In this paper, we address the scene segmentation task by capturing rich contextual dependencies based on the self-attention mechanism. Unlike previous wo