# Accumulation Knowledge Distillation for Conditional GAN Compression Supplementary Material

Tingwei Gao[1] Rujiao Long[1]
[1]Alibaba Group
tingwei.gtw@alibaba-inc.com, rujiao.lrj@alibaba-inc.com

## Overview

The supplementary material is organized as follows. In Sec. 1, the additional analysis of ACKD is introduced. Further qualitative results of ACKD are provided in Sec. 2.

## 1. Additional Analysis of Accumulation Knowledge Distillation

**The guidance of deep-stage knowledge in ACK-A.** Further detailed analysis concerning the guidance of deep-stage knowledge is described as follows. The accumulation knowledge serves two key purposes. The first is to help the student generator to learn the finer-grained structure, texture and detailed features with lower abstraction from several shallow stages of the teacher generator. The second is to enable the student generator to better learn the knowledge with higher abstraction from the corresponding deep stage of the teacher generator. Accordingly, for the second purpose, more targeted and effective attention operations can be achieved if the correlation between the accumulation knowledge and its corresponding deep-stage knowledge is constructed. In the cross attention module, the query is provided by the deep-stage knowledge, while the key and value are provided by the accumulation knowledge, then based on the dot-product attention, the correlation between the deep-stage knowledge and the corresponding accumulation knowledge can be constructed. In conclusion, by establishing the correlation, the student generator can achieve a more comprehensive, effective and accurate understanding of the knowledge being taught.

**The residual connection from deep-stage knowledge in ACK-F.** Further detailed analysis and ablation study concerning ACK-F is provided as follows. Since the learning focus of each layer evolves with the increase in the depth of the network, the deep-stage knowledge may be of greater significance to the corresponding stage with the same depth in the student generator, while the accumulation knowledge is an effective complement to the deep-stage knowledge. As such, a residual connection from deep-stage knowledge is introduced in ACK-F. Based on such design, the output of the conv block in ACK-F is the complement to the deep-stage knowledge. The ablation study on the edges→shoes is carried out to verify the above analysis and the experimental results can be found in Tab. 1. An observation can be made that removing the residual connection from deep-stage knowledge in ACK-F leads to a significant decrease in the performance of generator.

Table 1. Ablation study about the residual connection in ACK-F for self-distillation and compressing parameters 82.5× on the edges→shoes.

| Method | Self-Distillation | Compression |
|---|---|---|
| w/o res-connection | 19.63 | 23.67 |
| Ours | 18.47 | 22.10 |

## 2. Additional Qualitative Results

The additional visual results are provided. The comparison methods include the original teacher models provided by [2], GAN Compression [2], OMGD [3] and CAT [1]. The qualitative results on the horse→zebra are visualized in Fig. 1 and Fig. 2. The experimental results on the edges→shoes are illustrated in Fig. 3 and Fig. 4. The visual results on the map→aerial are shown in Fig. 5 and Fig. 6.

## References

[1] Qing Jin, Jian Ren, Oliver J Woodford, Jiazhuo Wang, Geng Yuan, Yanzhi Wang, and Sergey Tulyakov. Teachers do more than teach: Compressing image-to-image models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13600–13611, 2021.

[2] Muyang Li, Ji Lin, Yaoyao Ding, Zhijian Liu, Jun-Yan Zhu, and Song Han. Gan compression: Efficient architectures for interactive conditional gans. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5284–5294, 2020.

[3] Yuxi Ren, Jie Wu, Xuefeng Xiao, and Jianchao Yang. Online multi-granularity distillation for gan compression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6793–6803, 2021.

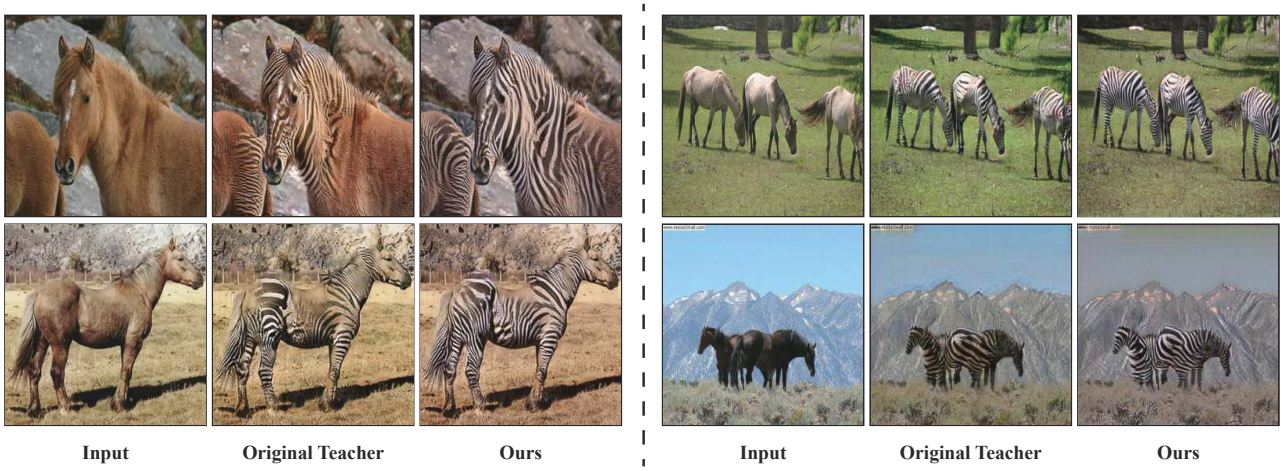


Figure 1. Qualitative comparisons of self-distillation using ACKD on the horse→zebra.

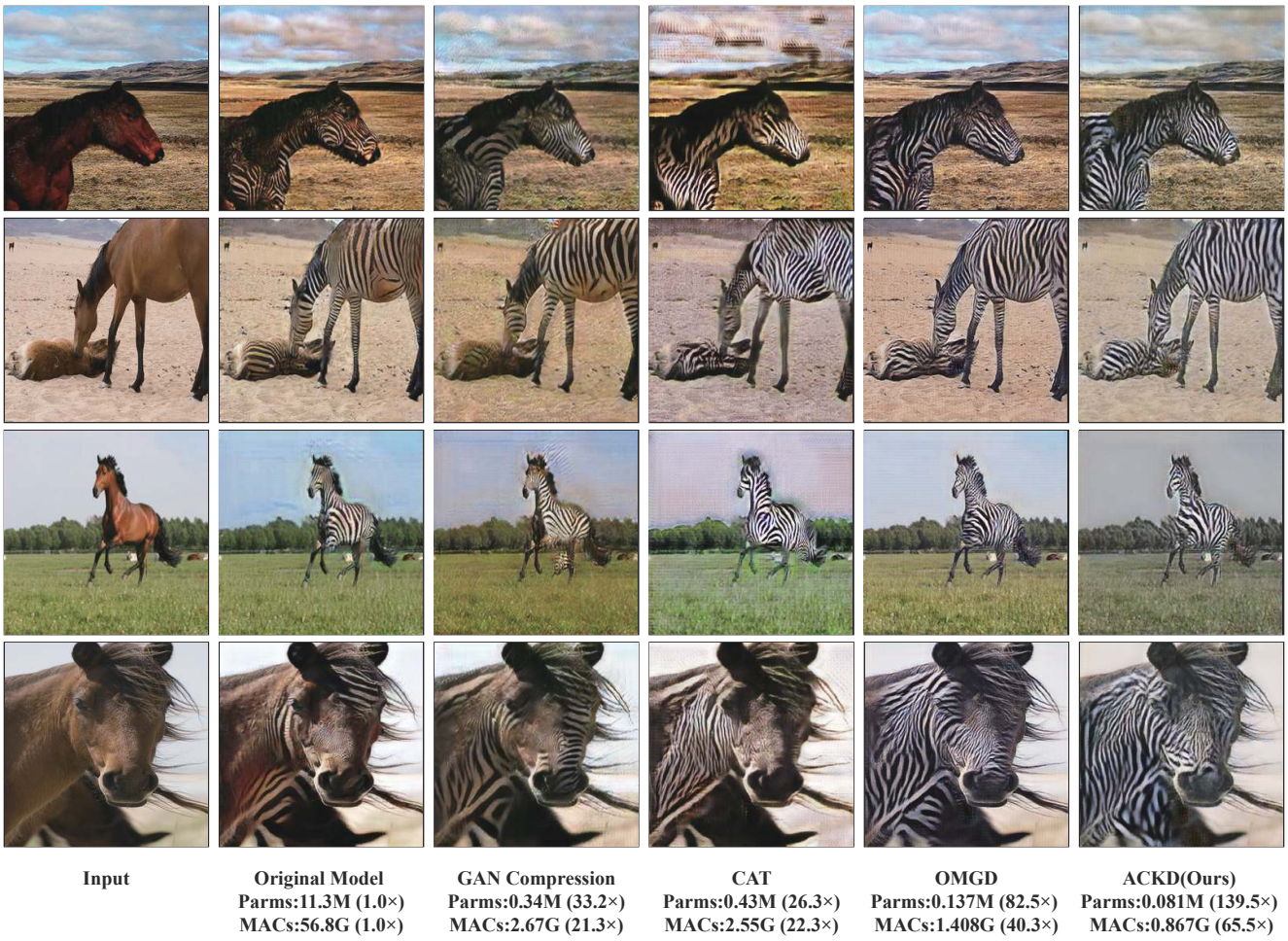


Figure 2. Qualitative comparisons with other cGANs compression methods on the horse→zebra.


Figure 3. Qualitative comparisons for self-distillation using ACKD on the edges→shoes.


Figure 4. Qualitative comparisons with other cGANs compression methods on the edges→shoes.

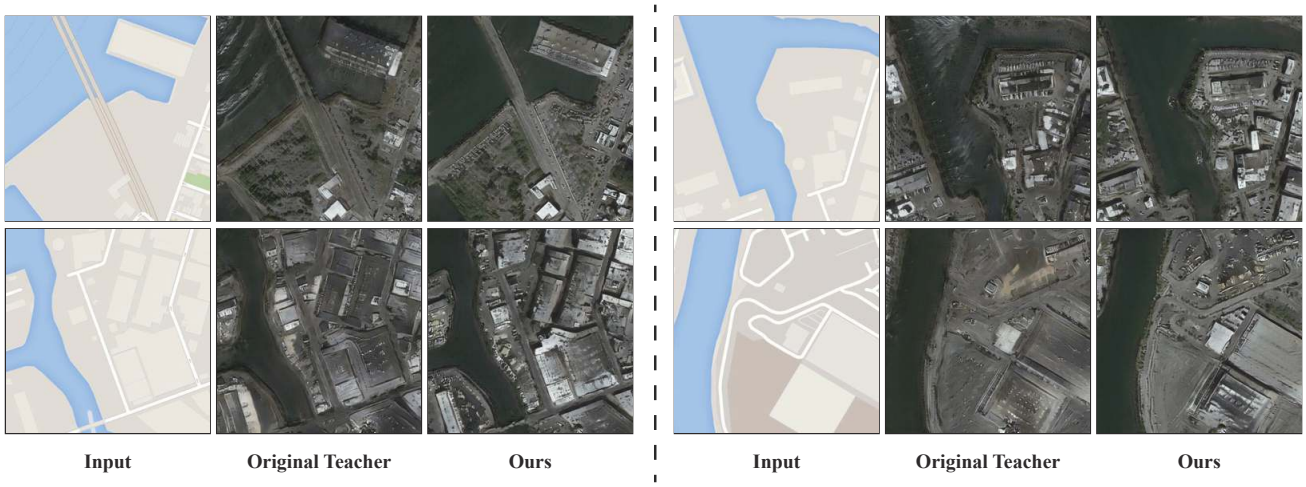


Figure 5. Qualitative comparisons for self-distillation using ACKD on the map→aerial.

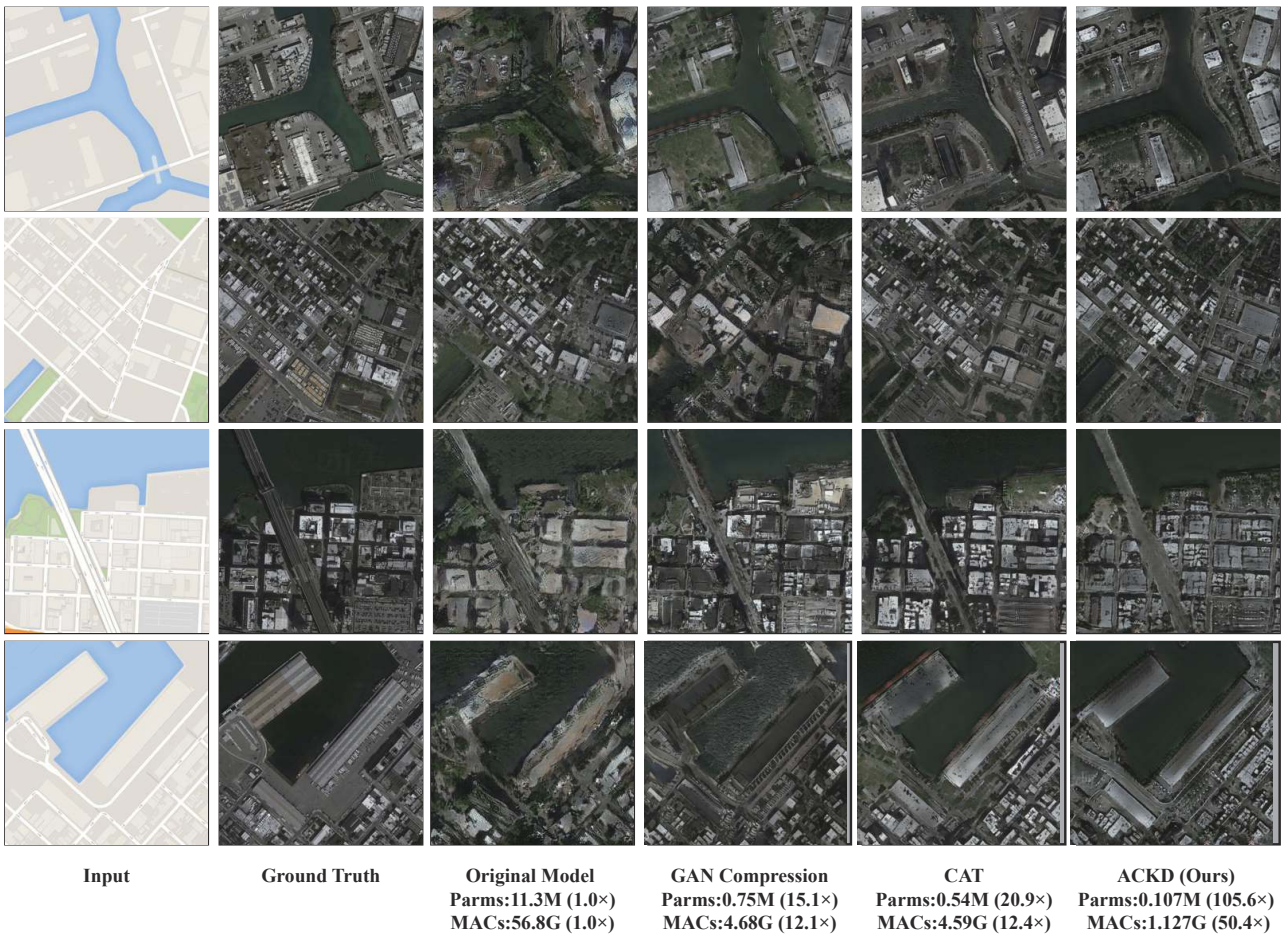


Figure 6. Qualitative comparisons with other cGANs compression methods on the map→aerial.