Mask-Guided Generative Sub-Network

From Local Embedding Sub-Network

Decoders $D_{sub-network}$

(8,12,256) *(8,12,256)* *(20,36,256)* *(64,64,256)* *(64,64,256)*

Embedding Layers

Target mask m^t

E_m

Concatenated masks
(64,64,256)

Concatenated masks and
Facial embeddings

G_m

(256,64,64)

(256,256,11)

To Background Fusing Sub-Network