

Input  $X$

Conv 0 ( $k=3, f=64$ , BN, leaky ReLU)

Conv 1 ( $k=3, f=128$ , BN, leaky ReLU)

Conv 2 ( $k=3, f=256$ , BN, leaky ReLU)

Conv 3 ( $k=3, f=512$ , BN, leaky ReLU)

Conv 4 ( $k=1, f=z_{dim}$ , no norm., no act.)

Output

Input  $F = C_\psi(X)$

Conv 0 ( $k=3, f=1024$ , no norm., leaky ReLU)

Concatenate [Flatten Conv 0,  $Z$ ]

Dense 0 (units=512, no norm., ReLU)

Dense 1 (units=512, no norm., ReLU)

Dense 2 (units=1, no norm., no act.)

Output

a) Encoder network  $E_\psi$

b) Global statistics network  $T_\theta$

Input  $F = C_\psi(X)$

Input  $Z = E_\psi(X)$

Concatenate [Input  $F$ , Tile  $Z$ ]

Conv 0 ( $k=1, f=512$ , no norm., ReLU)

Conv 1 ( $k=1, f=512$ , no norm., ReLU)

Conv 2 ( $k=1, f=1$ , no norm., no act.)

Output

c) Local statistics network  $T_\phi$