First-view Observations Depth inputs View Transformer Module top-down-view semantic map 4 **VRM** (simulation) Depth encoder full top-down-view **VRM** road mar feature map First-view road . car, Sensor , unknown Fusion Decoder RGB inputs car sidewalk **VRM RGB** Simulation environment encoder **VRM** Discriminator Real world (real or simulation?) Target domain semantic Real RGB segmentation Generator road Segmentation car & mapping Source Semantic **VPN** domain **u**nknown sidewalk  $\mathcal{L}_{seg}$ Synthetic Real masks with top-down-view masks synthetic style semantic map (real)