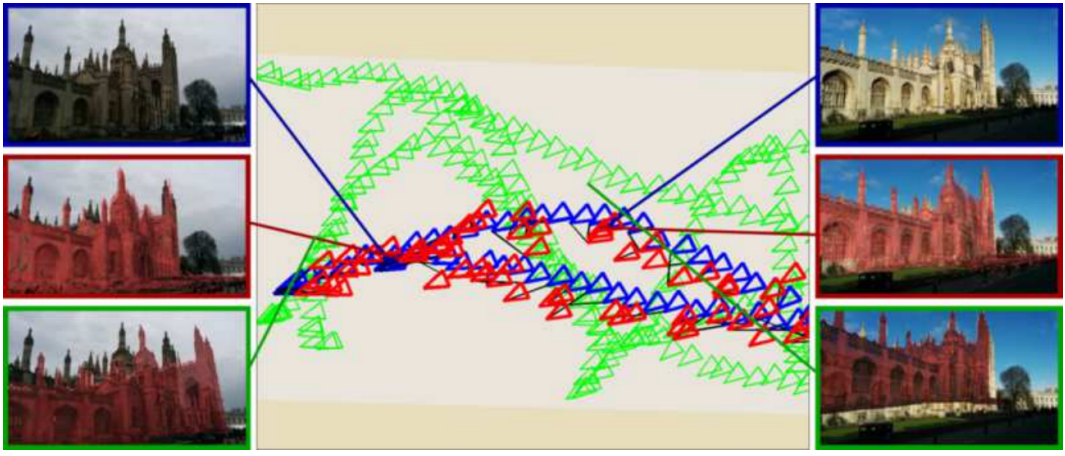


Image adopted from: Alex Kendall, Matthew Grimes, and Roberto Cipolla. “Posenet: A convolutional network for real-time 6-dof camera relocalization.” In: *Proceedings of the IEEE international conference on computer vision*. 2015, pp. 2938–2946

Neural Radiance Fields

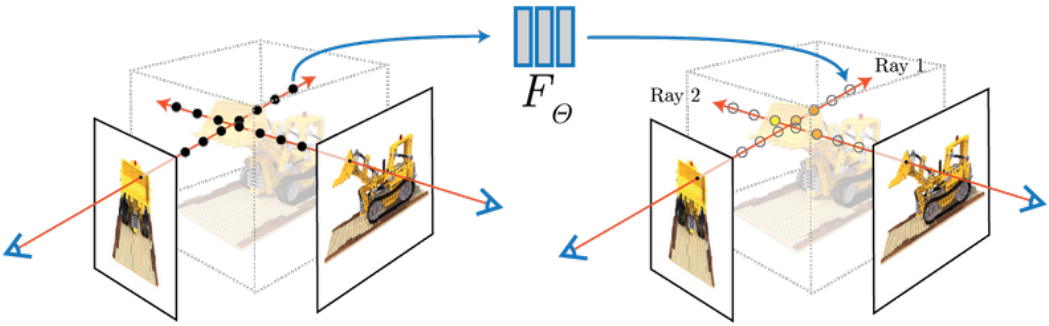


Image adopted from: Ben Mildenhall et al. “Nerf: Representing scenes as neural radiance fields for view synthesis.” In: *Communications of the ACM* 65.1 (2021), pp. 99–106

Neural Radiance Fields

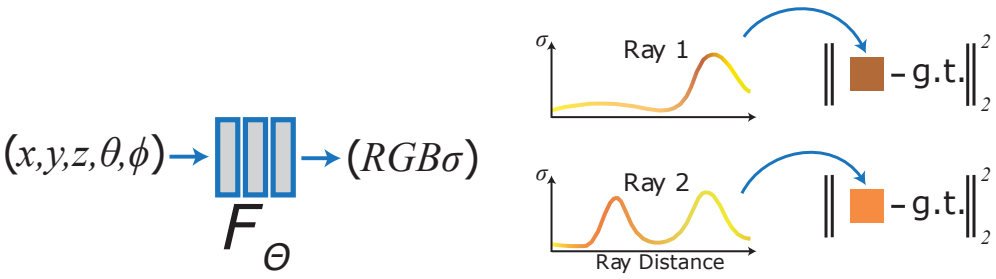


Image adopted from: Ben Mildenhall et al. “Nerf: Representing scenes as neural radiance fields for view synthesis.” In: *Communications of the ACM* 65.1 (2021), pp. 99–106