Detecting the Unexpected via Image Resynthesis

Krzysztof Lis, Krishna Nakka, Pascal Fua, Mathieu Salzmann

ICCV 2019

Presented By: Shane Rhoads

Presentation Overview

- Problem introduction
- Related works
- Motivation
- Proposed solution
- Experimental Results

Problem introduction

Unknown object detection

- Identify unknown classes, not seen in training



Input Image



Unknown Objects

Related Works

Utilizing uncertainty in labeling

- Intuitively unknown objects are difficult to classify
- Hard to classify object will have high uncertainty
- High uncertainty indicates unknown objects



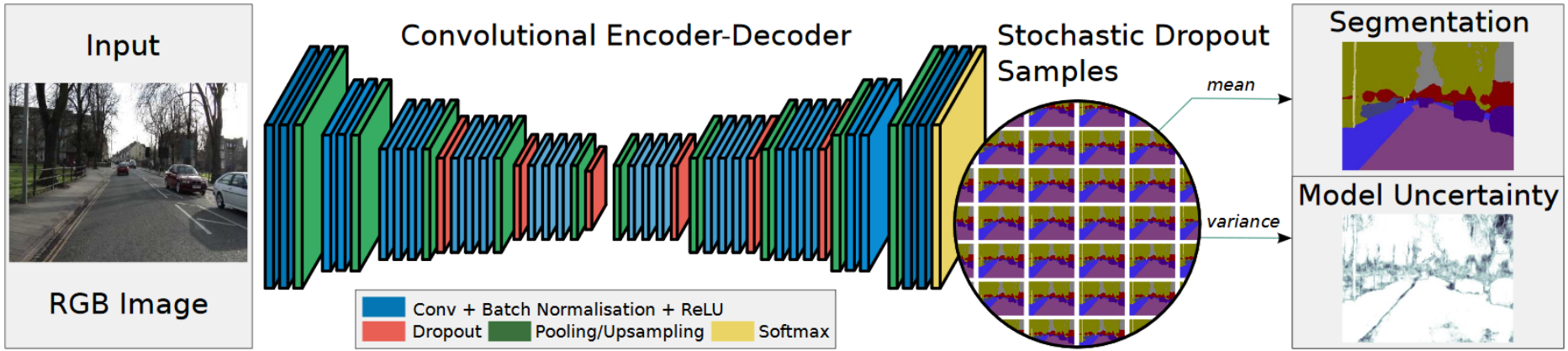


Source: Alex Kendall, Vijay Badrinarayana, and Roberto Cipolla. Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding. arXiv Preprint, 1511.02680, 2015.

Related Works

Utilizing uncertainty in labeling

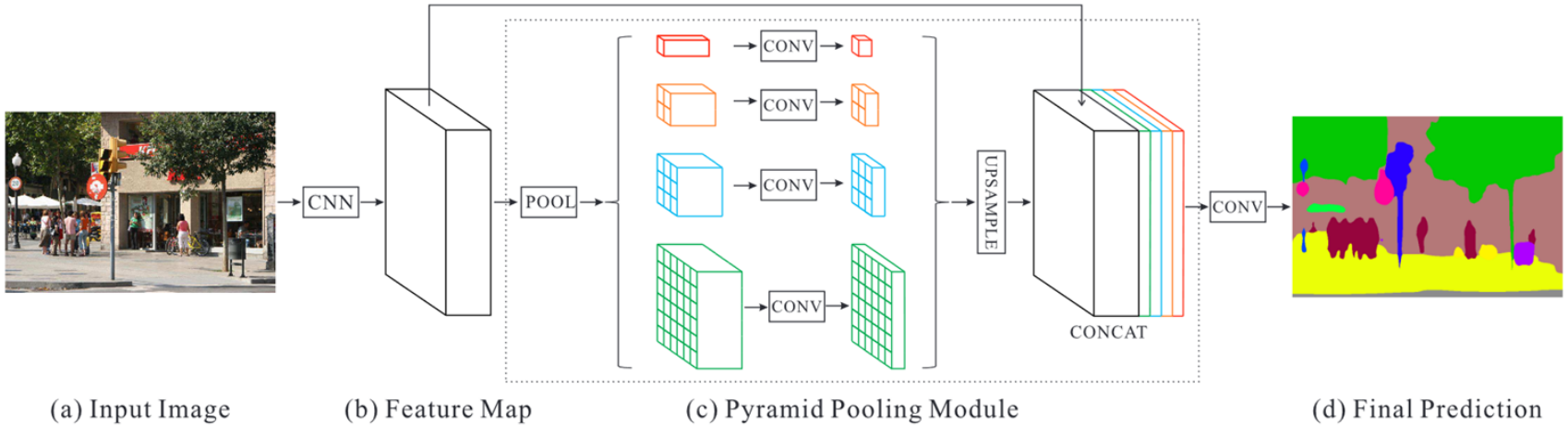
- Bayesian SegNet uses dropout layers to approximate uncertainty



Source: Alex Kendall, Vijay Badrinarayanan, and Roberto Cipolla. Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding.

Related Works

- PSP Net fuses the final layer feature map using pooling
- Multiple networks are used to approximate uncertainty



Source: Hengshaung Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid Scene Parsing Network. In Conference on Computer Vision and Pattern Recognition, 2017.

Related Works

Anomaly detection via Resynthesis

- Train an autoencoder to regenerate images
- Images with unknown objects will not regenerate correctly
- Segment regions of high error as unknown objects



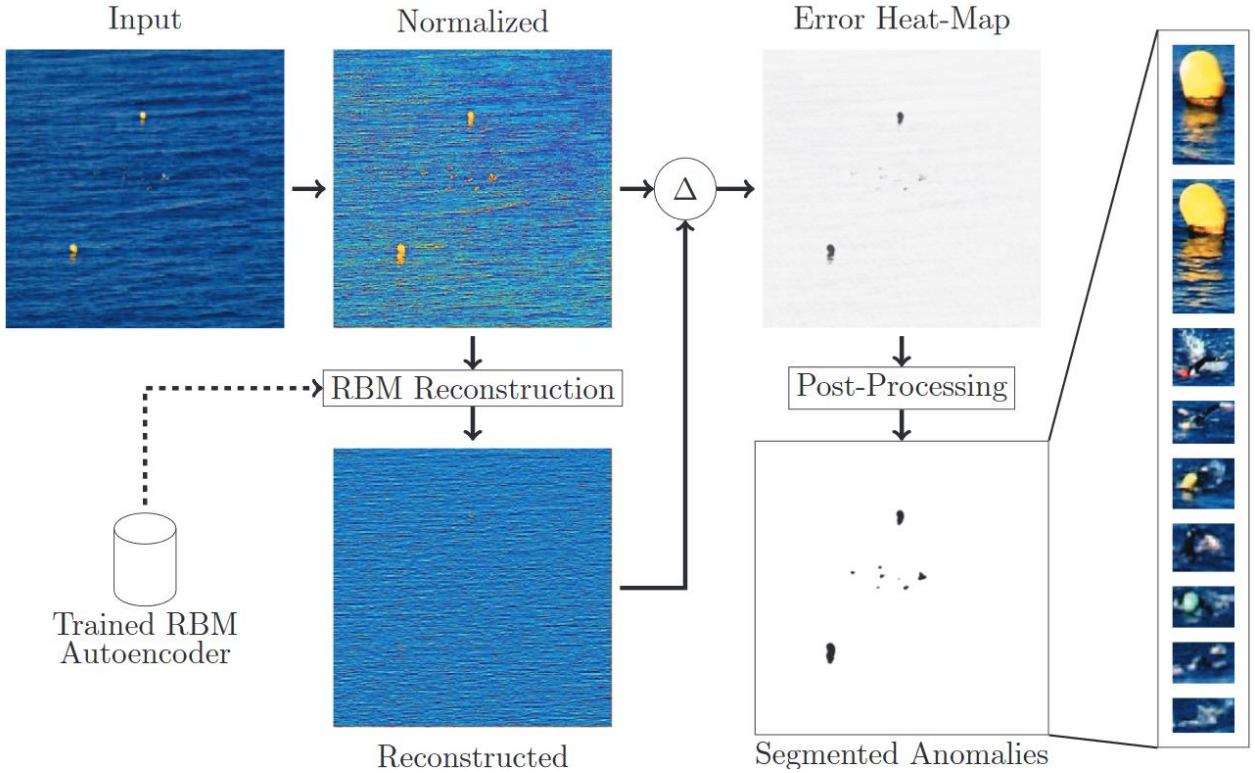


Source: Clement Creusot and Asim Munawar. Real-Time Small Obstacle Detection on Highways Using Compressive RBM Road Reconstruction.

Related Works

Anomaly detection via Resynthesis

- RBM reconstructs the texture of the road and looks for significant differences



Source: Clement Creusot and Asim Munawar. Real-Time Small Obstacle Detection on Highways Using Compressive RBM Road Reconstruction.

Motivation

- Uncertainty has high false positive rate and false negative rate

Ensemble
(PSP Net)

Dropout
(Bayesian SegNet)



Input Image

Uncertainty Map

Motivation

- Autoencoders tends to learn to simply compress the image
- Autoencoders do not generalize well



Input Image



Anomaly Map

Proposed Solution – Key idea

- Unknown objects will be misclassified as another class



Input Image



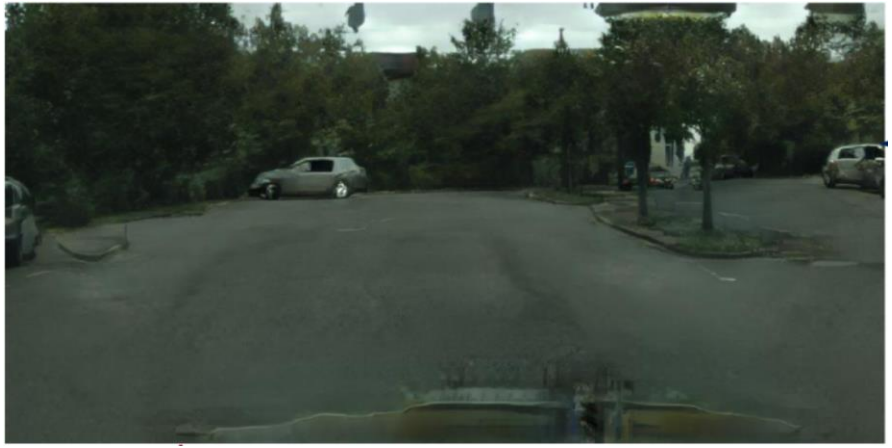
Semantic Map

Proposed Solution – Key idea

- Resynthesized images from a Semantic map with incorrect labels will have different features inside those labels



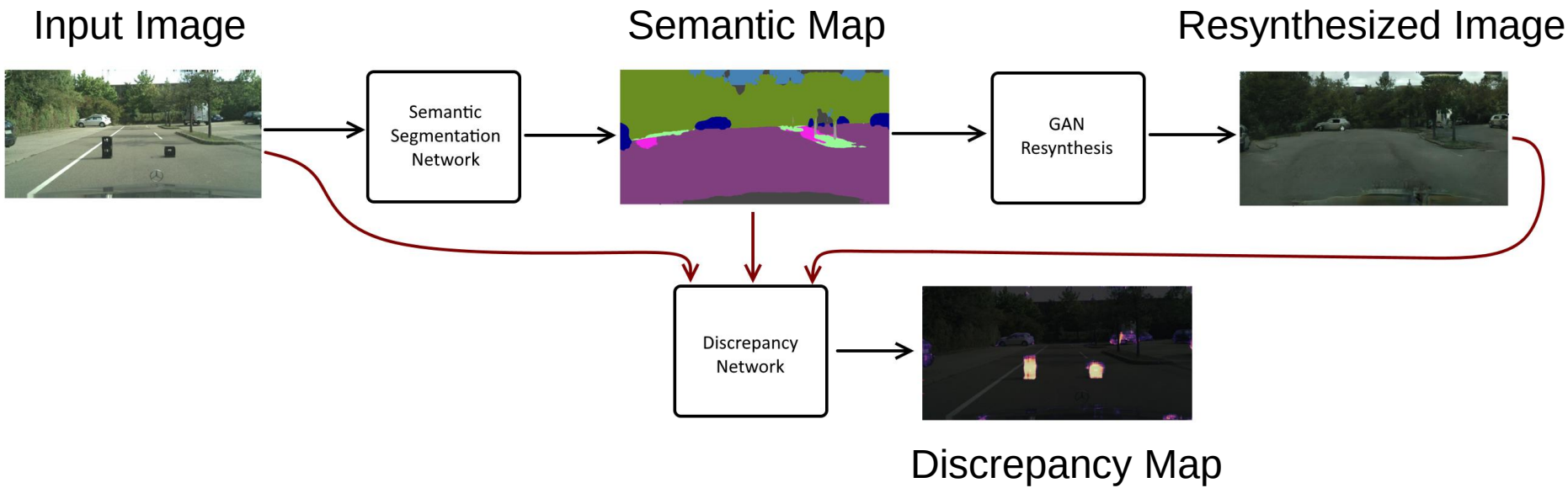
Semantic Map



Resynthesized Image

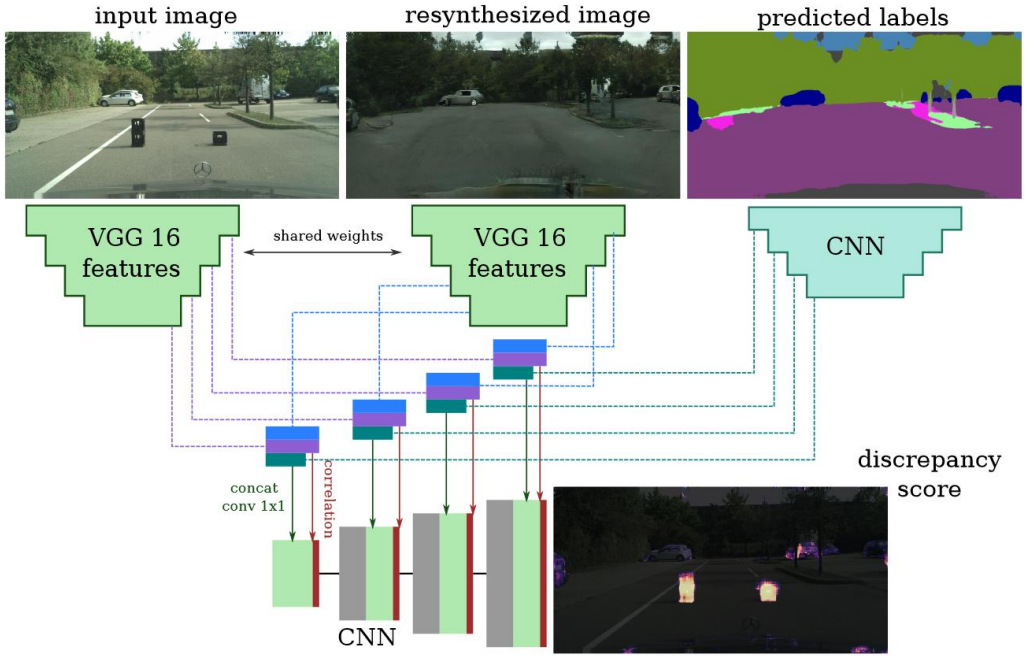
Proposed Solution – Network Overview

- Discrepancy network is the author's contribution

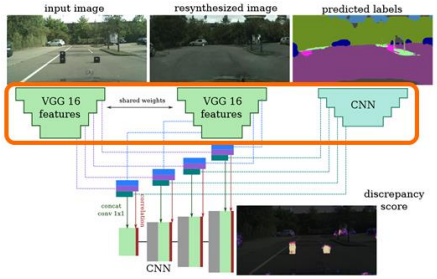


Proposed Solution – Network Architecture

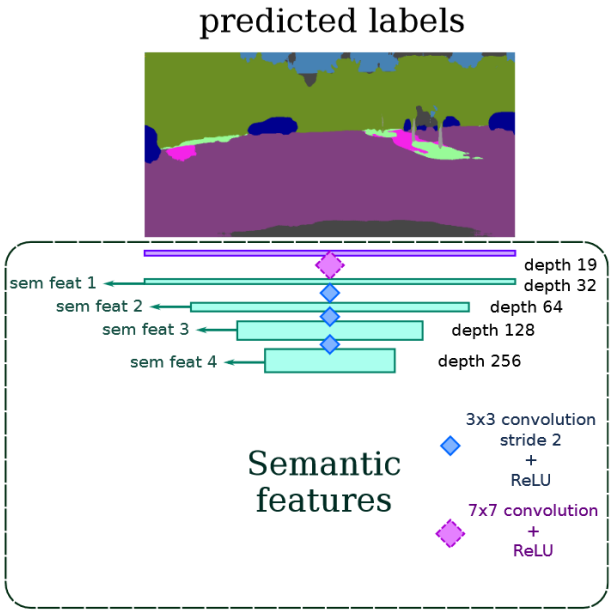
- Discrepancy network is a decoder network
- Output is a discrepancy map based on *feature* differences

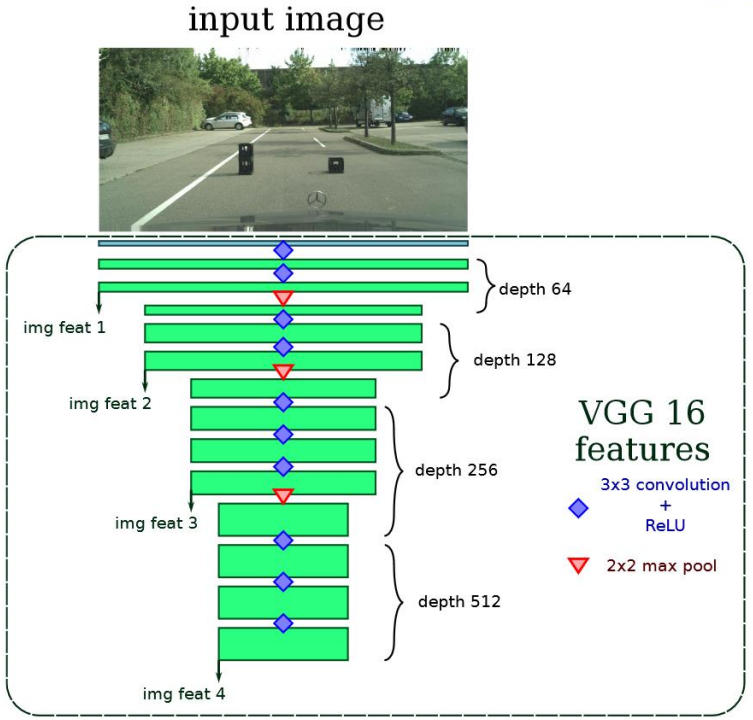


Proposed Solution – Feature Extraction

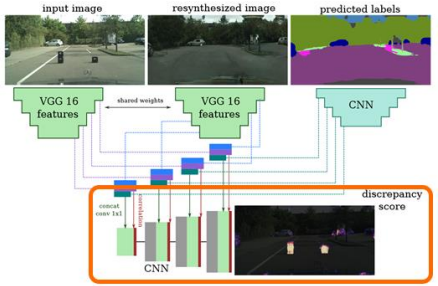


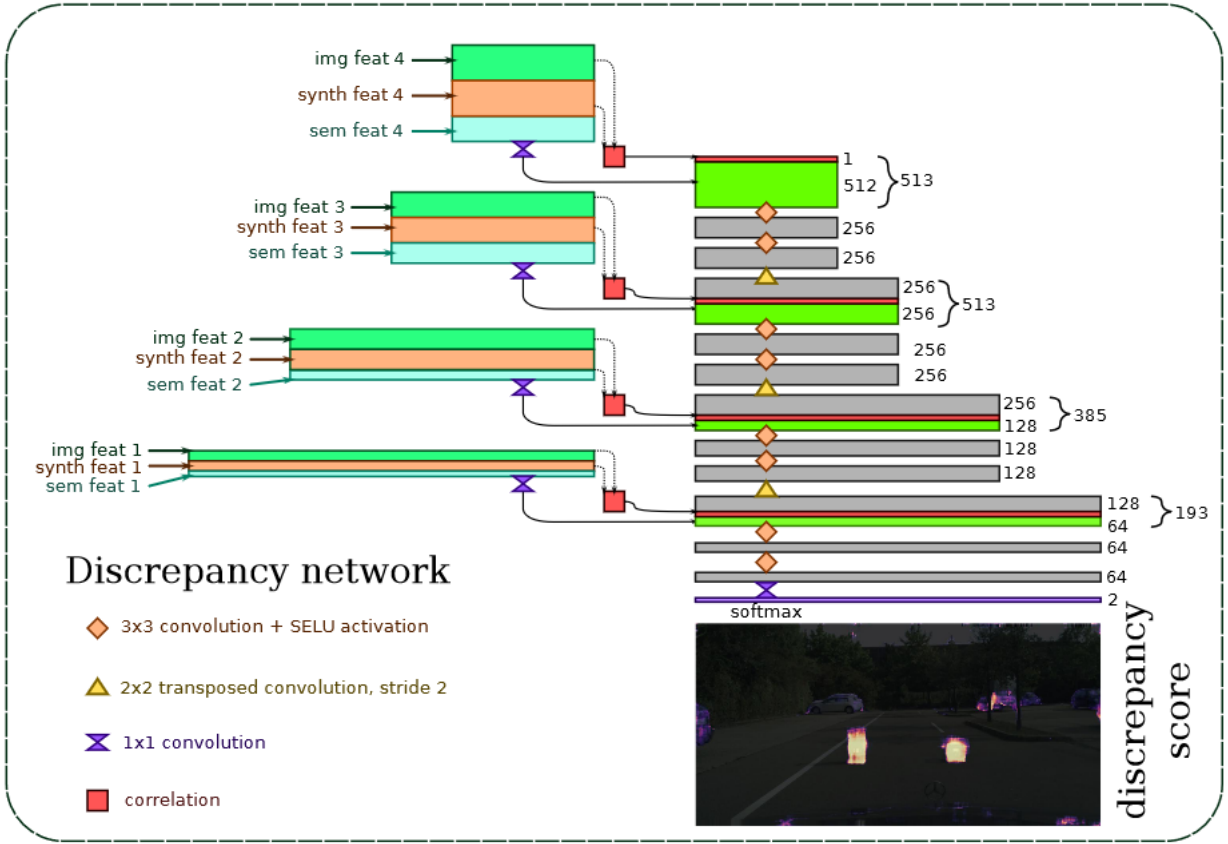
- Same VGG-16 network used for both images





Proposed Solution – Discrepancy Network

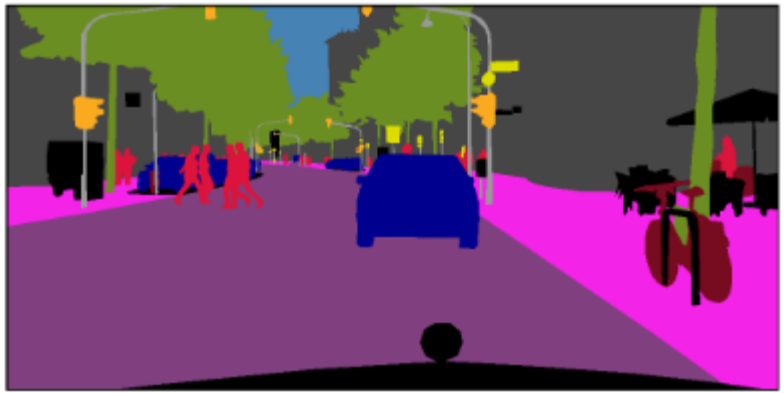




- Image and semantic features are combined
- Feature correlations between the images are also added

Proposed Solution – Discrepancy Network Training

- Model can be trained without unknown classes in data
- Randomly change some classes before resynthesizing image



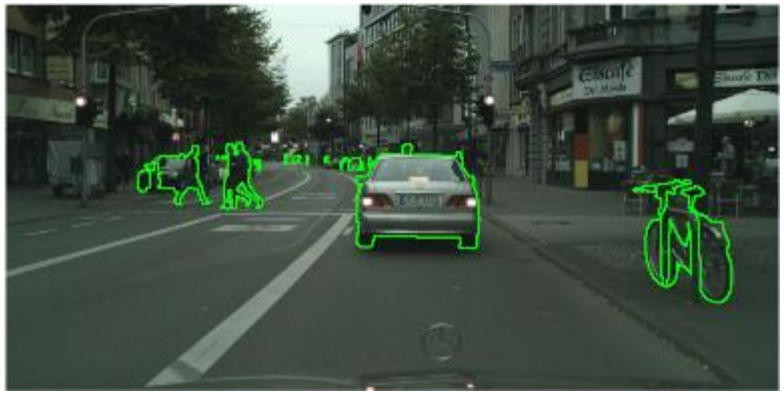
Original Semantic Map



Altered Semantic Map

Proposed Solution – Discrepancy Network Training

- Discrepancy network is trained to mark pixels within modified labels



Original Image Showing Altered Objects



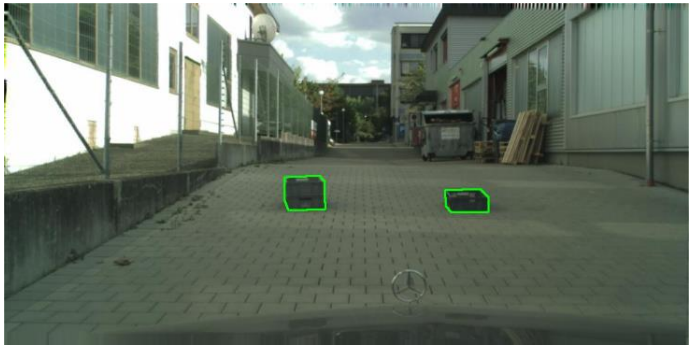
Resynthesized Image

Experimental Results

Lost and Found Dataset

- Dataset containing small items on the street
- 13 environments, 37 objects, 2104 annotated frames
- Provided as sequences of stereo videos
- 2048x1024 images downscaled and cropped to 1024x512





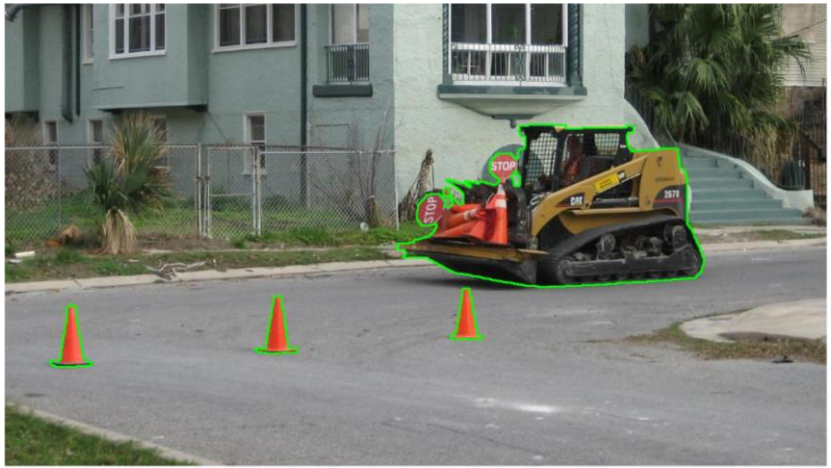
Source: Peter Pinggera, Sebastian Ramos, Stefan Gehrig, Uwe Franke, Carsten Rother, and Rudolf Mester. Lost and Found: Detecting Small Road Hazards for Self-Driving Vehicles

Experimental Results

Road Anomaly Dataset

- New dataset created from online images
- Contains unusal objects on or near roads
- 60 annotated images of size 1280x720



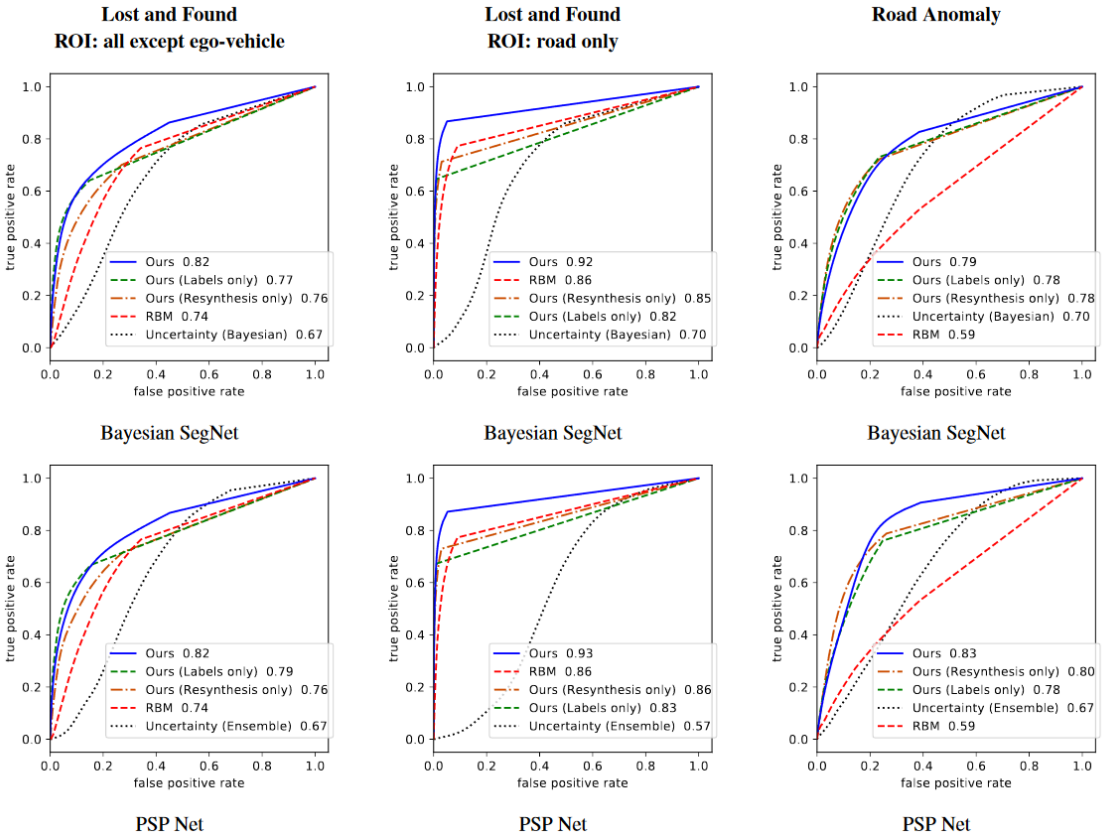


Experimental Results – Quantitative Metric

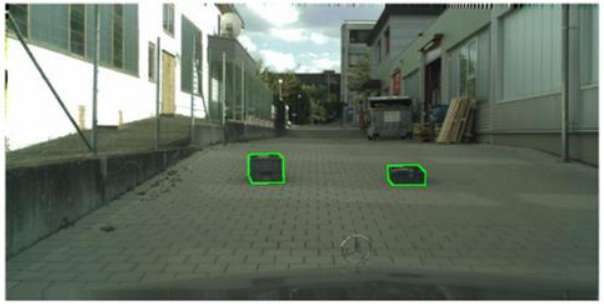
Receiver Operating Characteristic (ROC) curve:

- Plot of true positive rate vs false positive rate
- Visualization of the accuracy of the system

Experimental Results – Quantitative Results



Experimental Results – Lost and Found Examples



Input image with anomalies highlighted



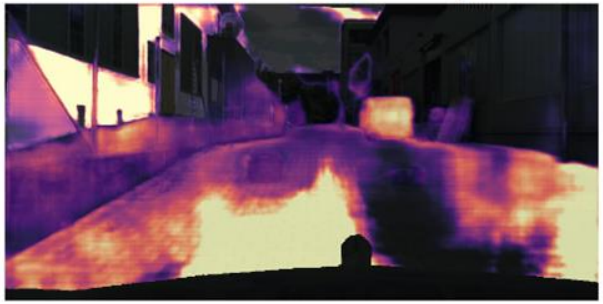
Predicted semantic map



Resynthesized image



Anomaly score - *Ours*



Anomaly score - *Uncertainty (Dropout)*

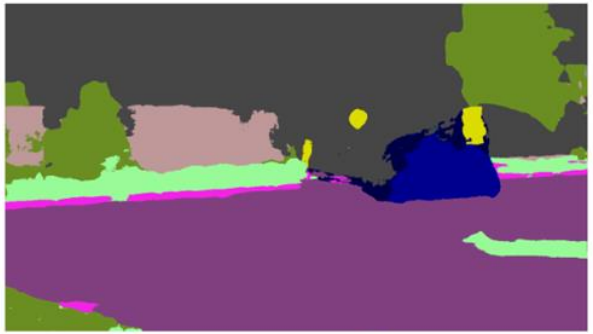


Anomaly score - *RBM*

Experimental Results – Road Anomalies Examples



Input image with anomalies highlighted



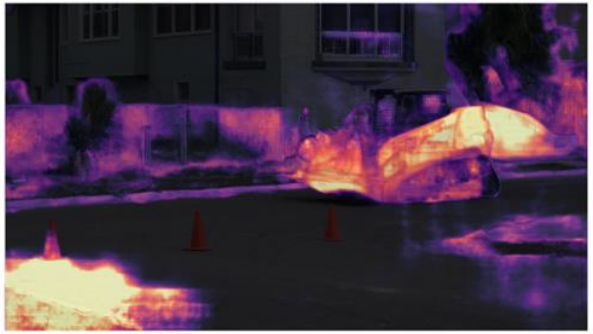
Predicted semantic map



Resynthesized image



Anomaly score - *Ours*



Anomaly score - *Uncertainty (Dropout)*



Anomaly score - *RBM*

Experimental Results

Adversarial Attack Algorithms

- Houdini
  - Attack based on maximizing task loss
  - Uses a margin factor to limit changes from the original image

- Dense Aversary Generation
  - Attack based on finding the perturbation that results in the greatest uncertainty among classes
  - Uses an iterative gradient-based method

Experimental Results

Target Semantic Maps for Adversarial Attacks

- Shift: Each predicted label is offset by a constant factor
- Pure: Every pixel is a single, randomly selected class



Original Semantic Map



Shifted Semantic Map



Pure Semantic Map

Experimental Results

Cityscapes Dataset

- 5000 pixel-level annotations of 2048x1024 images
- 30 classes within 8 categories in 50 different cities
- Each annotation is the 20th frame of a 30 frame video

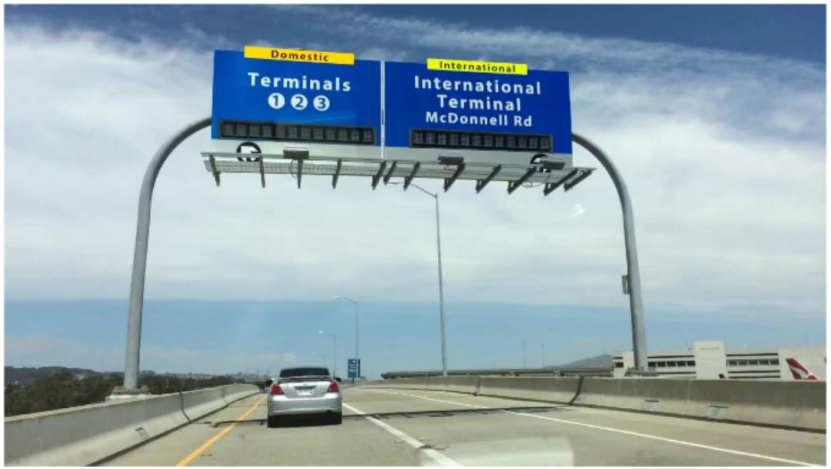


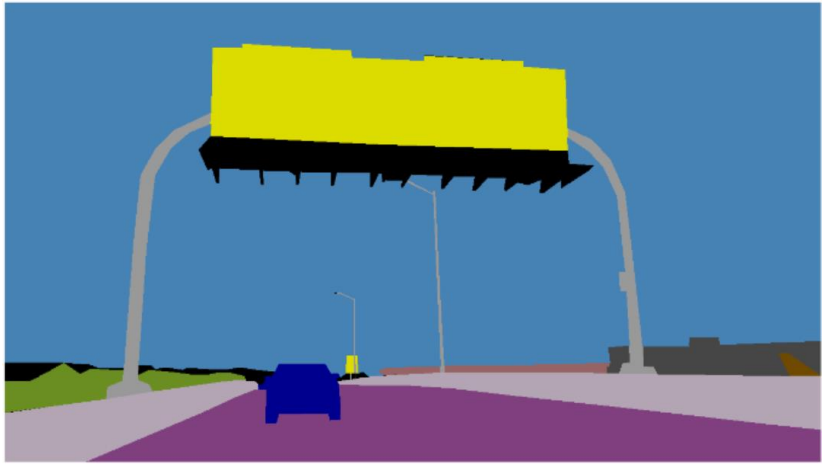


Experimental Results

BDD100K

- 5683 annotated 1280x720 images with 40 classes
- Supports multiple annotation styles
- Each image is a random frame of one of the 5683 videos





Experimental Results

Dataset	Model	Method	Detection			
			DAG		Houdini	
			Pure	Shift	Pure	Shift
Cityscapes	BSeg	SC	99%	98%	**100%**	**98%**
		Ours	**100%**	**100%**	**100%**	**98%**
	PSP	SC	98%	90%	98%	**100%**
		Ours	**100%**	**99%**	**99%**	**100%**
BDD	BSeg	SC	**100%**	**100%**	98%	**100%**
		Ours	98%	98%	**100%**	90%
	PSP	SC	92%	**100%**	96%	**100%**
		Ours	**100%**	96%	**98%**	95%

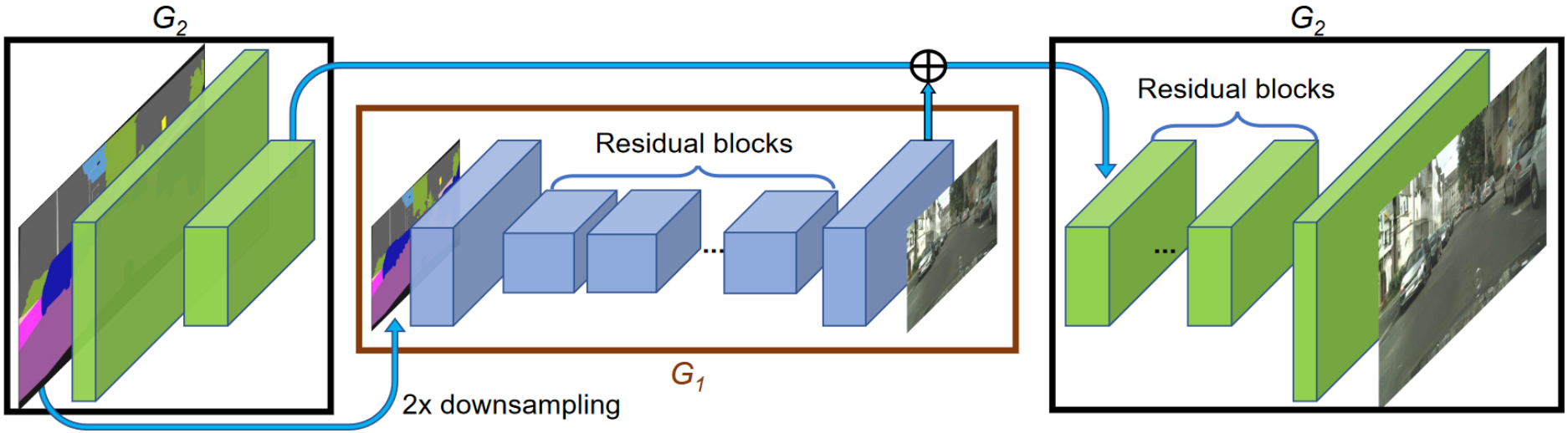
Conclusion

Resynthesized images from a semantic map with incorrect labels will contain significant differences with the original image. These differences can be detected by a decoder network and used to identify unknown objects or adversarial attacks.

Questions?

Related Works

- pix2pixHD is a conditional GAN using a feature-based loss



Source: Hengshaung Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid Scene Parsing Network. In Conference on Computer Vision and Pattern Recognition, 2017.