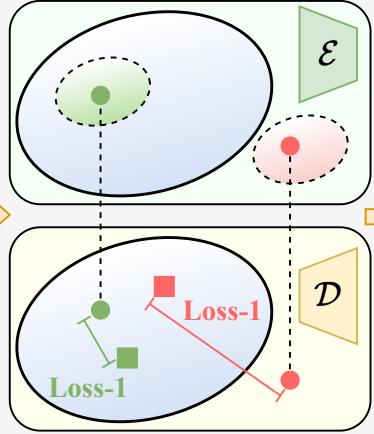**Double-Reconstruction Detailed Process**

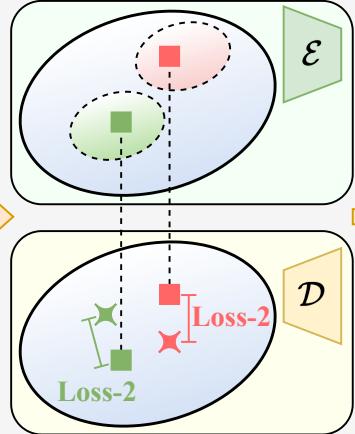Origin Images (x) | First Reconstruction | First Recon-Image ($x*$) | Second Reconstruction | Second Recon-Image ($x**$) | Loss Ratio (t)

$\mathcal{E}$ | $\mathcal{D}$ | Loss-1 | Loss-2 | $\mathcal{E}$ | $\mathcal{D}$

Loss-1 = 0.0035
Loss-2 = 0.0027
$t = 1.30$

Loss-1 = 0.0082
Loss-2 = 0.0025
$t = 3.28$

🟢 : Belonging Image   🔴 : Non-Belonging Image   🟢 : Origin Image Feature   🟩 : First Recon-Image Feature   ✖ : Second Recon-Image Feature

⬭ : VAE Feature Distribution   ⬭ : Encoded Feature Distribution (Belonging)   ⬭ : Encoded Feature Distribution (Non-Belonging)

Test Image

Double-Reconstruction

$\mathcal{E}$ $\mathcal{D}$ + $\mathcal{E}$ $\mathcal{D}$

Belonging ✅ — Yes

$t' < \tau$ ← $t' = t \times \mathcal{H}$

Non-Belonging ❌ — No

Origin Attribution   Calibration

t

τ

Example Cases of Origin Attribution (Online)

Sample Images

Double-Reconstruction

$\mathcal{E}$ $\mathcal{D}$ + $\mathcal{E}$ $\mathcal{D}$

t

$t' = t \times \mathcal{H}$

Threshold = τ KDE   Calibration

Threshold Estimation for the Target Model (Offline)