Active Learning for Pixel-wise Prediction

October 11th, 2019

Outline

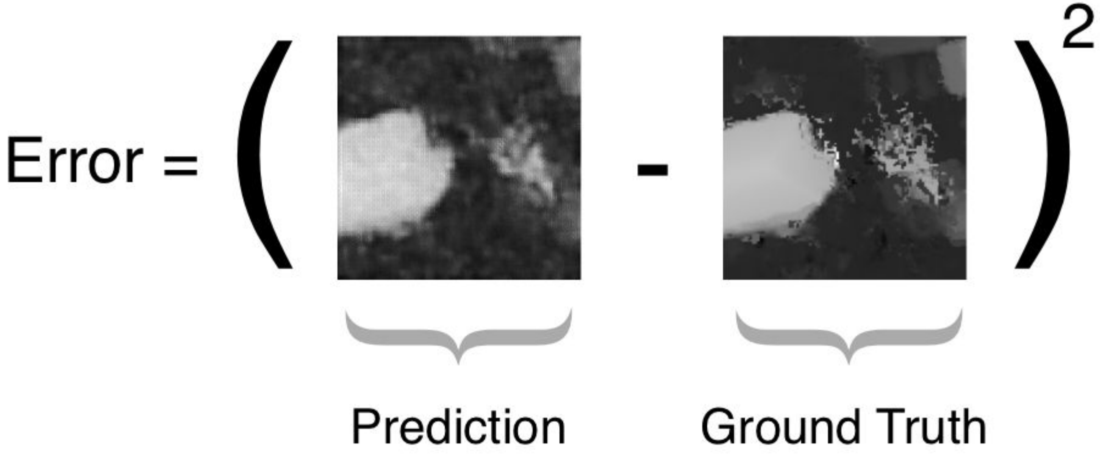
- Previous Work
- Generalization of Sparse Depth Completion
- Active Learning Methods
- Preliminary Results
- Future Work

Depth Completion Task

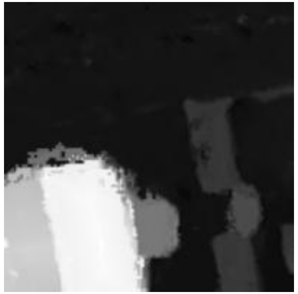
- Input: RGB aerial image and corresponding depth map with some percentage replaced with zeros.

Depth Completion Task

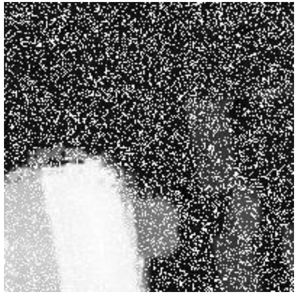
- Output: Predicted depth map



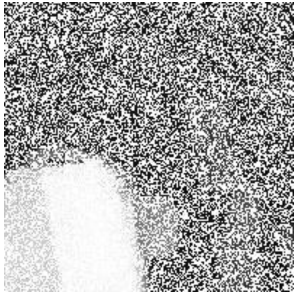
Sparsity Levels



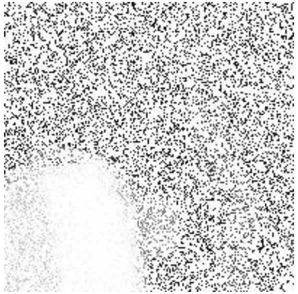
100%



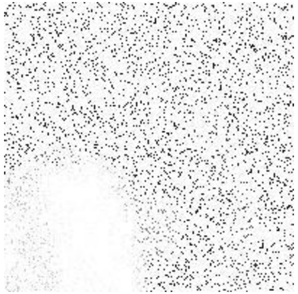
75%



50%

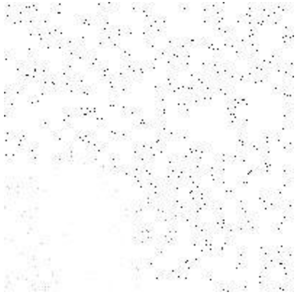


25%



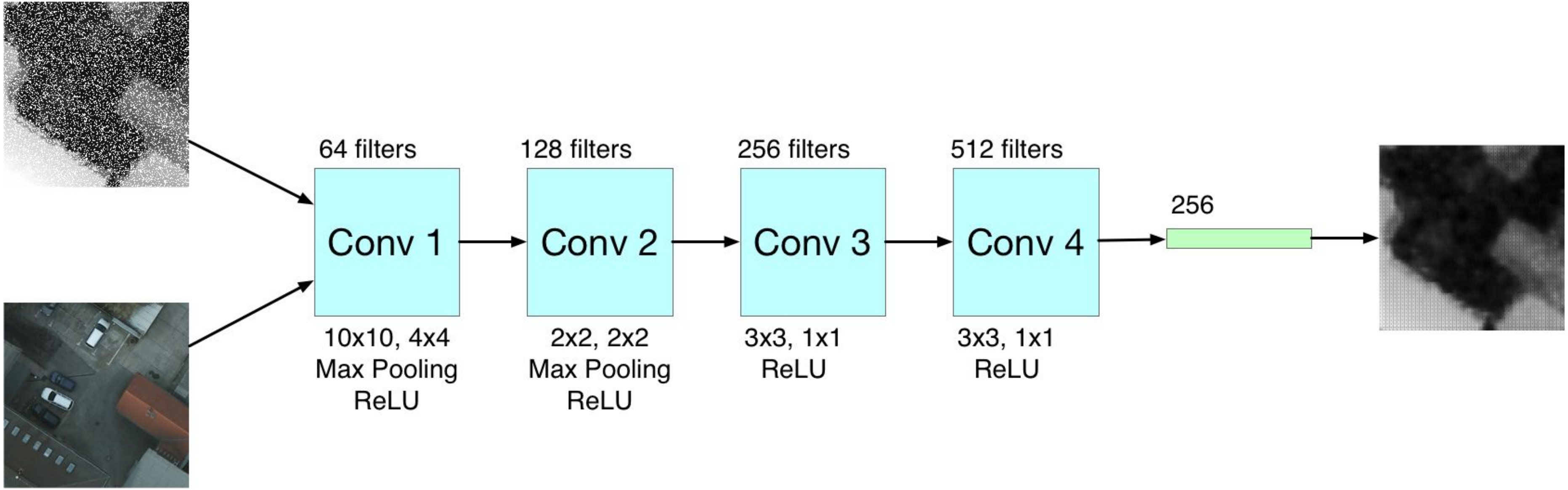
10%

5%



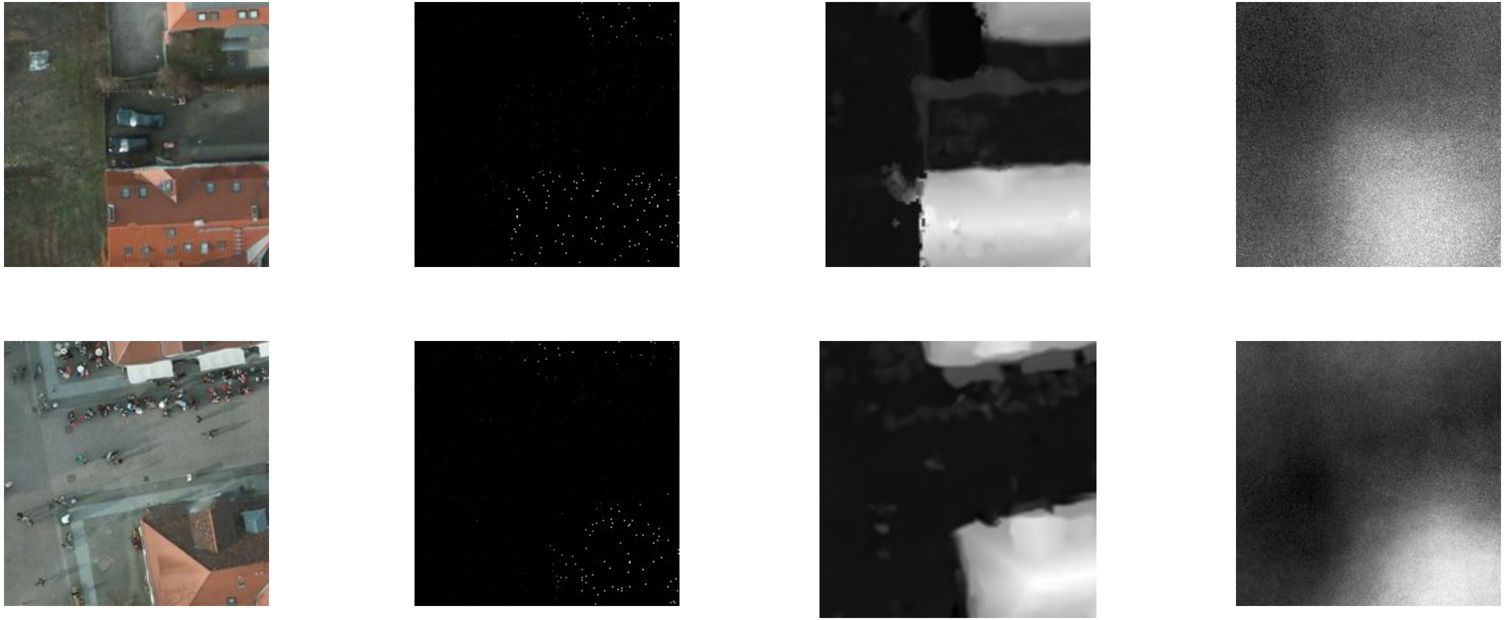
1%

Baseline Model

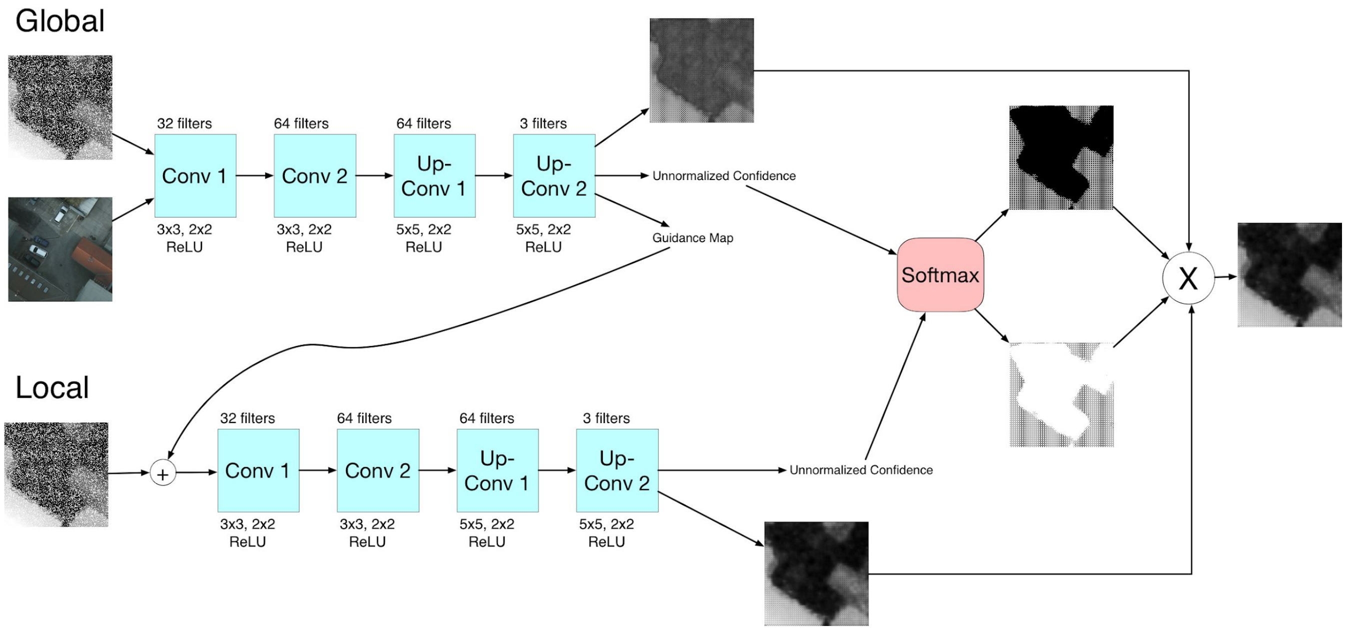


Baseline Results - 1% Sparse

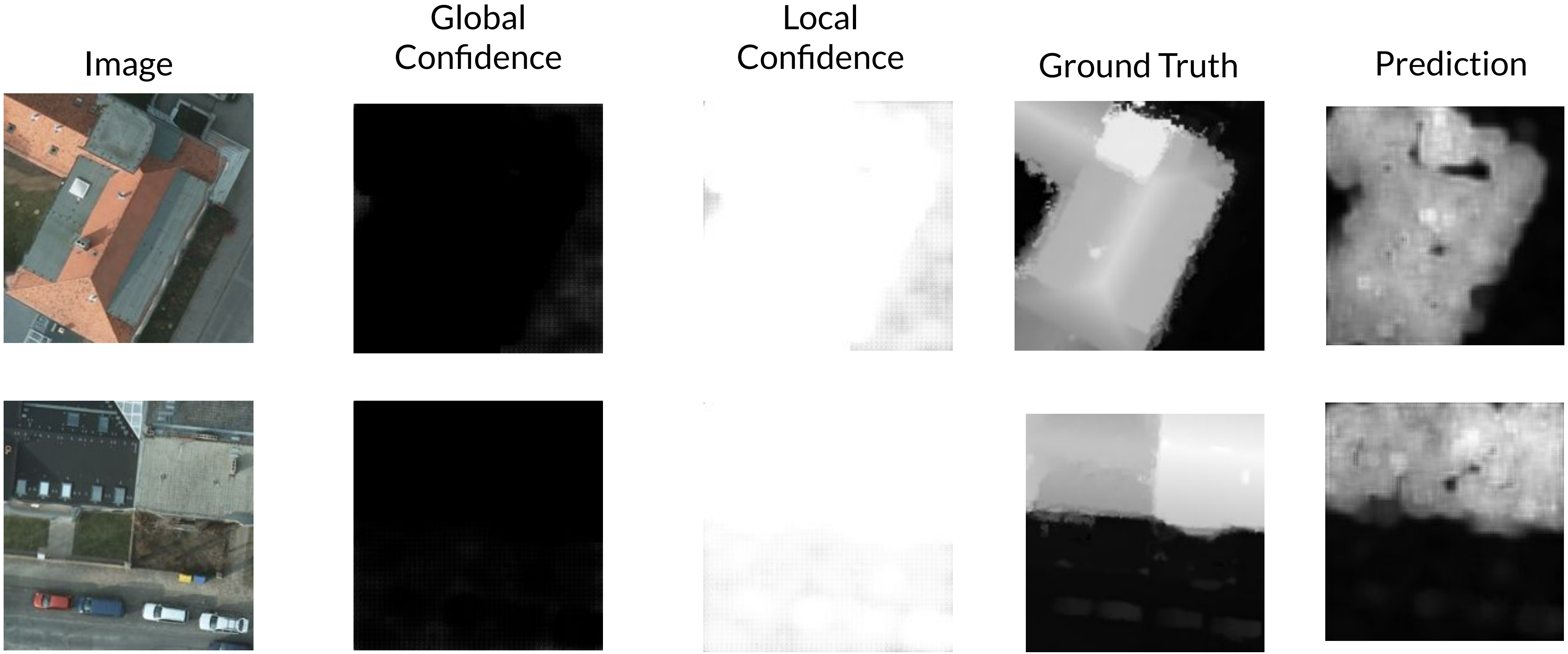
Image | Sparse | Ground Truth | Prediction



State of the Art Model



State of the Art - 1% Sparse



Limitations

While we’ve improved over the baseline for sparsities as low as 1%, we have been unable to produce quality results with lower sparsity. In order to move forward, we may need to choose the pixels we have associated heights for more carefully than random.

Depth Completion Generalization

Depth Completion is a specific form of a more general problem.

Given an image and a prediction task, select individual pixels for annotation so that given this information along with the image, your learning task is easiest.

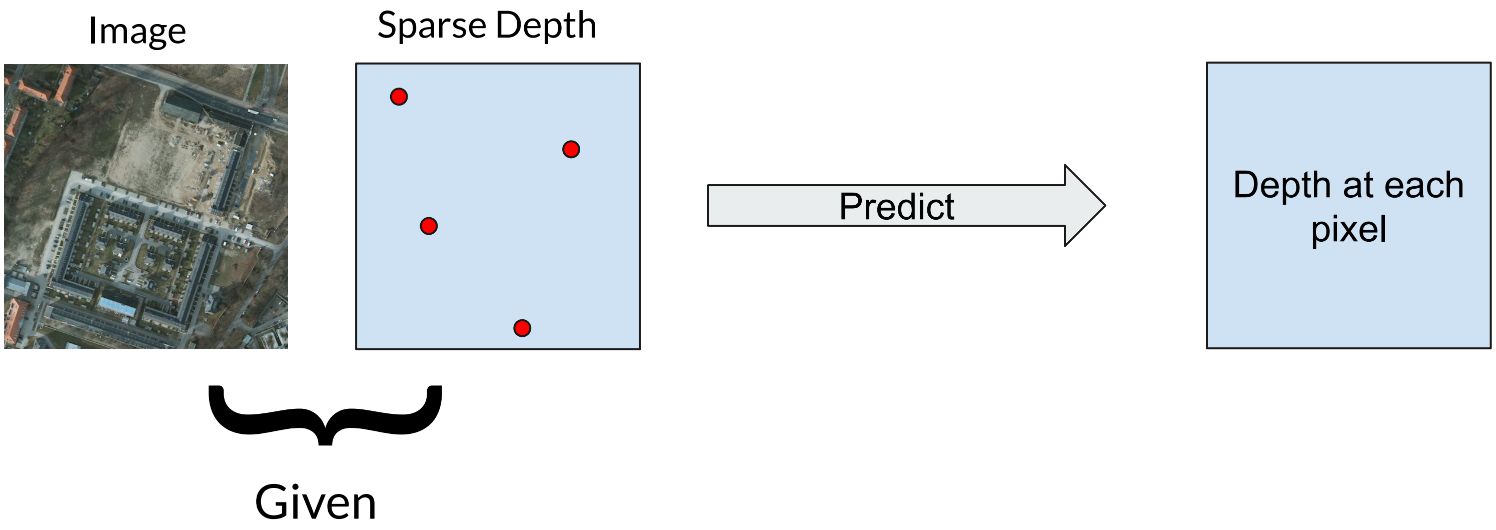
Depth Completion

Predict

Input Image

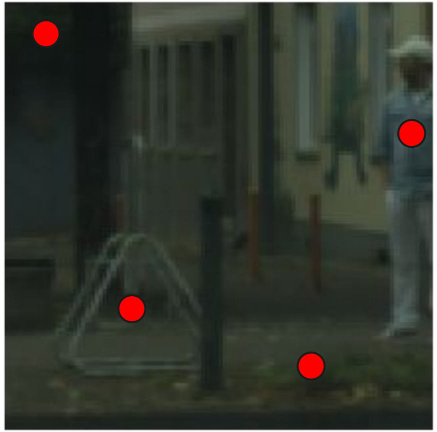
Suggested Annotation Locations

Depth Completion



Semantic Segmentation

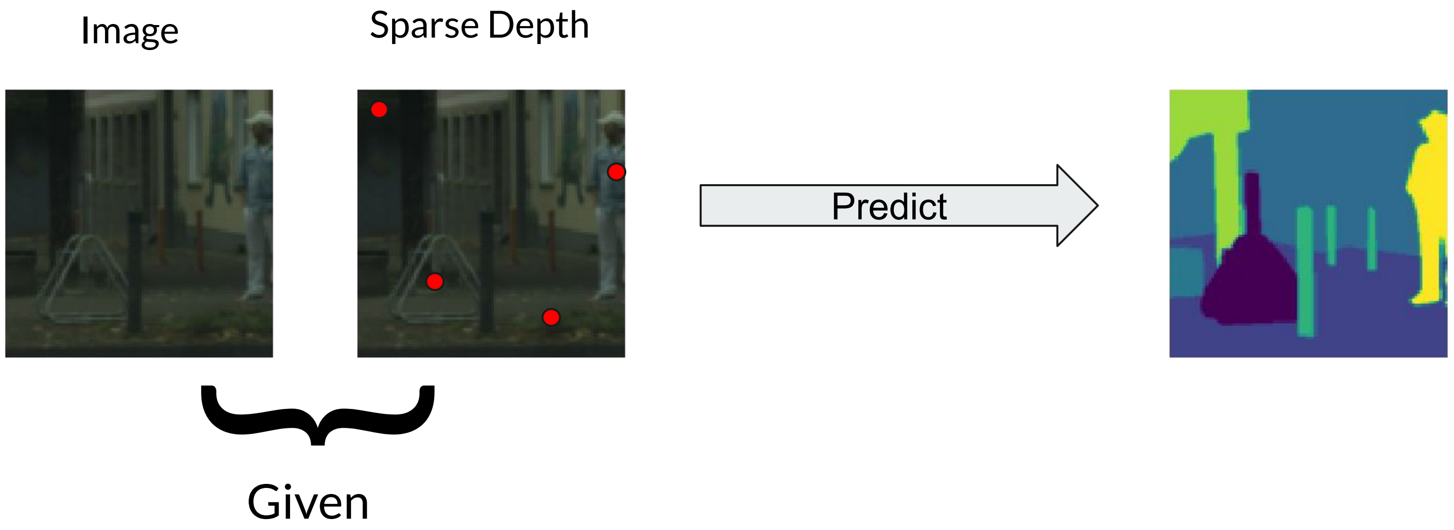
Predict



Input Image

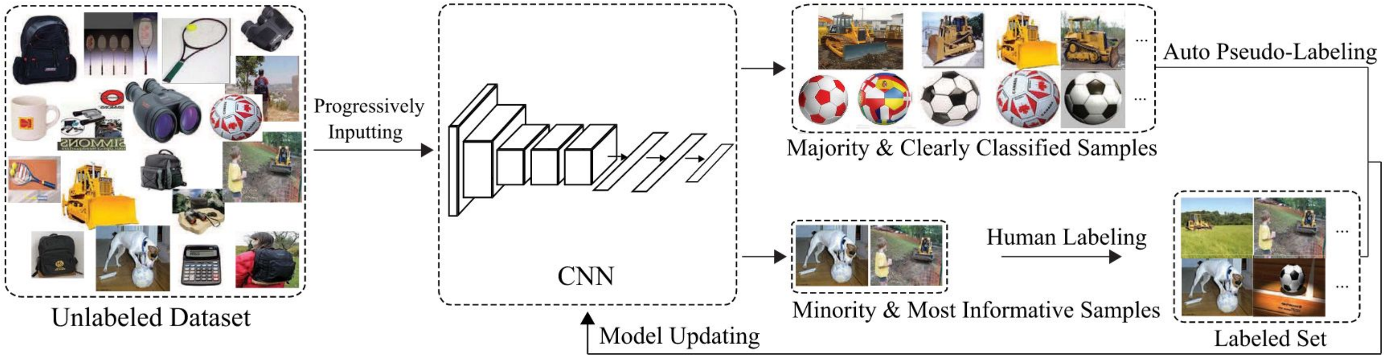
Suggested Annotation Locations

Semantic Segmentation



Active Learning and its Shortcomings

Active Learning techniques for image mostly focus on selecting a subset of *images* for annotation, not individual pixels for annotation.



Active Learning and its Shortcomings

Active Learning methods are predominantly based on selecting examples of maximum uncertainty.

Typical Acquisition Functions

$$x_H^* = \text{argmax}_x - \int_y P_\theta(y|x) \log P_\theta(y|x)$$

$$x_{LC}^* = \text{argmax}_x 1 - P_\theta(y|x)$$

Active Learning and its Shortcomings

This doesn’t capture what we really want with pixel-wise prediction. We want a small set of pixels that minimize the uncertainty in the *other* pixels.

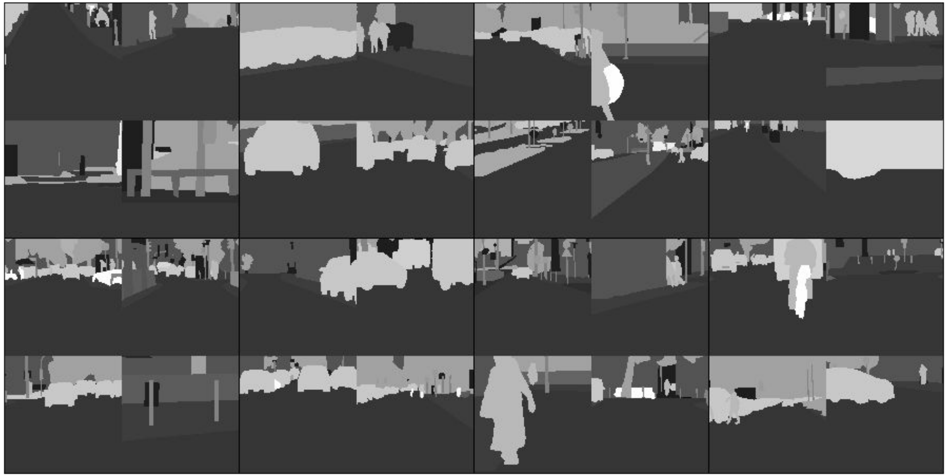
Semantic Segmentation Dataset

We will focus on the Cityscape image segmentation dataset due to its availability and its size, over 20,000 high quality images and their semantic annotation.

Cityscape Training Samples

Images

Segmentations

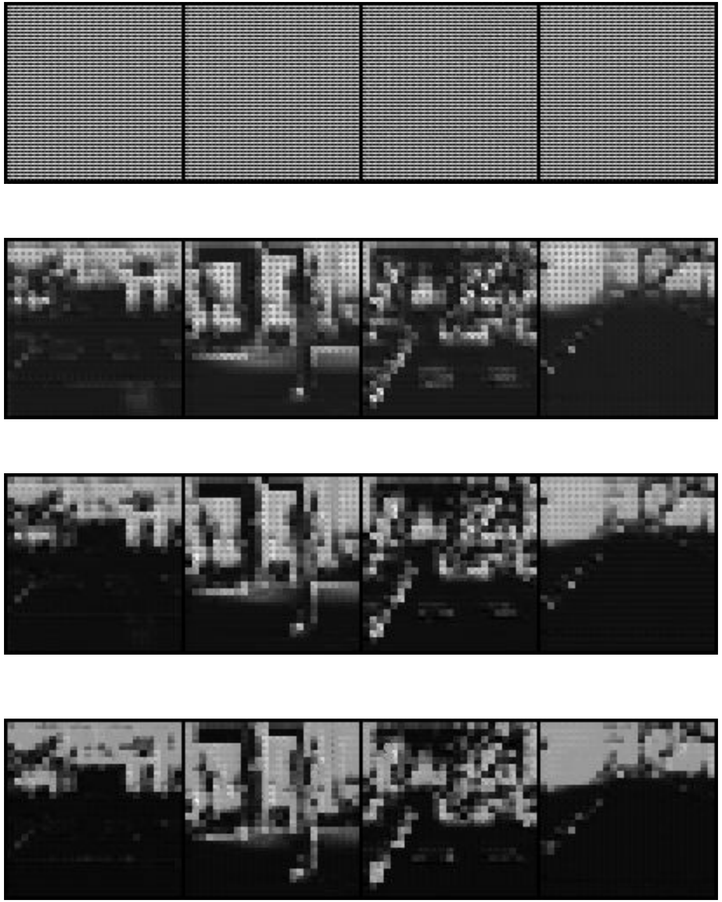


Preliminary Results

Baseline: 0% Sparse Input

Input Images

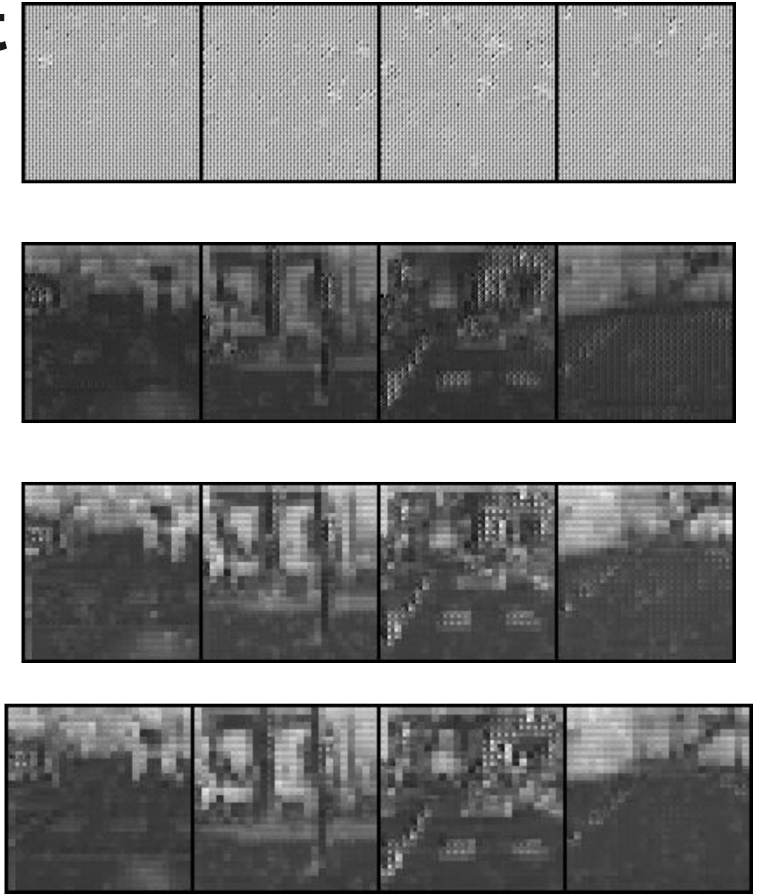
Ground Truth Segmentation



Baseline: .01% Sparse Input

Input Images

Ground Truth Segmentation



Naive Masking

$$\tilde{S} = M \odot S$$

Output from a neural net

Ground truth segmentation

$$\hat{S} = h_\theta(I, \tilde{S})$$

Baseline: Naive Masking

Input Images

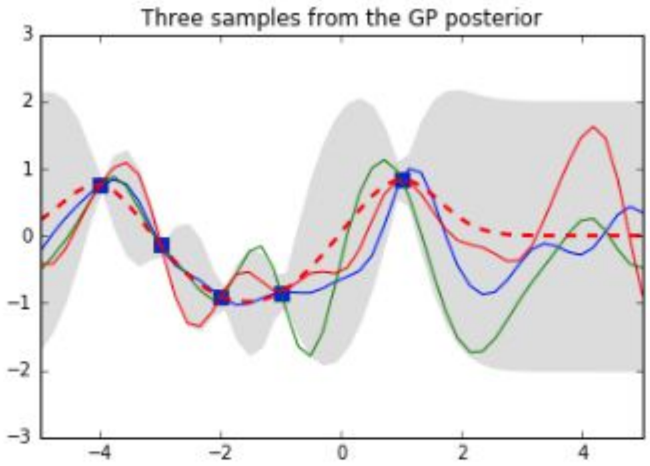
Ground Truth Segmentation

What went wrong?

One explanation is that it can be difficult for neural networks to learn multiplication-based outputs. Another may just be the dissimilarity between this active learning approaches and existing successful ones that rely on probability.

Probabilistic Models

Gaussian Processes



Given: $x_1, x_2, \ldots x_n$

Assume:

$$p(f(x_1), \ldots, f(x_n)) = \mathcal{N}(\mu(x), \Sigma(x))$$

Gaussian Processes for Active Learning

Journal of Machine Learning Research 9 (2008) 235-284 Submitted 9/06; Revised 9/07; Published 2/08

Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies

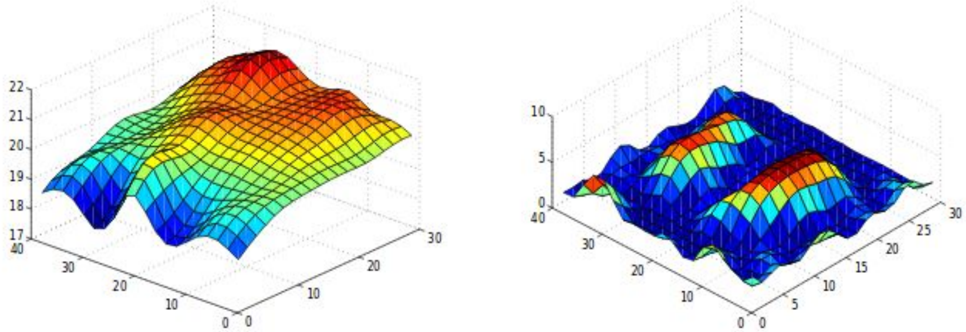
Andreas Krause KRAUSEA@CS.CMU.EDU
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

Ajit Singh AJIT@CS.CMU.EDU
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213

Carlos Guestrin GUESTRIN@CS.CMU.EDU
Computer Science Department and Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213

Editor: Chris Williams

Abstract



(a) *Temperature prediction using GP* (b) *Variance of temperature prediction*

Figure 2: Posterior mean and variance of the temperature GP estimated using all sensors: (a) Predicted temperature; (b) predicted variance.