Generating Top-down View from 6 Stereo Images

Muhammad Osama Khan
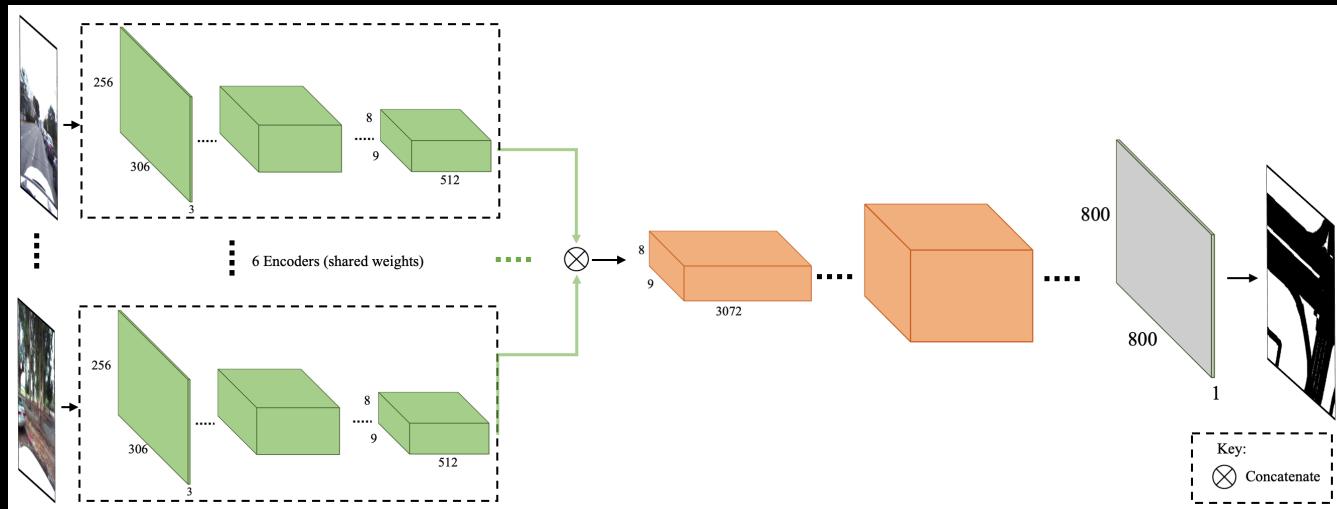
Muhammad Muneeb Afzal

Divya Juneja

Road Segmentation – Architectures

- 6 Encoders (shared weights) + Decoder

- Combine feature maps → concatenate, mean, attention?
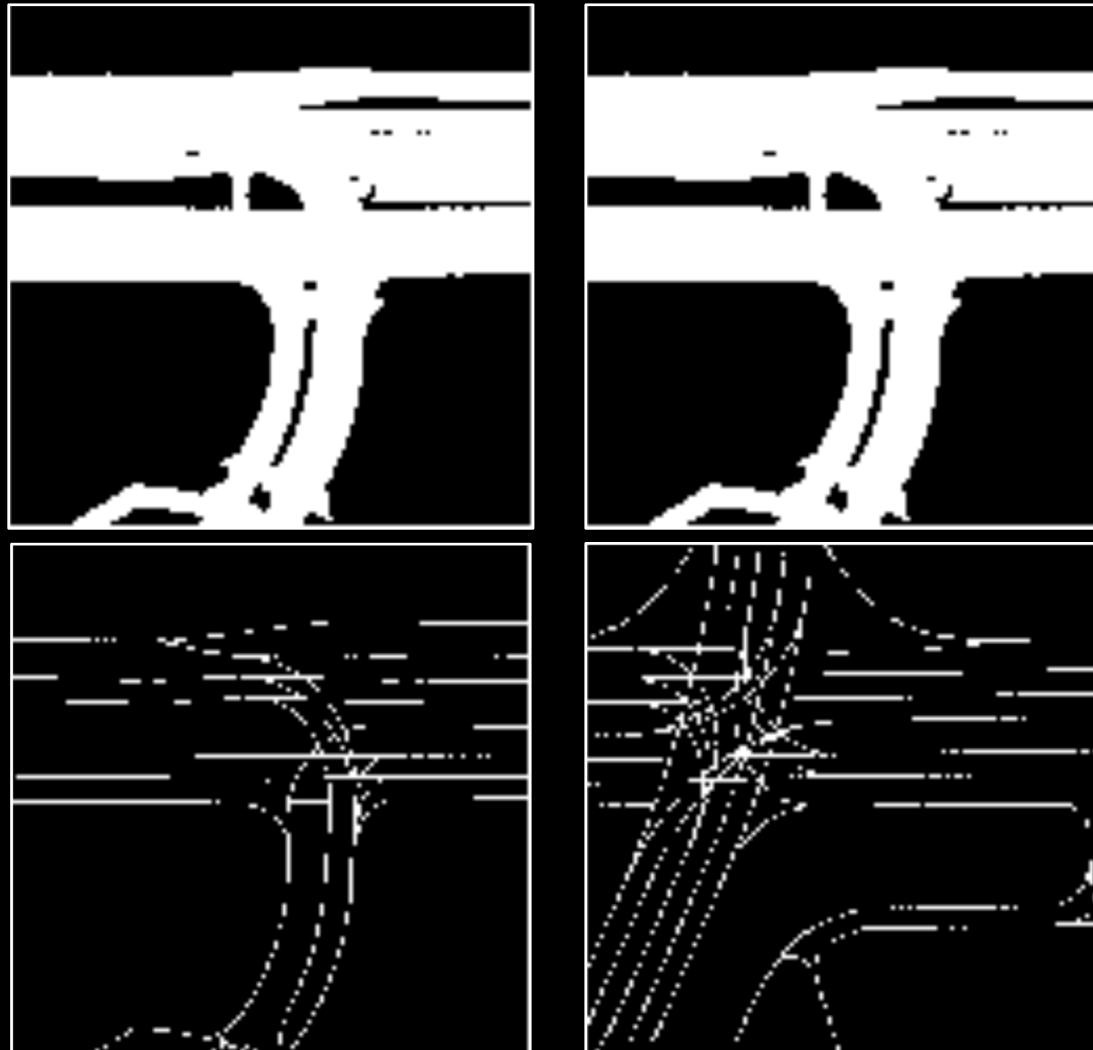
Upsampling + Convolutions	Transposed Convolutions
0.733	0.741

Using Extra Info (Lane Masks)

USING LANE MASKS

No Pretrain	Lane + Road Segmentation	Lane → Road Correspondence
0.741	0.743	0.746

positive

negative

Self-Supervised Learning

PRETEXT TASKS

No Pretrain	Jigsaw (700)*	Stereo (700)*	Jigsaw (1000)*
0.741	0.750	0.753	0.762
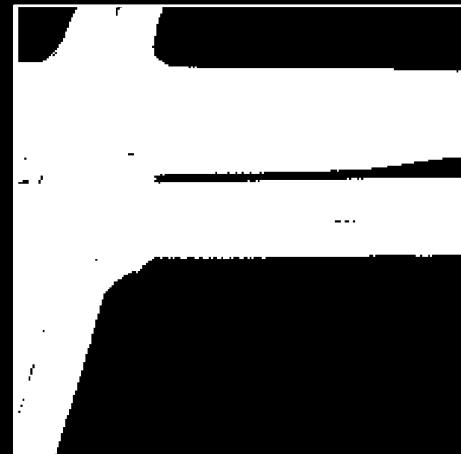
- Stereo produces better results for equal number of permutations

* represents number of permutations

Visualizations
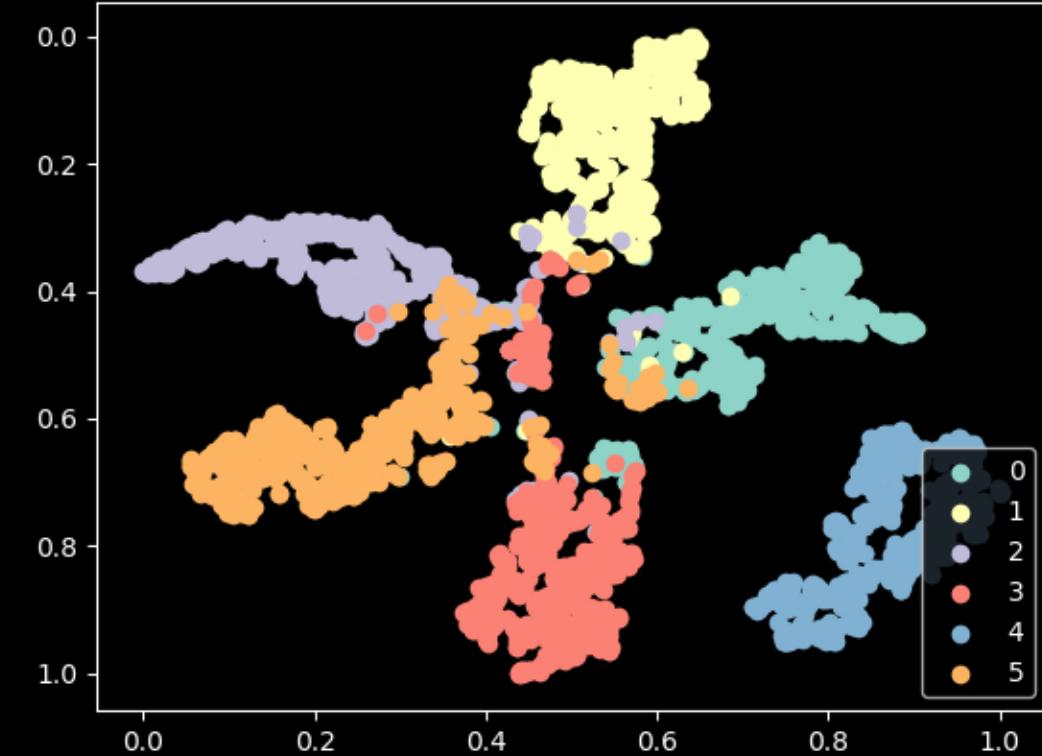
Predictions

Ground Truth

Object Detection

Modified YOLOv3

YOLOv3: Double Targets
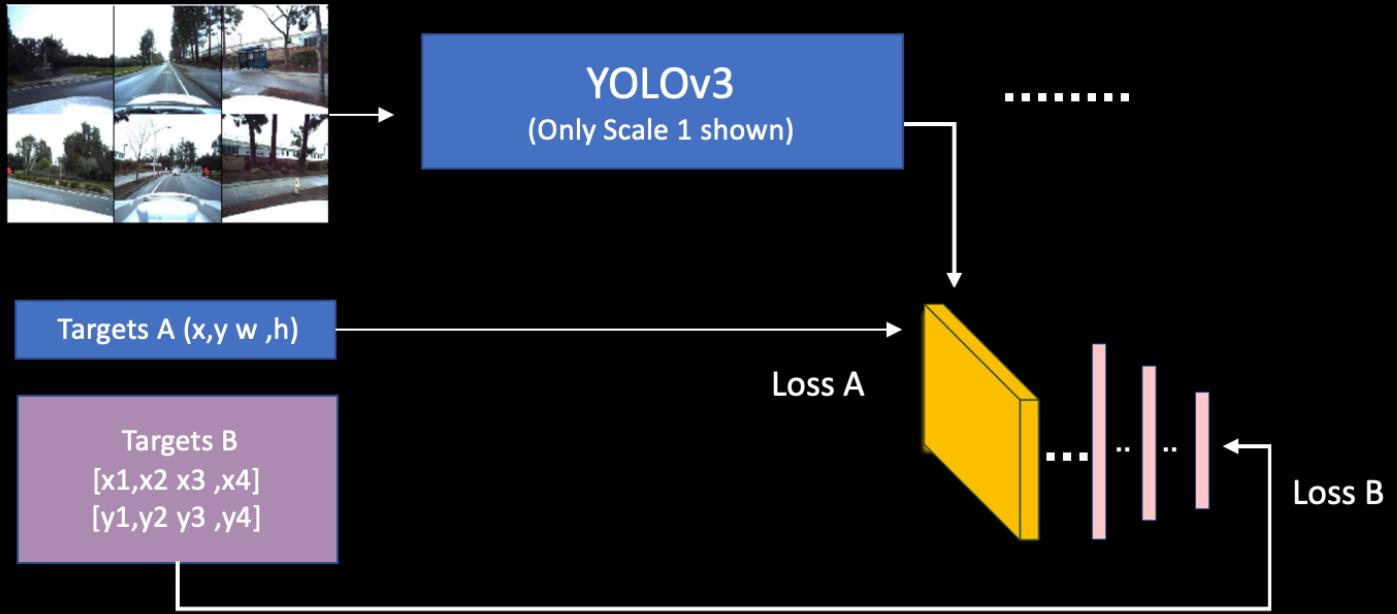
INPUT TYPES

Concatenated	Tiled
0.000143	0.000498

Total Loss = Loss A + Loss B

ARCHITECTURE

Normal	Double Targets
0.000298	0.000498

Objection Detection: Faster R-CNN

LOSS FUNCTION WEIGHTS

Coordinates (Regression)	Objectness	Classifier	RPN Regression	Score
1.0	1.0	1.0	1.0	0.0096
5.0	1.0	1.0	1.0	0.0143