Input: $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^D$

γ

$\mathbb{R}^D \to \mathbb{R}^{D_1}$

$\mathbb{R}^{D_1} \to \mathbb{R}$

Linear

Batch normalization

ReLU

$\mathbb{R}^D \to \mathbb{R}^{D_1} : \mathbf{x} \mapsto \underbrace{\mathbf{A}\mathbf{x} + b}_{\in \mathbb{R}^{D_1}}$

D ... 4096
D_1 ... 64

Output: $\hat{t} \in \mathbb{R}_+$