

In this paper, we propose Text2Scene, a model that generates various forms of compositional scene representations from natural language descriptions. Unlike