

Implementation Setups for “CCDM: Continuous Conditional Diffusion Models for Image Generation”

TABLE S.1: Detailed Implementation Setups for Compared Methods in Table I (Part I).

Dataset	Method	Setup
RC-49 (64×64)	ReACGAN	big resnet, hinge loss, steps=30K, batch size=256, $\text{lr}_D = 4 \times 10^{-4}$, $\text{lr}_G = 1 \times 10^{-4}$
	ADCGAN	big resnet, hinge loss, steps=20K, batch size=128, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$
	ADM-G	Classifier: steps=20K, batch size=128, $\text{lr}=3 \times 10^{-4}$ Diffusion: steps=50K, batch size=32, $\text{lr}=1 \times 10^{-4}$, diffusion steps=4K
	CFG	steps=50K, $\text{lr}=10^{-4}$, batch size=128, time steps (train/sampling)=1000/100
	CcGAN	SAGAN, hinge loss, SVDL+ILI, $\sigma = 0.047$, $\nu = 50625$, use DiffAugment, steps=30K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step
	Dual-NDA	SAGAN, hinge loss, SVDL+ILI, $\sigma = 0.047$, $\nu = 50625$, $\kappa = 0.0044$, use DiffAugment, steps=40K, Dual-NDA starts at 30K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step, $\lambda_1 = 0.15$, $q_1 = 0.9$, $\lambda_2 = 0.15$, $q_2 = 0.5$, $N_{\text{II}} = 18K$
	CcDPM	ϵ -prediction, steps=50K, soft vicinity, $\sigma = 0.047$, $m_\kappa = 2$, $\nu = 50625.00$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 250$, batch size=128, $\text{lr}=10^{-4}$
	CCDM	α^0 -prediction, steps=50K, hard vicinity, $\sigma = 0.047$, $m_\kappa = 2$, $\kappa = 0.0044$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 250$, batch size=128, $\text{lr}=10^{-4}$
UTKFace (64×64)	ReACGAN	big resnet, hinge loss, steps=40K, batch size=256, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$
	ADCGAN	big resnet, hinge loss, steps=20K, batch size=128, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$
	ADM-G	Classifier: steps=20K, batch size=128, $\text{lr}=3 \times 10^{-4}$ Diffusion: steps=65K, batch size=64, $\text{lr}=1 \times 10^{-5}$, diffusion steps=1K
	CFG	steps=100K, $\text{lr}=10^{-4}$, batch size=1024, time steps (train/sampling)=1000/100
	CcGAN	SNGAN, the vanilla loss, SVDL+ILI, $\sigma = 0.041$, $\nu = 3600$, use DiffAugment, steps=40K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step
	Dual-NDA	SNGAN, the vanilla loss, SVDL+ILI, $\sigma = 0.041$, $\nu = 3600$, $\kappa = 0.017$, use DiffAugment, steps=60K, Dual-NDA starts at 40K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step, $\lambda_1 = 0.05$, $q_1 = 0.9$, $\lambda_2 = 0.15$, $q_2 = 0.9$, $N_{\text{II}} = 60K$ (1K per age value)
	CcDPM	ϵ -prediction, steps=100K, soft vicinity, $\sigma = 0.041$, $m_\kappa = 1$, $\nu = 3600$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 250$, batch size=128, $\text{lr}=10^{-4}$
	CCDM	α^0 -prediction, steps=100K, hard vicinity, $\sigma = 0.041$, $m_\kappa = 1$, $\kappa = 0.017$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 250$, batch size=128, $\text{lr}=10^{-4}$
Steering Angle (64×64)	ReACGAN	big resnet, hinge loss, steps=20K, batch size=256, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$
	ADCGAN	big resnet, hinge loss, steps=20K, batch size=128, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$
	ADM-G	Classifier: steps=20K, batch size=128, $\text{lr}=3 \times 10^{-4}$ Diffusion: steps=50K, batch size=32, $\text{lr}=3 \times 10^{-4}$, diffusion steps=4K
	CFG	steps=80K, $\text{lr}=10^{-4}$, batch size=128, time steps (train/sampling)=1000/100
	CcGAN	SAGAN, the hinge loss, SVDL+ILI, $\sigma = 0.029$, $\nu = 1000.438$, use DiffAugment, steps=20K, batch size=512, $\text{lr}=10^{-4}$, update D twice per step
	Dual-NDA	SAGAN, the hinge loss, SVDL+ILI, $\sigma = 0.029$, $\nu = 1000.438$, $\kappa = 0.032$, use DiffAugment, steps=20K, Dual-NDA starts at 0, batch size=512, $\text{lr}=10^{-4}$, update D twice per step, $\lambda_1 = 0.1$, $q_1 = 0.5$, $\lambda_2 = 0.2$, $q_2 = 0.9$, $N_{\text{II}} = 17740$ (10 Type II negative images for 1774 distinct training angle values)
	CcDPM	ϵ -prediction, steps=50K, soft vicinity, $\sigma = 0.029$, $m_\kappa = 5$, $\nu = 1000.438$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 250$, batch size=128, $\text{lr}=10^{-4}$
	CCDM	α^0 -prediction, steps=50K, hard vicinity, $\sigma = 0.029$, $m_\kappa = 5$, $\kappa = 0.032$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 250$, batch size=128, $\text{lr}=10^{-4}$
Cell-200 (64×64)	ReACGAN	big resnet, hinge loss, steps=40K, batch size=256, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$
	ADCGAN	big resnet, hinge loss, steps=20K, batch size=128, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$
	ADM-G	Classifier: steps=20K, batch size=128, $\text{lr}=3 \times 10^{-4}$ Diffusion: steps=65K, batch size=64, $\text{lr}=1 \times 10^{-5}$, diffusion steps=1K
	CFG	steps=100K, $\text{lr}=10^{-4}$, batch size=1024, time steps (train/sampling)=1000/100
	CcGAN	SNGAN, the vanilla loss, SVDL+ILI, $\sigma = 0.041$, $\nu = 3600$, use DiffAugment, steps=40K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step
	Dual-NDA	SNGAN, the vanilla loss, SVDL+ILI, $\sigma = 0.041$, $\nu = 3600$, $\kappa = 0.017$, use DiffAugment, steps=60K, Dual-NDA starts at 40K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step, $\lambda_1 = 0.05$, $q_1 = 0.9$, $\lambda_2 = 0.15$, $q_2 = 0.9$, $N_{\text{II}} = 60K$ (1K per age value)
	CcDPM	ϵ -prediction, steps=50K, soft vicinity, $\sigma = 0.077$, $m_\kappa = 2$, $\nu = 2500$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 250$, batch size=128, $\text{lr}=5 \times 10^{-5}$
	CCDM	α^0 -prediction, steps=50K, hard vicinity, $\sigma = 0.077$, $m_\kappa = 2$, $\kappa = 0.02$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 250$, batch size=128, $\text{lr}=5 \times 10^{-5}$

TABLE S.2: Detailed Implementation Setups for Compared Methods in Table I (Part II).

Dataset	Method	Setup
UTKFace (128×128)	ReACGAN	big resnet, hinge loss, steps=20K, batch size=128, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$
	ADCGAN	big resnet, hinge loss, steps=20K, batch size=128, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$, use DiffAugment
	ADM-G	Classifier: steps=20K, batch size=64, $\text{lr}=3 \times 10^{-4}$ Diffusion: steps=50K, batch size=24, $\text{lr}=1 \times 10^{-5}$, diffusion steps=1K
	CFG	steps=50K, $\text{lr}=10^{-5}$, batch size=64, time steps (train/sampling)=1000/100
	CcGAN	SAGAN, the hinge loss, SVDL+ILI, $\sigma = 0.041$, $\nu = 900$, use DiffAugment, steps=20K, batch size=256, $\text{lr}=10^{-4}$, update D four times per step
	Dual-NDA	SAGAN, the hinge loss, SVDL+ILI, $\sigma = 0.041$, $\nu = 900$, $\kappa = 0.033$, use DiffAugment, steps=22500, Dual-NDA starts at 20K, batch size=256, $\text{lr}=10^{-4}$, update D four times per step, $\lambda_1 = 0.05$, $q_1 = 0.9$, $\lambda_2 = 0.15$, $q_2 = 0.9$, $N_{\text{II}} = 60K$ (1K per age value)
	CcDPM	ϵ -prediction, steps=200K, soft vicinity, $\sigma = 0.041$, $m_\kappa = 1$, $\nu = 3600$, $p_{\text{drop}} = 0.1$, $\gamma = 2.0$, timesteps in training $T = 1000$, timesteps in sampling $T' = 150$, batch size=64, $\text{lr}=10^{-5}$
	CCDM	α^0 -prediction, steps=200K, hard vicinity, $\sigma = 0.041$, $m_\kappa = 1$, $\kappa = 0.017$, $p_{\text{drop}} = 0.1$, $\gamma = 2.0$, timesteps in training $T = 1000$, timesteps in sampling $T' = 150$, batch size=64, $\text{lr}=10^{-5}$
Steering Angle (128×128)	ReACGAN	big resnet, hinge loss, steps=20K, batch size=128, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$
	ADCGAN	big resnet, hinge loss, steps=20K, batch size=128, update D twice per step, $\text{lr}_D = 2 \times 10^{-4}$, $\text{lr}_G = 5 \times 10^{-5}$, use DiffAugment
	ADM-G	Classifier: steps=20K, batch size=64, $\text{lr}=3 \times 10^{-4}$ Diffusion: steps=50K, batch size=24, $\text{lr}=1 \times 10^{-5}$, diffusion steps=1K
	CFG	steps=50K, $\text{lr}=10^{-5}$, batch size=64, time steps (train/sampling)=1000/100
	CcGAN	SAGAN, the hinge loss, SVDL+ILI, $\sigma = 0.029$, $\nu = 1000.438$, use DiffAugment, steps=20K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step
	Dual-NDA	SAGAN, the hinge loss, SVDL+ILI, $\sigma = 0.029$, $\nu = 1000.438$, $\kappa = 0.032$, use DiffAugment, steps=20K, Dual-NDA starts at 15K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step, $\lambda_1 = 0.2$, $q_1 = 0.5$, $\lambda_2 = 0.3$, $q_2 = 0.9$, $N_{\text{II}} = 17740$ (10 Type II negative images for 1774 distinct training angle values)
	CcDPM	ϵ -prediction, steps=200K, soft vicinity, $\sigma = 0.029$, $m_\kappa = 5$, $\nu = 1000.438$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 150$, batch size=64, $\text{lr}=5 \times 10^{-5}$
	CCDM	α^0 -prediction, steps=200K, hard vicinity, $\sigma = 0.041$, $m_\kappa = 5$, $\kappa = 0.032$, $p_{\text{drop}} = 0.1$, $\gamma = 1.5$, timesteps in training $T = 1000$, timesteps in sampling $T' = 150$, batch size=64, $\text{lr}=5 \times 10^{-5}$
UTKFace (192×192)	CcGAN	SAGAN, the hinge loss, SVDL+ILI, $\sigma = 0.029$, $\nu = 1000.438$, use DiffAugment, steps=20K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step
	Dual-NDA	SAGAN, the hinge loss, SVDL+ILI, $\sigma = 0.029$, $\nu = 1000.438$, $\kappa = 0.032$, use DiffAugment, steps=20K, Dual-NDA starts at 15K, batch size=256, $\text{lr}=10^{-4}$, update D twice per step, $\lambda_1 = 0.2$, $q_1 = 0.5$, $\lambda_2 = 0.3$, $q_2 = 0.9$, $N_{\text{II}} = 17740$ (10 Type II negative images for 1774 distinct training angle values)
	CcDPM	ϵ -prediction, steps=300K, soft vicinity, $\sigma = 0.041$, $m_\kappa = 1$, $\nu = 3600$, $p_{\text{drop}} = 0.1$, $\gamma = 2.0$, timesteps in training $T = 1000$, timesteps in sampling $T' = 150$, batch size=64, $\text{lr}=10^{-5}$
	CCDM	α^0 -prediction, steps=300K, hard vicinity, $\sigma = 0.041$, $m_\kappa = 1$, $\kappa = 0.017$, $p_{\text{drop}} = 0.1$, $\gamma = 2.0$, timesteps in training $T = 1000$, timesteps in sampling $T' = 150$, batch size=64, $\text{lr}=10^{-5}$

TABLE S.3: Detailed Implementation Setups for DMD2-M in Table II.

Dataset	Setup
RC-49	SNGAN, hinge loss, steps=50K, batch size=128, $m_\kappa = 0$, $\kappa = 0$, update D twice per step, $w_D = 10.0$, $w_G = 1.0$, $\text{lr}_D = 10^{-4}$, $\text{lr}_G = 10^{-4}$, DiffAugment
Steering Angle	SNGAN, hinge loss, steps=200K, batch size=128, $m_\kappa = 1$, $\kappa = 0.011$, update D twice per step, $w_D = 10.0$, $w_G = 1.0$, $\text{lr}_D = 10^{-4}$, $\text{lr}_G = 10^{-4}$, DiffAugment