INFORMATIK

augmented VISION DFKI

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Final Presentation
Bachelor Thesis

Deep Learning for Scene Flow Estimation Using Monocular Camera and Sparse LiDAR

Name : Rishav

Email: f2016108@pilani.bits-pilani.ac.in, rishav.rishav@dfki.de

Supervisors : Ramy Battrawy, René Schuster

Outline

- Introduction
- Motivation
- Previous Work
- Frameworks
  - Novel feature Extractor
  - DeepLiDARFlow
- Summary
- Future Work

Introduction

- **Captures 3-D motion and 3-D geometry of the scene between two frames.**
- **3-dimensional displacement vector of each image point**
- **Scene Flow : 4-D Vector at each pixel : $\{u,v,d_0,d_1\}$ if the image are calibrated & rectified.**

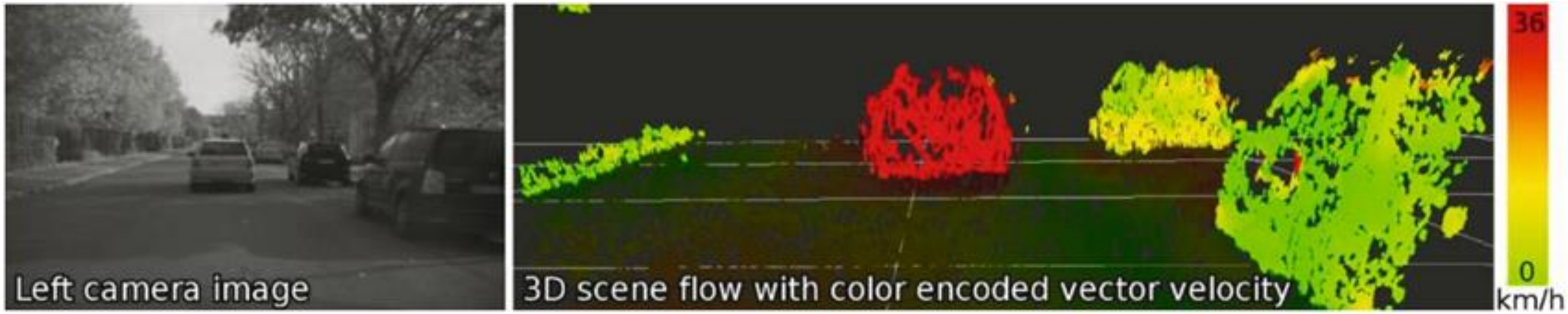


Image credits: Wedel, Andreas, et al. "Efficient dense scene flow from sparse or dense stereo data." ECCV 2008.

Motivation

- **Motivation**
  - Image based methods depend on image quality.
  - LiDAR measurements are robust
  - Mutual Improvement by fusion

- **Goal**
  - End-to-End Deep Learning architecture for Scene flow
  - Uses monocular images and LiDAR measurements as input.

- **Challenges**
  - Sparse LiDAR measurements
  - Fusion of RGB and LiDAR data is non trivial.

Related Work

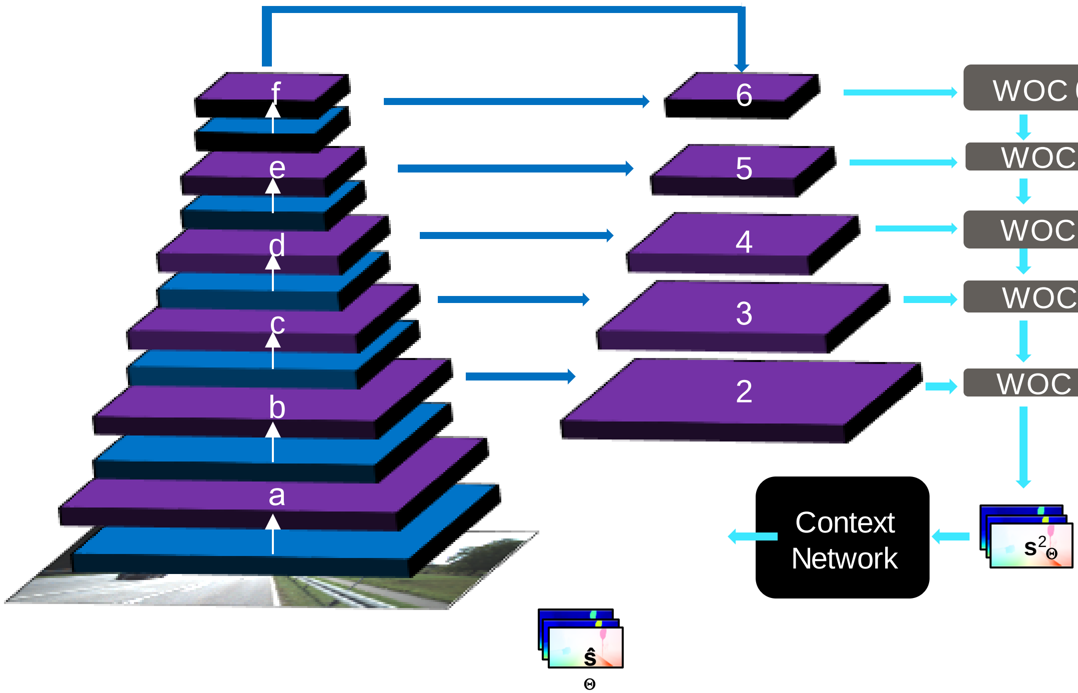
- **Monocular (Image based)**
  - Mono-SF [1]
  - Mono-Stixels [2]
- **Stereo (Image based)**
  - DWARF [3]
  - PWOC-3D [4]
  - SceneFlowNet [5]
- **LiDAR only**
  - FlowNet3D [6]
  - HPLFLowNet [7]
- **LiDAR + RGB (monocular & stereo)**
  - LiDARFlow [8]

References

[1] F. Brickwedde, et. Al. "Mono-SF: Multi-View Geometry Meets Single-View Depth for Monocular Scene Flow Estimation of Dynamic Traffic Scenes" (CVPR 2019)
[2] F. Brickwedde, et Al. "Mono-Stixels: monocular depth reconstruction of dynamic street scenes (ICRA 2019)
[3] Aleotti, et Al. "Learning end-to-end scene flow by distilling single tasks knowledge" (AAAI, 2020)
[4] Saxena et Al, "PWOC-3D: Deep Occlusion-Aware End-to-End Scene Flow Estimation" (IV, 2019)
[5] Ilg et. Al "Occlusions, motion and depth boundaries with a generic network for disparity, optical flow or scene flow estimation" (ECCV, 2018)
[6] Liu et. Al "Flownet3d: Learning scene flow in 3d point clouds" (CVPR 2019)
[7] Gu et. Al, "HPLFlowNet: Hierarchical Permutohedral Lattice FlowNet for Scene Flow Estimation on Large-scale Point Clouds" (CVPR, 2019)
[8] Battrawy et. Al, "LiDAR-Flow: Dense Scene Flow Estimation from Sparse LiDAR and Stereo Images", (IROS 2019)

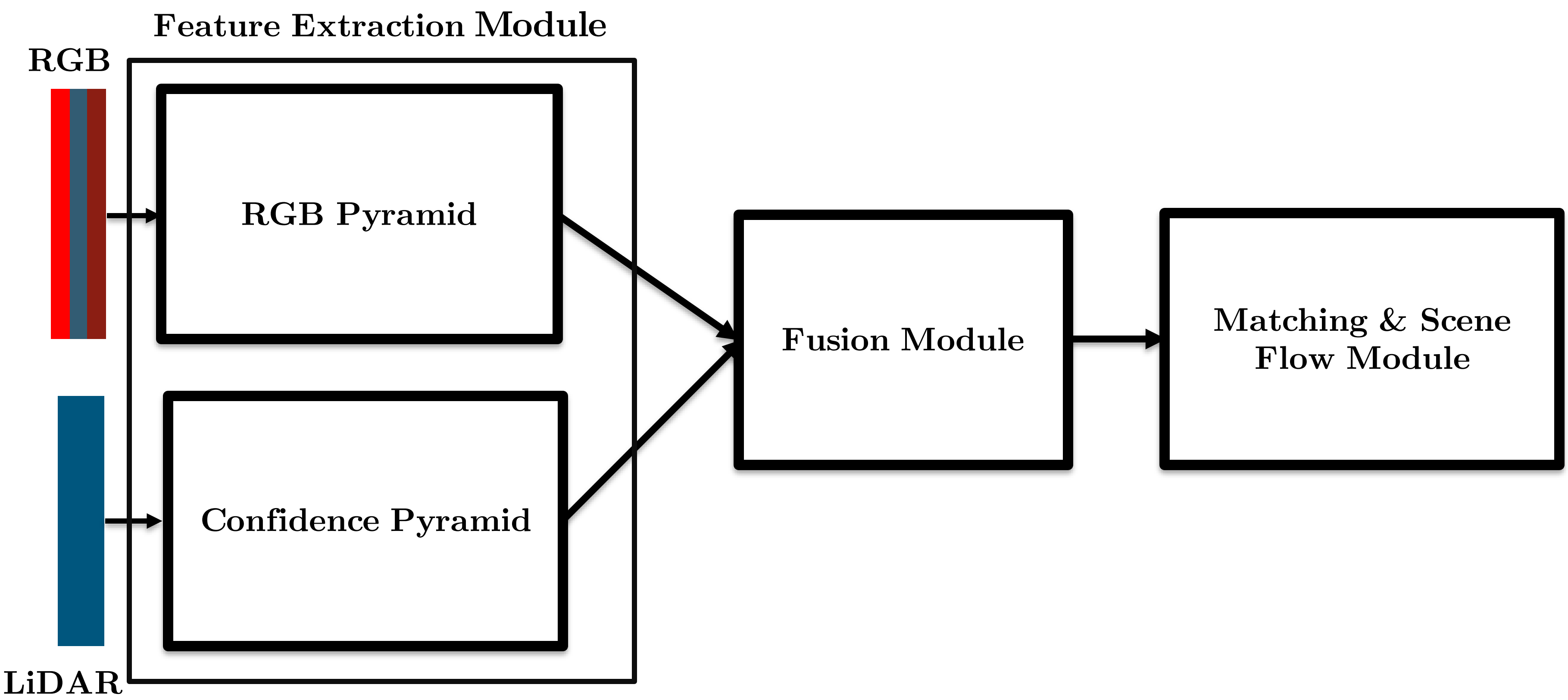
Related Work

- **PWOC-3D [1]**



[1] Saxena et Al, "PWOC-3D: Deep Occlusion-Aware End-to-End Scene Flow Estimation" (IV, 2019)

An overview



Feature Extraction Module
RGB
RGB Pyramid
LiDAR
Confidence Pyramid
Fusion Module
Matching & Scene Flow Module

Novel Feature Extractor (DenseFPN)

- **Motivation**
  - Feature maps are basic cues of computer vision tasks
  - Strong feature representations have significantly improved results.
  - (Feature Pyramid Networks) FPN [1] improved results.

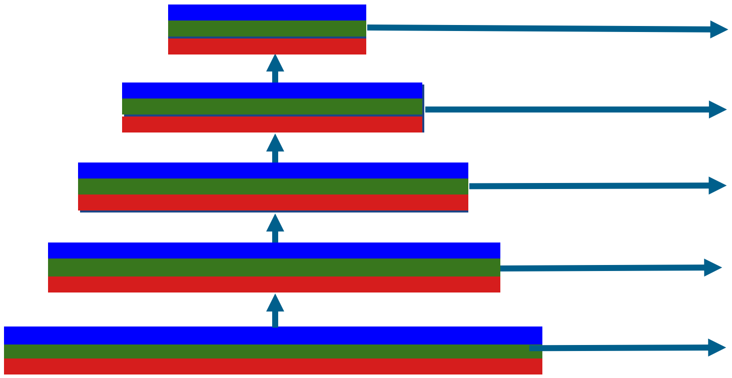
- **Dense pixel-wise matching and features**
  - Demands high spatial accuracy in features.
  - Localization is very important

- **Feature pyramids**
  - Use information from multiple scales
  - Helps in handling general problems like large motion.

[1] Tsung-Yi Lin et al. Feature pyramid networks for object detection". In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR 2017)

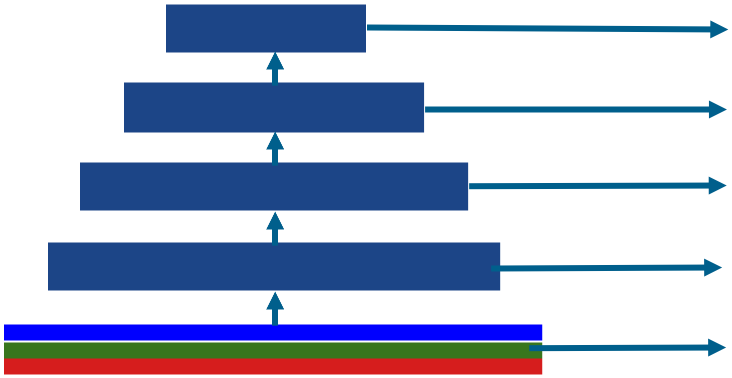
DenseFPN

- **Image Pyramids**



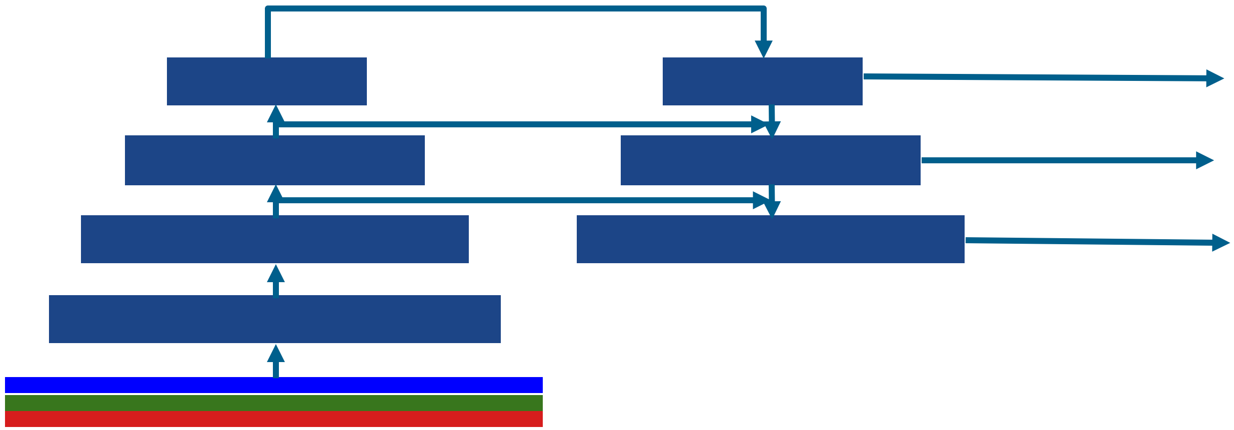
DenseFPN

- **Feature Pyramids**



DenseFPN

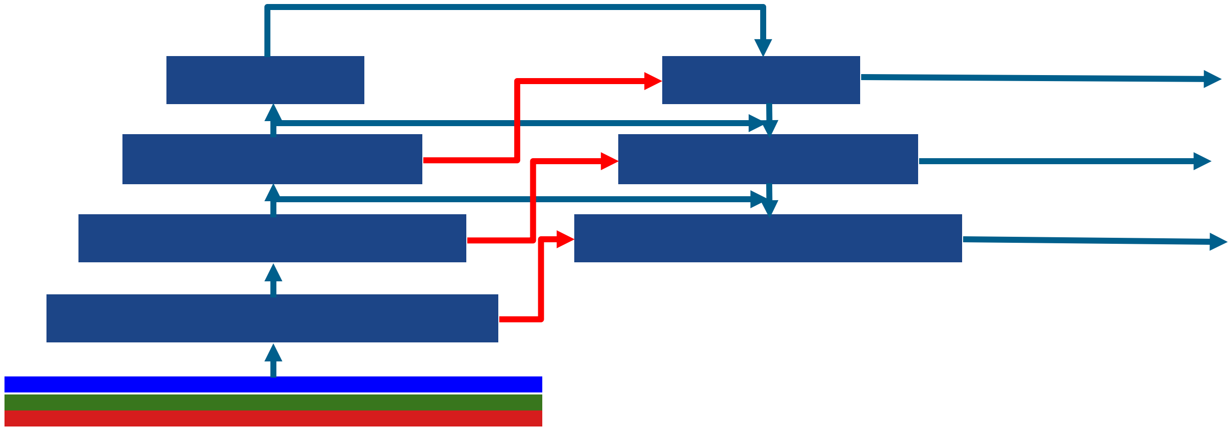
- **Feature Pyramid Network [1]**



[1] Tsung-Yi Lin et al. Feature pyramid networks for object detection". In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR 2017)

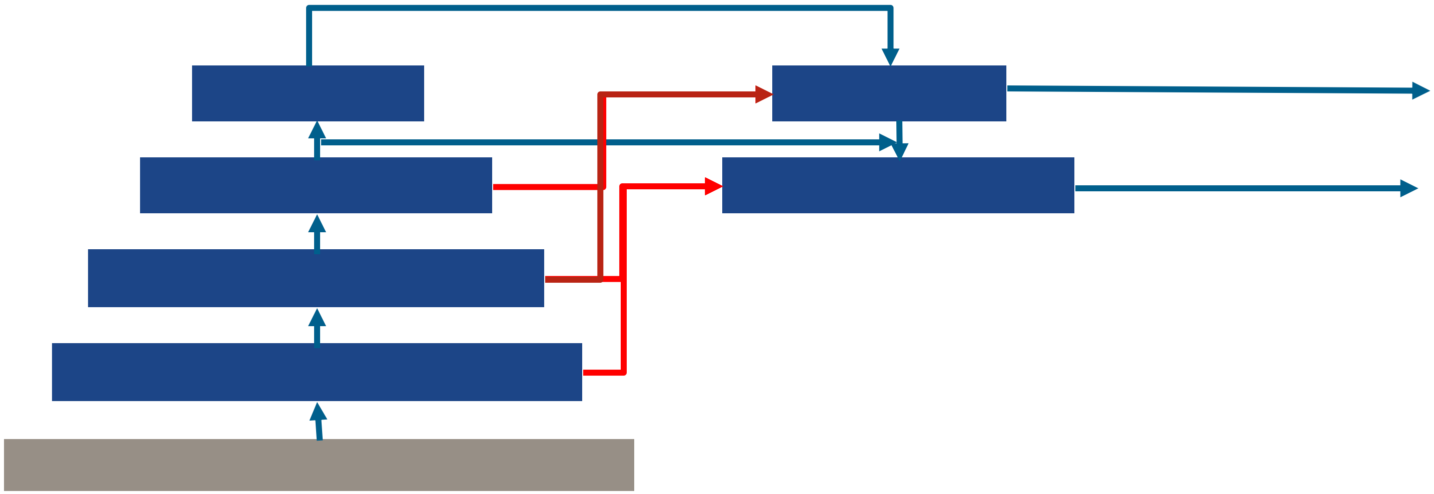
DenseFPN

- **Idea of DenseFPN**

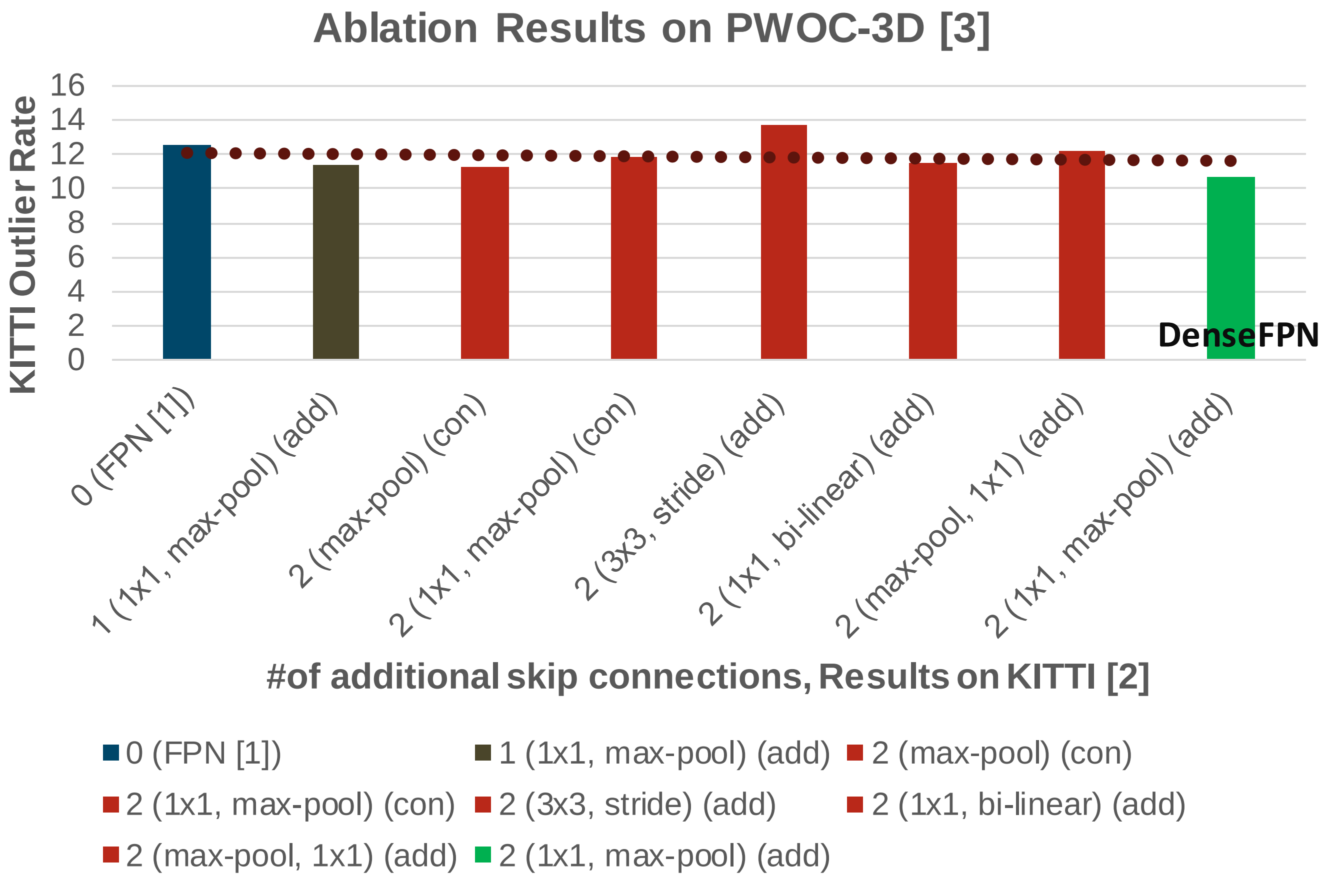


DenseFPN

- **Final Design (Ablation Next!)**

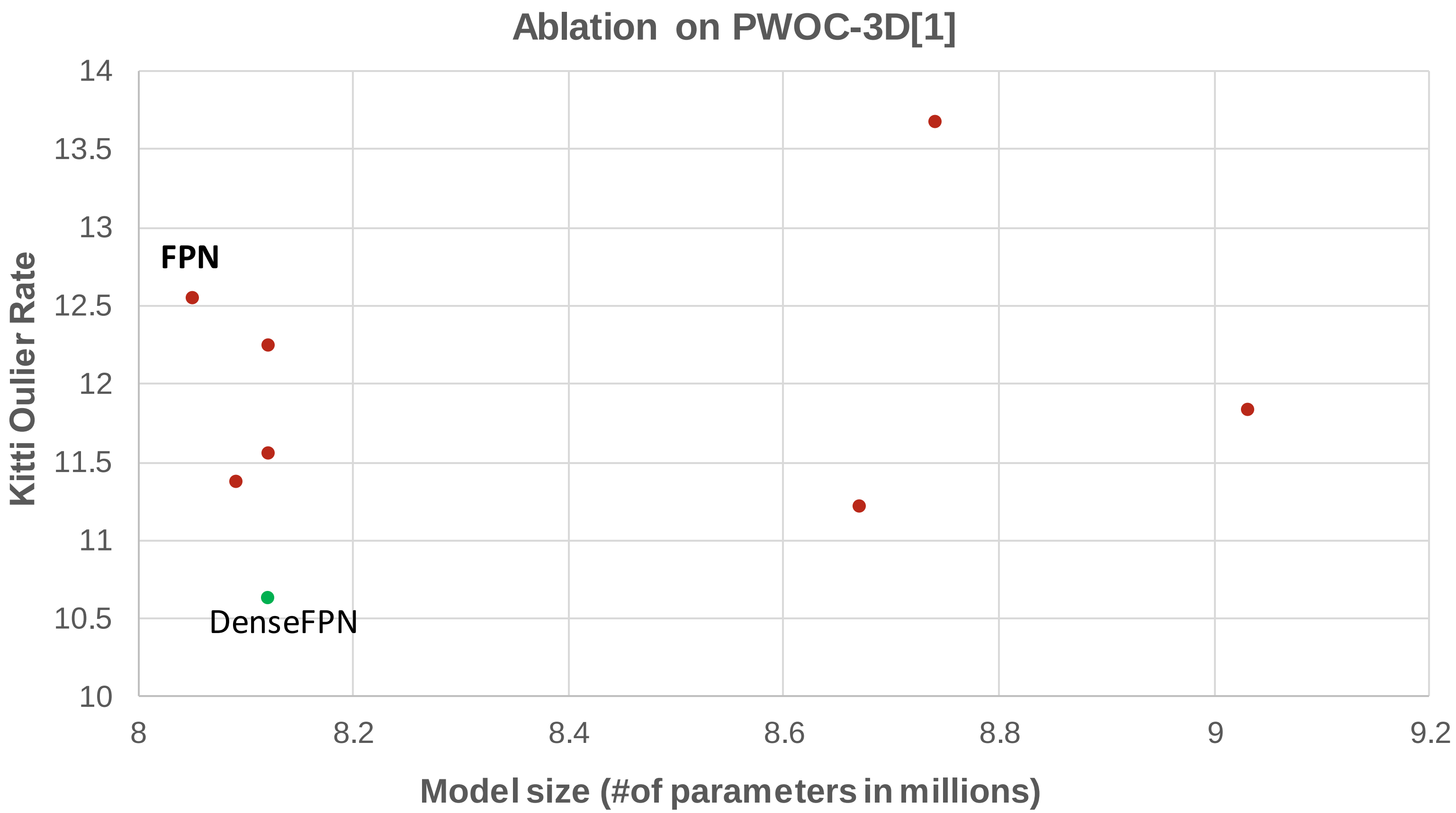


DenseFPN (Ablation)



[1] Tsung-Yi Lin et al. Feature pyramid networks for object detection". In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR 2017)
[2] Geiger et. Al, “Are we ready for autonomous driving? The KITTI vision benchmark suite” (CVPR 2015)
[3] Saxena et Al, “PWOC-3D: Deep Occlusion-Aware End-to-End Scene Flow Estimation” (IV, 2019)

DenseFPN (Ablation)



[1] Saxena et Al, “PWOC-3D: Deep Occlusion-Aware End-to-End Scene Flow Estimation” (IV, 2019)

DenseFPN

- **Results on State-of-the-Art Algorithms**

Results on FlyingThings3D

Task	Algorithm	%change KOE	%change in EPE
Stereo Disparity	PSM-Net [1]	30 %	21 %
Optical Flow	PWC-Net [2]	6 %	2.3 %
Optical Flow	LiteFlowNet [3]	8 %	8 %
Scene Flow	PWOC-3D [4]	12 %	7.5 %

References

[1] Jia-Ren Chang et. Al, "Pyramid stereo matching network" (CVPR 2018)
[2] Deqing Sun et Al, "CNNs for optical flow using pyramid, warping, and cost volume" (CVPR 2018)
[3] Tak-Wai Hui, "Liteflownet: A lightweight convolutional neural network for optical flow estimation" (CVPR 2018)
[4] Saxena et Al, "PWOC-3D: Deep Occlusion-Aware End-to-End Scene Flow Estimation" (IV, 2019)

DenseFPN

- **Results on State-of-the-Art Algorithms**

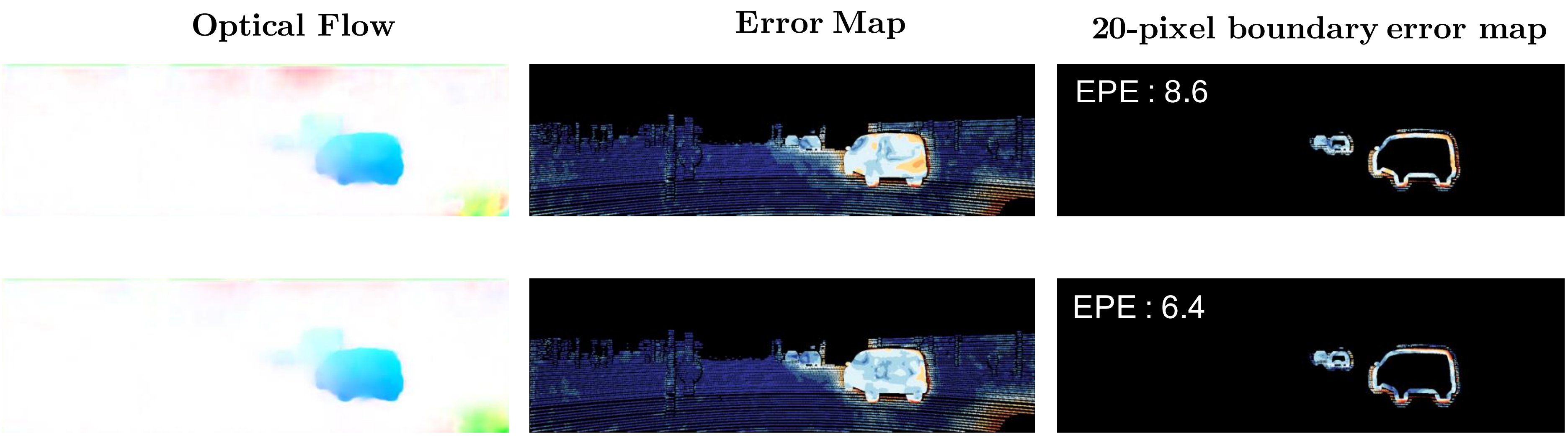
Results on KITTI

Task	Algorithm	%change KOE	%change in EPE
Stereo Disparity	PSM-Net [1]	15 %	0 %
Optical Flow	PWC-Net [2]	5 %	13 %
Optical Flow	LiteFlowNet [3]	8 %	5.4 %
Scene Flow	PWOC-3D [4]	15.8 %	6.2 %

References
[1] Jia-Ren Chang et. Al, “Pyramid stereo matching network“ (CVPR 2018)
[2] Deqing Sun et Al, “CNNs for optical flow using pyramid, warping, and cost volume” (CVPR 2018)
[3] Tak-Wai Hui, “Liteflownet: A lightweight convolutional neural network for optical flow estimation” (CVPR 2018)
[4] Saxena et Al, “PWOC-3D: Deep Occlusion-Aware End-to-End Scene Flow Estimation” (IV, 2019)

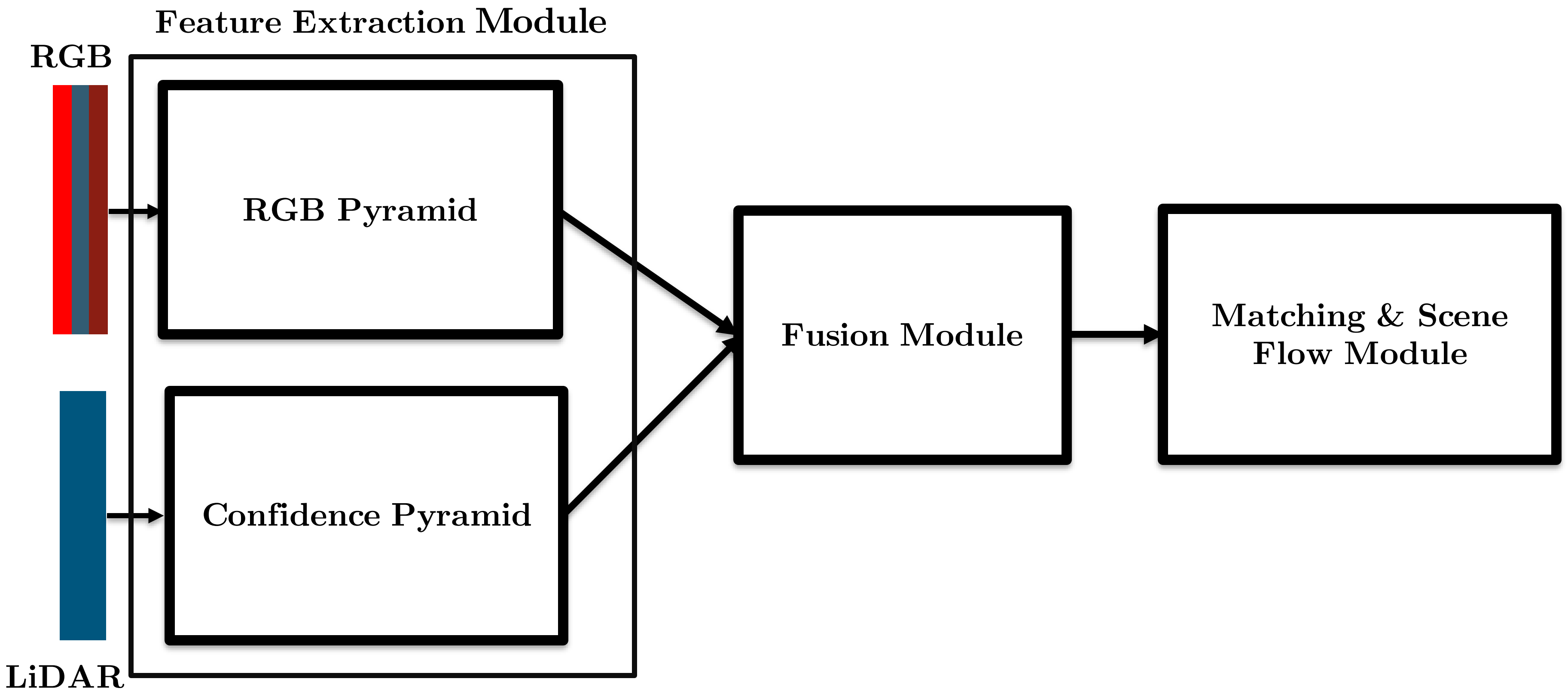
DenseFPN

- **Localization! (Output from LiteFlowNet [1] , top row BASELINE and bottom row DFPN)**

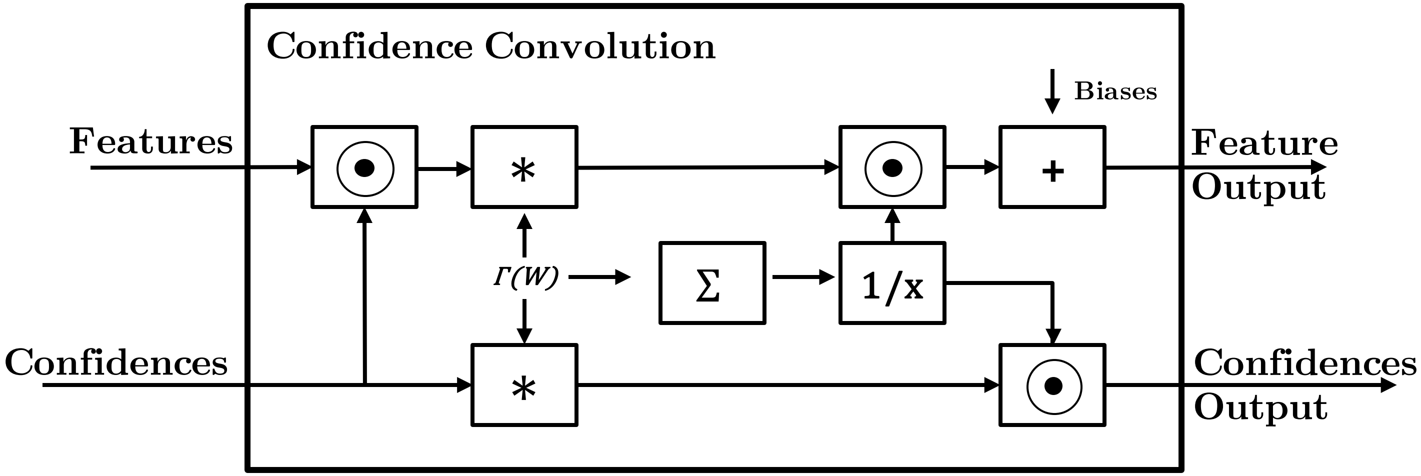


[1] Tak-Wai Hui, “Liteflownet: A lightweight convolutional neural network for optical flow estimation” (CVPR 2018)

An overview


Feature Extraction Module
RGB
RGB Pyramid
Fusion Module
Matching & Scene Flow Module
Confidence Pyramid
LiDAR

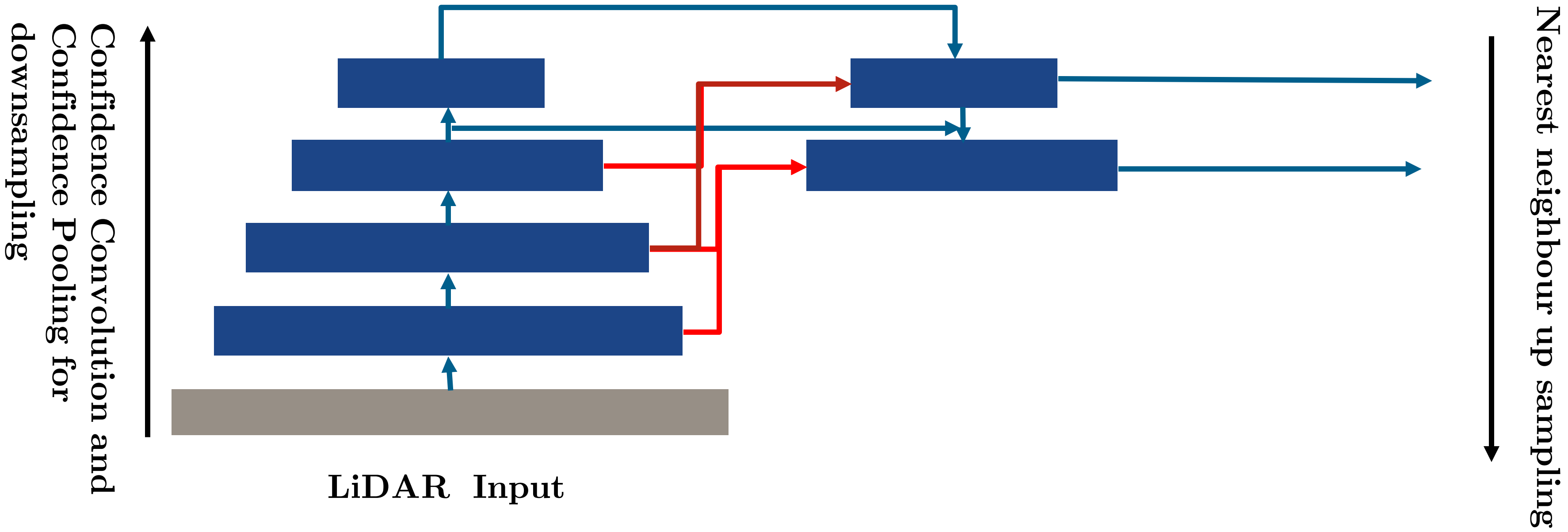
Confidence Convolution [1]



[1] Eldesokey et.Al, “Confidence propagation through cnns for guided sparse depth regression”, (PAMI 2019)

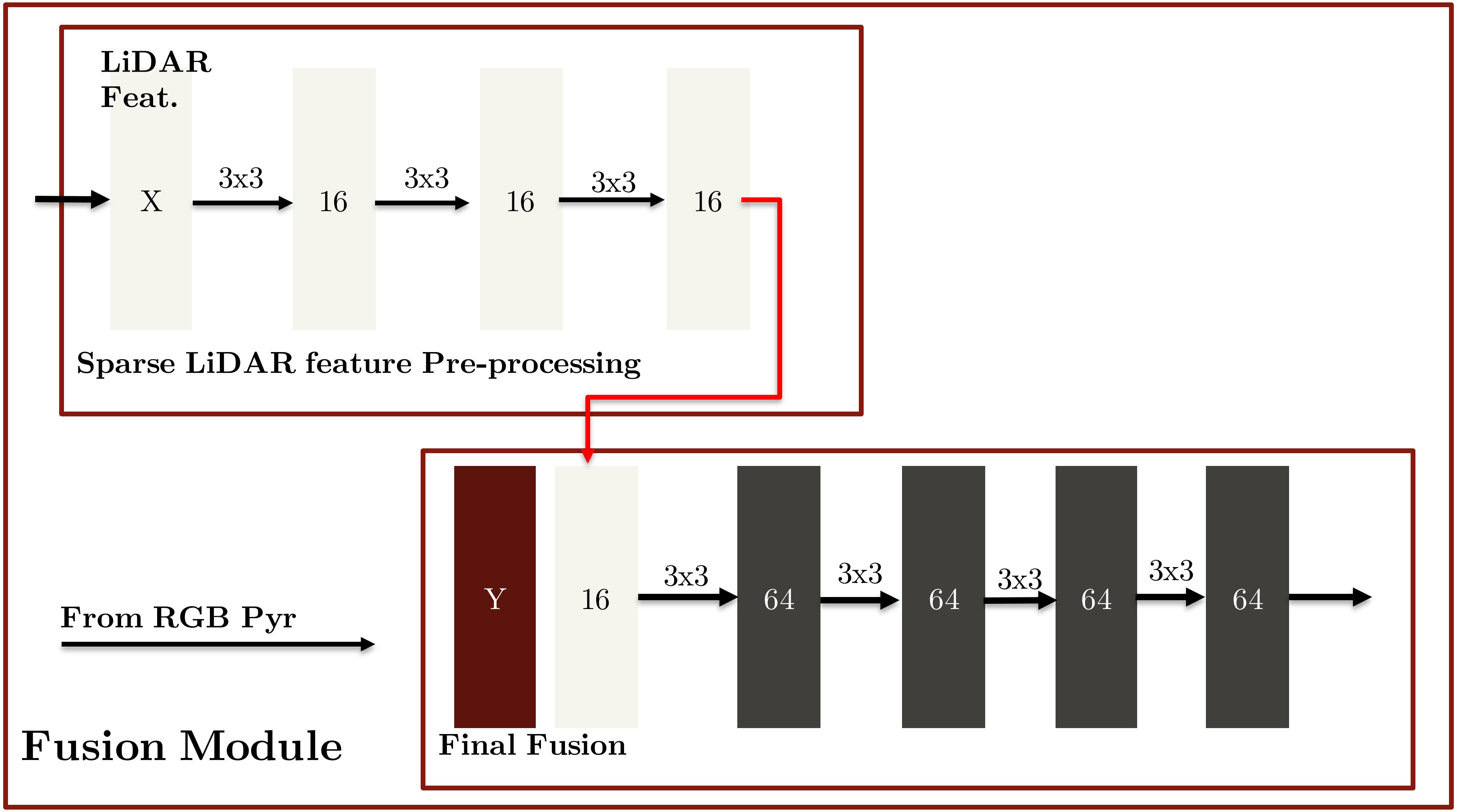
DeepLiDARFlow

- **Confidence Pyramid**



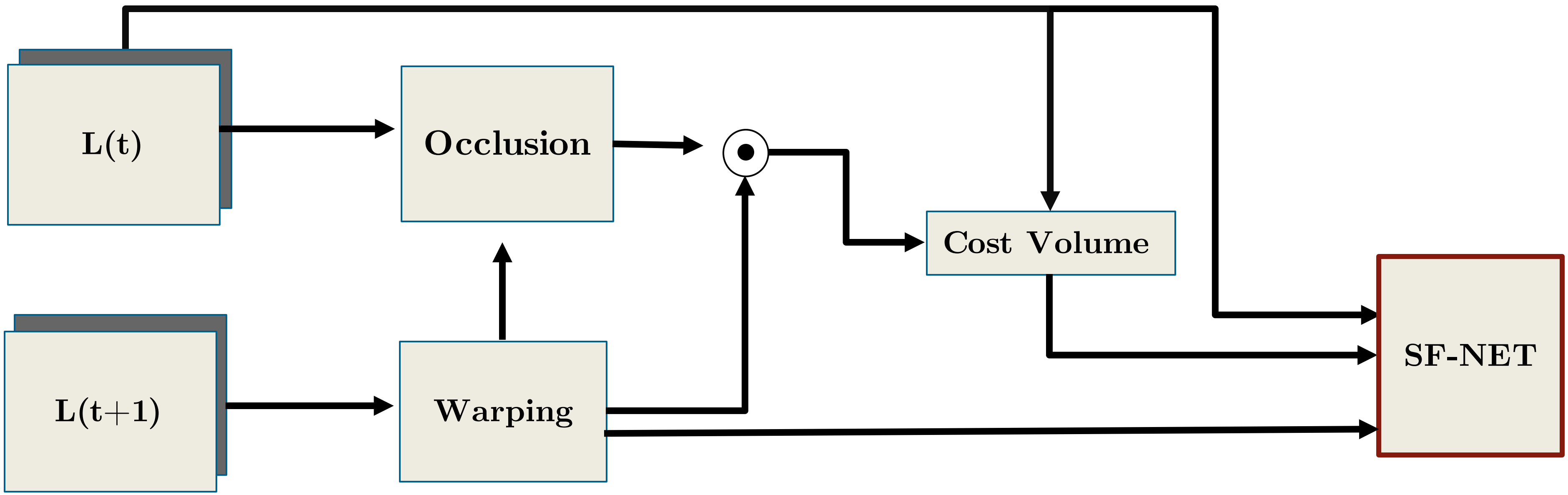
DeepLiDARFlow

Every scale



DeepLiDARFlow

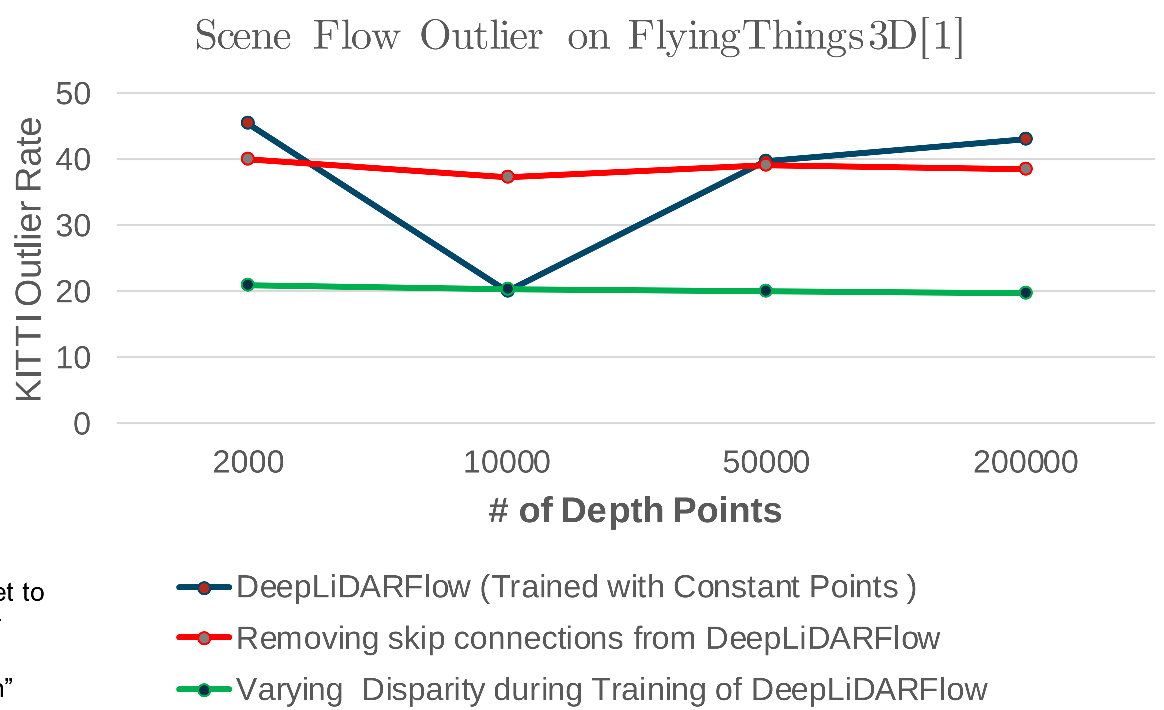
Matching and Scene Flow Estimation Module



L(t+1) and L(t) are fused RGB and LiDAR features. ⊙ Element wise Multiplication

Invariance to sparsity?

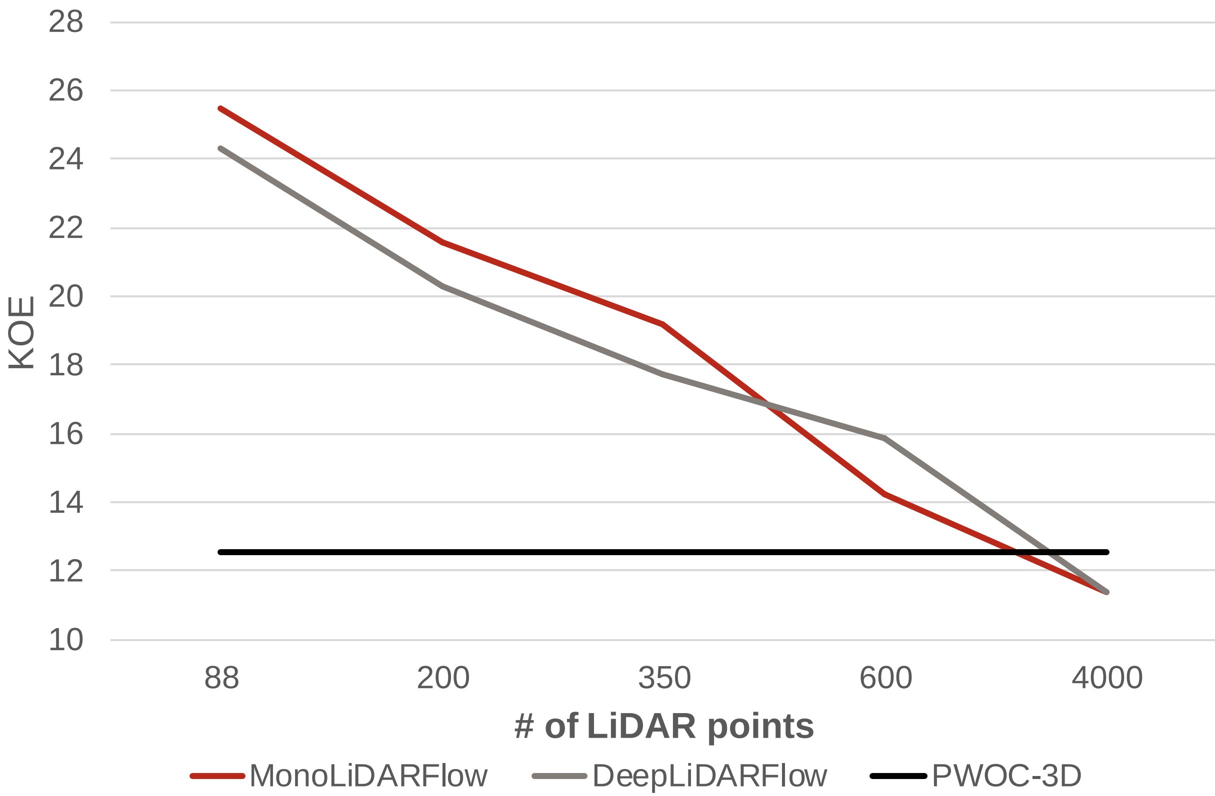
- **Important for generalization**
- **Two Strategies:**
  - Removing skip connections!
    - Invariance achieved but errors too high
  - Training with variable number of points
    - Achieved Similar performances across a wide density



[1] Mayer et al., “A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation” (CVPR 2016)

DeepLiDARFlow

- **DeepLiDARFlow vs MonoLiDARFlow* vs PWOC-3D [2] (KITTI 2015)**



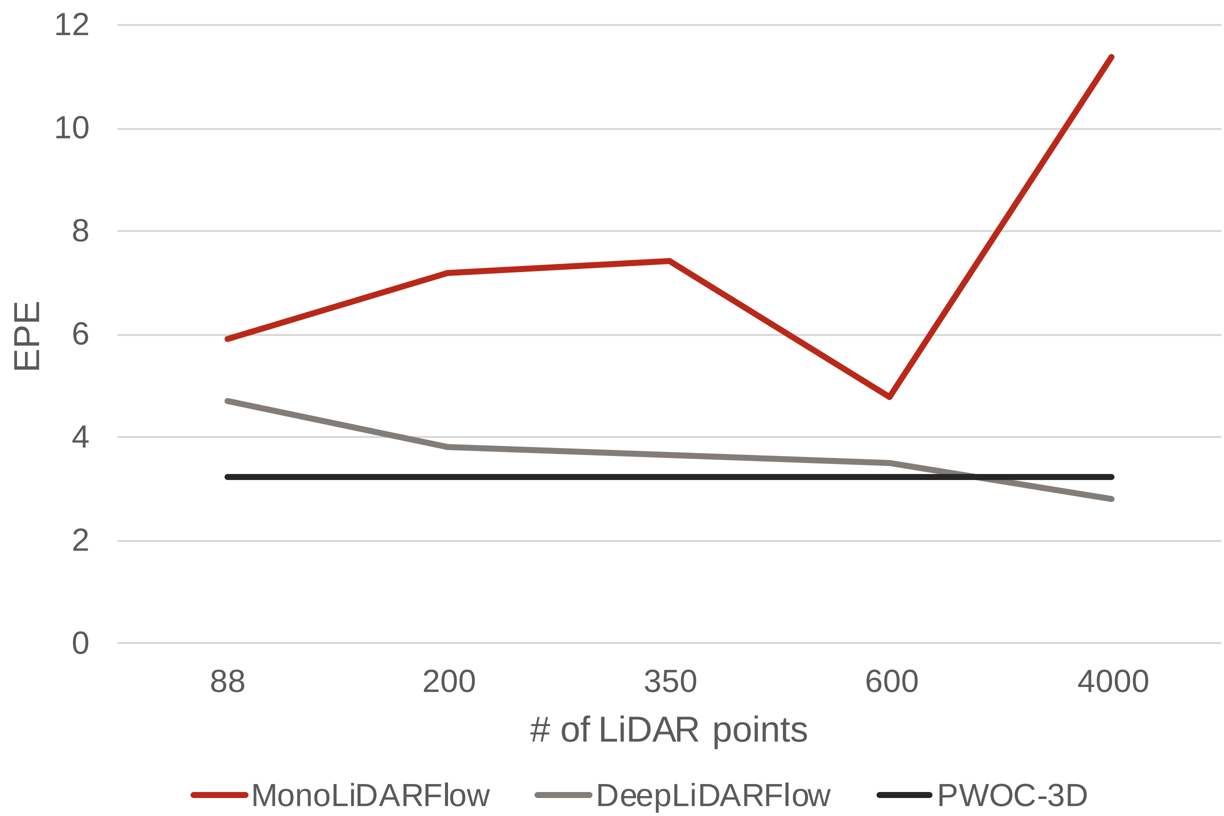
***Monocular version of LiDARFlow[1], Not yet published**

[1] Battrawy et al, "LiDAR-Flow: Dense Scene Flow Estimation from Sparse LiDAR and Stereo Images", (IROS 2019)

[2] Saxena et Al, "PWOC-3D: Deep Occlusion-Aware End-to-End Scene Flow Estimation" (IV, 2019)

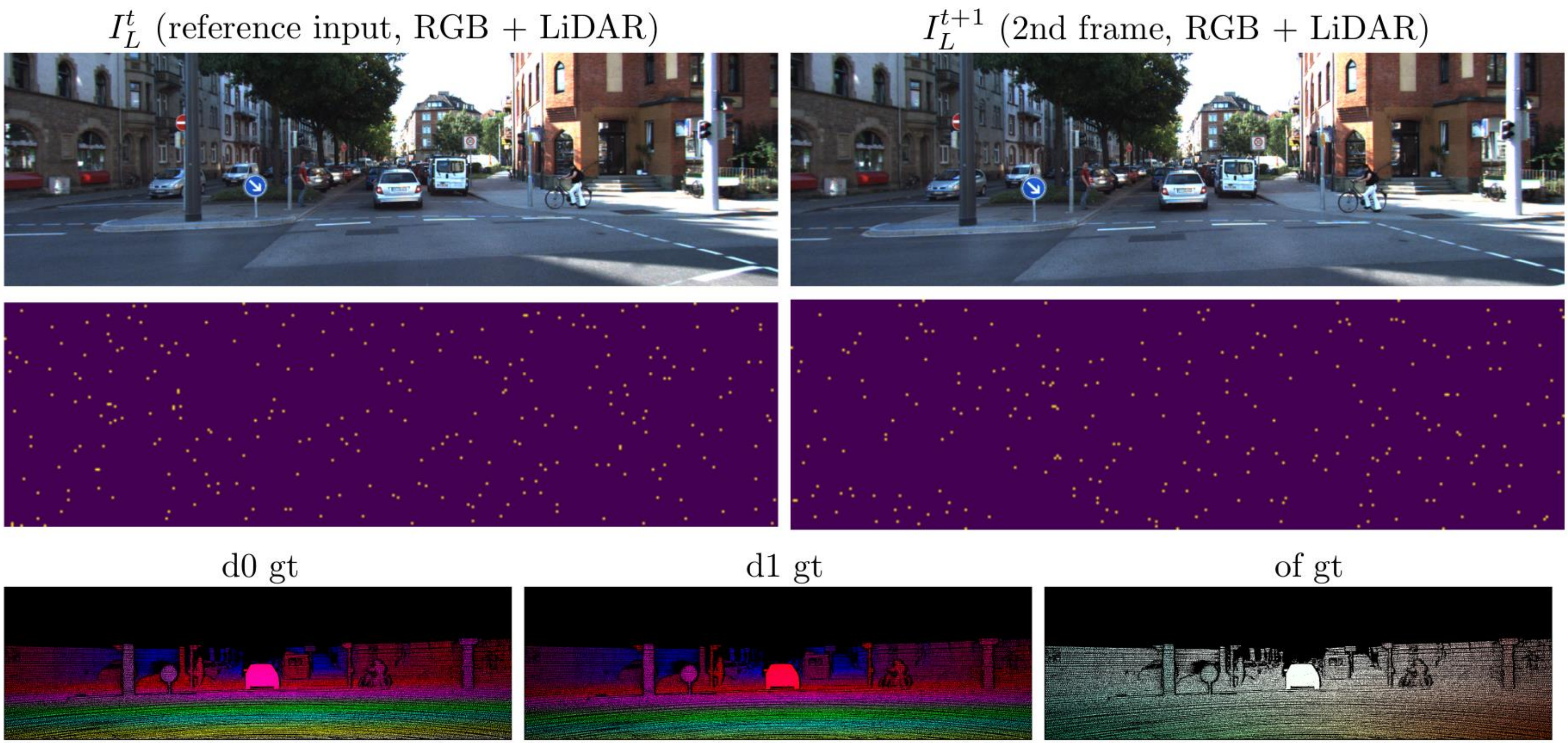
DeepLiDARFlow

- **DeepLiDARFlow vs MonoLiDARFlow vs PWOC-3D (KITTI 2015)**



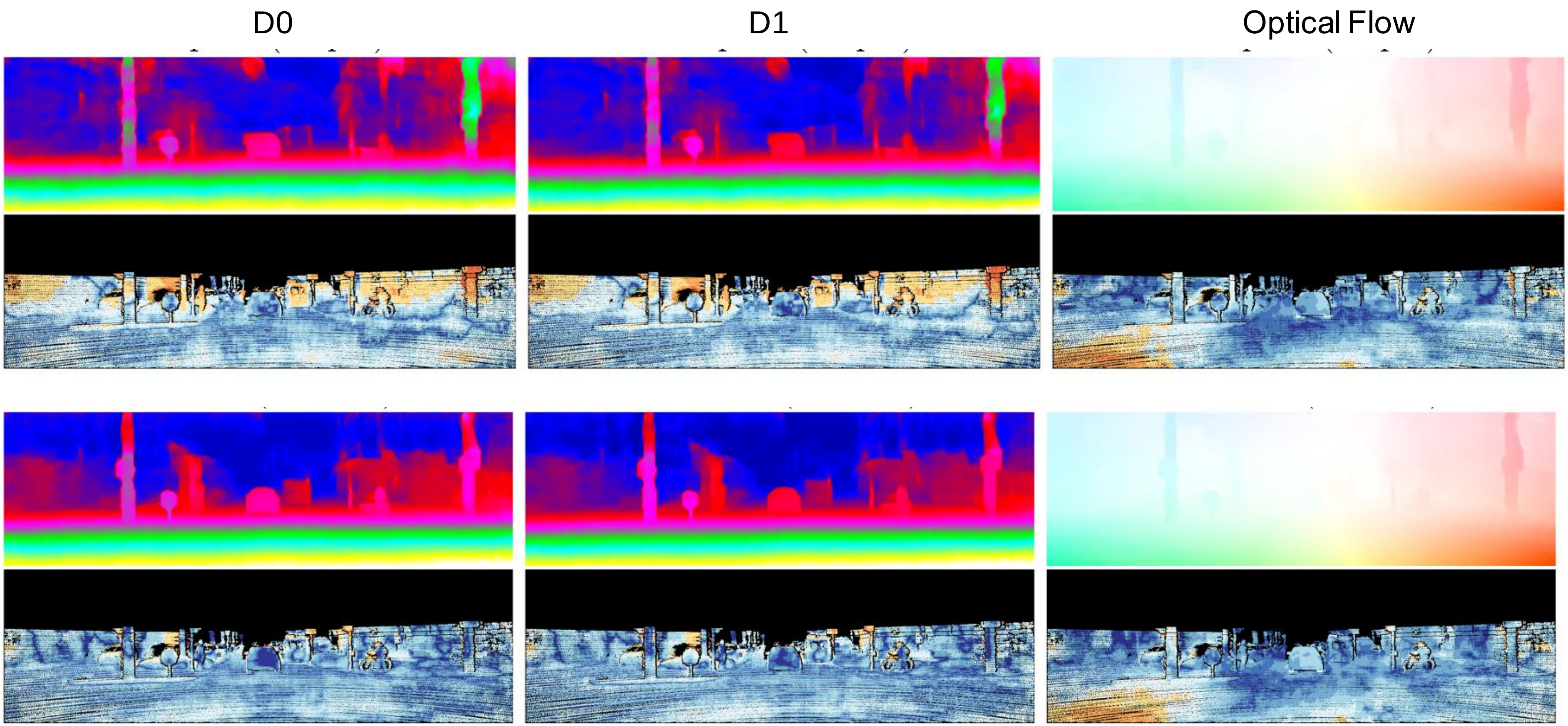
DeepLiDARFlow

- **Visuals (GT)**



DeepLiDARFlow

- **Visuals (KITTI with 88 LiDAR points (top) and 4000 points (bottom))**



Conclusion

DenseFPN

- Improved the accuracies of various dense matching tasks across datasets.
- Improved the localization of features.

DeepLiDARFlow

- Novel deep learning architecture for Scene flow using Monocular camera and sparse LiDAR.
- Outperformed MonoLiDARFlow for very sparse LiDAR points
- Outperformed PWOC-3D with monocular images and 4000 LiDAR points (6% points of total density)

Future Work

- Stage wise guidance from RGB pyramid
- Depth Representation
- Fusion Strategies
- Occlusion Estimator
- Confidence Based loss

Q&A