



基於VAE圖像生成

指導教授：余松年 教授 學生：黃建齊

摘要

本專題的目的是為了研究深度學習在圖像生成任務的應用。在專題中首先探討現有的相關工作，以了解其優勢和不足之處，並著重於擴大生成圖像的尺寸，提高圖像品質，以及優化生成過程的速度這三個方面進行比較。

研究過程

在專題中使用的資料集為Celeba與Celeba-Hq兩個，分別用於訓練生成64*64、128*128以及256*256三種尺寸。

對於生成圖像的評分是依據PSNR與FID，分別於重建的過程中，以及在最後生成結果要能夠愈接近真實圖像的分布。

專題的模型架構是使用Variational AutoEncoder[1]以及在這之後基於此方向所提出的改進，在專題中依序復現其結果。

以下是改進的過程

- 加入 perceptual loss 提升重建的效果。[2]
- 將潛變數離散化處理後驗崩潰的問題，在透過自回歸模型學習其分布。[3]
- 加入 discriminator提升重建的效，自回歸模型用transformer。[4]
- 為了提升生成的效率使用residual vector quantization的方法提高壓縮率以及重建的品質，並將第二步驟的自回模型改善與之對應。[5]

研究結果

以下為實驗結果。

Celeba 64*64				
	params	Psnr dB	Fid	Generate time
Vae_pixel_wise_loss	5.15M	21.28	27.926	0.14 minutes
Vae_perceptual_loss	5.15M	22.76	18.455	
vqvae_v0 + Pixel_cnn	1.50M+3.45M	29.26	16.345	
Celeba 128*128				
	params	Psnr dB	Fid	Generate time
vqvae_v1	1.50M	28.94	22.167	0.71 minutes
vqvae_v2+ Pixel_cnn	2.02M+21.42M	23.28		
vqgan_v0+ Pixel_cnn	5.49M+21.42M	24.35	21.385	0.71 minutes
vqgan_v0+transformer	5.49M+85.15M	24.35	10.753	2.02 minutes
vqgan_v1+transformer	7.01M+85.15M	20.54	10.678	0.30 minutes
Celeba -hq 256*256				
	params	Psnr	Fid	Generate time
vqgan_v2+transformer	7.52M +85.15M	22.65	10.691	2.06 minutes
vqgan_v3	8.64M	20.43	13.282	0.15 minutes
rqvae_v0+ rq transformer	8.64M +121.84M	23.08		
rqvae_v1+ rq transformer	9.82M +121.84M	22.30	12.018	0.15 minutes



以上為生成圖像

結論

經過專題的學習使我對深度學習的實作更加的熟練，我認為深度學習的方法相當靈活，未來若能應用在不同領域的工作中，加上對於問題的理解，說不定能找到其他的解法。