NÜWA: Visual Synthesis Pre-training for Neural visUal World creAtion

Chenfei Wu[1*] Jian Liang[2*] Lei Ji[1] Fan Yang[1] Yuejian Fang[2†]
Daxin Jiang[1] Nan Duan[1†]

[1] Microsoft Research Asia [2] Peking University

Fig. 1: More samples of Text-to-Image (T2I) task generated by NÜWA.

* Both authors contributed equally to this research. † Corresponding author.

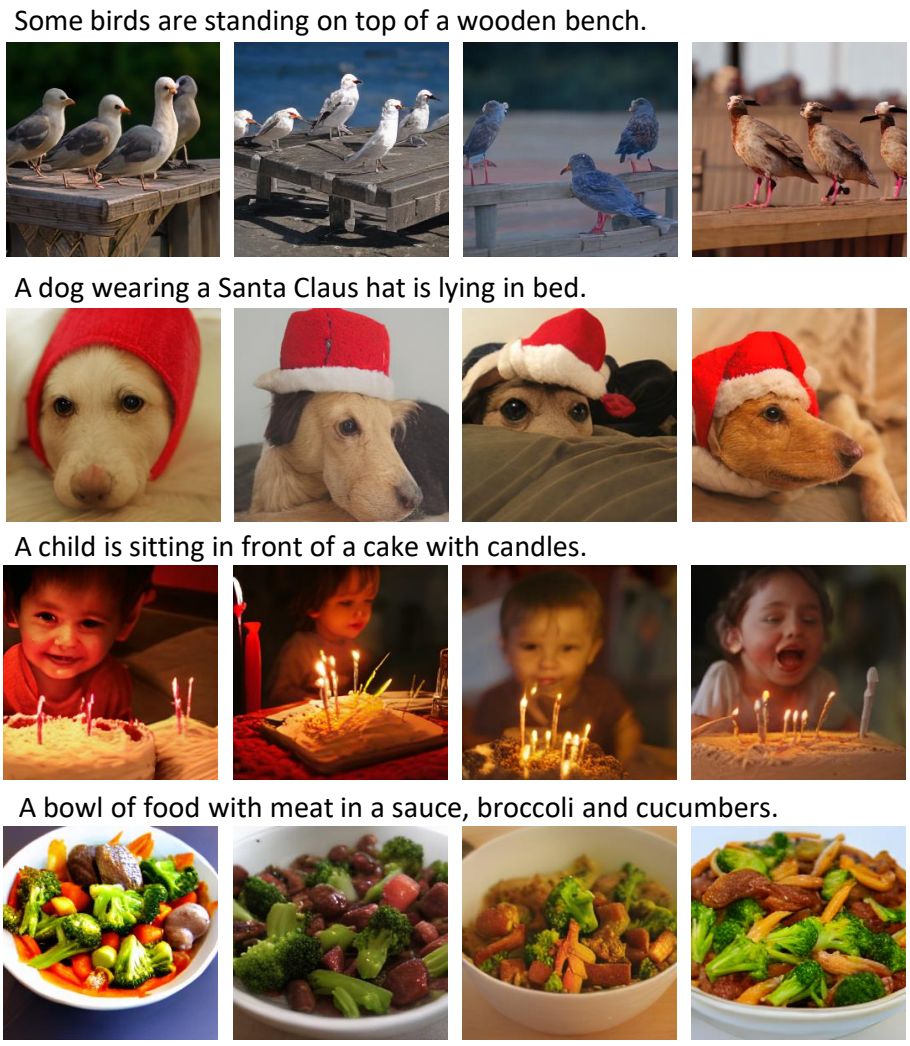

Fig. 2: More samples of Text-to-Image (T2I) task generated by NÜWA.

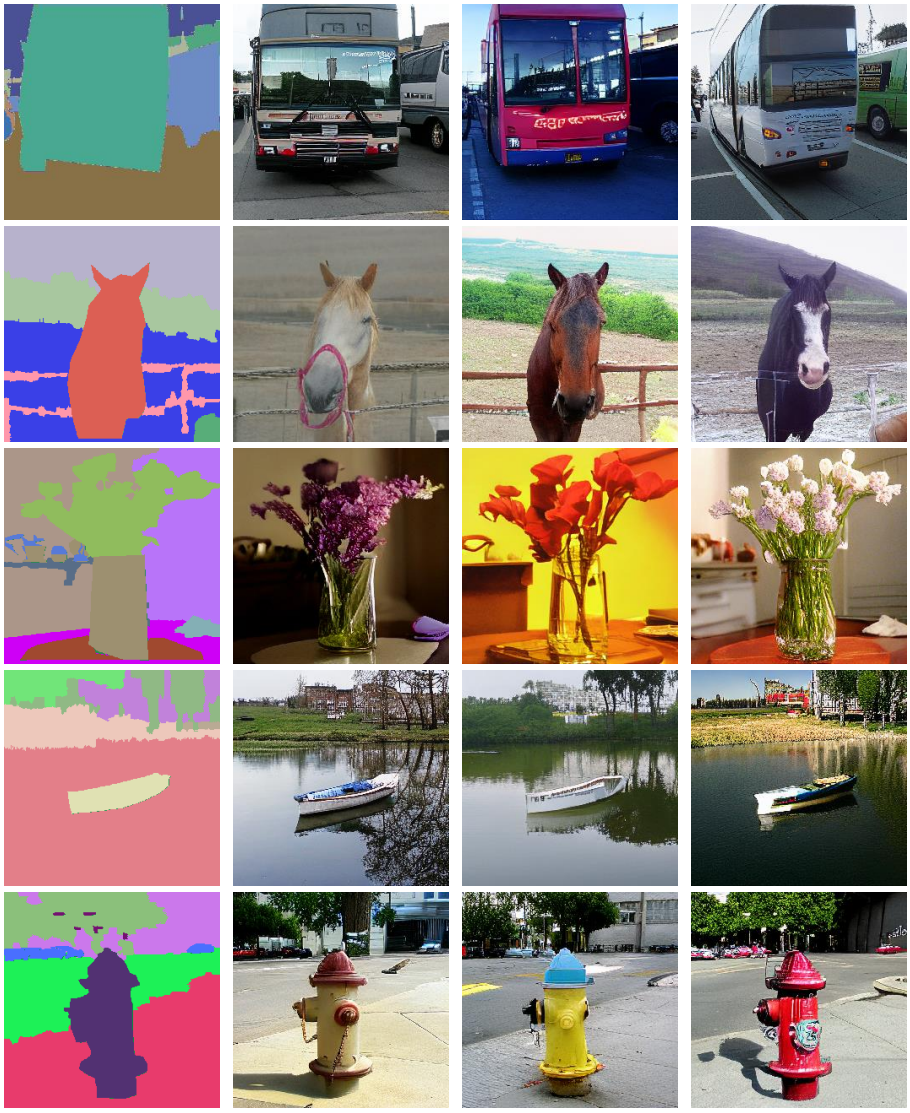

Fig. 3: More samples of Sketch-to-Image (S2I) task generated by NÜWA.

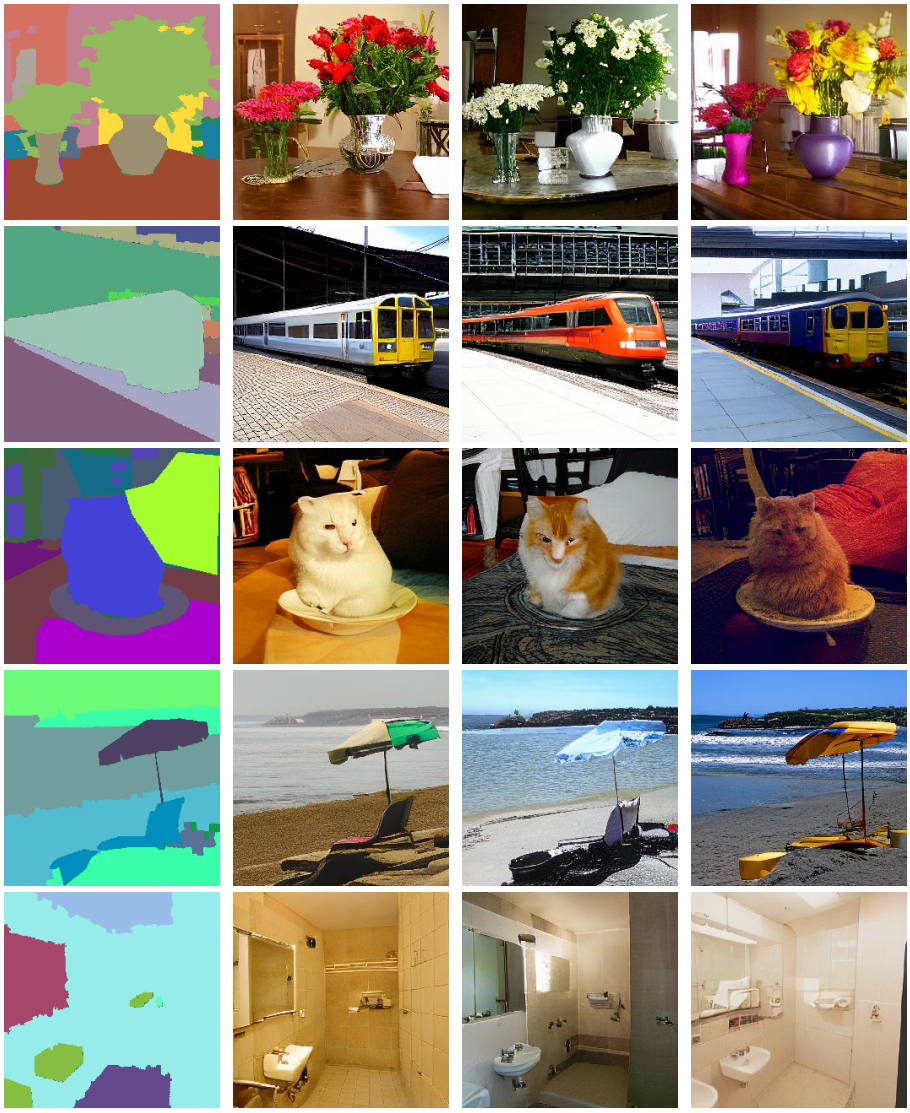

Fig. 4: More samples of Sketch-to-Image (S2I) task generated by NÜWA.

Fig. 5: More samples of the Image Completion (I2I) task generated by NÜWA.

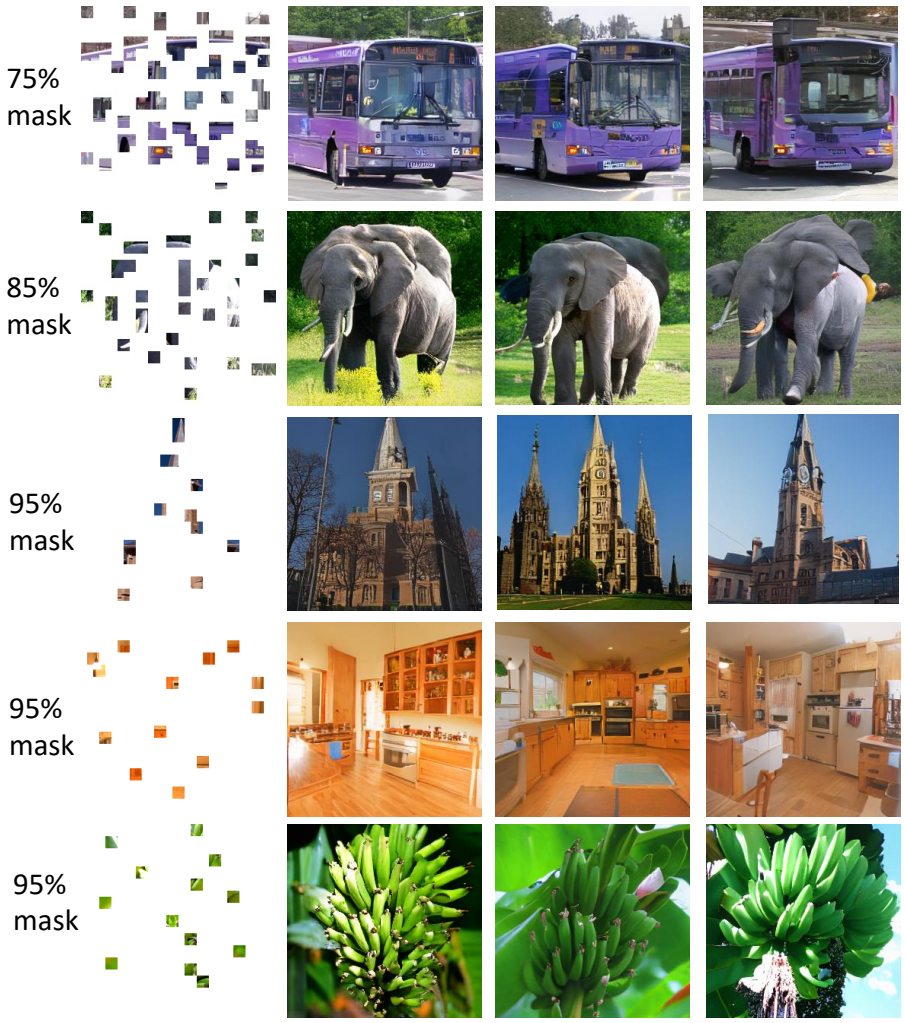

Fig. 6: More samples of the Image Completion (I2I) task generated by NÜWA.

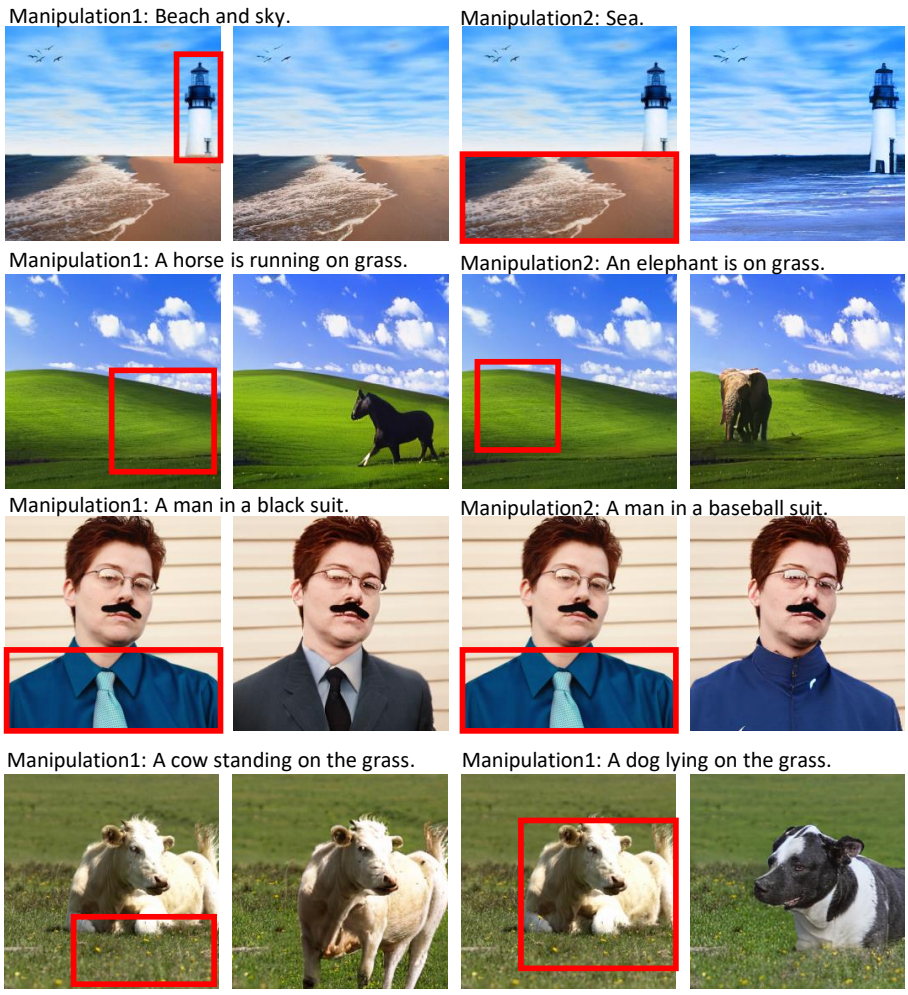

Fig. 7: More samples of the Text-Guided Image Manipulation(TI2I) task generated by NÜWA.

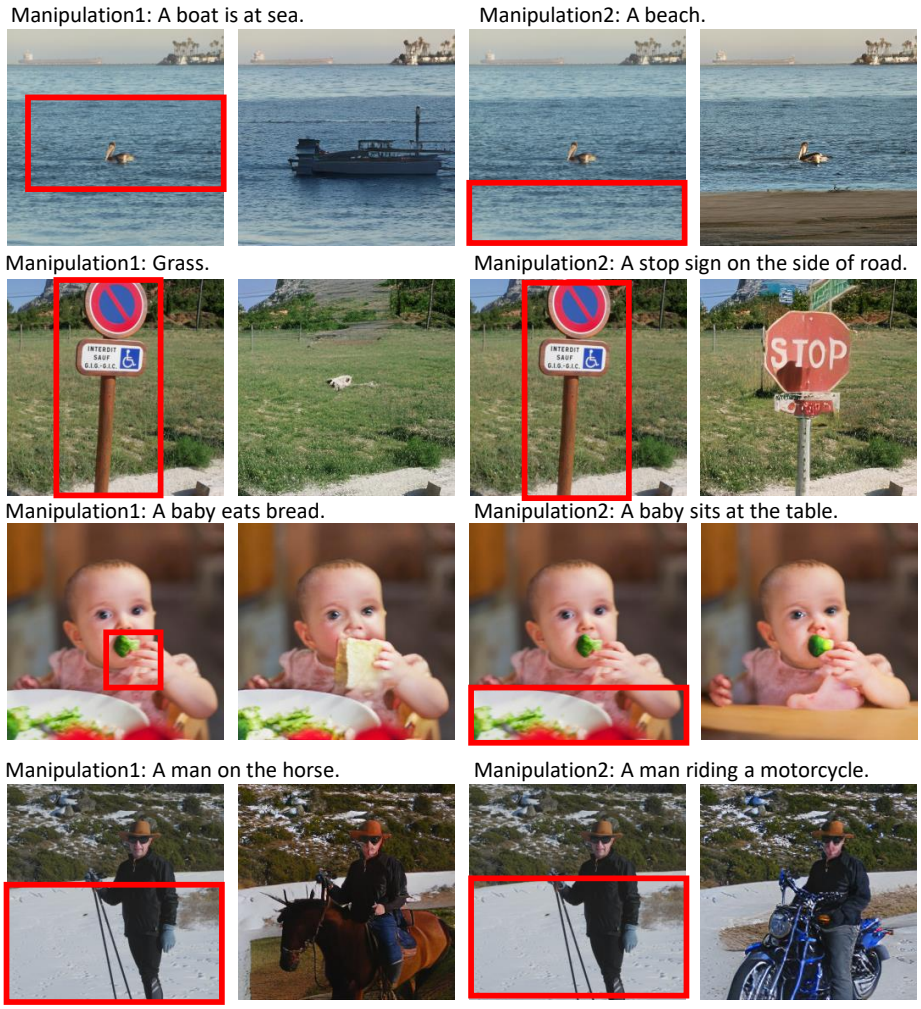

Fig. 8: More samples of the Text-Guided Image Manipulation(TI2I) task generated by NÜWA.

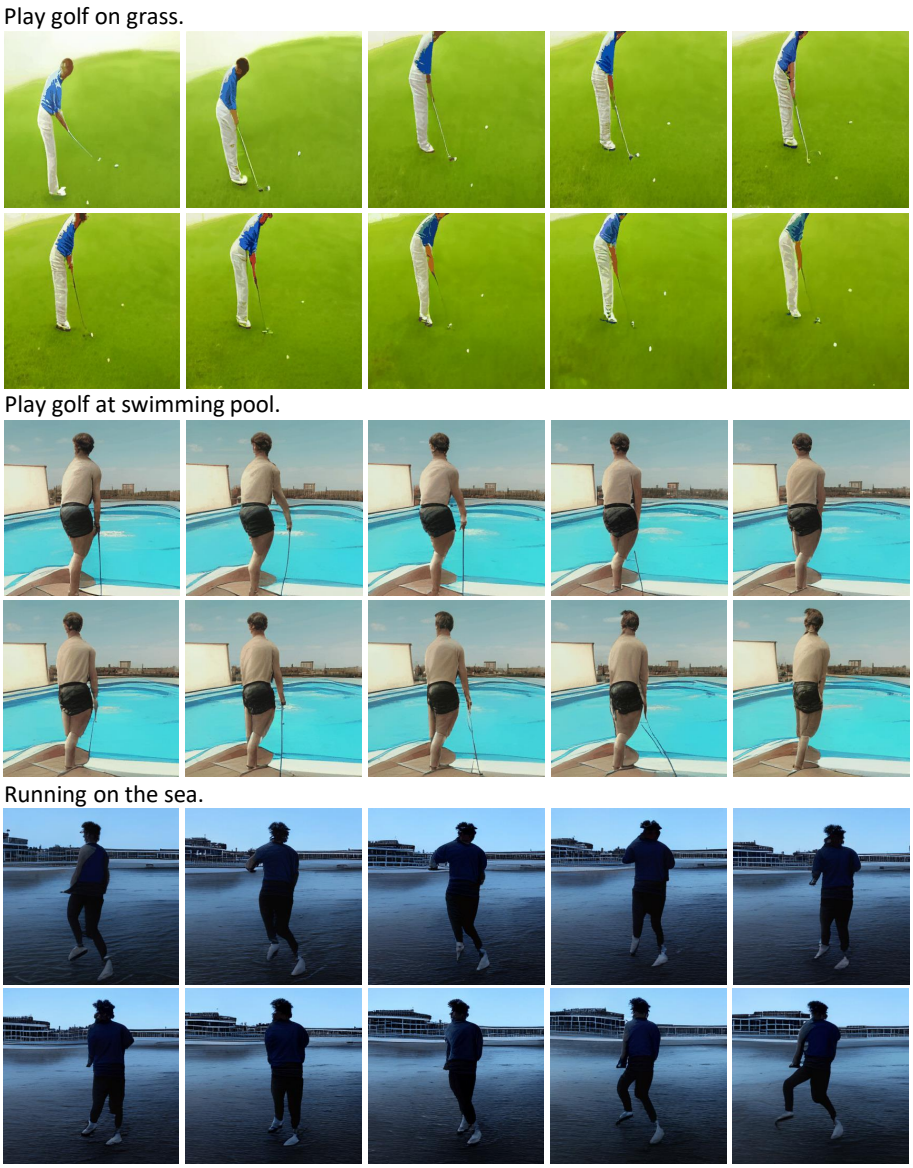

Fig. 9: More samples of the Text-to-Video (T2V) task generated by NÜWA.

Fig. 10: More samples of the Text-to-Video (T2V) task generated by NÜWA.

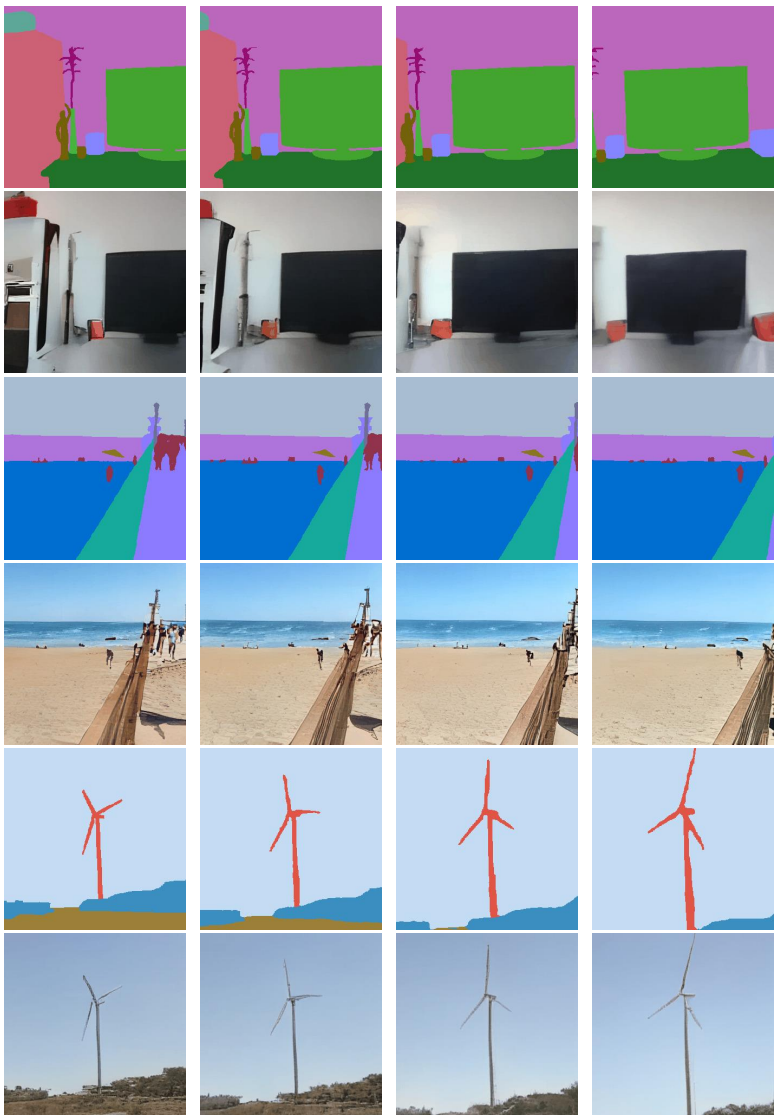

Fig. 11: More samples of Sketch-to-Video (S2V) task generated by NÜWA.

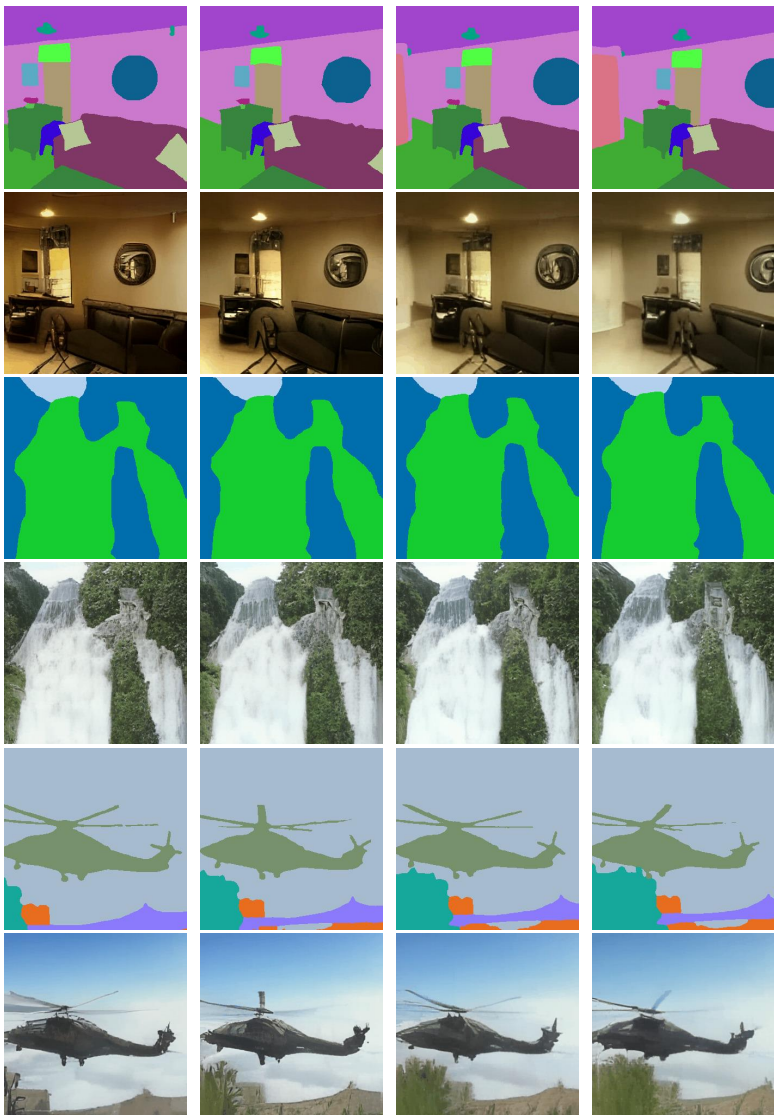

Fig. 12: More samples of Sketch-to-Video (S2V) task generated by NÜWA.

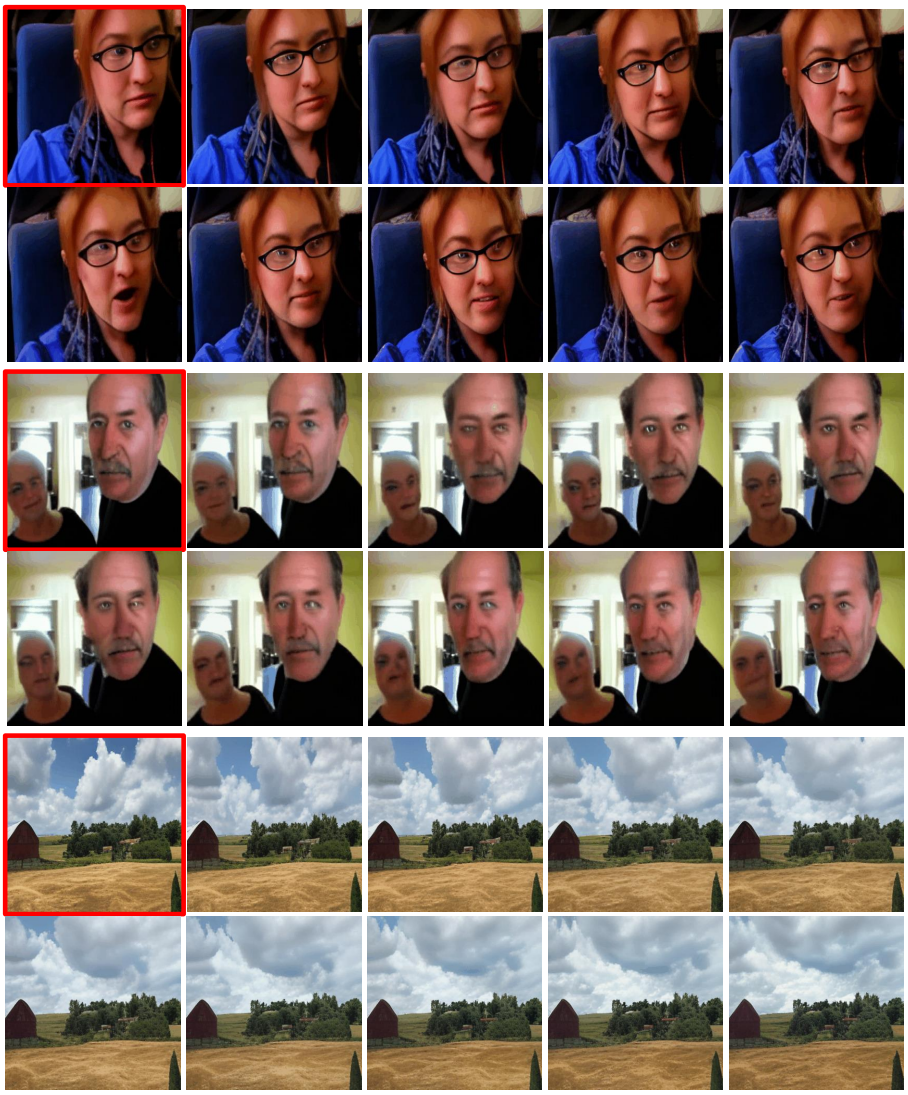

Fig. 13: More samples of the Video Prediction (V2V) task generated by NÜWA. Only one frame (see red box) is used as condition.

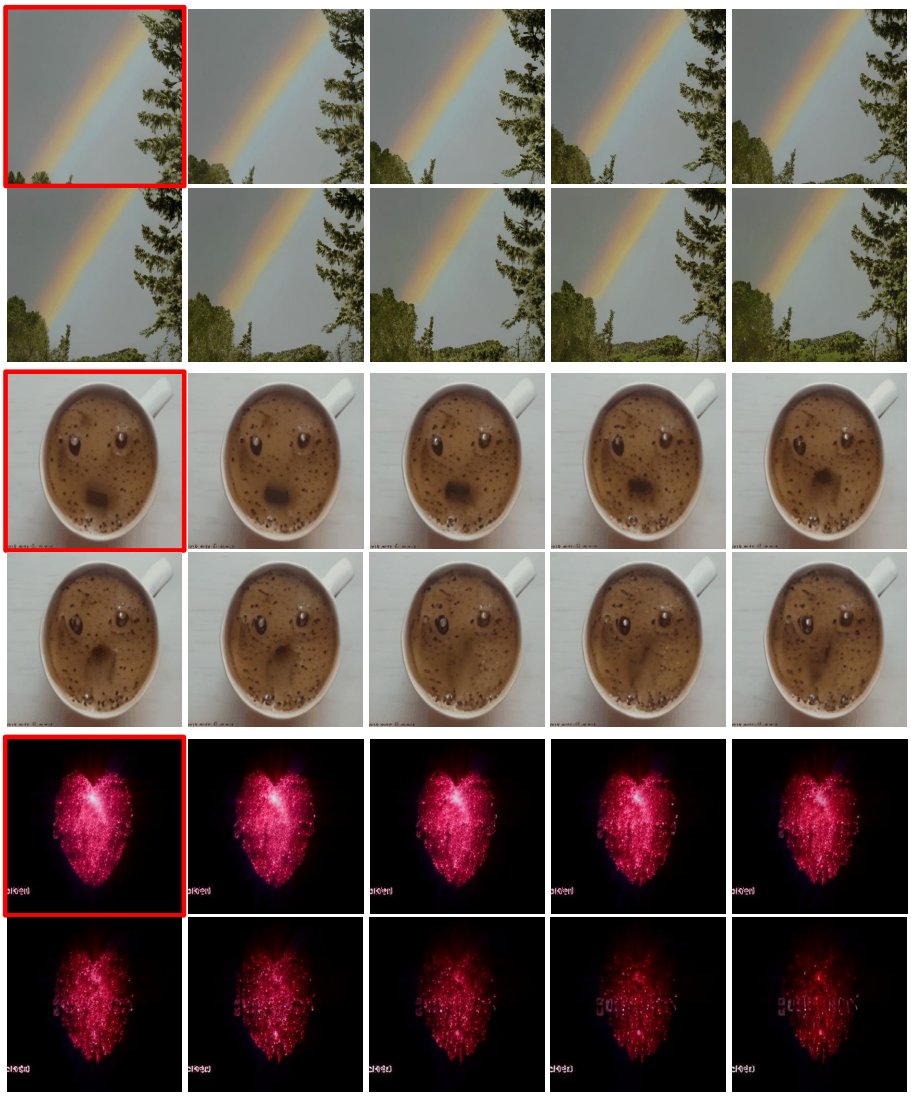

Fig. 14: More samples of the Video Prediction (V2V) task generated by NÜWA.Only one frame (see red box) is used as condition.

Fig. 15: More samples of Text-Guided Video Manipulation (TV2V) task generated by NÜWA.

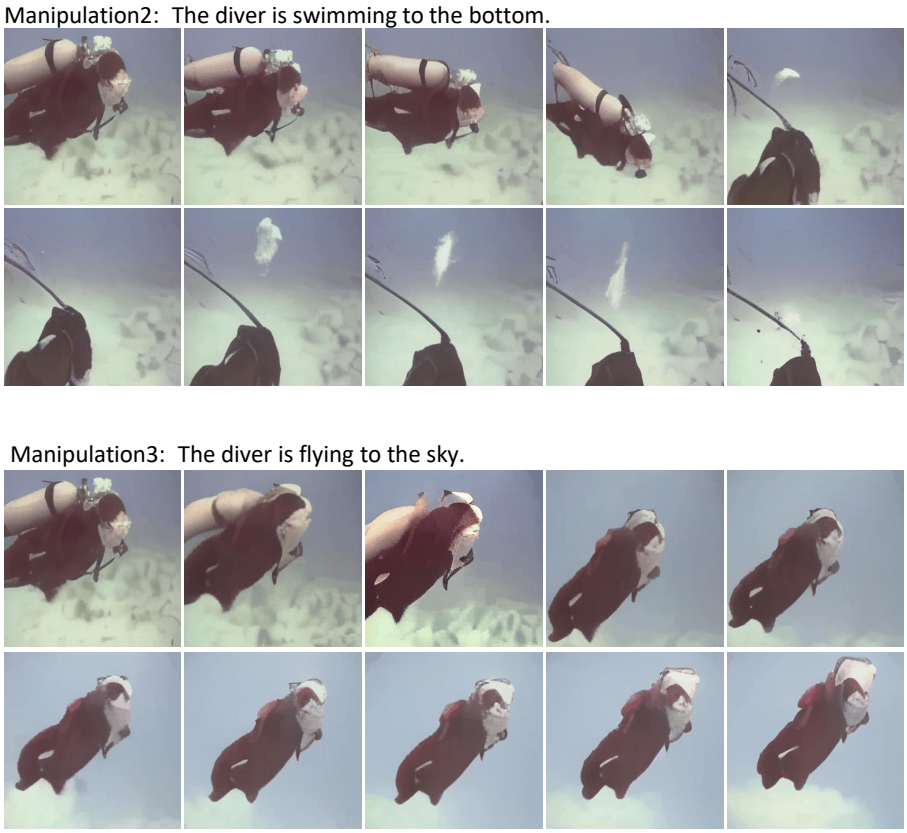

Fig. 16: More samples of Text-Guided Video Manipulation (TV2V) task generated by NÜWA.