Adaptive Fusion using Convoluted Mixture of Deep Experts for Robust Semantic Segmentation

Ankit Dhall

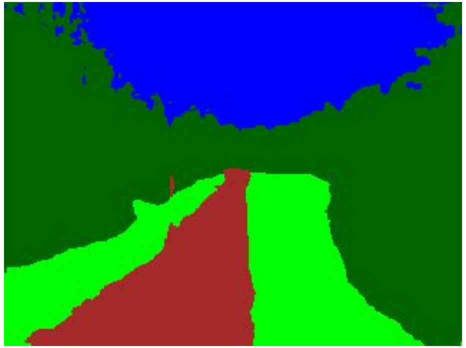
Supervisors: Abhinav Valada and Wolfram Burgard

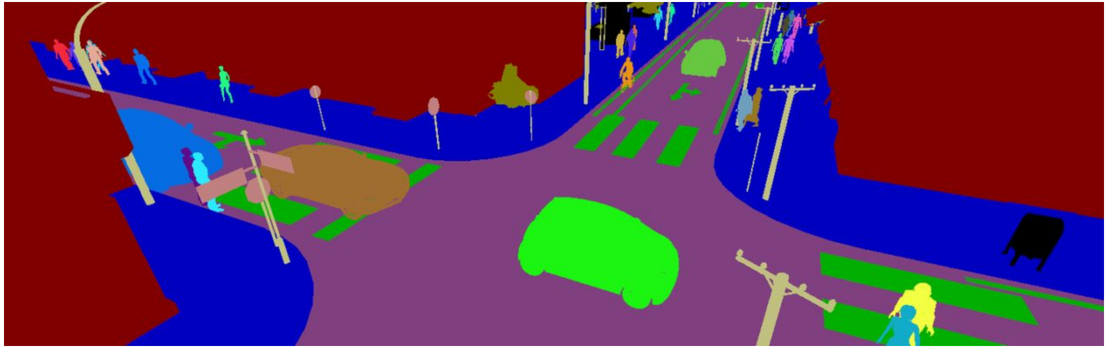
UNI FREIBURG

AIS Autonomous Intelligent Systems

Scene Understanding and Segmentation

- Understand dynamic unstructured environments
- Maps for off-roads change frequently
- Changing foliage and seasons, conditions
- Build robotics applications like navigation on top





Source: Synthia dataset

Motivation for Fusion with Probabilities

- Prone to low-lighting, snow, glare and motion blur
- Changing conditions over time
- Overcome modality weaknesses

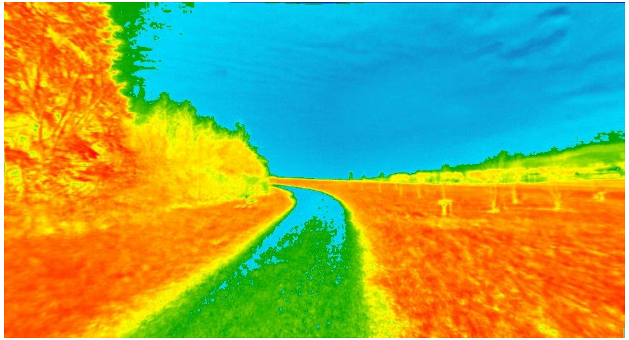
Motivation for Fusion with Probabilities

Source: KITTI dataset

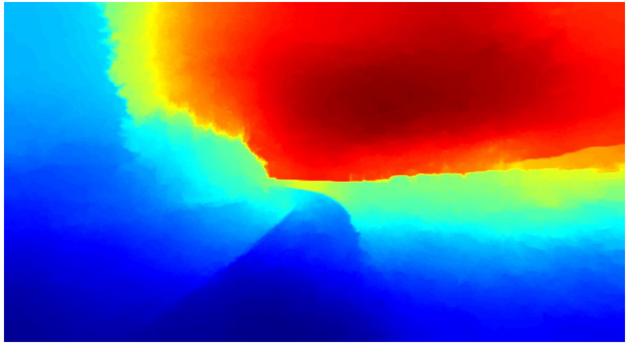
- Prone to low-lighting, snow, glare and motion blur
- Changing conditions over time
- Overcome modality weaknesses

Freiburg Multi-spectral Forest Dataset (ISER'16)

RGB



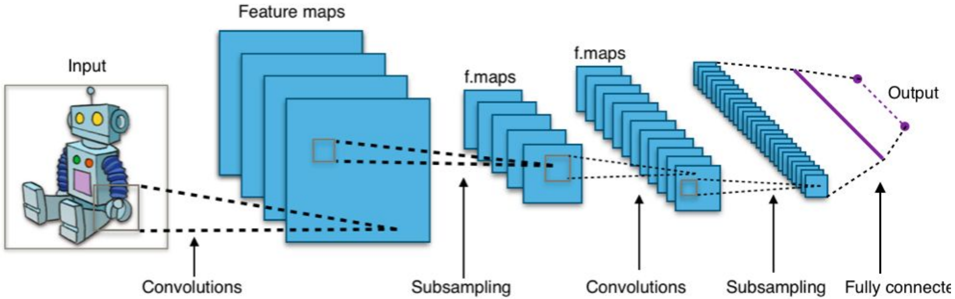
EVI (Enhanced vegetation index)



Depth

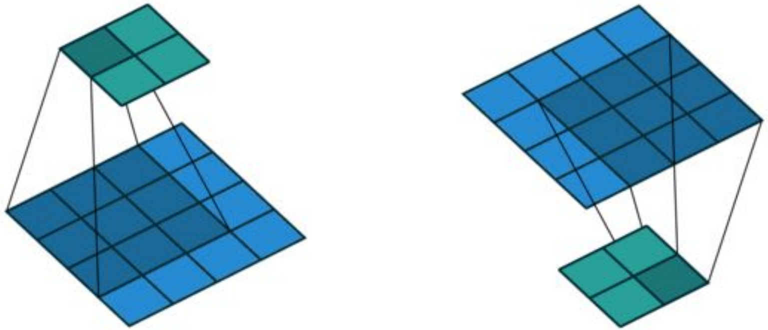
Segmentation output

What is a Convolutional Neural Networks (CNN)?

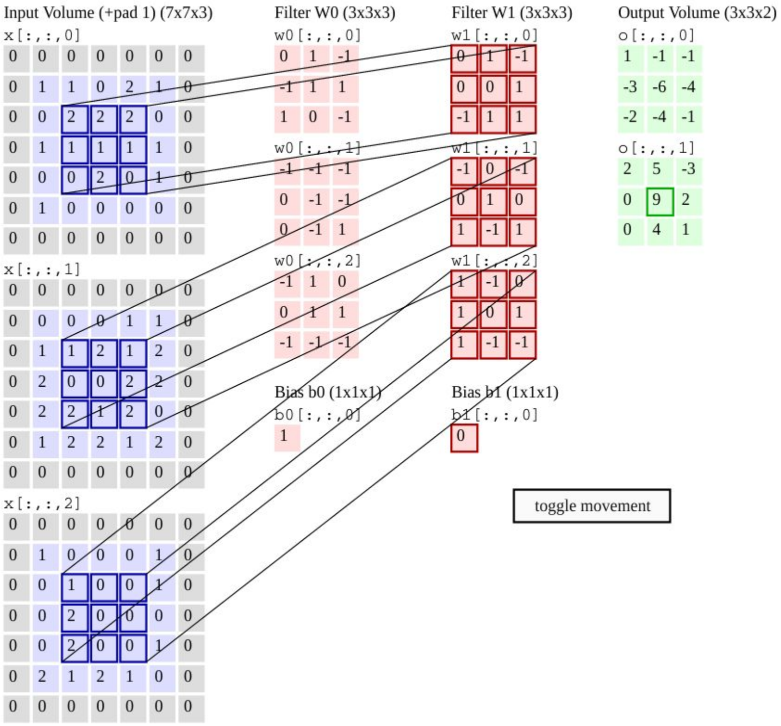


Source: Wikipedia

Fig. Typical CNN for classification

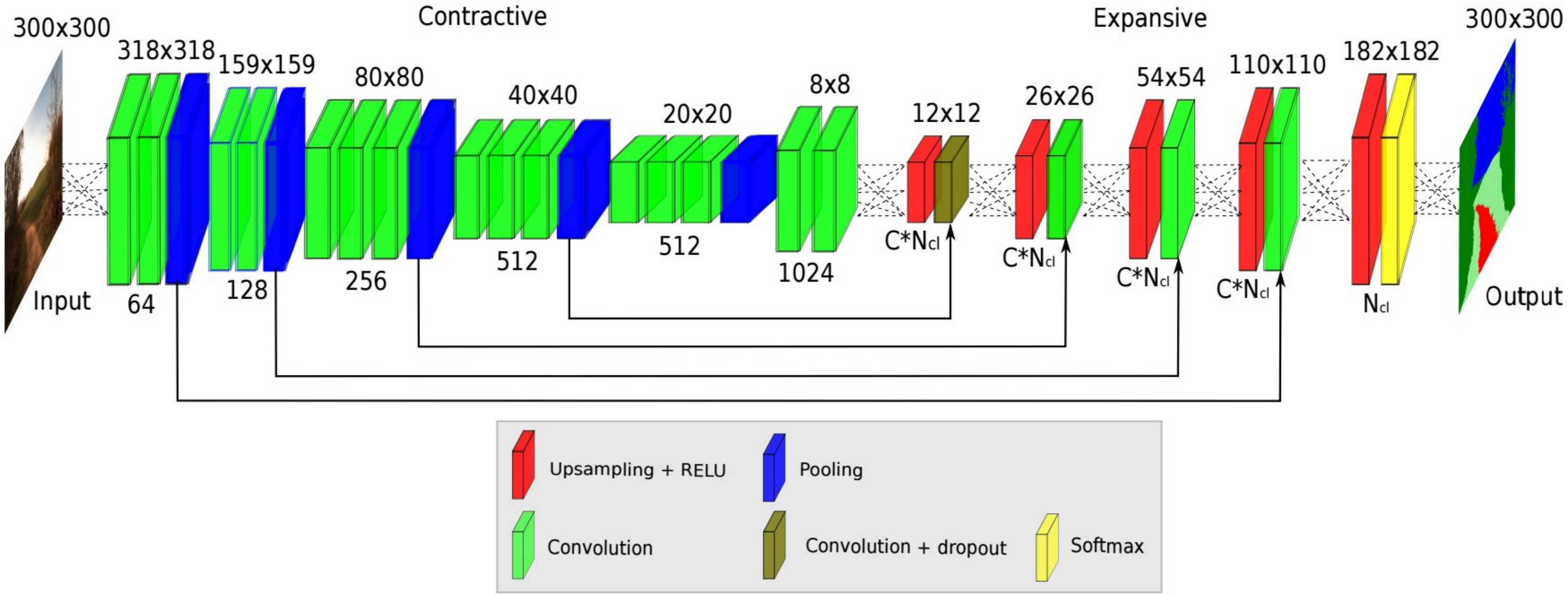


Source:github.com/vdumoulin/conv_arithmetic



Source: karpathy.io

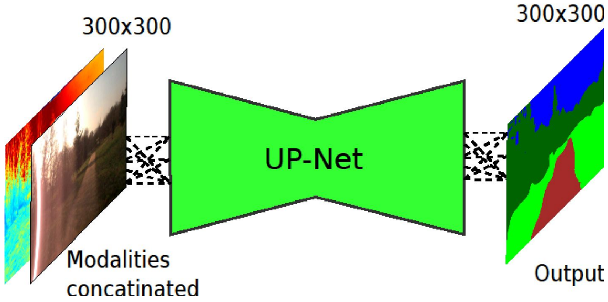
Network Architecture: FCN Experts



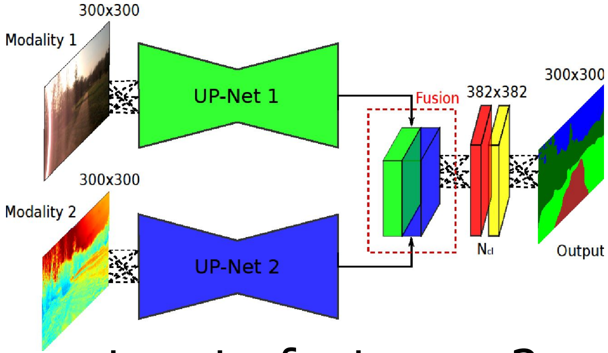
Each expert is trained exclusively on a particular modality and in parallel with other experts using the UpNet (Oliveira et al.) FCN (fully-convolutional net) architecture

Previous Approaches: Learning to Fuse

- Most convenient: concatenate channels, single net
  - Problem: vanishing gradients
- Concatenate or sum individual network features
  - Problem: learning on outputs of weak modalities

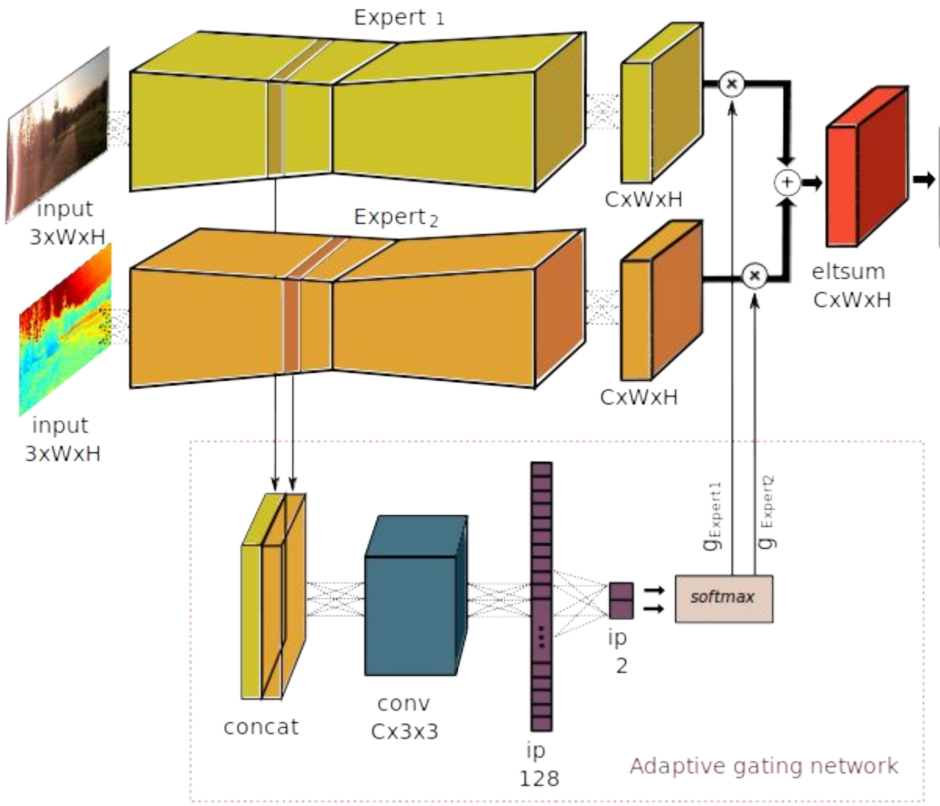


Concatenate channels, single net



Concatenate features, 2 nets

Learning Distributions before Fusion



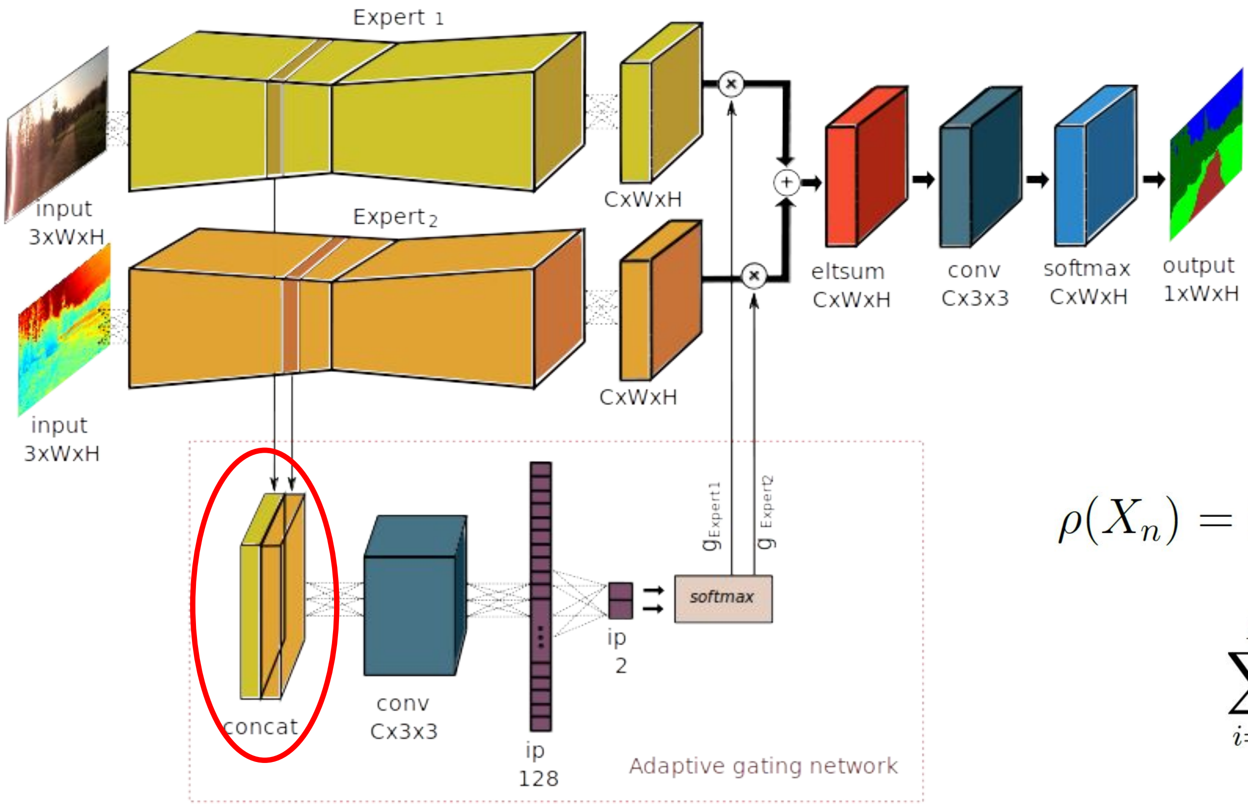
- 2-part training
- Experts shown exclusive data
- Extend for arbitrary experts

$$S = \{(X_n, y_n), n = 1, 2, \ldots, N\}$$

$$(x_1, x_2, \ldots, x_E)$$

$$y_n(r, c) \in \{0, 1, \ldots, C\}$$

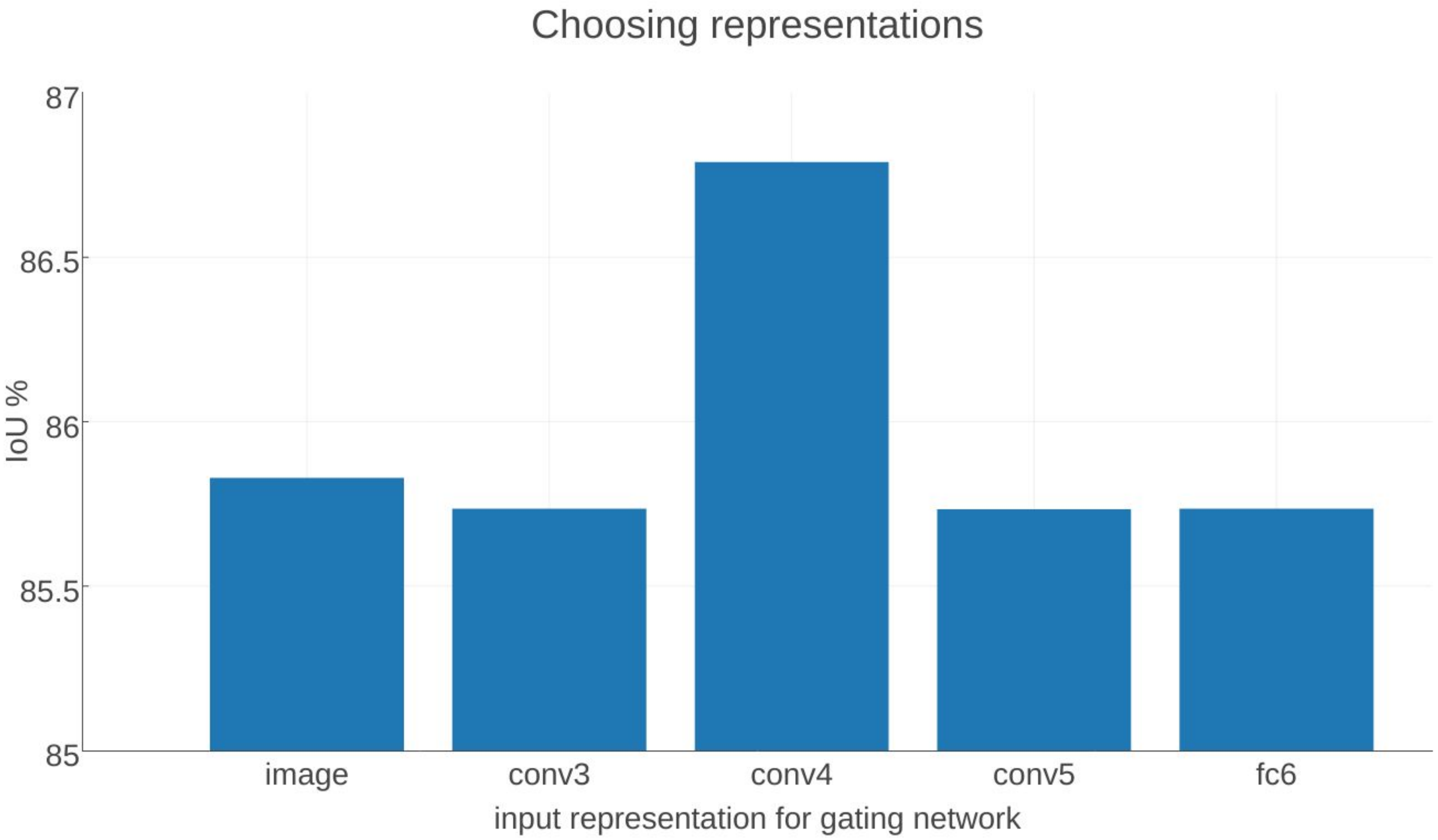
Input Representations



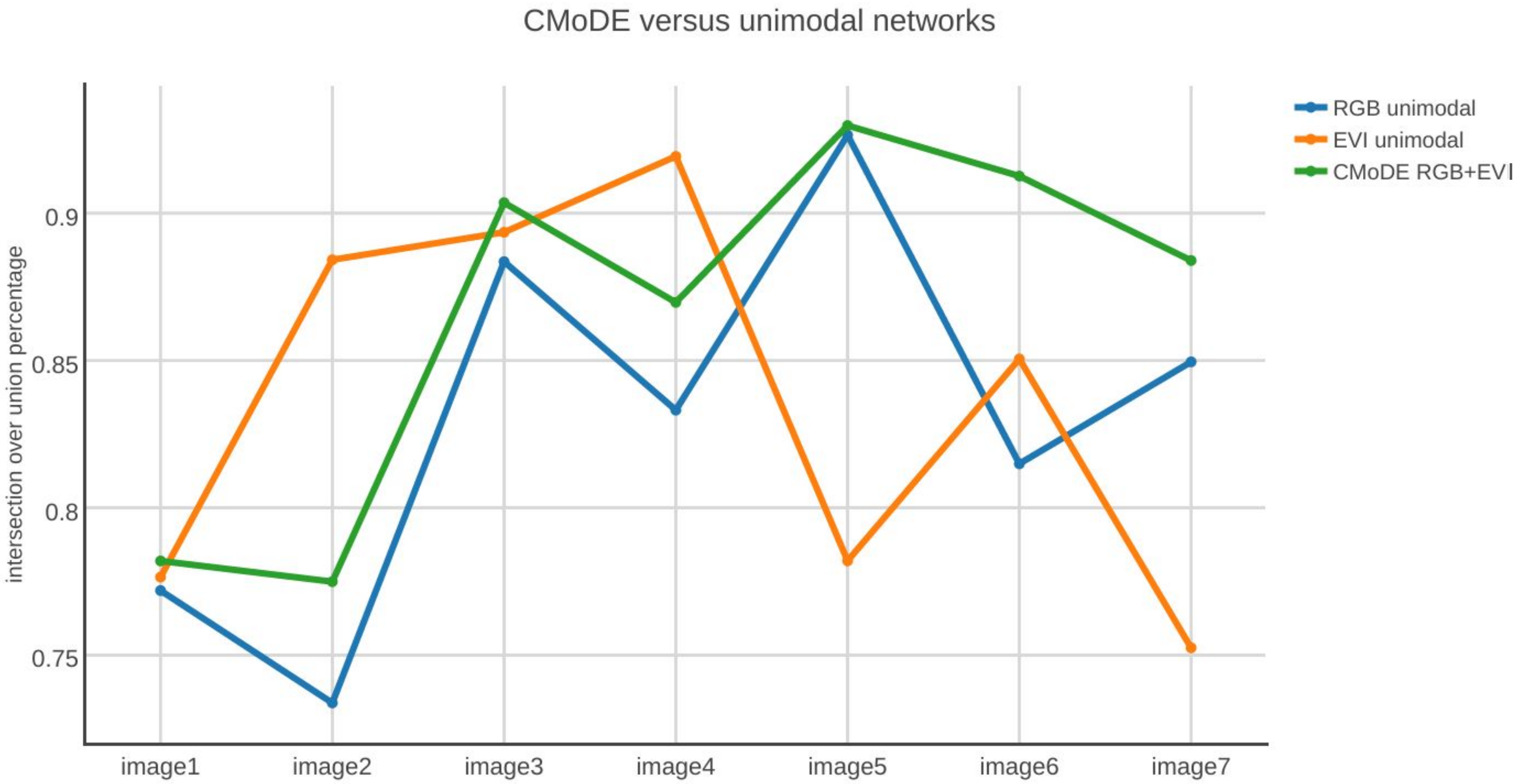
$$\rho(X_n) = (r(x_1), r(x_2), \ldots, r(x_E))$$

$$\sum_{i=1}^{E} g_i(\rho(X_n)) \cdot h_i(x_i)$$

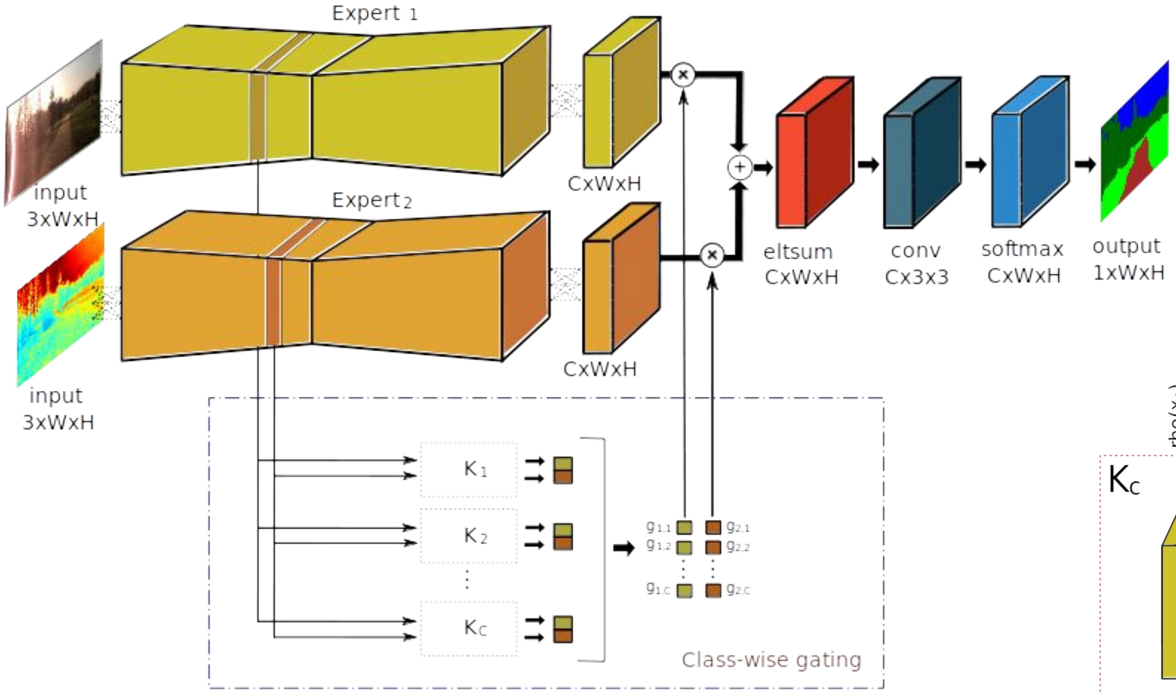
Experiments – Input Representations

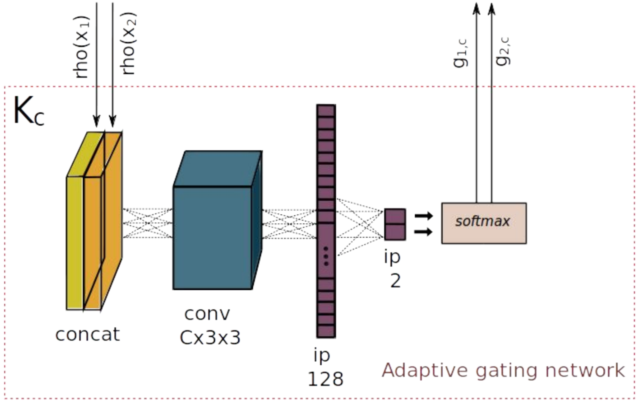


Fused Using Probabilities is Better



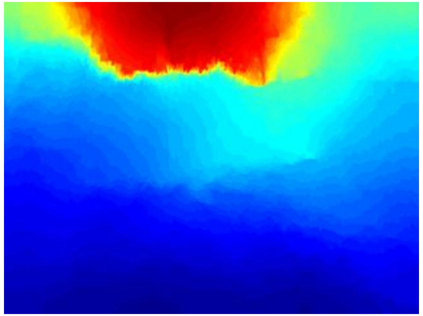
Gating Network for C-classes

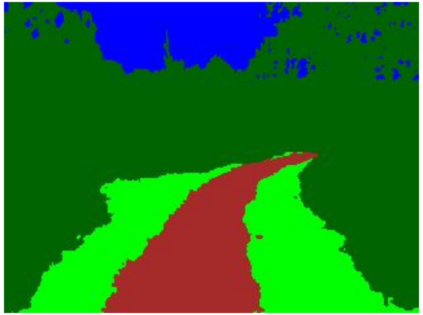




Gating Network for C-classes


RGB


DEPTH

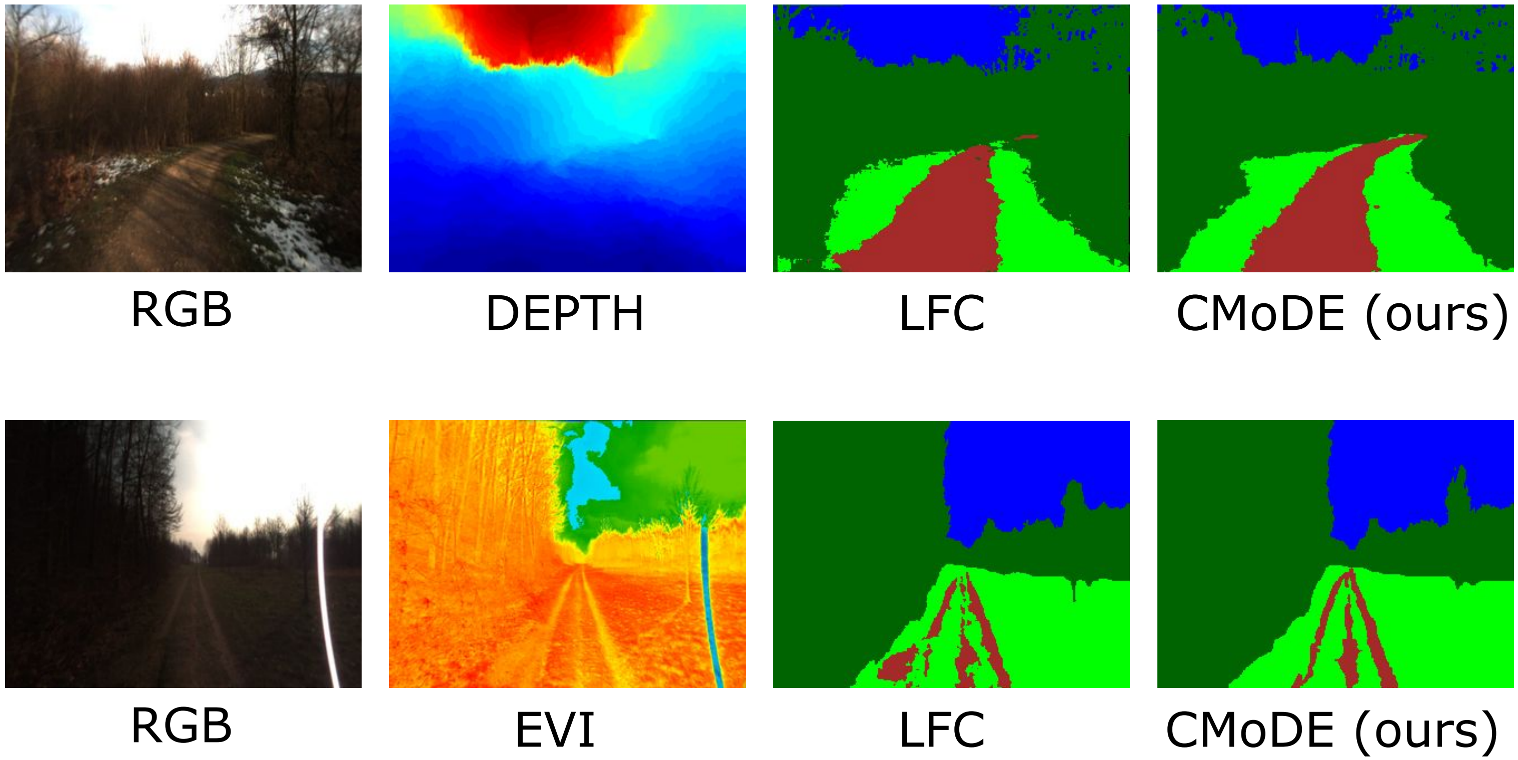

CMoDE

	RGB	DEPTH
Sky	0.44	0.56
Road	0.80	0.20

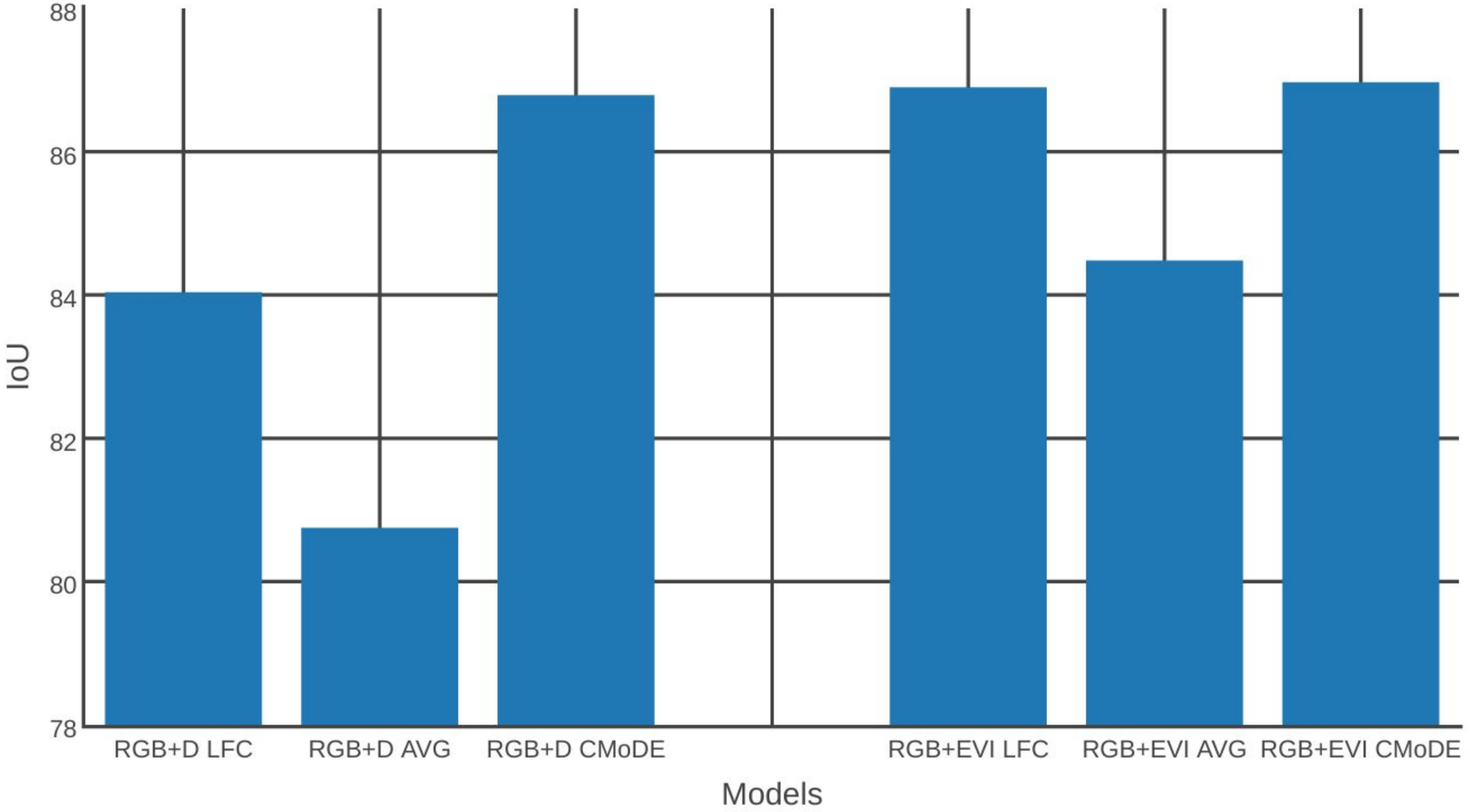
Results and Observations

- A CMoDE of RGB and EVI gives an IoU of 86.97%
- A CMoDE with RGB and Depth gives an IoU of 86.79%, 2.75% points higher than the previous best
- The gating prefers the EVI expert when RGB images contain glare, snow or low-light
- A CMoDE performs better than 50-50 fusion ratio, concatenation of channels and element-wise sum(Late Fused Convolution)

Comparison with State-of-the-art(LFC)

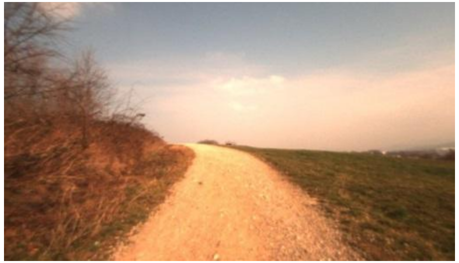


Comparison with State-of-the-art(LFC)

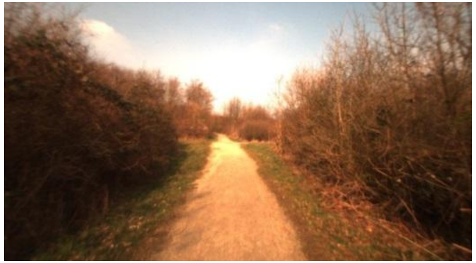


Results – CMoDE RGB+EVI

0.72

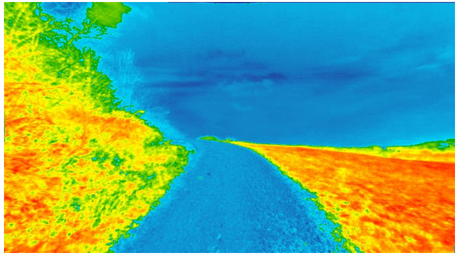


0.76

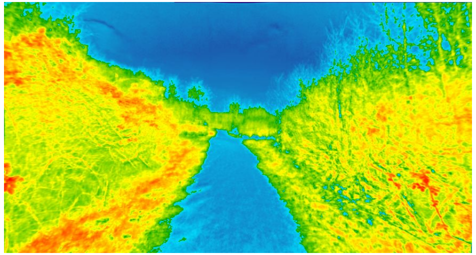


0.75

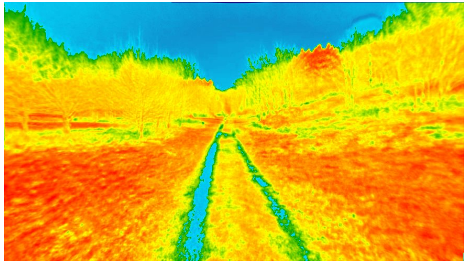
0.28

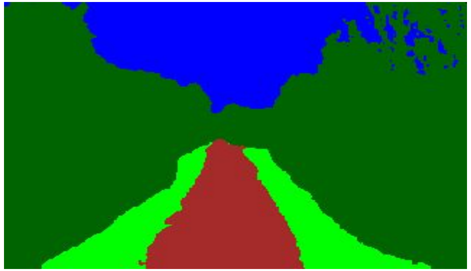


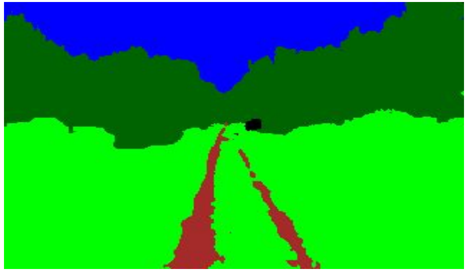
0.24



0.25







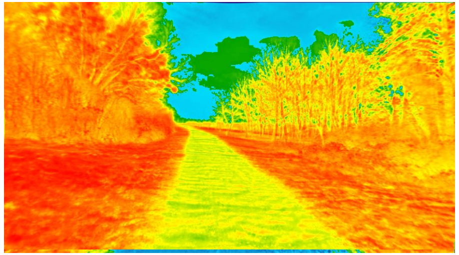
Results – CMoDE RGB+EVI

0.55

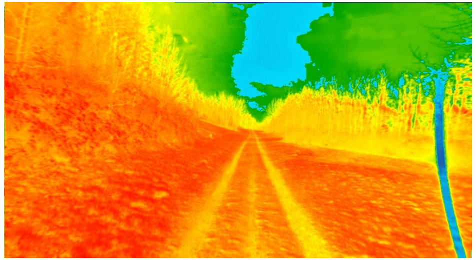
0.49

0.51

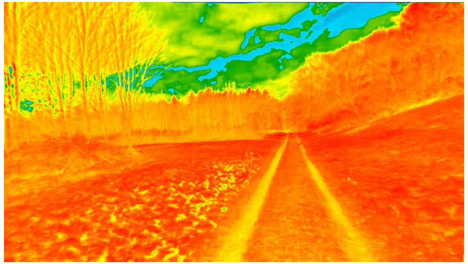
0.45

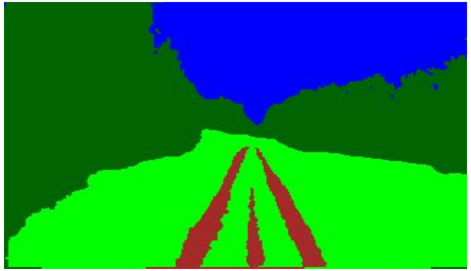


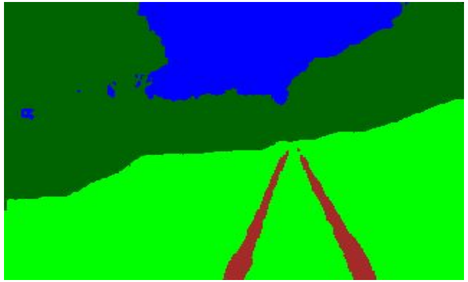
0.51



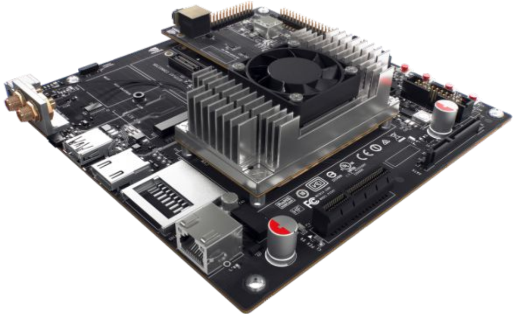
0.49

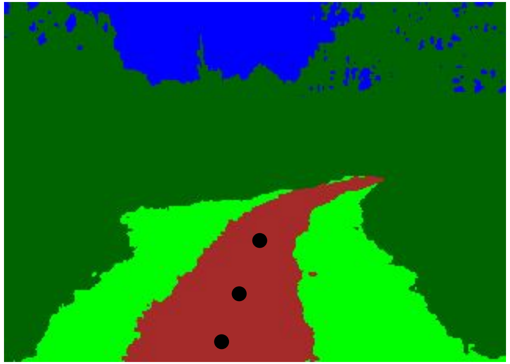






Experiment with VIONA





- Used segmentation to determine traversable terrain
- Provide waypoints to the planner
- DCNN runs as a ROS node on TX1

Conclusions

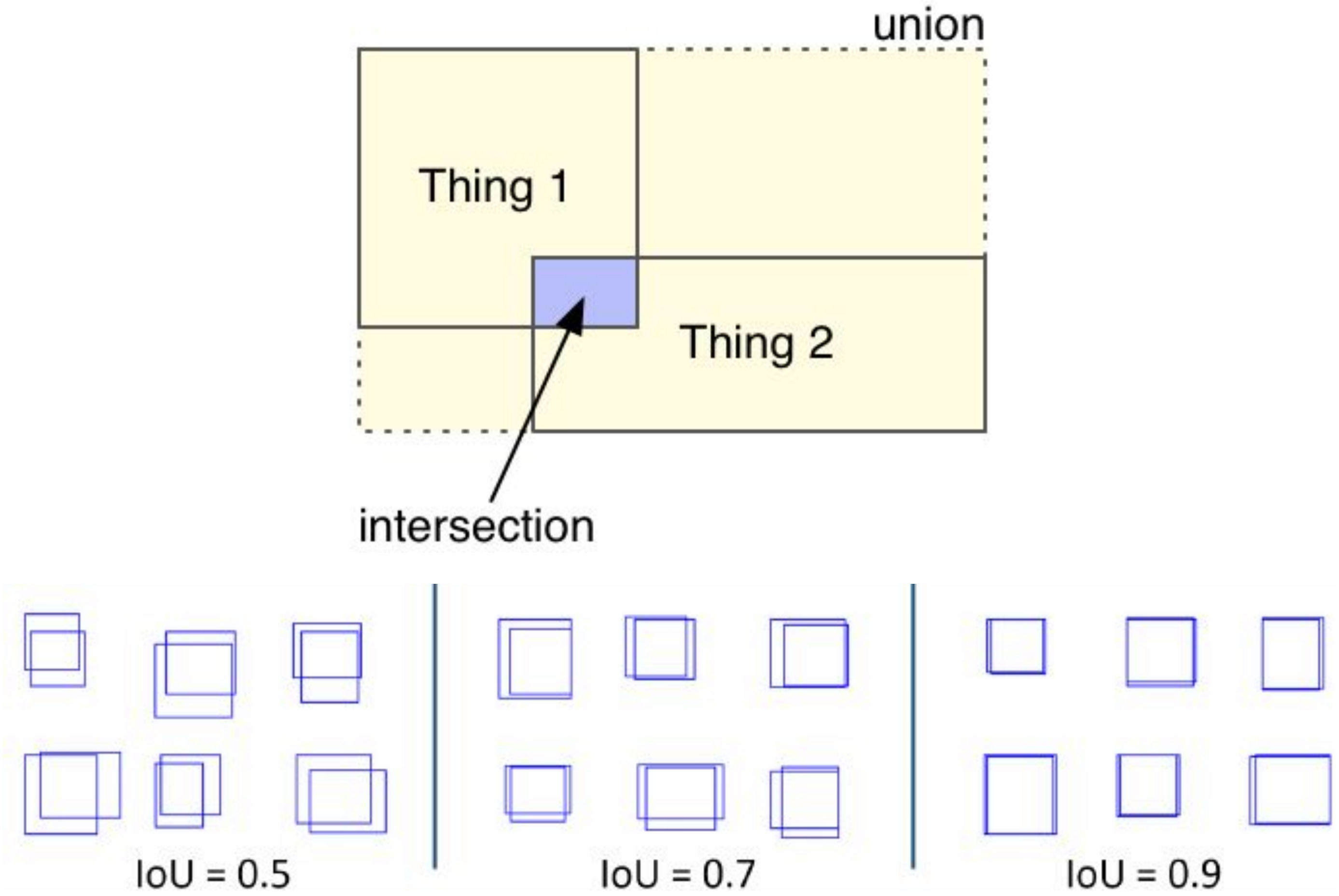
- Merge networks with probability distribution
- More competitive experts form a better mixture
- Using blurred and noisy images helps to generalize
- Overcome weaknesses– use complementary modality
- Extend models to produce per-class probabilities
- Performs state-of-the-art segmentation on the Freiburg multi-spectral forest dataset
- Significantly increase learnable parameters by using parallel networks without causing computational burden

Future work

- Training on Cityscapes and Synthia datasets
- Currently working to create uncertainty using dropout
- Improve speed by adding convolutions before inner products
- Train experts for seasons and pass same inputs, fusing using the adaptive gating

Thank you for your attention!

Live demo with VIONA


union
Thing 1
Thing 2
intersection
IoU = 0.5
IoU = 0.7
IoU = 0.9