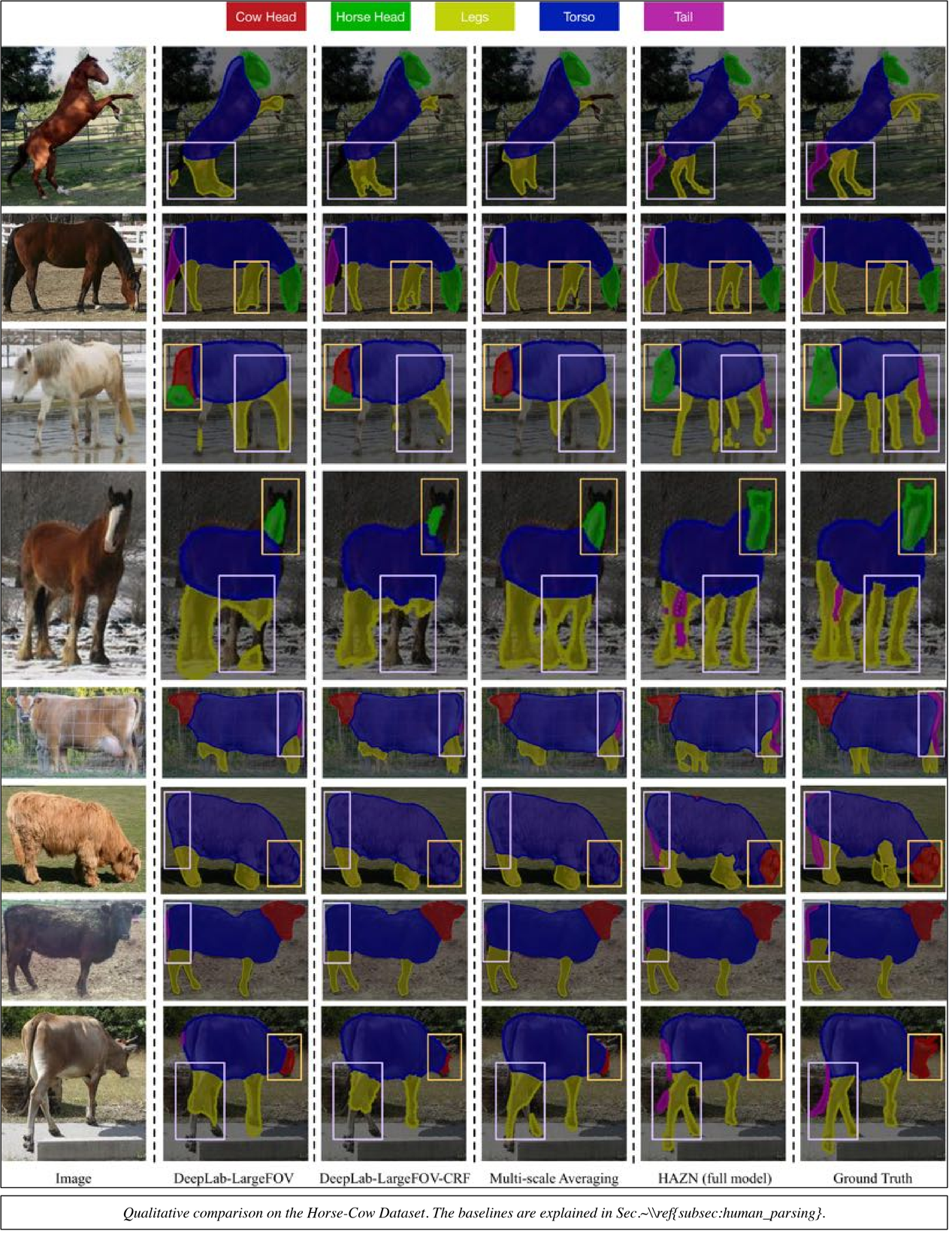
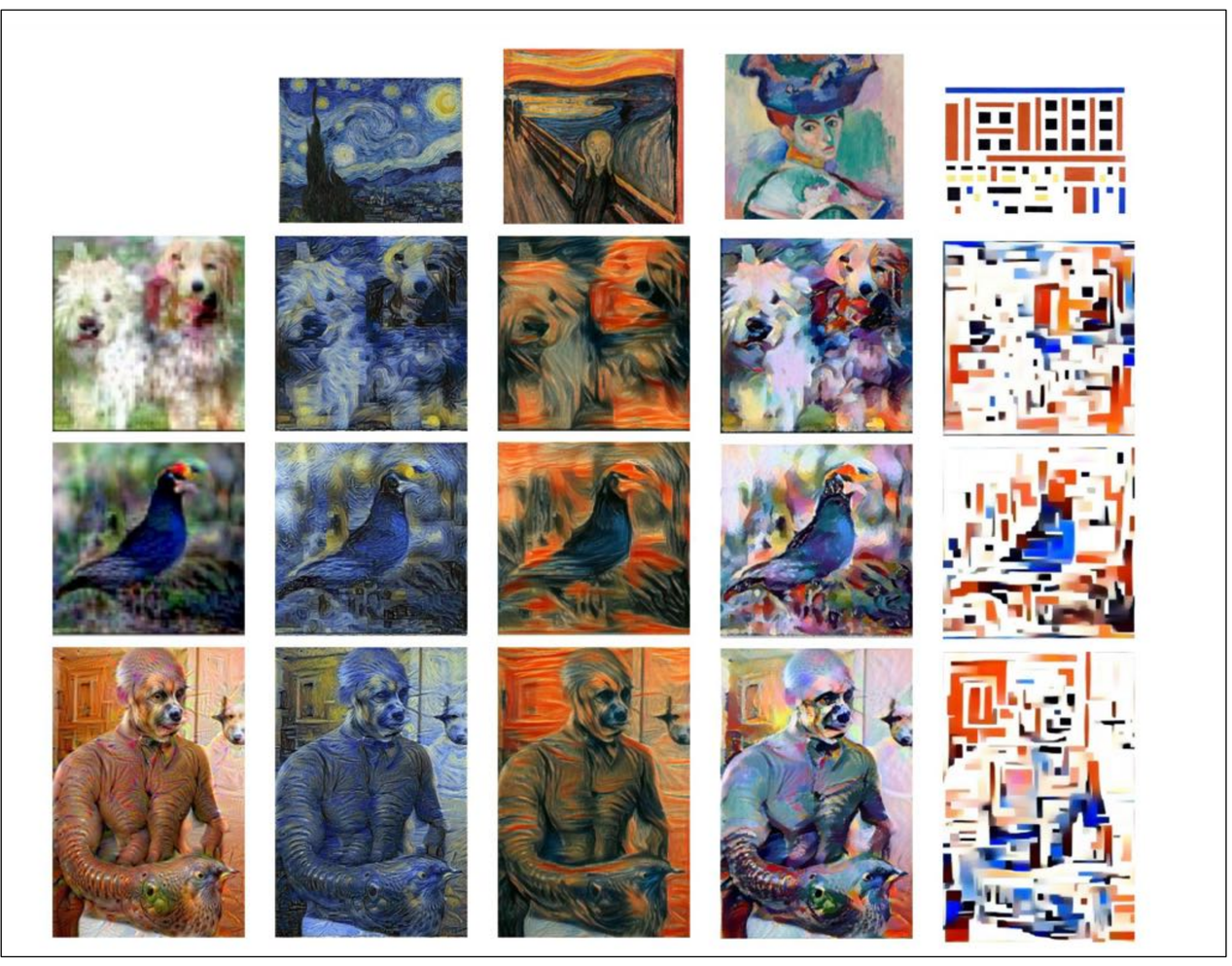


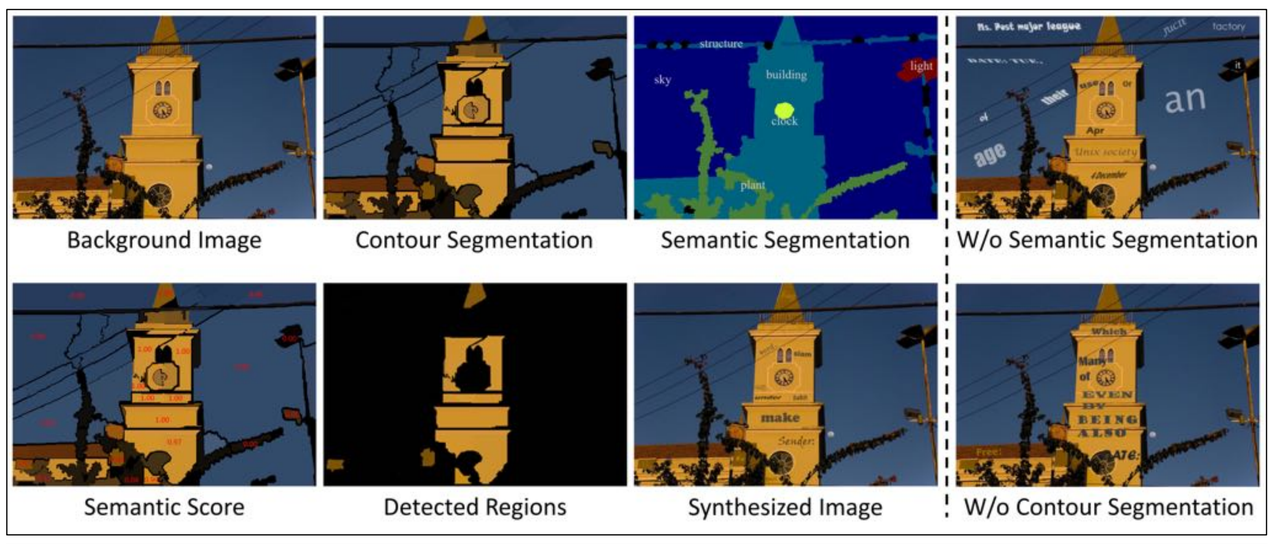
Qualitative comparison on the Horse-Cow Dataset. The baselines are explained in Sec.~\\ref{subsec:human_parsing}.

Actor-action segmentation of all models when full model with label costs is considered. Best view zoomed and in color.

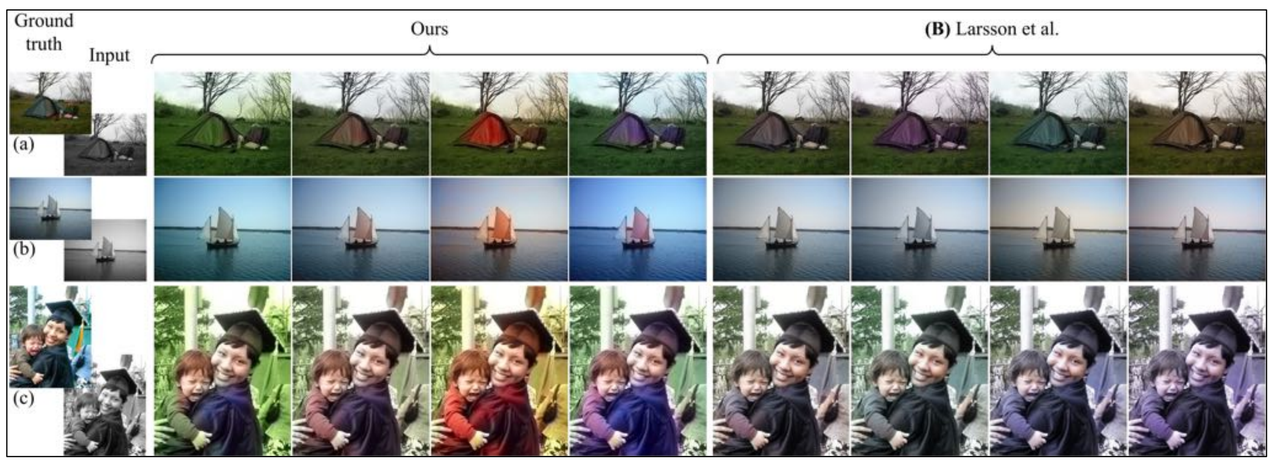
Sketch-based photo synthesis of hand-drawn test sketches. Despite the diverse sketch styles, our network usually produces high quality, diverse results. Notice the skin tone and hair color variations in the output. Some outputs are non-photorealistic because they are being somewhat faithful to caricatured input sketches. Unfortunately, some results have unrealistically high low-frequency contrast and appear unnaturally lit.



Combining ST and FV. The images in left column are content images generated by FV (first two rows) and DD (third row). Fusion images generated by the author using ST as in Fig. 9. Zoom in to see details.



The overview of the proposed image synthesis system: With background images and foreground texts as inputs, the proposed system identifies suitable text embedding regions and places the foreground texts into the background images with realistic geometry and appearance.



Colorization Results Comparison.

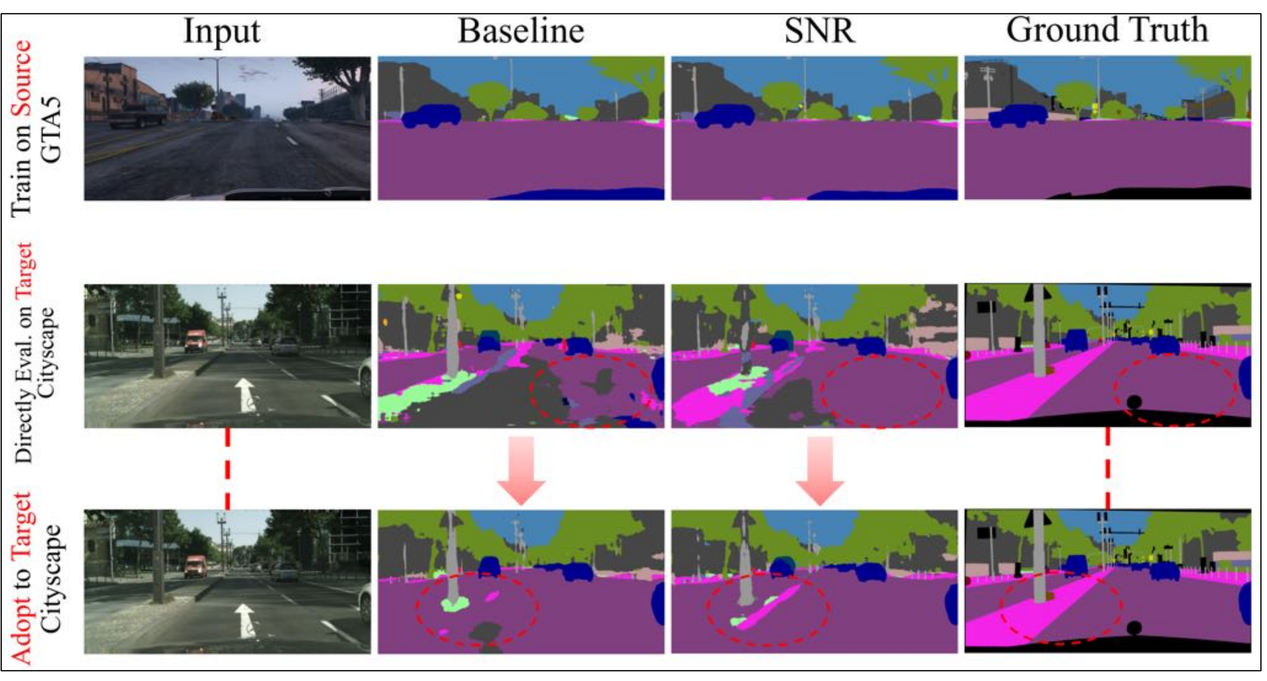
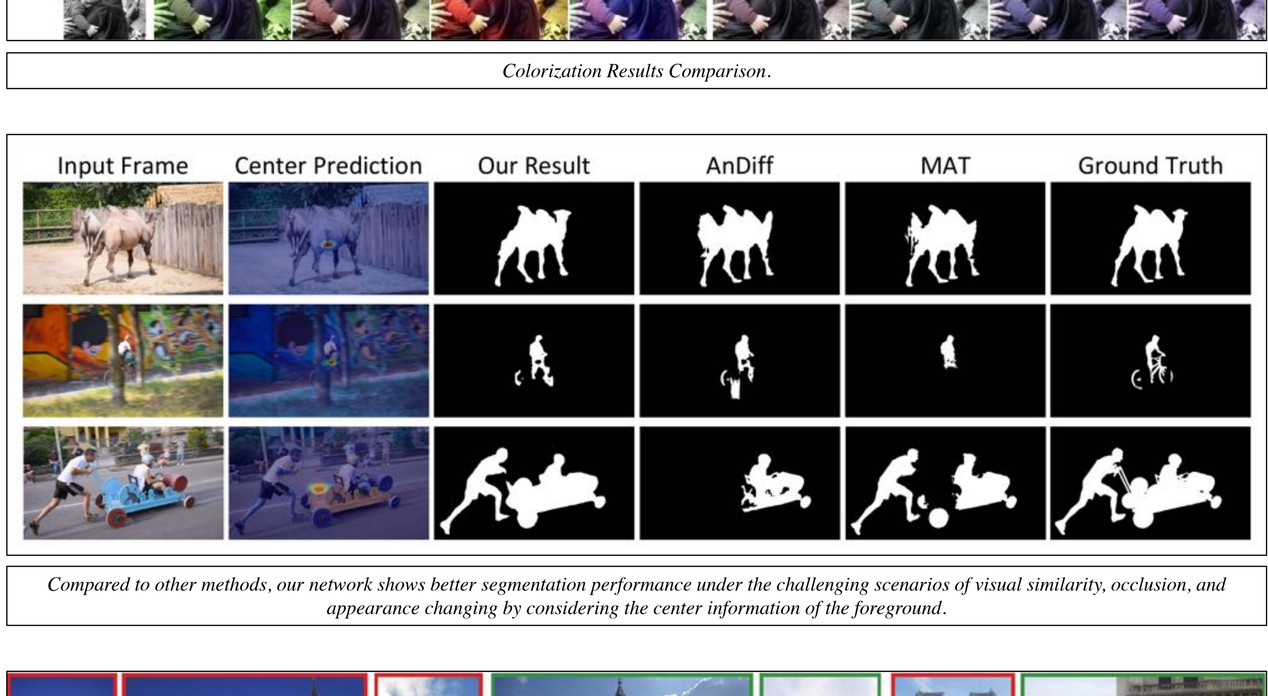
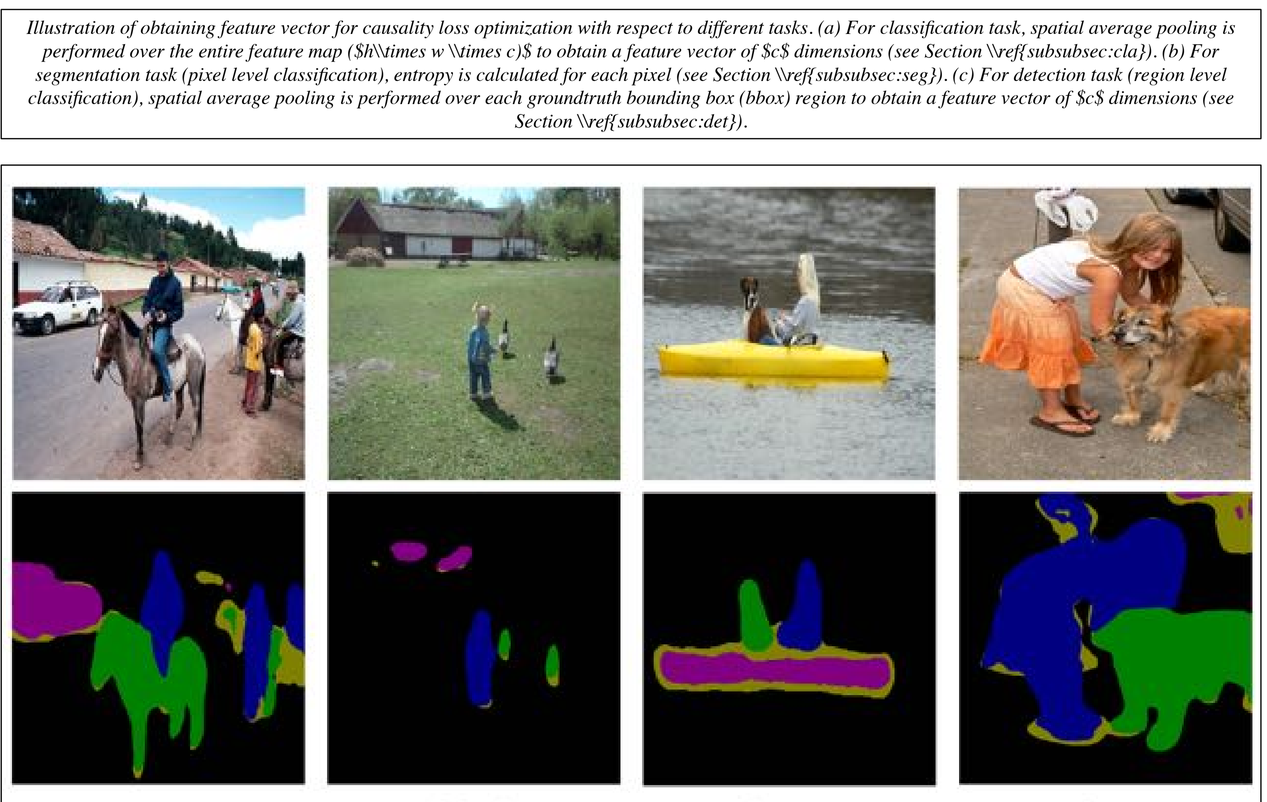


Illustration of obtaining feature vector for causality loss optimization with respect to different tasks. (a) For classification task, spatial average pooling is performed over the entire feature map ($h\\times w \\times c)$ to obtain a feature vector of c dimensions (see Section \\ref{subsubsec:cla}). (b) For segmentation task (pixel level classification), entropy is calculated for each pixel (see Section \\ref{subsubsec:seg}). (c) For detection task (region level classification), spatial average pooling is performed over each groundtruth bounding box (bbox) region to obtain a feature vector of c dimensions (see Section \\ref{subsubsec:det}).



Compared to other methods, our network shows better segmentation performance under the challenging scenarios of visual similarity, occlusion, and appearance changing by considering the center information of the foreground.



Visualization of pseudo masks on PASCAL-5^i. Note that, the colors in pseudo masks only stand for abstract semantic categories of sub-clusters rather than any concrete semantics. The last line illustrates the mined latent novel classes.

Exemplar photos with good (marked with red frames) and bad (marked with green frames) viewpoints evaluated by our multi-view learner SVM-2K.