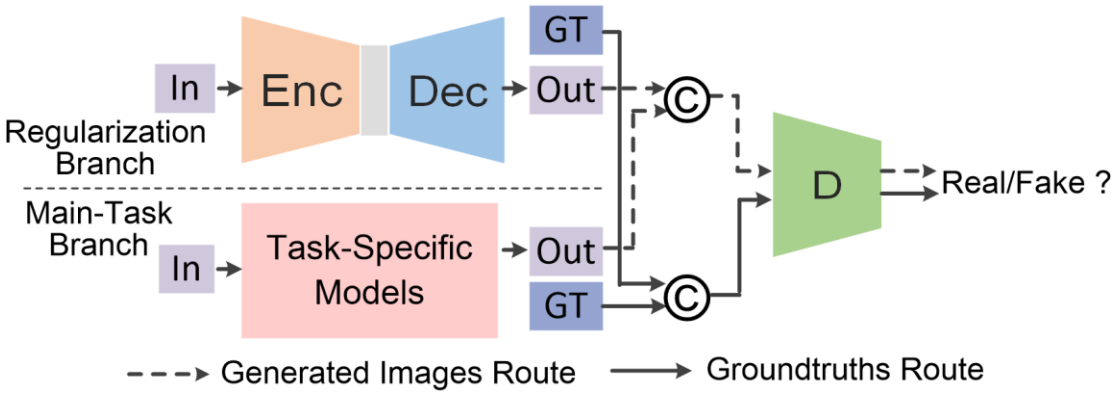

Figure1: Illustration of a general generative model design with the improved adversarial loss setup.

As shown in Fig.1, by only combining the "main-task" branch (MTB) with discriminator (D), it is a traditional GAN and the D gets too successful that the generator gradient vanishes and learns nothing. However, by adding the regularization branch (RB), which shares the same task as MTB but take the groundtruth as input, the output from RB will be much more like the groundtruth. By concatenating the output from MTB and RB together as input, the D will not easily recognize the combined input as fake, thus can be trained in a more stable and easier manner, resulting in better image generation by the MTB. Such mechanism is also helpful for the proposed CERLD-Net because it can help guarantee a more stable knowledge transfer procedure, thus resulting in better factor disentanglement. Despite the simplicity of the novel discriminator design, elaborate experiments have been carried out (Sec4.3 and 4.5 of the paper) to demonstrate its effectiveness, and we will add the above analysis to the paper in the camera-ready version for better understanding our motivation.