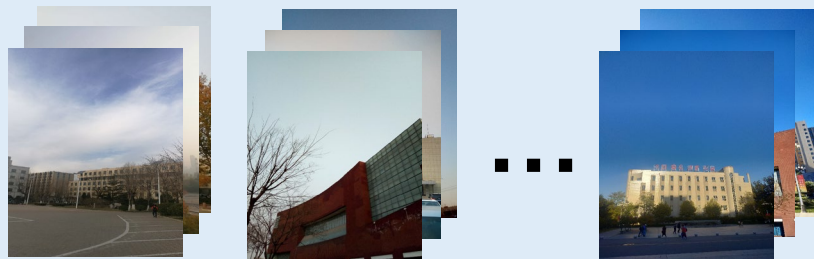


Pre-training

Data Stream



D_0

D_1

D_n

Encoder

$\varphi(\theta_o)$

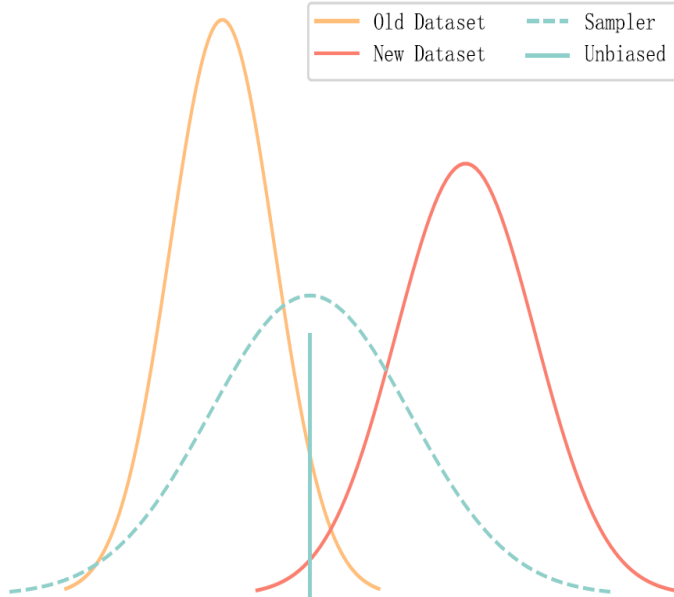
$\varphi(\theta_1)$

...

$\varphi(\theta_n)$

Distillation

Gaussian Sampler



Learnable



Frozen

Target Model

Encoder

$\varphi(\theta)$