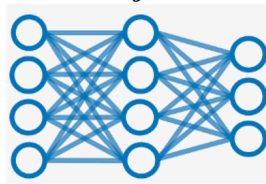Original data $\mathcal{T} = \{(x_i, y_i)\}_{i=1}^{|\mathcal{T}|}$

Parameter of network

$\theta^{\mathcal{T}}$
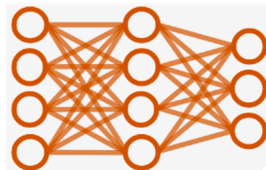
Train

Dataset distillation

$\theta^{\mathcal{S}}$

$\mathcal{L}_{\mathcal{T}}(\theta^{\mathcal{S}}) \simeq \mathcal{L}_{\mathcal{T}}(\theta^{\mathcal{T}})$

Train

Synthetic data $\mathcal{S} = \{(\hat{x}_j, \hat{y}_j)\}_{j=1}^{|\mathcal{S}|}$