
Algorithm : GAN Training For Text-to-Image Generation

Input: Image batch X , matching text T , mismatching text T_{hat} , batch size B , learning rate η

Output: Trained generator and discriminator

```
1: for  $n = 1$  to  $B$  do
2:   Encode matching text:  $H \leftarrow \text{EncodeText}(T)$ 
3:   Encode mismatching text:  $H_{\text{hat}} \leftarrow \text{EncodeText}(T_{\text{hat}})$ 
4:   Generate noise:  $Z \sim \text{Gaussian}(0, I)$ 
5:   Generate fake images:  $G(Z, H)$ 
6:   Compute discriminator scores:  $D(X, H)$  (real image with correct text)
7:   Compute discriminator scores:  $D(X, H_{\text{hat}})$  (real image with incorrect
    text)
8:   Compute discriminator scores:  $D(X_{\text{hat}}, H)$  (fake image with correct
    text)
9:   Compute discriminator loss:  $L_D \leftarrow \log(D(X, H)) + (\log(1 -$ 
     $D(X_{\text{hat}}, H)) + \log(1 - D(X, H_{\text{hat}})))/2$ 
10:  Update discriminator parameters:  $\rho_D \leftarrow \rho_D - \eta \cdot \frac{\partial L_D}{\partial \rho_D}$ 
11:  Compute generator loss:  $L_G \leftarrow \log(S_F)$ 
12:  Update generator parameters:  $\rho_G \leftarrow \rho_G - \eta \cdot \frac{\partial L_G}{\partial \rho_G}$ 
13: end for
```
