Language-driven diversified Image Retargeting

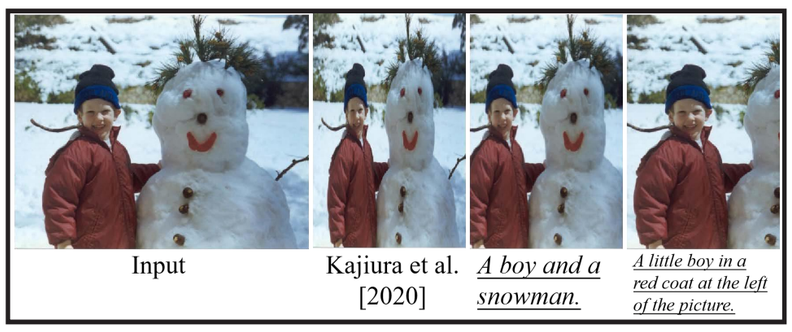
Rui Wang[1], Nisha Huang[2], FanTang[3], Weiming Dong[2], Tong-Yee Lee[4]

[1]Jilin University, [2]School of Artificial Intelligence, UCAS, [3]ICT, Chinese Academy of Sciences,
[4]National Cheng Kung University

SIGGRAPH ASIA 2022 DAEGU

Problem

Existing retargeting approaches focus on improving the output quality in terms of visual appearance and similarity preservation with the input image but neglect the personal control of the results.



Related Work

Liu et al. [1] uses semantic components as supplements of the inputs to guide the retargeting process.

Kajiura et al.[2]. utilize the resinforcement method to retarget original image.

Different previous retargeting works, we utilize texts as descriptions for the output images. By using different texts, we can acquire different corresponding image contents.

ACKNOWLEDGMENTS

This research was supported in part by National Natural Science Foundation of China under Nos. 62102162, 61832016, by Ministry of Science and Technology, Taiwan (no. 111-2221-E-006-112-MY3) and by Open Projects Program of NLPR.

References

[1]. Si Liu, Zhen Wei, Yao Sun, Xinyu Ou, Junyu Lin, Bin Liu, and Ming-Hsuan Yang. 2018. Composing Semantic Collage for Image Retargeting. IEEE Transactions on Image Processing 27, 10 (2018), 5032–5043.

[2].Nobukatsu Kajiura, Satoshi Kosugi, Xueting Wang, and Toshihiko Yamasaki. 2020. Self-Play Reinforcement Learning for Fast Image Retargeting. In Proceedings of the 28th ACM International Conference on Multimedia. Association for Computing Machinery, New York, NY, USA, 755–1763.

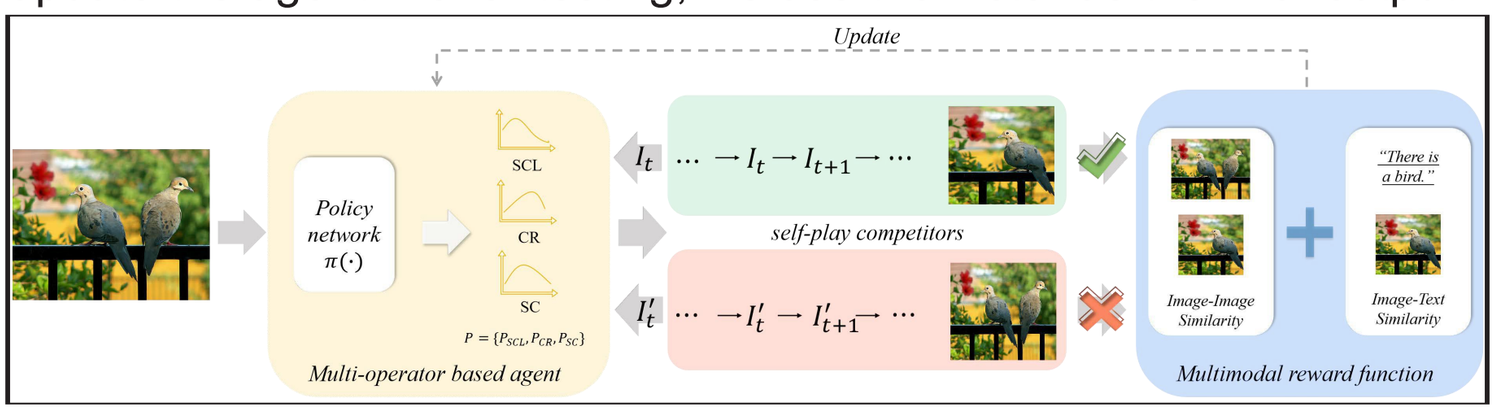
Method

We propose language-driven diversified image retargeting method (LDIR) based on a self-play pipeline. The agent works as a decision worker and interacts with the reward to select the operator step by step. The reward as a assessment to evaluate the selected sequence and update the agent.

For each step, the agent receives the retargeted image and text and calculates the possibilities of adopting different operators from action space to operate the retargeted image. The agent works tmax steps until the output ratio satisfies the target ratio.

The reward is consisted of BDW which is used to measure similarity of original and retargeted image and consine which measure the similarity of text and retargeted image.

To guarantee the selected operator sequence is optimal, we introduce the self-play procedure. The final retargeted results are judged by reward. When training, we return the victor's reward to update the agent. As for testing, we use the victor as the final output.



Results