Visual Scene Understanding for Self-Driving Cars using Deep Learning and Stereovision

The aim of the project is to give awareness to self-driving cars regarding their surroundings using an image-based approach.

Description:

- Using pair of colour cameras
- Simultaneous recognition and detection of objects in images: instance segmentation
  - Pixel-wise semantic segmentation + disparity map
- Localisation of each instance w.r.t. the ego-vehicle

Objective:

- Improve segmentation and detection rate
  - Differentiate overlapping instances

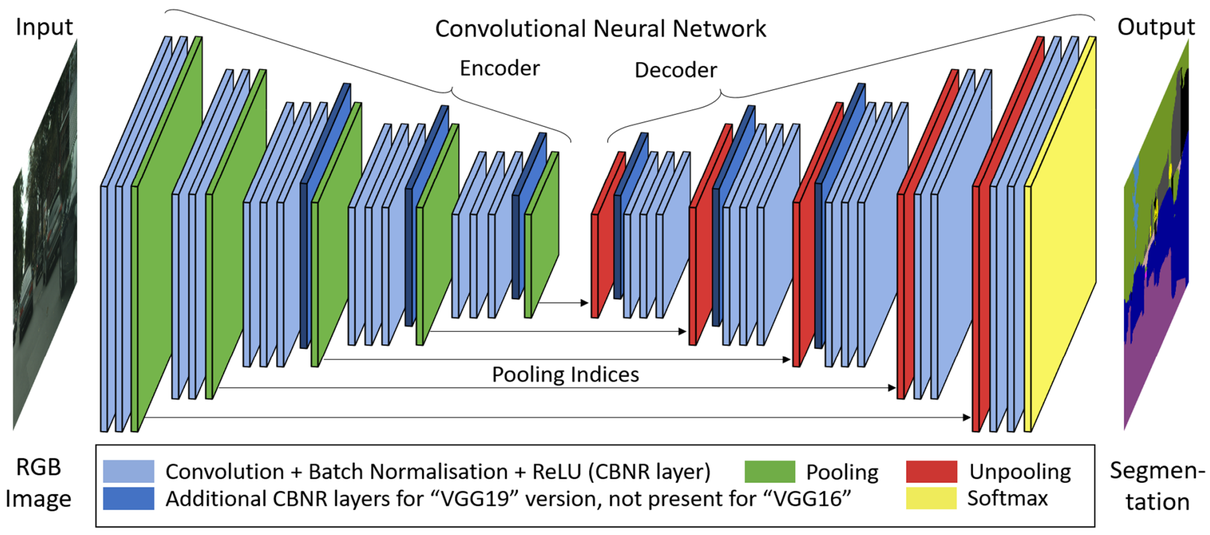

Fig. 1 - Architecture of the SegNet network, with depth variation. Adapted from [1].

Method:

- Convolutional Encoder-Decoder architecture: SegNet
- Modification of the encoder depth: VGG16 → VGG19

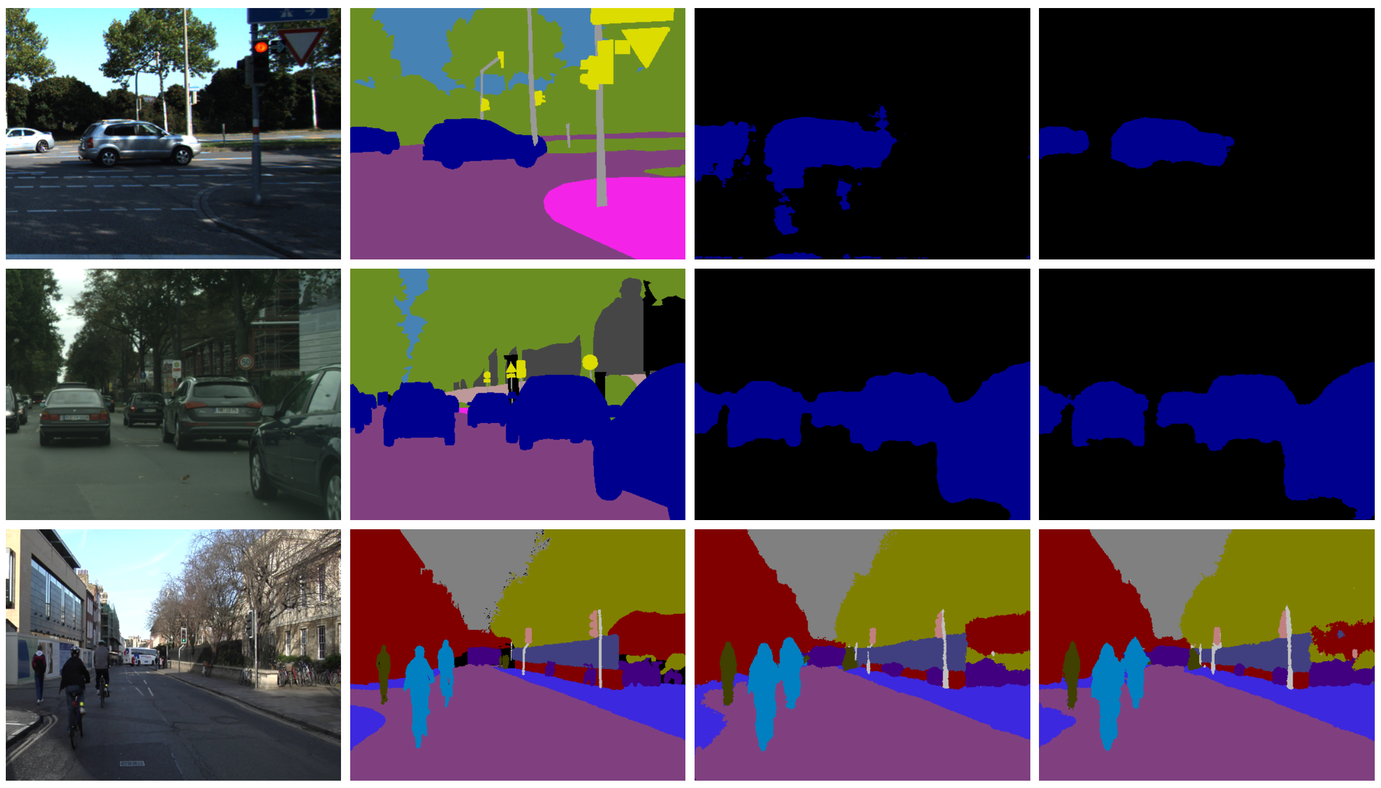

Fig. 2 – From left to right: original image, its corresponding ground truth, the segmentation prediction from SegNet whose encoder is initialised with VGG16, and the one initialised with VGG19.

Trained on	Encoder	Global Accuracy	Mean Accuracy	Mean IoU	Weight IoU	Mean BF score	Class Accuracy		Class IoU		Class BF score	
							Vehicle	Other	Vehicle	Other	Vehicle	Other
KITTI	16	92.8	**86.6**	70.8	85.3	47.3	**78.8**	94.4	52.3	89.2	27.0	63.3
CamVid	16	92.9	**90.1**	71.8	85.9	49.5	**86.6**	93.7	54.0	89.6	29.5	65.7
Cityscapes	16	97.6	**97.5**	86.6	92.7	72.1	**97.4**	97.7	79.0	94.3	62.5	80.0
KITTI	19	**94.5**	86.0	**74.4**	**87.5**	**49.6**	75.3	96.7	**57.7**	91.0	**31.3**	63.5
CamVid	19	**95.9**	87.9	**79.0**	**89.6**	**58.2**	77.7	98.0	**65.6**	92.4	**42.5**	69.9
Cityscapes	19	**98.5**	95.5	**90.4**	**94.0**	**80.7**	91.8	99.3	**85.8**	95.0	**78.7**	82.4

Table 1 – Quantitative comparison between SegNet's encoder initialised with VGG16 and VGG19. Trained and tested on different datasets for a binary classification.

What's next?

- Better boundary prediction = Fewer outliers in the disparity map per class
- Clustering methods will be used to identify each vehicle's instance and estimate its position relative to the ego-vehicle.

[1] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495.

Data, Classes and Training:

- Binary classification: Vehicle detector
- 11 classes of interest for autonomous driving applications considered
- Class balancing to reduce biases

- Pre-trained on ImageNet dataset
- CamVid, KITTI and Cityscapes

Adding more layers to a network, and creating more complex architectures allow the network to extract more complex features. However, the addition of parameters also poses a greater risk of overfitting the data and therefore not able to generalise on unseen images.

Preliminary results:

- Overall accuracy increase: + 2%
- Enhancement of the contour definition (BF score): up to + 16%
- Increase if the number of correctly classified pixels (IoU): up to + 8%
- The gain in boundary accuracy is higher when the model is trained on a larger dataset.

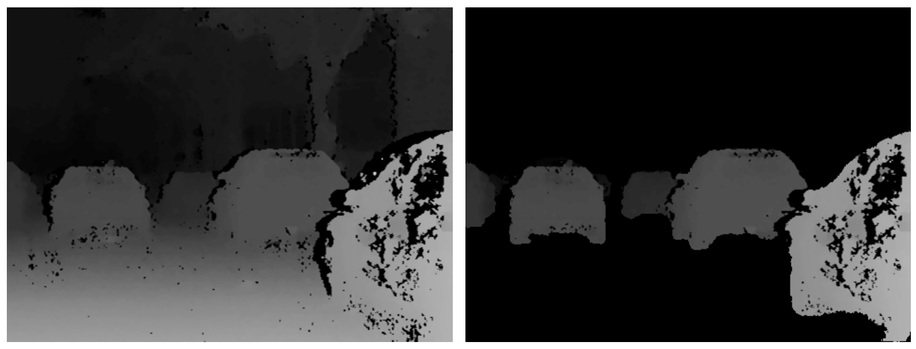

Fig. 3 – Left: disparity map, Right: disparity map for the vehicle class.

Amélie Grenier*, Dr D Nam, Dr L Feetham, Dr P Yoo**, Prof M A Richardson and Dr L Chermak

Centre for Electronic Warfare, Information and Cyber
Cranfield University, Defence Academy of the UK
Shrivenham, SN6 8LA

*a.grenier@cranfield.ac.uk, **paul@dcs.bbk.ac.uk

www.cranfield.ac.uk