

Supplementary: End-to-end Training for Text-to-Image Synthesis using Dual-Text Embeddings

Yeruru Asrar Ahmed

*Department of Computer Science and Engineering
Indian Institute of Technology Madras*

asrar@cse.iitm.ac.in

Anurag Mittal

*Department of Computer Science and Engineering
Indian Institute of Technology Madras*

amittal@cse.iitm.ac.in

Reviewed on OpenReview: <https://openreview.net/forum?id=gJ10knHV5e>

1 More Qualitative Results

Additional qualitative results are illustrated of DTE-GAN on CUB Welinder et al. (2010) dataset in Figures 1 and 2, on Oxford-102 Nilsback & Zisserman (2008) dataset in Figures 3 and 4, and on MS-COCO Lin et al. (2014) dataset in Figures 5 and 6 . Comparing these results with those of AttnGANXu et al. (2018) and DM-GANZhu et al. (2019) on CUB dataset, HDGANZhang et al. (2018) on Oxford-102 dataset, and AttnGAN for MS-COCO dataset. Additional qualitative results are provided for image manipulations using LightWeight GANLi et al. (2020) bootstrapped with DTE-GAN pre-trained embeddings in Figures 7 and 8.

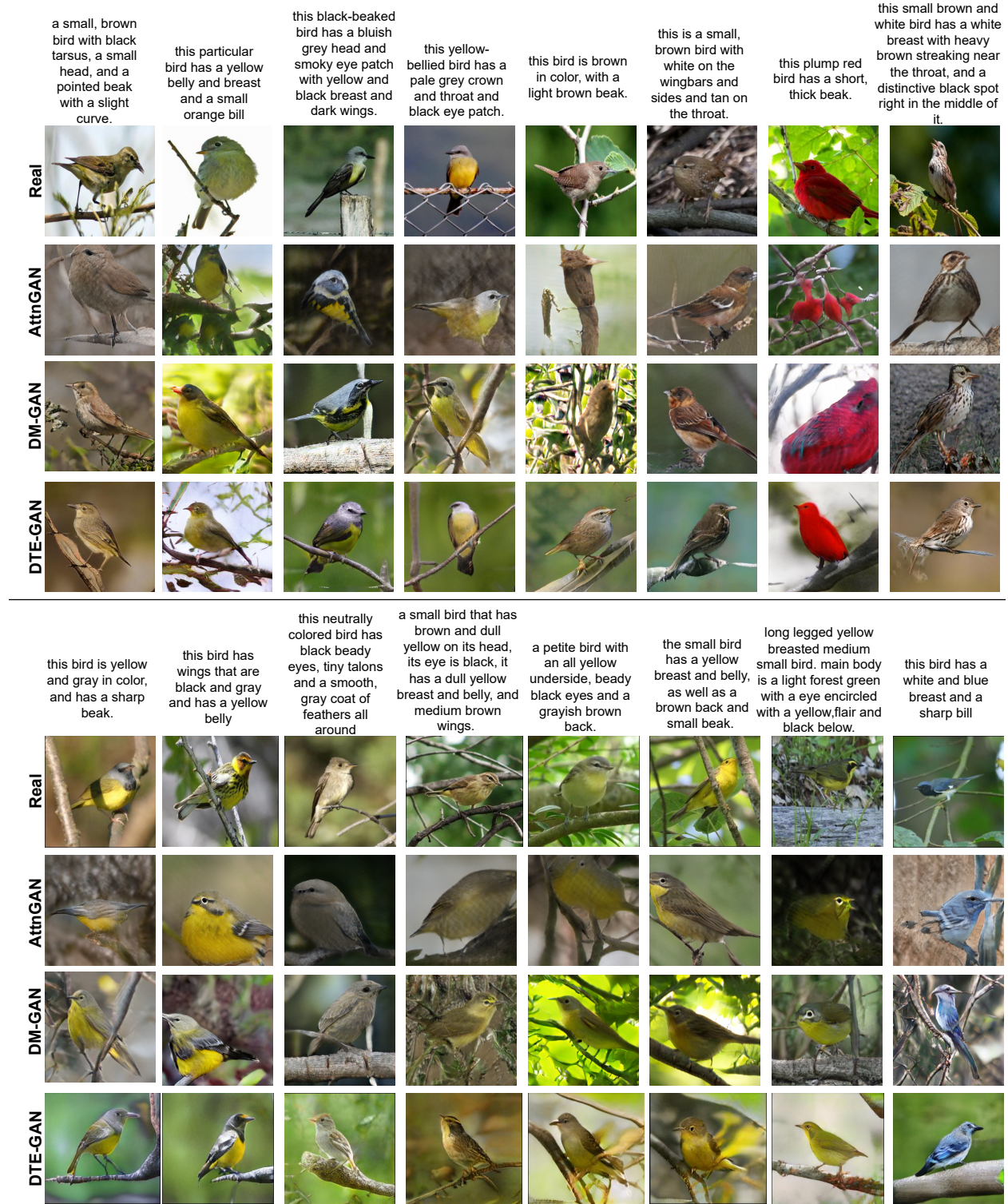


Figure 1: Images generated by DTE-GAN (ours) on CUB dataset and compared with AttnGAN Xu et al. (2018) & DM-GAN Zhu et al. (2019).

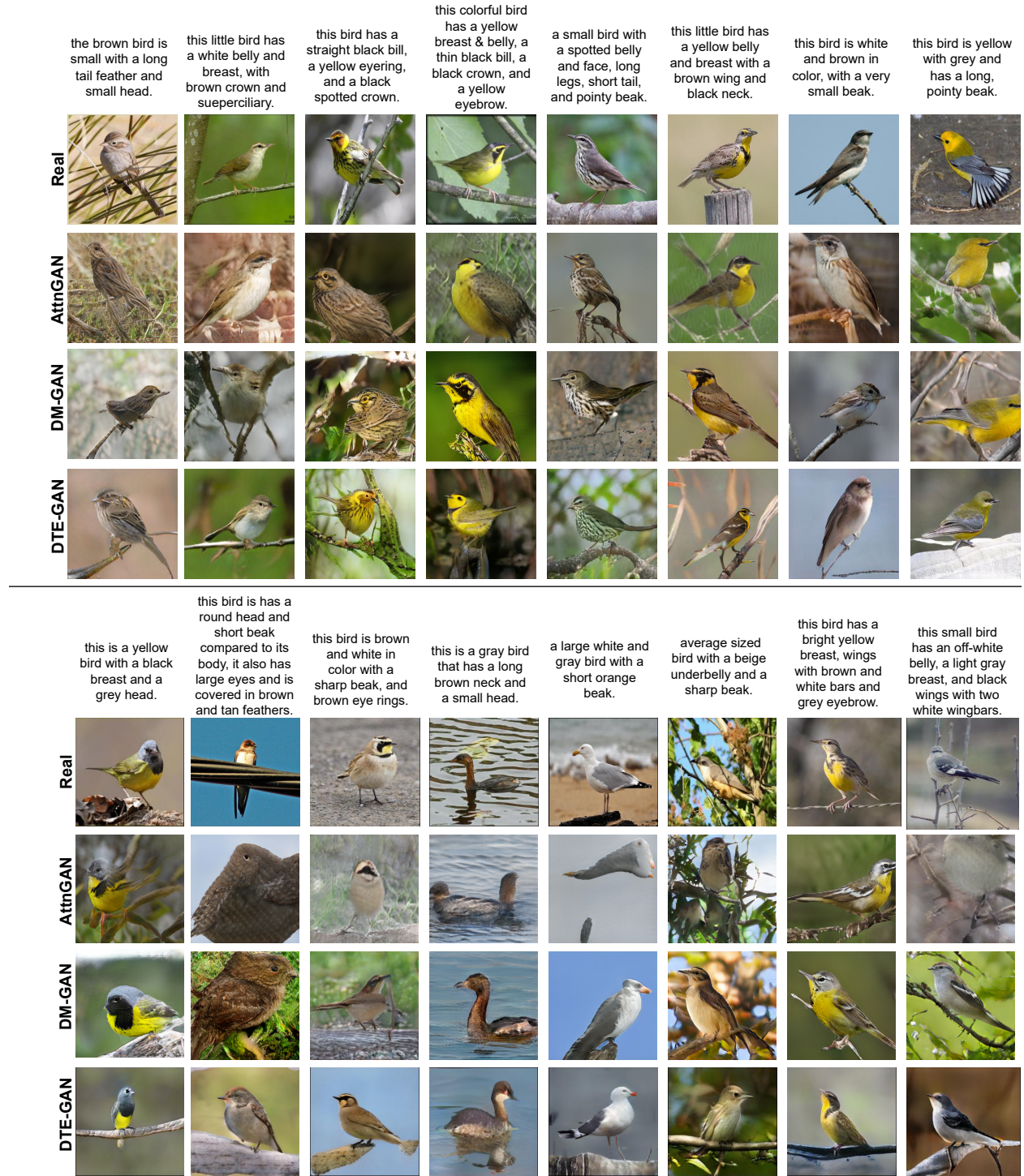


Figure 2: Images generated by DTE-GAN (ours) on CUB dataset and compared with AttnGAN Xu et al. (2018) & DM-GAN Zhu et al. (2019).

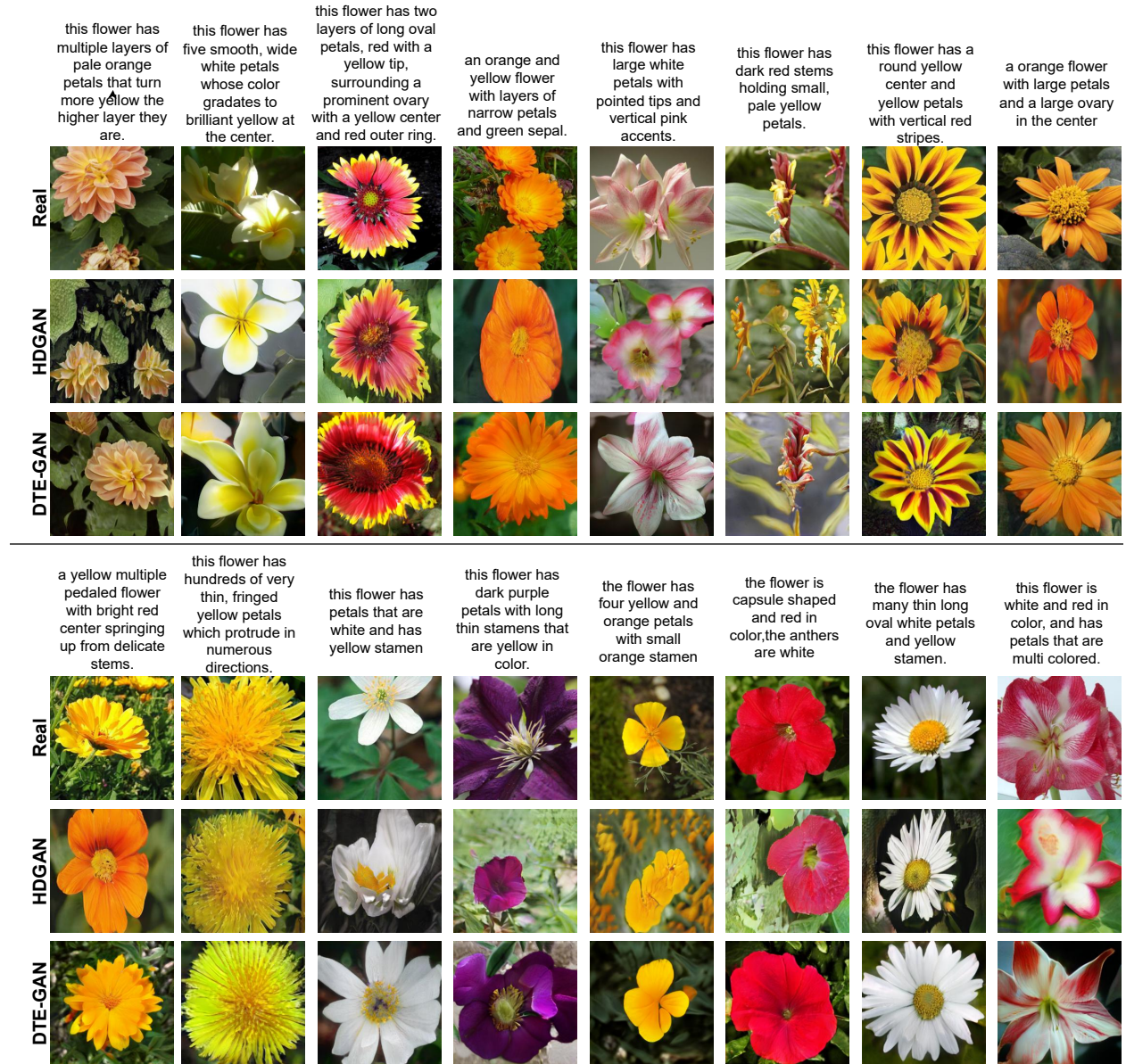


Figure 3: Images generated by DTE-GAN (ours) on Oxford-102 dataset and compared with HDGAN Zhang et al. (2018).

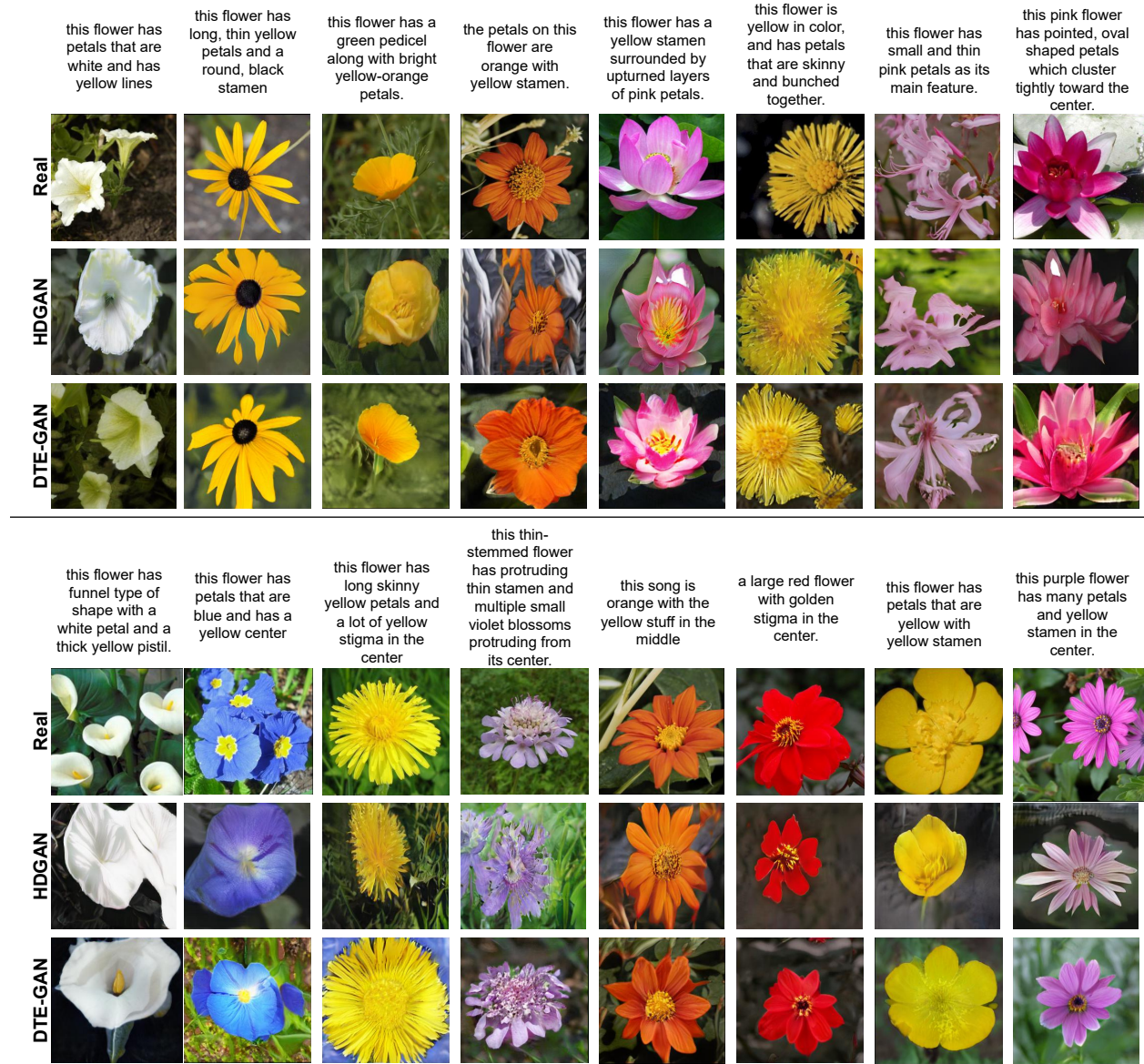


Figure 4: Images generated by DTE-GAN (ours) on Oxford-102 dataset and compared with HDGAN Zhang et al. (2018).

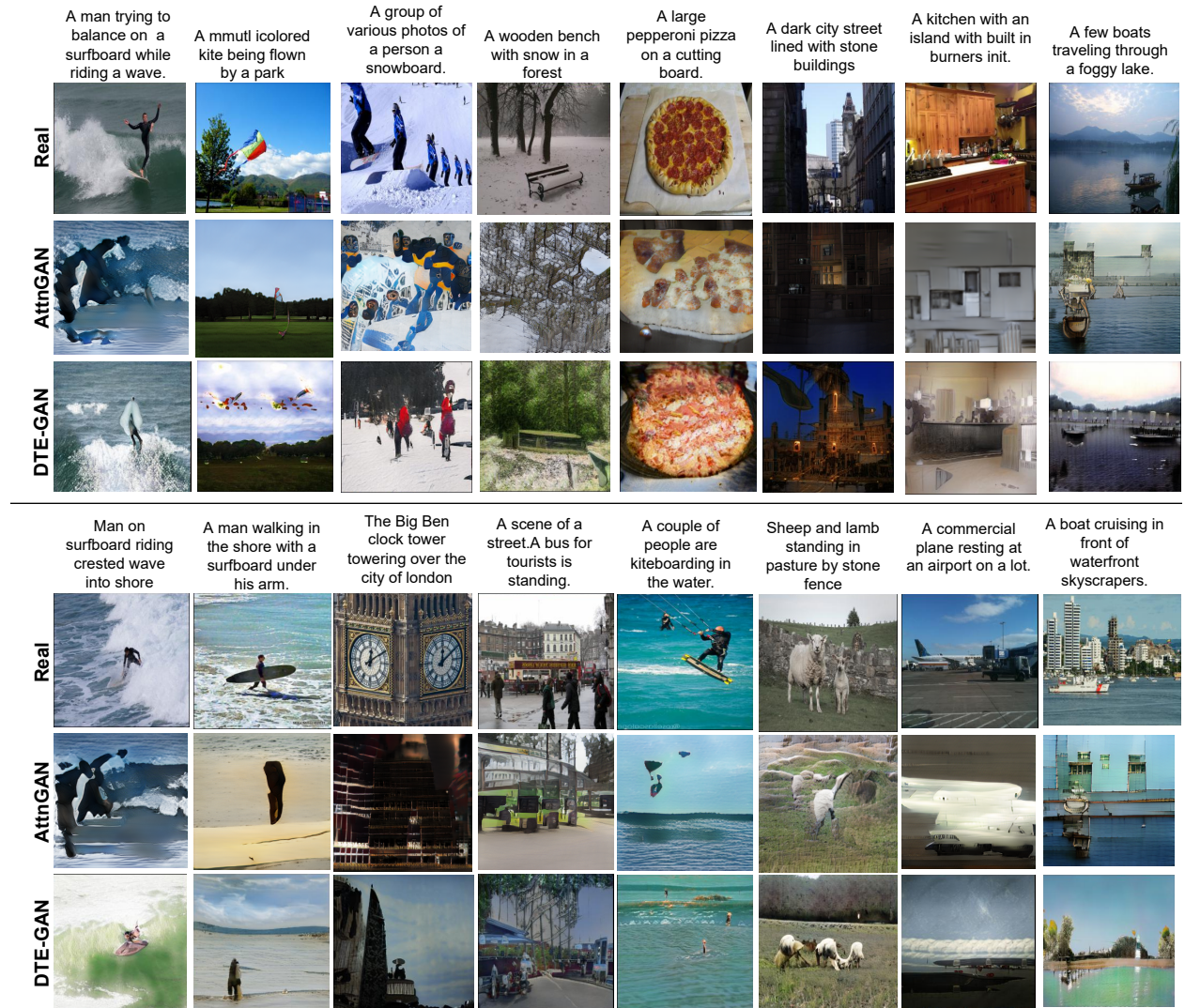


Figure 5: Images generated by DTE-GAN (ours) on COCO dataset and compared with AttnGAN Xu et al. (2018).

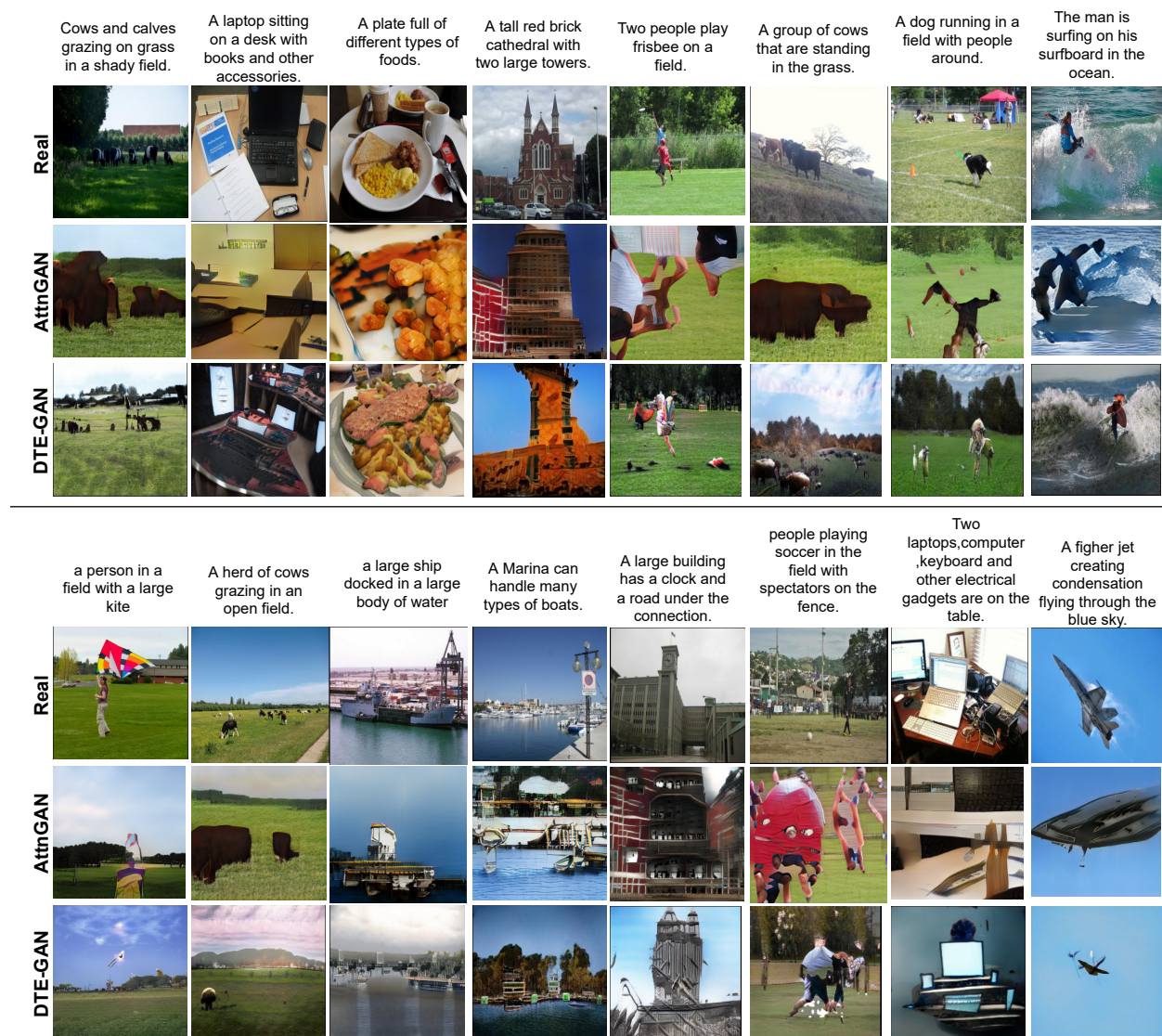


Figure 6: Images generated by DTE-GAN (ours) on COCO dataset and compared with AttnGAN Xu et al. (2018).



Figure 7: Examples of manipulated images generated by LightWeight GAN Li et al. (2020) using DTE-GAN pre-trained embeddings on CUB dataset. Source images (row 1) are manipulated by the text from concept images (row 2) and results in manipulated images (row 3).



Figure 8: Examples of manipulated images generated by Light Weight GAN Li et al. (2020) using DTE-GAN pre-trained embeddings on CUB dataset. Source images (row 1) are manipulated by the text from concept images (row 2) and results in manipulated images (row 3).



Figure 9: Random images generated on CUB dataset. (best viewed in full zoom and color)

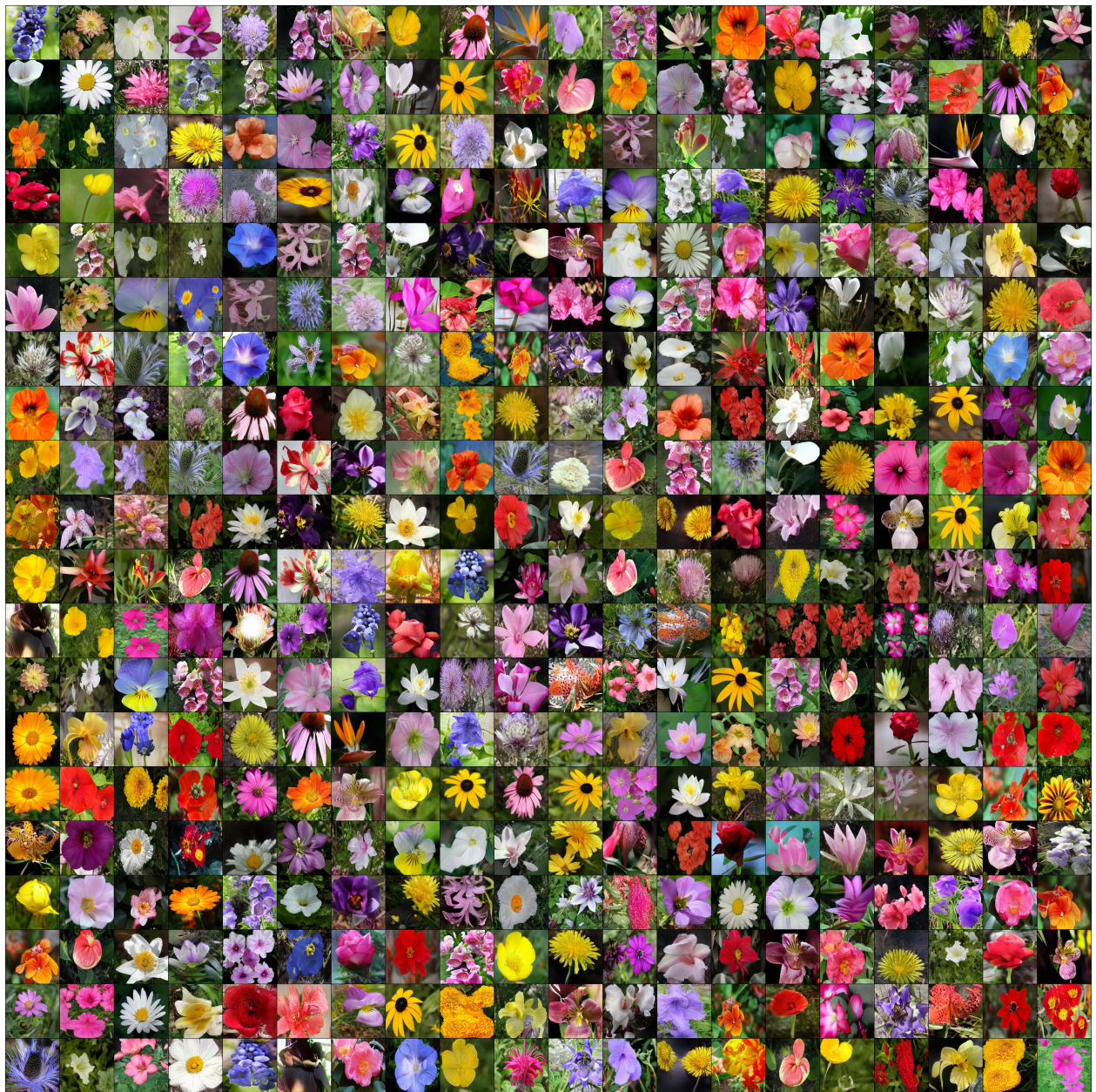


Figure 10: Random images generated on Oxford-102 dataset. (best viewed in full zoom and color)

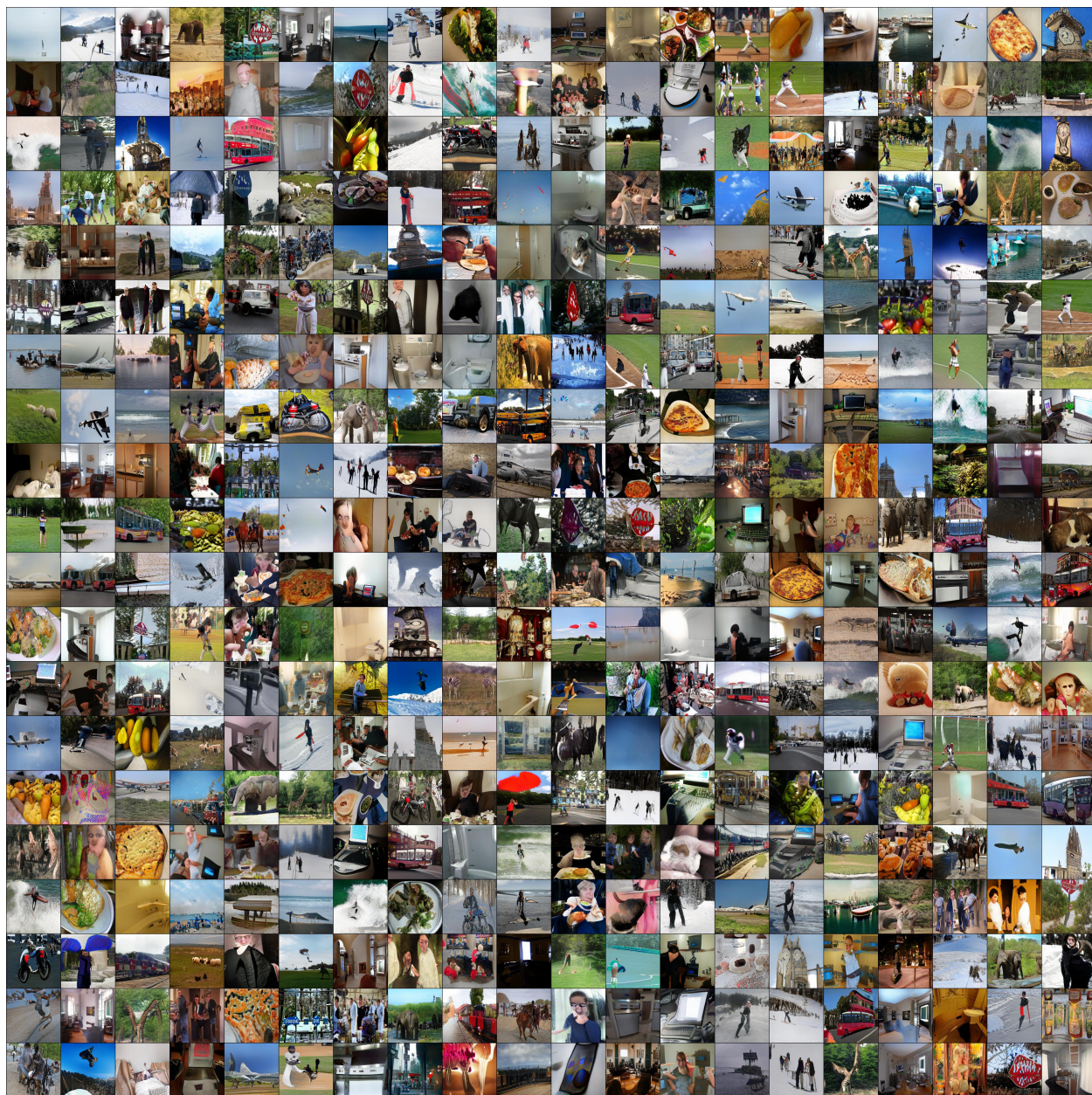


Figure 11: Random images generated on COCO dataset. (best viewed in full zoom and color)

References

- Bowen Li, Xiaojuan Qi, Philip Torr, and Thomas Lukasiewicz. Lightweight generative adversarial networks for text-guided image manipulation. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, pp. 22020–22031. Curran Associates, Inc., 2020.
- Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (eds.), *Computer Vision – ECCV 2014*, 2014.
- Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *Indian Conference on Computer Vision, Graphics and Image Processing*, Dec 2008.
- P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona. Caltech-UCSD Birds 200. Technical Report CNS-TR-2010-001, California Institute of Technology, 2010.
- Tao Xu, Pengchuan Zhang, Qiuyuan Huang, Han Zhang, Zhe Gan, Xiaolei Huang, and Xiaodong He. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- Zizhao Zhang, Yuanpu Xie, and Lin Yang. Photographic text-to-image synthesis with a hierarchically-nested adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. Dm-gan: Dynamic memory generative adversarial networks for text-to-image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.