



在求导的过程中,可以省略从z到o的这一步,因为从z到 x_{n-1} 的求导过程等于,从z到o,再从o到 x_{n-1} ,因为o里面的其他unit在z到o的过程中全为常数,可以忽略