MirrorGAN: Learning Text-to-image Generation by Redescription

Tingting Qiao[1,3], Jing Zhang[2,3,*], Duanqing Xu[1,*], and Dacheng Tao[3]

[1]College of Computer Science and Technology, Zhejiang University, China
[2]School of Automation, Hangzhou Dianzi University, China
[3]UBTECH Sydney AI Centre, School of Computer Science, FEIT, The University of Sydney, Australia
qiaott@zju.edu.cn, jing.zhang@uts.edu.au, xdq@zju.edu.cn, dacheng.tao@sydney.edu.au
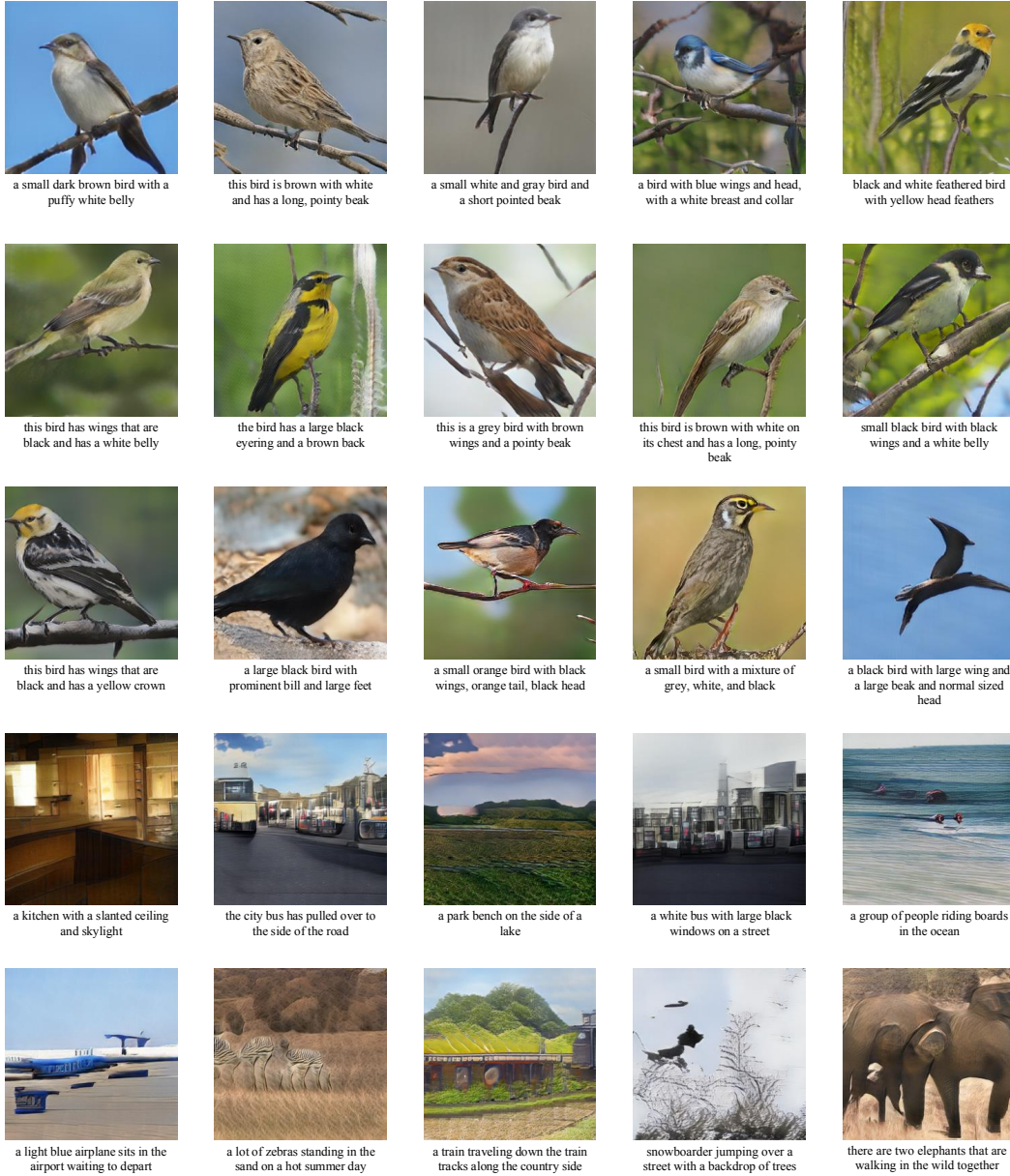
Figure 1: Examples of images generated by MirrorGAN conditioned on text descriptions from CUB and COCO test sets (the supplymentary material for Figure 3). As mentioned in [33,37], the COCO dataset is known to be much more challenging than the CUB dataset as it consists of images with more complex scenarios and objects.
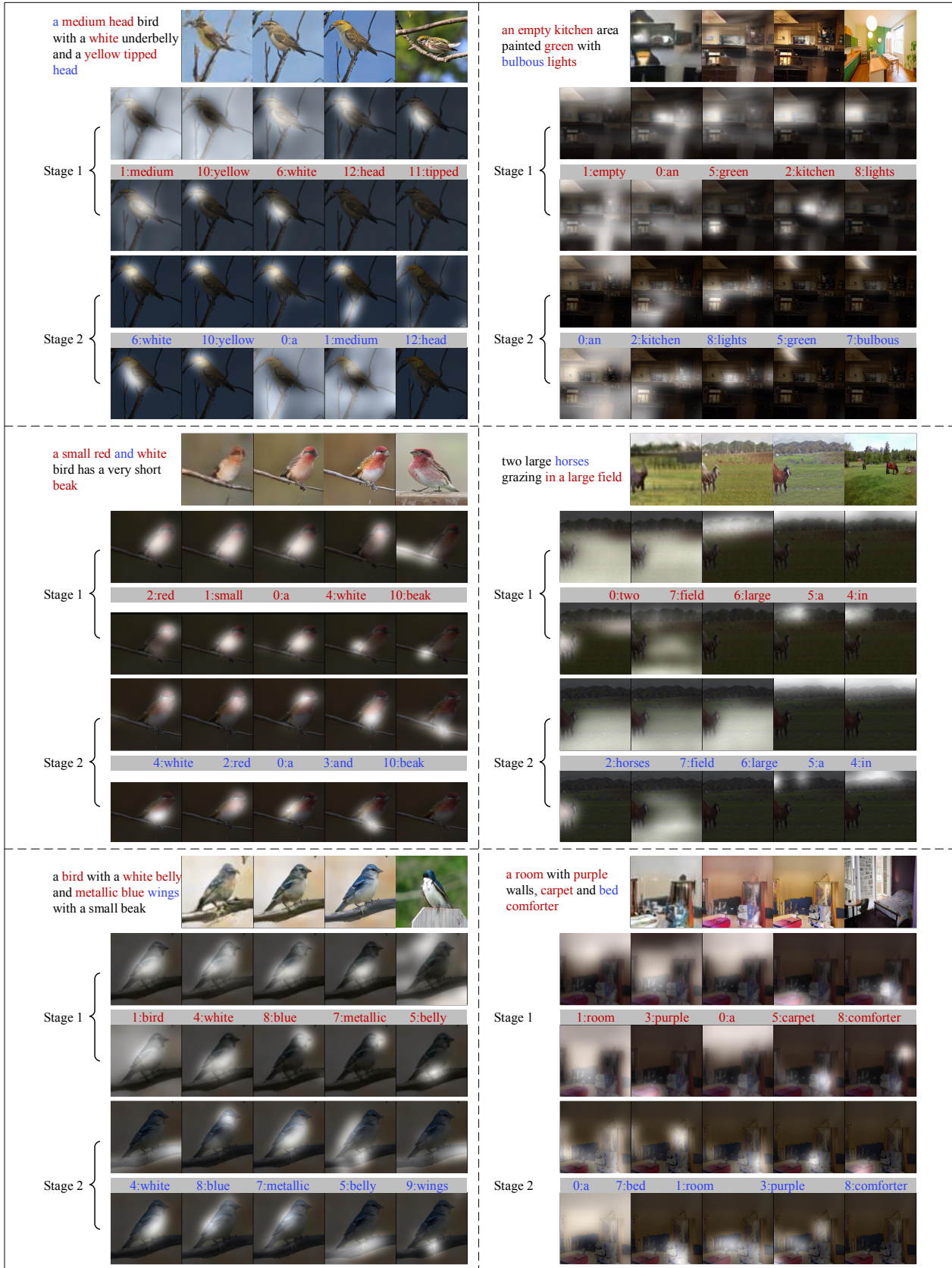
Figure 2: Attention visualization on the CUB and the COCO test sets (the supplymentary material for Figure 5).
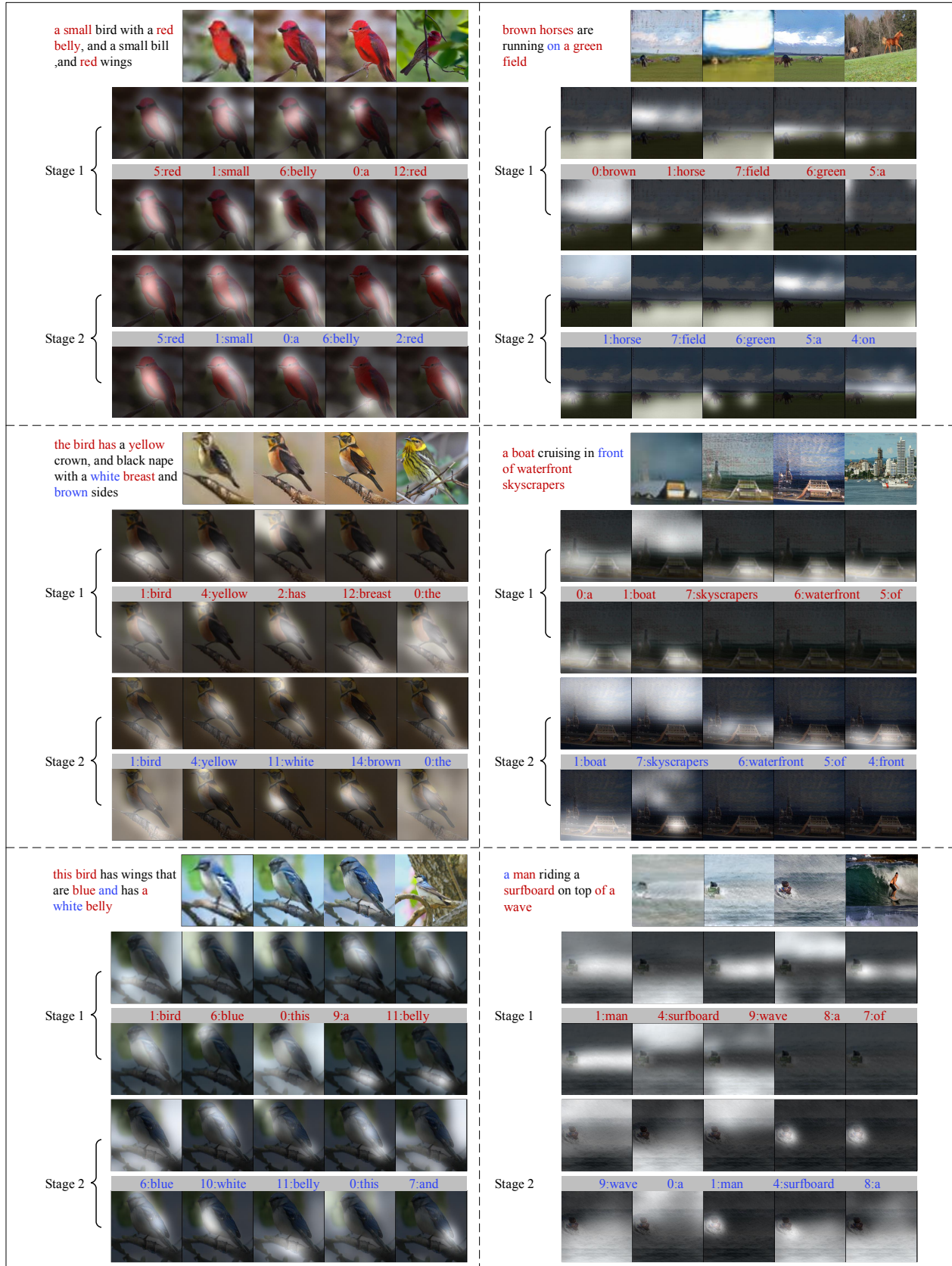
Figure 3: Attention visualization on the CUB and the COCO test sets (the supplymentary material for Figure 5).
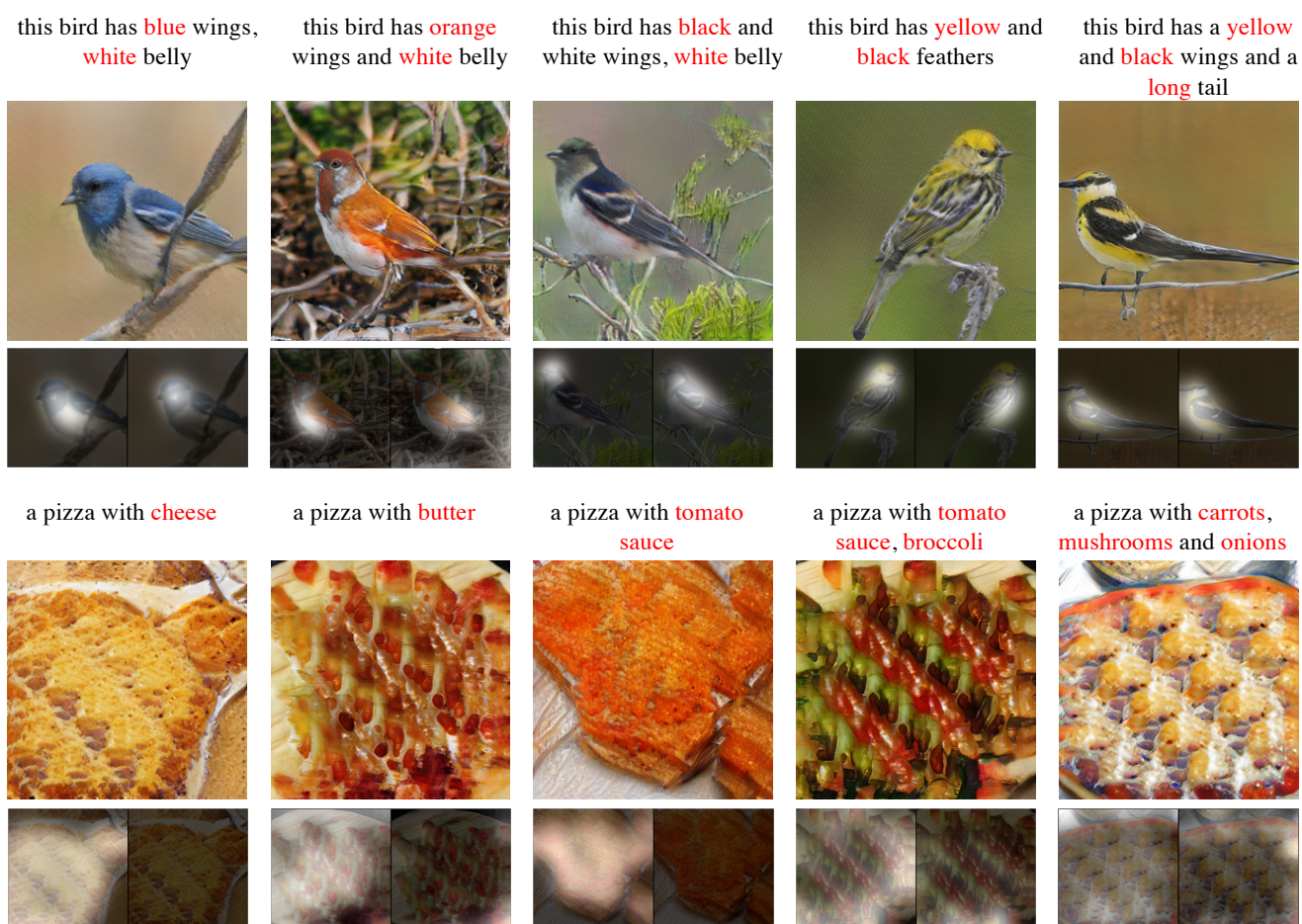
Figure 4: Images generated by MirrorGAN by modifying some words of the text descriptions and the corresponding top-2 attention maps in the last stage (the supplementary material for Figure 6). These examples show again that the COCO dataset is a very challenging dataset, however, MirrorGAN still captures the main textual feature of the input text and the generated images keep the semantic similarity.