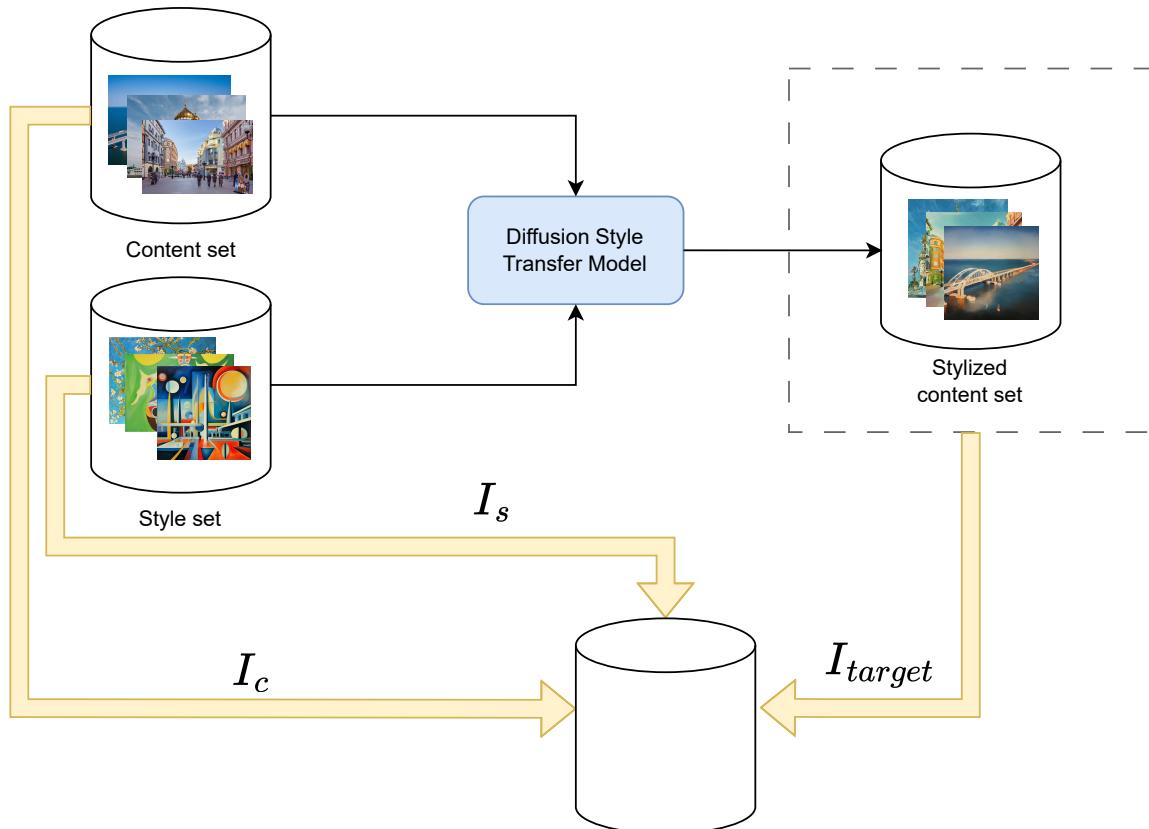


Distillation of knowledge from the diffusion model



I_c - content image

I_s - style image

Knowledge distil
supervised set(large)

I_{target} - image stylized by a diffusion model

$f_\theta(\cdot)$ - the result of the light model with the θ parameter

$\Phi(\cdot)$ - intermediate representation in the form of a multidimensional tensor

$$Loss = \|\Phi(f_\theta(I_c, I_s)) - \Phi(I_{target})\|^2$$

We use the Knowledge distil supervised set to train lighter models on supervised learning task, for example, small transformers.