As deep learning continues to make progress for challenging perception tasks, there is increased interest in combining vision, language, and decision-making. S