Research Progress

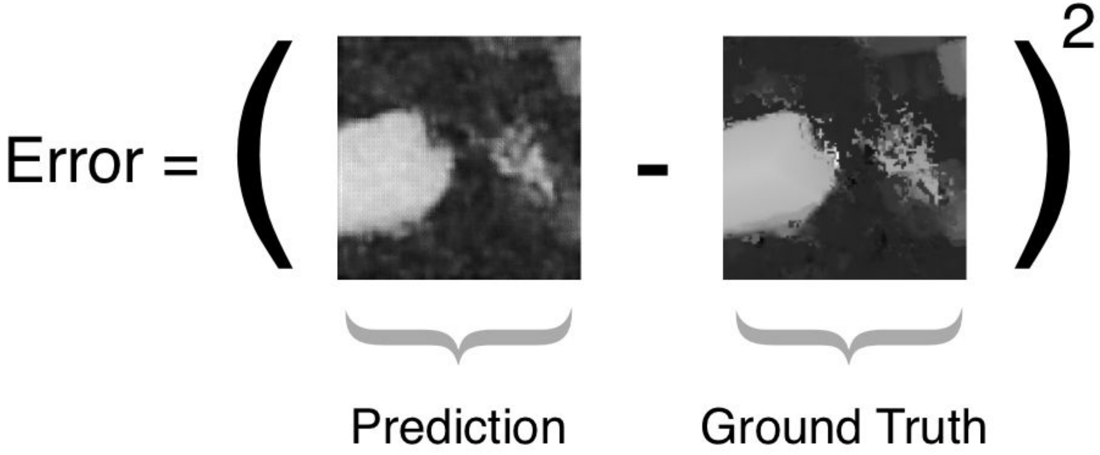
August 2nd

Depth Completion Task

- Input: RGB aerial image and corresponding depth map with some percentage replaced with zeros.

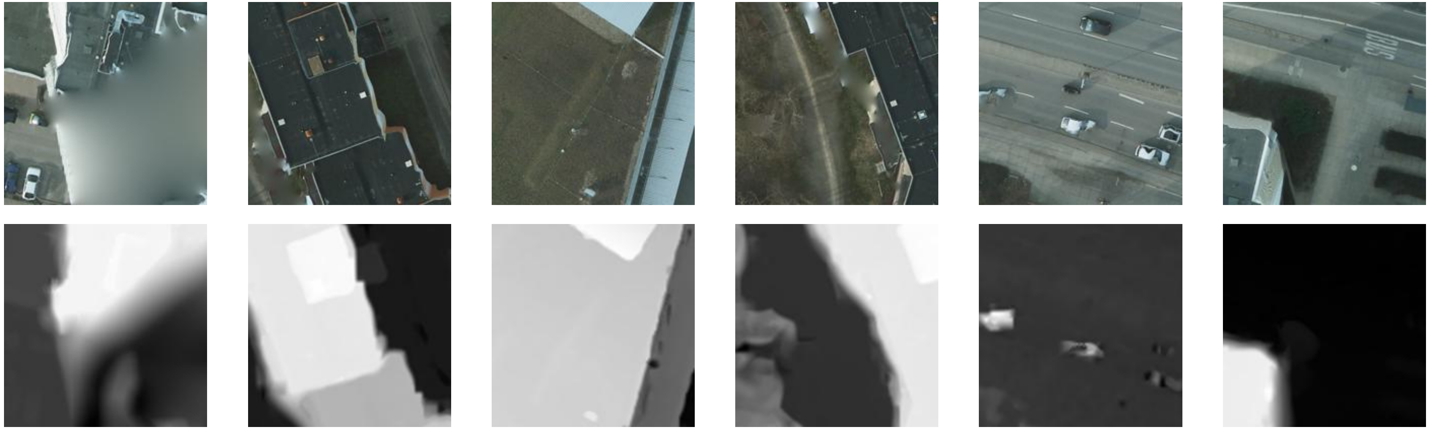
Depth Completion Task

- Output: Predicted depth map



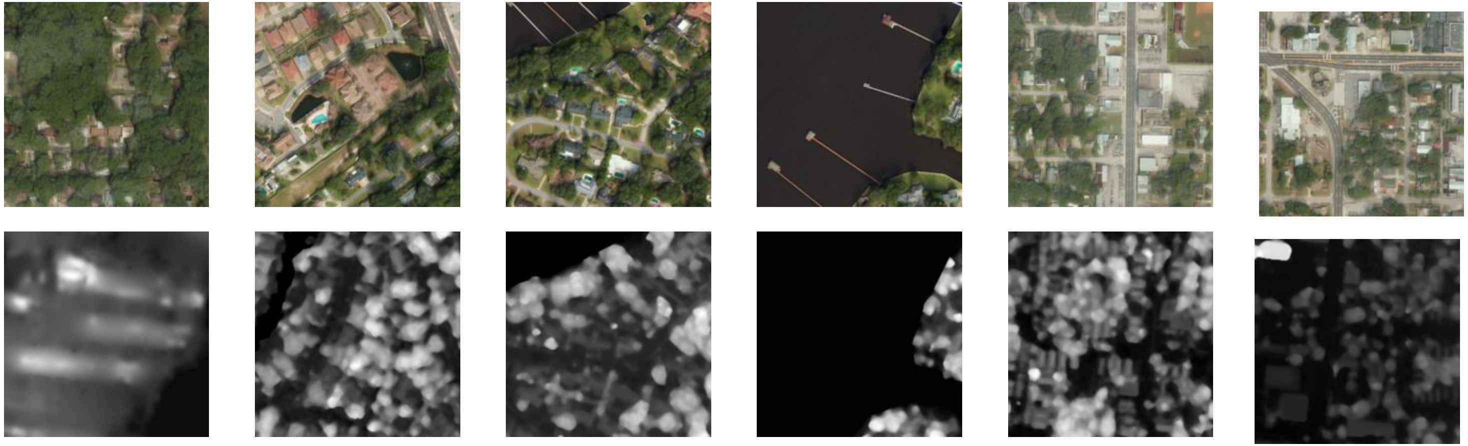
Dataset Overview - ISPRS

280 training images and 100 testing images of size 200x200.

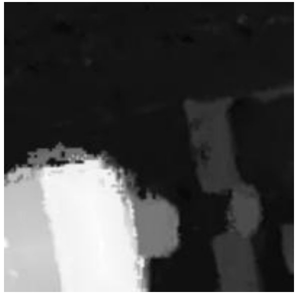


Dataset Overview - Urban 3D

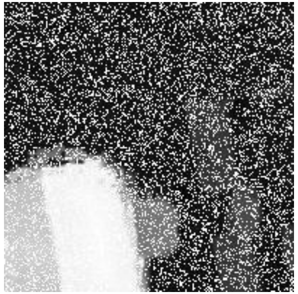
2,610 training images and 870 testing images of size 200x200.



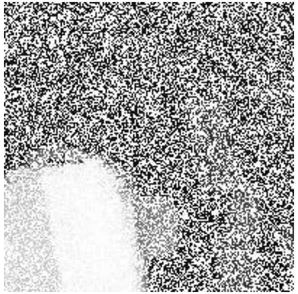
Sparsity Levels



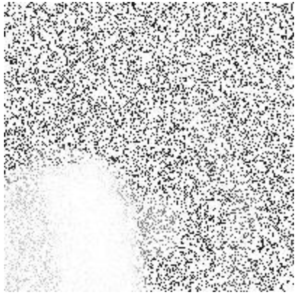
100%



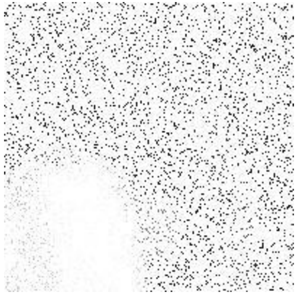
75%



50%

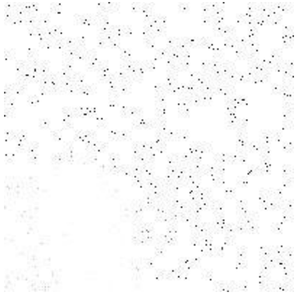


25%



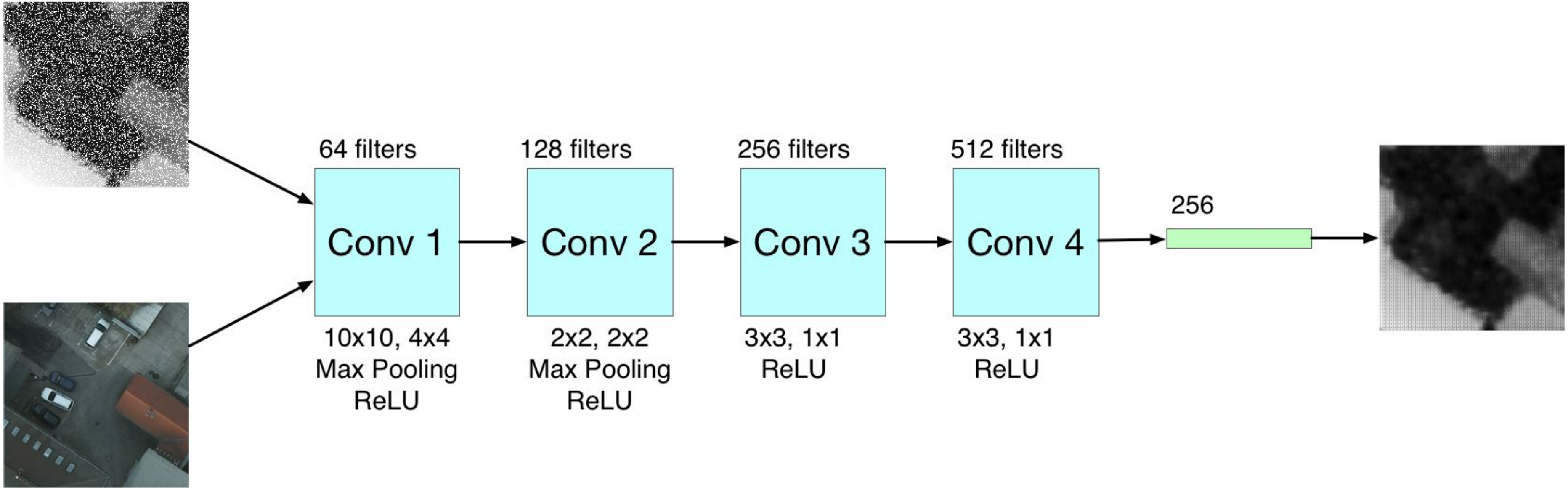
10%

5%



1%

Baseline Model

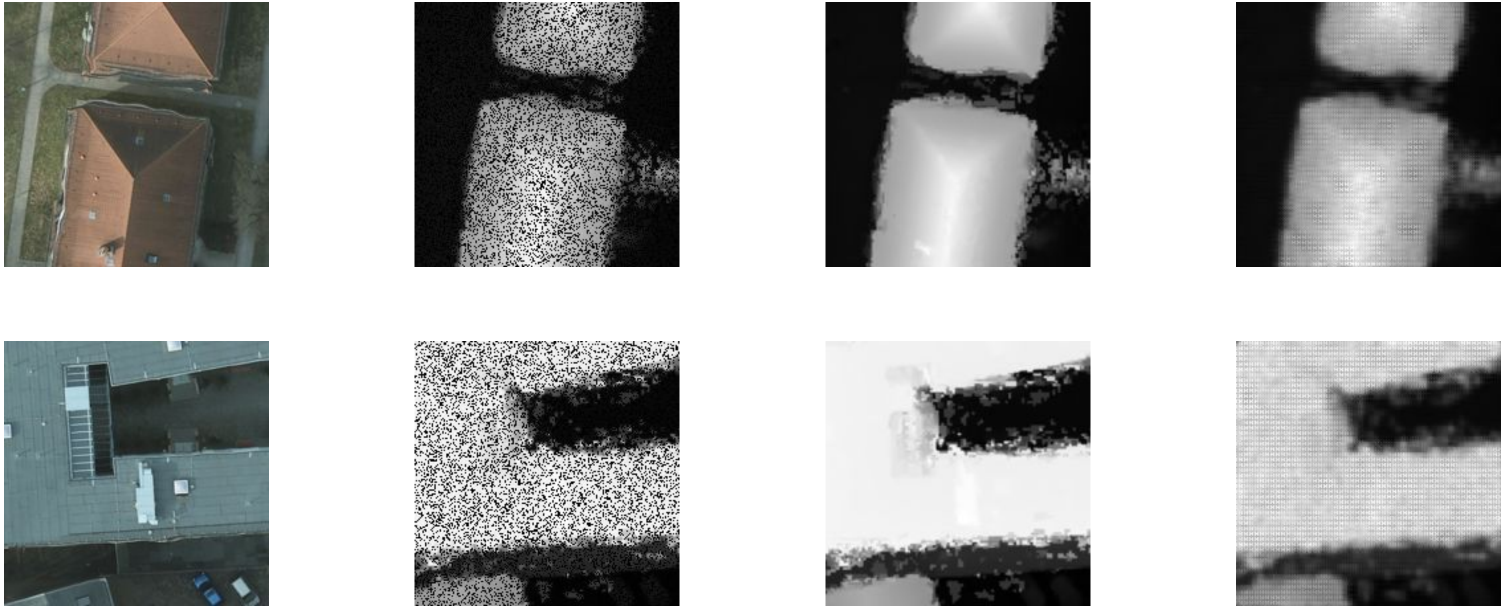


Baseline Performance

Sparsity	Training	Testing
1%	88.21612	88.81636
5%	92.60367	91.13361
10%	88.59271	97.71786
25%	73.41143	72.48245
50%	37.572323	38.391655
75%	24.560465	23.6261
100%	18.67808	22.053225

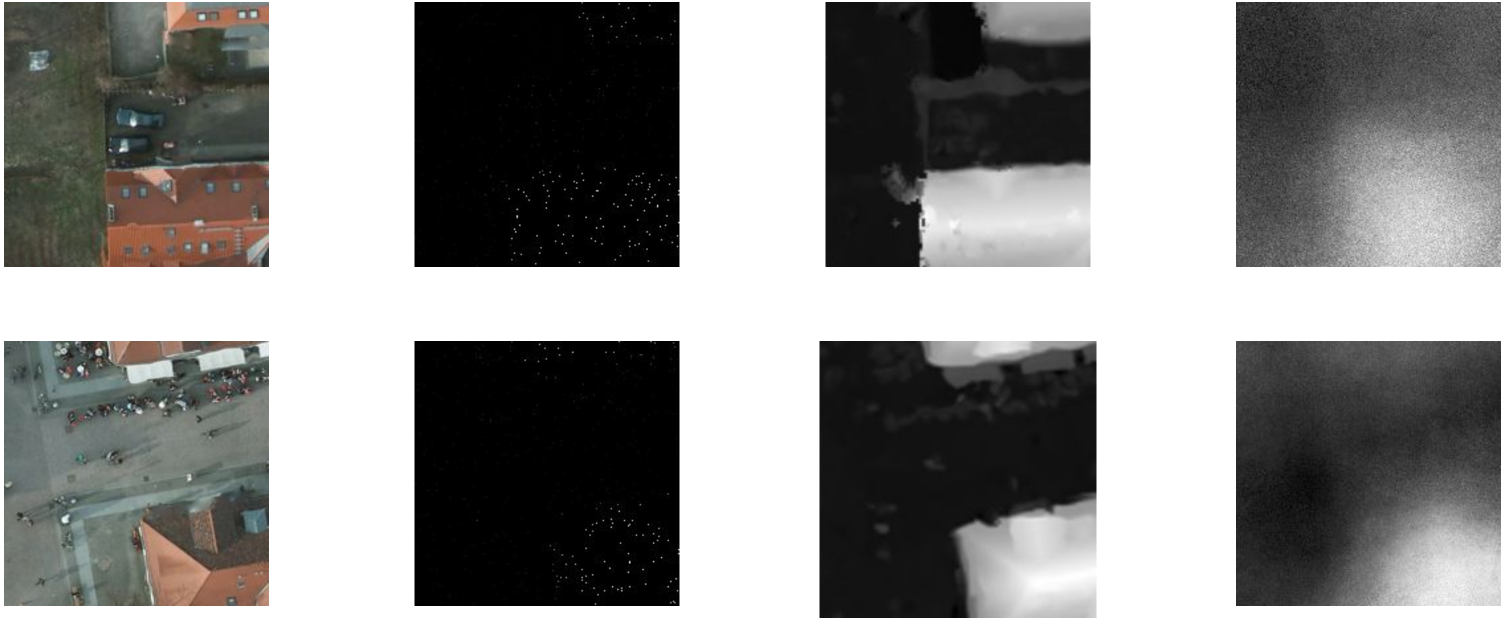
Sampling of Results - 75% Sparse

Image Sparse Ground Truth Prediction



Sampling of Results - 1% Sparse

Image | Sparse | Ground Truth | Prediction



State of the Art Model

Sparse and noisy LiDAR completion with RGB guidance and uncertainty

Wouter Van Gansbeke Davy Neven Bert De Brabandere Luc Van Gool
ESAT-PSI, KU Leuven
firstname.lastname@esat.kuleuven.be

Abstract

This work proposes a new method to accurately complete sparse LiDAR maps guided by RGB images. For autonomous vehicles and robotics the use of LiDAR is indispensable in order to achieve precise depth predictions. A multitude of applications depend on the awareness of their surroundings, and use depth cues to reason and react accordingly. On the one hand, monocular depth prediction methods fail to generate absolute and precise depth maps. On the other hand, stereoscopic approaches are still significantly outperformed by LiDAR based approaches. The goal of the depth completion task is to generate dense depth predictions from sparse and irregular point clouds which are mapped to a 2D plane. We propose a new framework which extracts both global and local information in order to produce proper depth maps. We argue that simple depth completion does not require a deep network. However, we additionally propose a fusion method with RGB guidance from a monocular camera in order to leverage object information and to correct mistakes in the sparse input. This improves the accuracy significantly. Moreover, confidence masks are exploited in

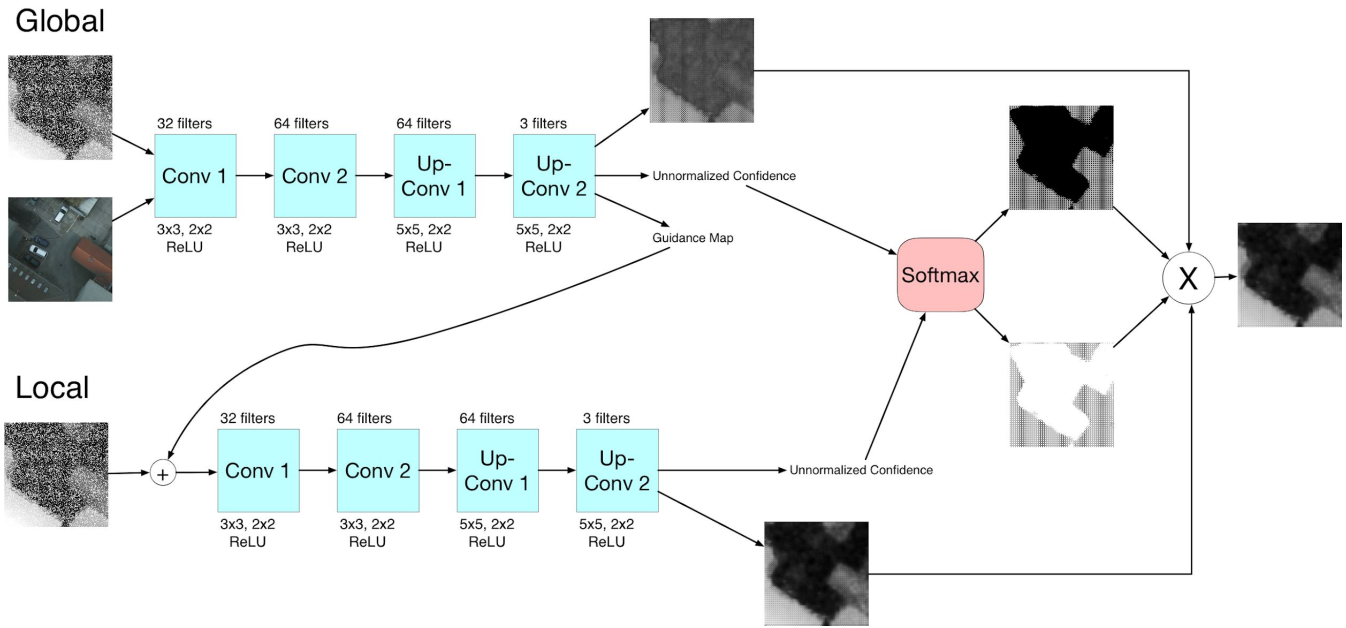
sparse and irregular spaced input points make this task stand out from others. Since a vast amount of applications use LiDAR with a limited amount of scan lines, the industrial relevance is indisputable, currently leading to a very active research domain. The reason why this task is challenging is threefold. Firstly, the input is randomly spaced which makes the usage of straightforward convolutions difficult. Secondly, the combination of multiple modalities is still an active area of research, since multiple combinations of sensor fusion are possible, namely early and/or late fusion. This paper will focus on the fusion between RGB info and the LiDAR points. Thirdly, the used annotations are only partially completed. The construction of the pixel-wise ground truth annotations is expensive after all. Our method needs to cope with this constraint.

The contributions of this paper are:

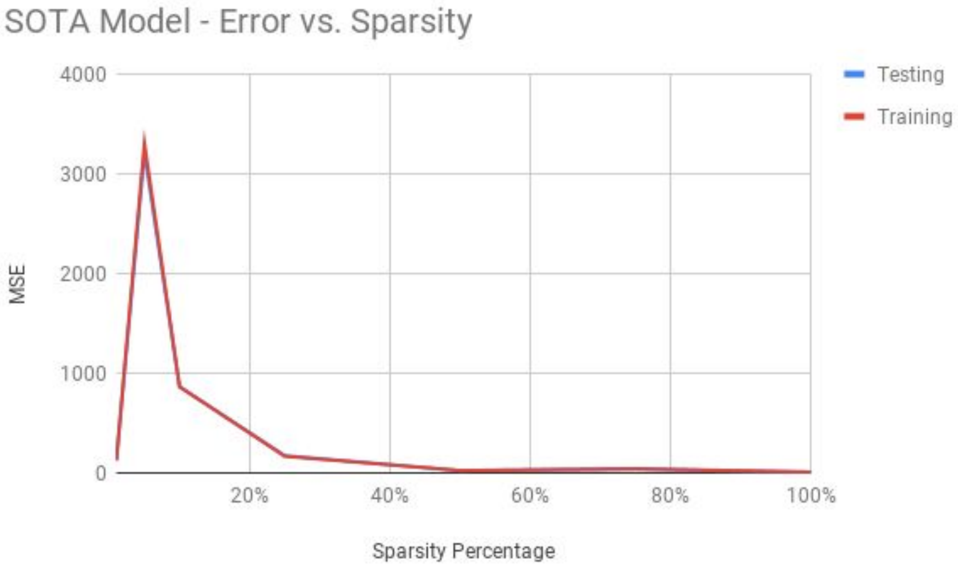
(1) Global and local information are combined in order to accurately complete and correct the sparse input. Monocular RGB images can be used as guidance for this depth completion task.

(2) Confidence maps are learned for both the global

State of the Art Model

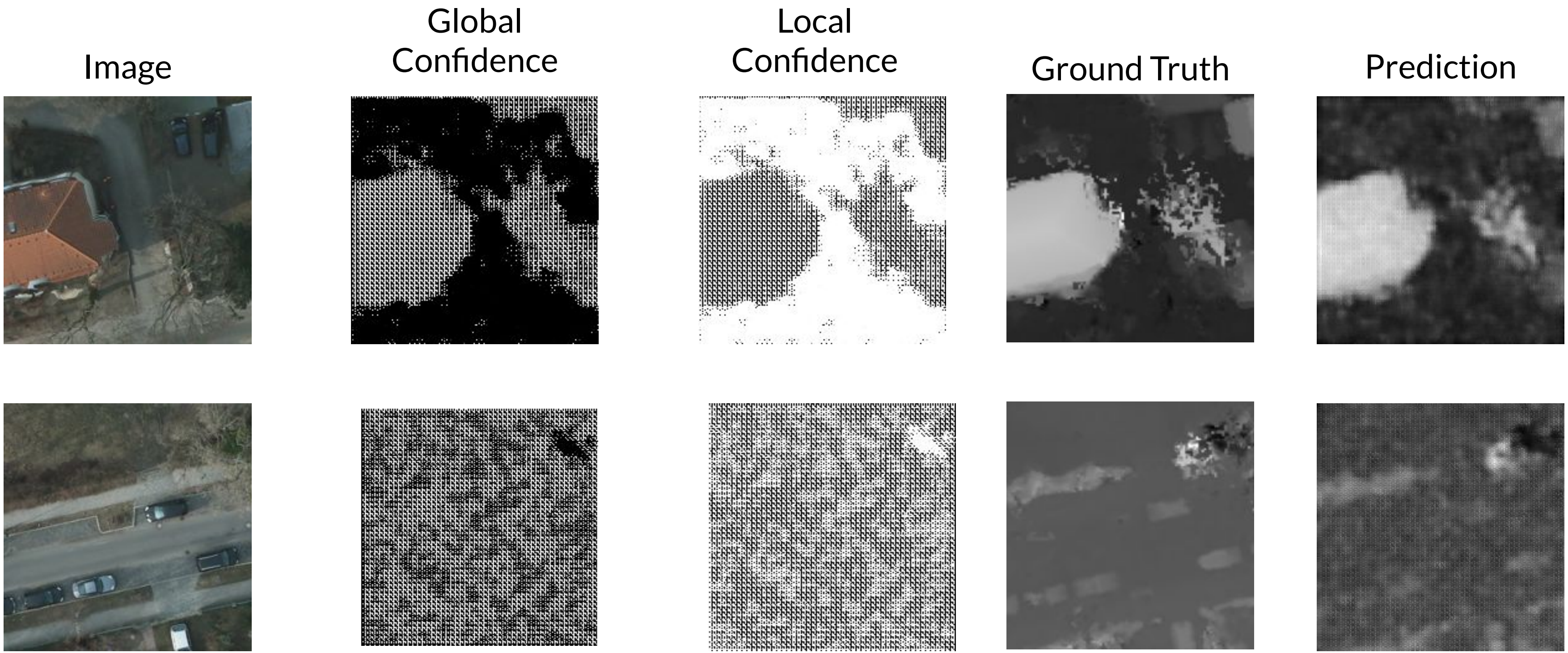


State of the Art Performance



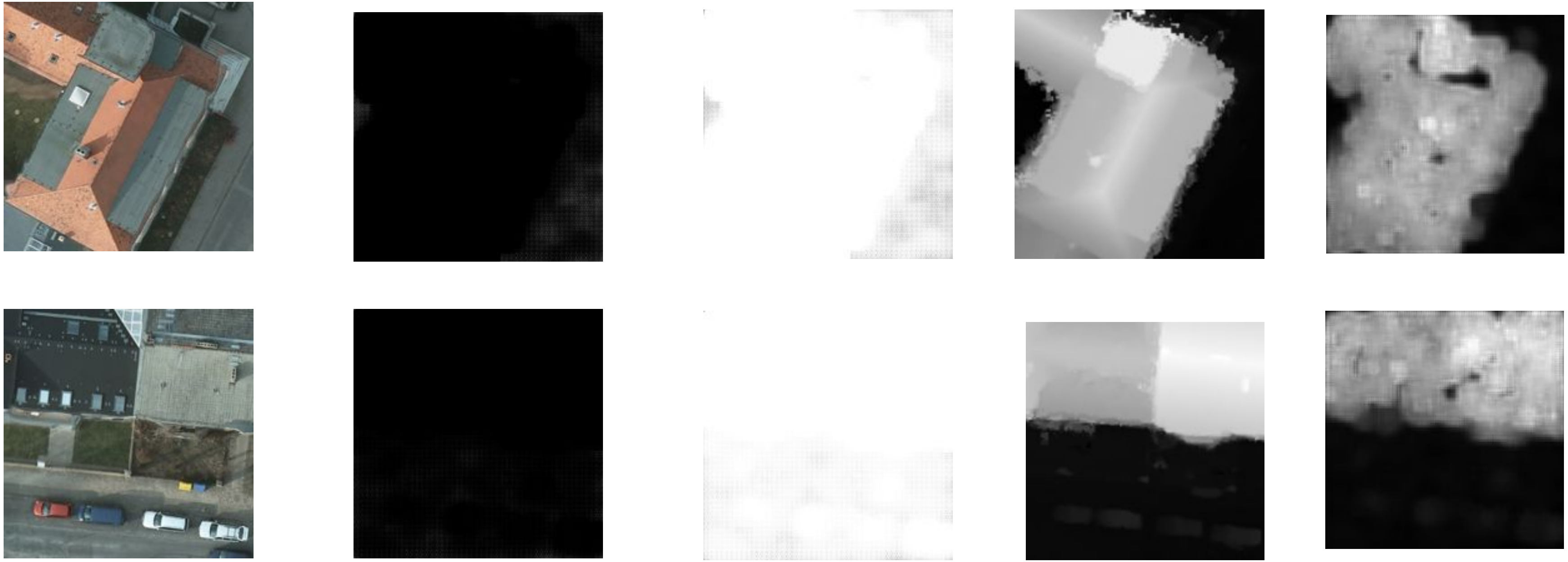
Sparsity	Training	Testing
1%	126.62895	124.65079
5%	3306.9077	3212.0188
10%	864.1157	865.7612
25%	167.64038	173.15616
50%	24.166473	23.901941
75%	43.312	37.389217
100%	8.802126	8.66935

Sampling of Results - 75% Sparse



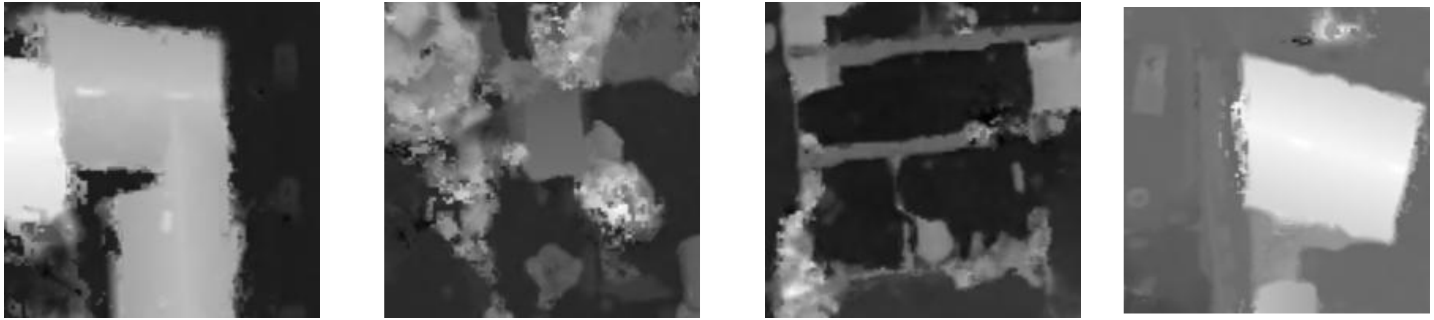
Sampling of Results - 1% Sparse

Image | Global Confidence | Local Confidence | Ground Truth | Prediction



Alternative - Pixel-wise Ranked Depth

Original Depth



Ranked Depth