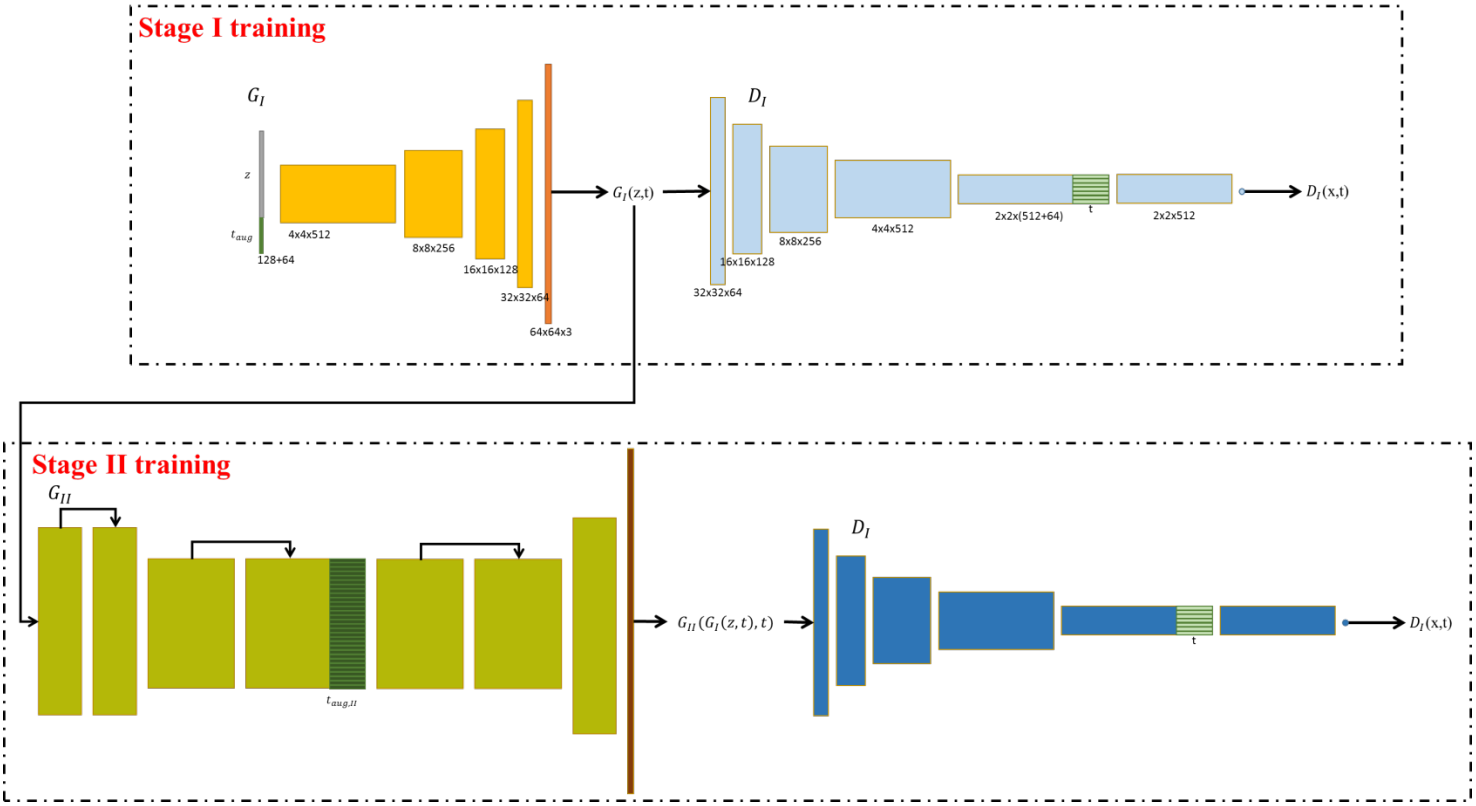
Environment

Linux, Intel i7, GTX 960M(4G memory) & GTX 1060M(6G memory)

Package: tensorflow, numpy

Model description

Ensemble of Conditional Stack Least-Square GAN



$$\mathcal{L}_D = 1.2\,\mathcal{L}_{D,r} + \mathcal{L}_{D,w} + \mathcal{L}_{D,f}$$

$$where \quad \mathcal{L}_{D,r} = \frac{1}{2}\,(D(x_{real},t)-1)^2$$

$$\mathcal{L}_{D,w} = \frac{1}{2}\,\left(D(x_{wrong},t)\right)^2$$

$$\mathcal{L}_{D,f} = \frac{1}{2}\,\left(D(G(z,t_{real}),t)\right)^2$$

$$\mathcal{L}_G = |D(G(z,t)-1)|$$

(StackGAN: https://arxiv.org/abs/1612.03242)

(LSGAN: https://arxiv.org/abs/1611.04076)

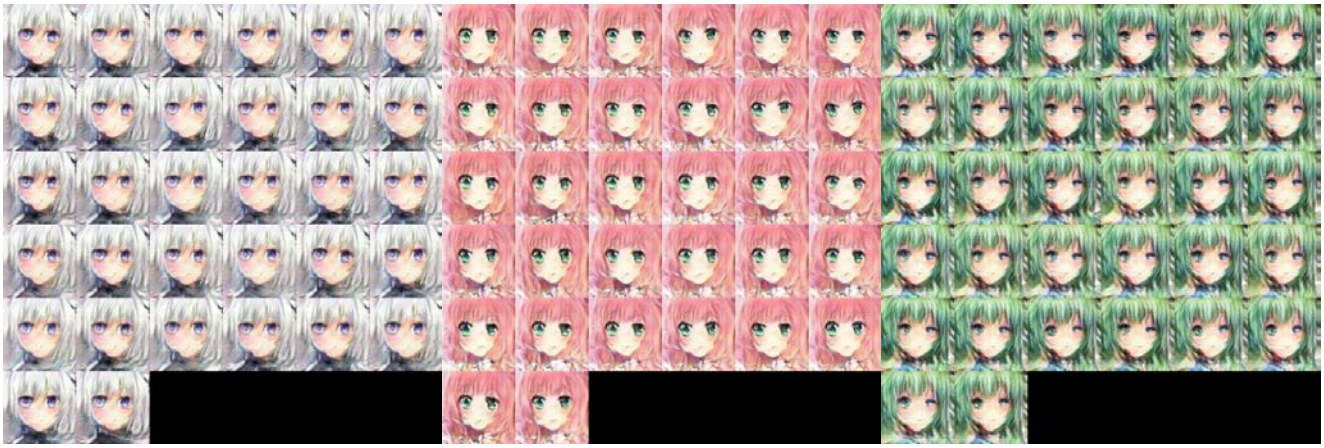
Performance

1. 使用 L1 loss 更新 G 可防止過多訓練後 G 有可能爆掉
2. 不使用 skip thought，使用單純的 attribute embedding
3. Stack GAN 可得到 96x96 生成影像，再縮小為 64x64，可增加細節
4. 生成影像時(testing 時)限制 noise 高斯分佈的 std 為 0.35(training 時為 1.0)可產生較為真實的影像
5. 低 std noise 產生影像均較類似，以 ensemble 選擇最後 5 個 epoch 的生成模型來產生，可較多樣性

white hair, blue eyes pink hair, green eyes green hair, blue eyes

Stage I, std = 0.432, iters=51000

Stage II, std = 0.123, iters=129000



Stage II, std = 0.432, iters=129000

Stage II, std = 0.950, iters=129000

Experiment settings and results

('image_size_I',64),
('image_size_II',96),
('flip',True),
('z_dim',128),
('z_normal',True),
('t_dim',64),
('t_aug',True),
('t_aug_penalty',True),
('g_conv_depth',256),
('d_conv_depth',256),
('d_loc_step_I',1),
('g_loc_step_I',1),
('d_loc_step_II',1),
('g_loc_step_II',1),
('AdamOpt',True),
('batch_size',64)])