

ADLxMLDS HW4 Report

B03902086 資工四 李鈺昇

Model Description

首先，由於髮色和眼色共有23種，因此我對於所有「至少有出現一種髮色或眼色」的圖片，建立一個23維的向量：如果有那個tag就設為1否則為0（類似one-hot的概念，不過因為data可能包含好幾種顏色，因此可能有複數個1），配上這些圖片作為training data。

這麼做的好處是，不需要用上skip-thought vector或是其他語言模型。使用類似one-hot vector的另一個好處會在下一個部分說明。

- Structure

Input: noise (100), tags (23)

Generator (

Concatenate noise and tags, and then resize(view) to 4D

ConvTranspose2d(128, 512, kernel_size=(4, 4), stride=(1, 1))

ConvTranspose2d(512, 256, kernel_size=(4, 4), stride=(2, 2), padding=(1, 1))

BatchNorm2d(256, eps=1e-05, momentum=0.1, affine=True)

LeakyRelu(0.05)

ConvTranspose2d(256, 128, kernel_size=(4, 4), stride=(2, 2), padding=(1, 1))

BatchNorm2d(128, eps=1e-05, momentum=0.1, affine=True)

LeakyRelu(0.05)

ConvTranspose2d(128, 64, kernel_size=(4, 4), stride=(2, 2), padding=(1, 1))

BatchNorm2d(64, eps=1e-05, momentum=0.1, affine=True)

LeakyRelu(0.05)

ConvTranspose2d(64, 3, kernel_size=(4, 4), stride=(2, 2), padding=(1, 1))

Tanh

)

Output: image (3, 64, 64)

Input: image (3, 64, 64), tags (23)

Discriminator (

Conv2d(3, 64, kernel_size=(4, 4), stride=(2, 2), padding=(1, 1))

LeakyRelu(0.05)

Conv2d(64, 128, kernel_size=(4, 4), stride=(2, 2), padding=(1, 1))

LeakyRelu(0.05)

Conv2d(128, 256, kernel_size=(4, 4), stride=(2, 2), padding=(1, 1))

LeakyRelu(0.05)

Conv2d(256, 512, kernel_size=(4, 4), stride=(2, 2), padding=(1, 1))

LeakyRelu(0.05)

Replicate and permute tags, and then concatenate with the previous output

Conv2d(535, 1, kernel_size=(4, 4), stride=(1, 1))

)

Output: real number (1)

- Objective function
 - Generator
 - WGAN的loss，多了text input
 - Discriminator
 - WGAN-GP的loss，，多了text input以及有wrong text的部分

Improving the Performance

原本generate的結果很差，後來我把tag的類one-hot vector乘上50（或任意大數字）之後就會變得很好

Experiments and Results

我有嘗試 `batch_size = 4`，但是表現沒變好。