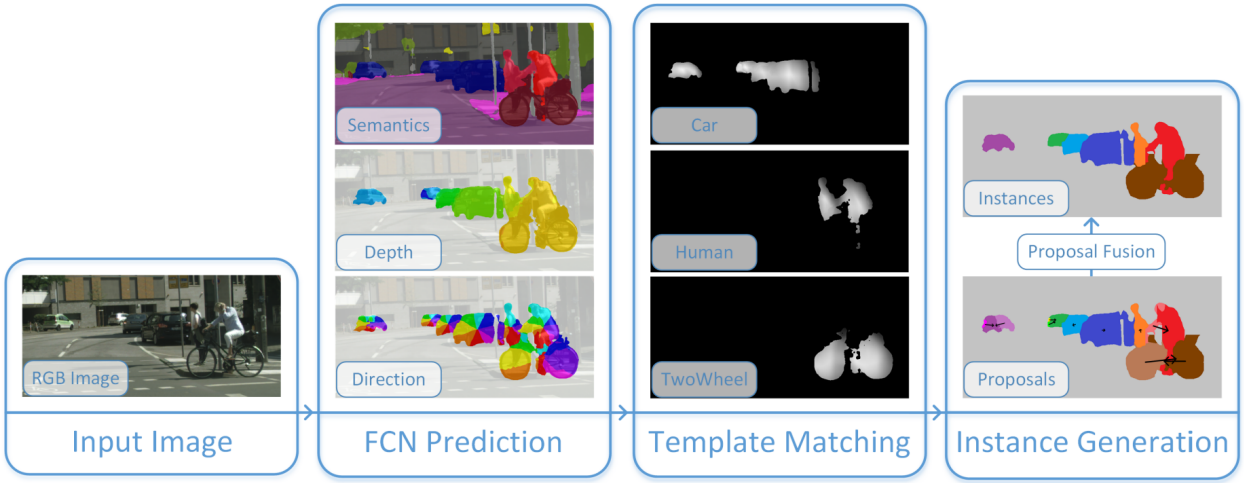
Pixel-level Encoding and Depth Layering for Instance-level Semantic Labeling

J. Uhrig, M. Cordts, U. Franke, T. Brox (GCPR 2016)



- Existing state-of-the-art methods have augmented convolutional neural networks (CNNs) with complex multitask architectures or computationally expensive graphical models

- Contributions:
  - Presents a fully convolutional network that predicts pixel-wise depth, semantics, and instance-level direction cues for holistic scene understanding
  - Instead of complex architectures or graphical models this performs post-processing using only standard computer vision techniques applied to the networks 3 output channels
  - This approach does not depend on region proposals and scales for arbitrary numbers of object instances in an image

- Evaluates KITTI and Cityscapes instance segmentation datasets