Supplement to Generative Adversarial Text to Image Synthesis

In this supplement we include additional material that did not fit into the main paper.

1. Additional text-to-image examples for birds, flowers and COCO

Many additional text-to-image examples are contained in the following .pdf files included with the supplement:

- `Flowers-GAN-INT-CLS.pdf`.
- `Flowers-GAN-E2E.pdf`: Flowers End2End.
- `CUB-GAN-INT-CLS.pdf`:
- `CUB-GAN-E2E.pdf`: Birds End2End.
- `coco_compare.pdf`: Comparison to AlignDraw on COCO examples chosen for the ICLR paper.

2. Robustness of GAN variants

When training the baseline GAN on CUB, we found that simply choosing a different random seed (affecting network initialization and also minibatch selection) could yield results of dramatically varying quality. The classification and interpolation regularizers improved the robustness, and GAN-CLS-INT consistently yielded good results regardless of the random seed. To quantify this, we trained 10 instances each (varying only the random seed) of GAN, GAN-CLS, GAN-INT and GAN-CLS-INT on the 100 CUB training classes for 200 epochs. Using samples from each of these GAN models, we trained a zero-shot image classifier from scratch, following the same protocol described in section 5.5.

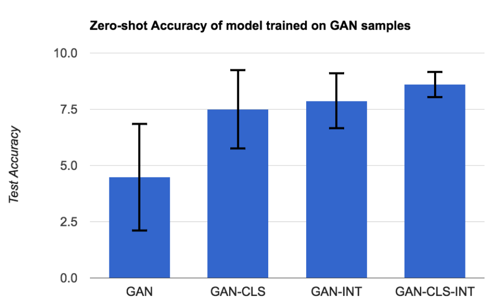

Figure 1. Zero-shot accuracies for image encoders trained on text-conditional GAN samples.

Figure 1 shows the result. All variants perform better than the GAN baseline, and GAN-CLS-INT has the highest average performance and lowest variance. Our impression is that the classification (-CLS) and especially interpolation regularizer (-INT) stabilize the training, significantly reducing the incidence of "failed" GANs.