SUPPLEMENTARY MATERIAL FOR DISCRIMINATOR MODIFICATION IN GAN FOR TEXT-TO-IMAGE GENERATION

Fei Fang\textsuperscript{1}, Ziqing Li\textsuperscript{1}, Fei Luo\textsuperscript{1∗}, and Chunxia Xiao\textsuperscript{1∗}

1. School of Computer Science, Wuhan University, Wuhan 430072, China.
fangfei369@163.com, thalialee@163.com, luofei_whu@126.com, cxxiao@whu.edu.cn

ABSTRACT

In this supplementary material, we present more qualitative comparison results as complementary results of the main paper. Moreover, we describe some implementation details and parameter values.

1. QUALITATIVE COMPARISONS WITH SOME STATE-OF-THE-ART METHODS

We compare the images generated using AttnGAN \cite{1}, DMGAN \cite{2}, ObjGAN \cite{3}, CPGAN \cite{4} and our improved versions of AttnGAN and DMGAN on the MSCOCO14 dataset. As shown in Fig. 1, our improved methods can generate better quality images with realistic details. Moreover, the images generated by our methods are more consistent with the input text and more similar with the real images, as shown in Fig. 1.

2. IMPLEMENTATION DETAILS

In Section 2.3.1 of the main paper, we set the number of the positive and negative samples $m = 10$. For the positive samples of the current real image $I_i^{R}$, we first transform $I_i^{R}$ to $I_i^{R'}$ with a small range of 5\% using the tool \texttt{transforms.ColorJitter} of PyTorch. Then both $I_i^{R}$ and $I_i^{R'}$ are merged with the current fake image $I_i^{F}$ with five parameters. The parameters are randomly selected from $[0.85, 0.98]$.

For the negative samples of $I_i^{R}$, we first randomly select five real images from other $b − 1$ real images, where $b$ is the batchsize. Then we merge $I_i^{R}$ with the random noise image (described in Section 2.2.1 of the main paper) with five parameters. The parameters are randomly selected from $[0.5, 0.1]$.

In Section 2.3.2 of the main paper, the number of channels $C = 576$. In Equ.13 of the main paper, we set $\lambda_3 = 0.5$, $\lambda_4 = 0.9$, $\lambda_5 = 0.05$ in our experiments. In Equ.14 of the main paper, we set $\lambda_6 = 50.0$, $\lambda_7 = 0.5$ in our experiments. Note that for the generated $64 \times 64 \times 3$ and $128 \times 128 \times 3$ images, we use the traditional unconditional discriminator as \cite{5}.

3. REFERENCES


Fig. 1. Comparison of the original images and the generated images from AttnGAN [1], DMGAN [2], ObjGAN [3], CPGAN [4] and our improved versions of AttnGAN and DMGAN on the MSCOCO14 dataset.