

The recent advances in instance-level detection tasks lay a strong foundation for automated visual scenes understanding. However, the ability to fully comprehend