

target visual prediction

Visual loss

visual prediction

W_2, b_2

hidden layer

W_1, b_1

$\begin{bmatrix} 0 & 1 & \dots & 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix}$

giraffe

a
eating

tree