Appendix

A. Analysis on Training Time

In comparison to Ens, our SVRE introduces an extra loop, which brings additional costs for crafting adversarial examples. The complexity is proportional to the total number of queries, so SVRE is $(2M+n)/n$ times of Ens where M ($M = 16$) is the internal update frequency and n ($n = 4$) is the number of ensemble models. Note that other efforts for improving the adversarial transferability also introduce additional costs. *E.g.*, SIM [16] makes $m = 5$ copies of the input for querying, VT [31] samples $N = 20$ neighborhoods for variance tuning, and Admix [32] randomly sample $m_2 = 3$ images from other categories and copy each image for $m_1 = 5$ times. In the light of the improved performance, the additional time cost is acceptable.

B. SVRE with other Advanced Method

To show how SVRE compares to the state-of-the-art black-box adversarial attacks, we further incorporate SVRE with Admix [32], the most recent black-box attack method, and show how SVRE help promote its performance. Specifically, we use SVRE-Admix-TI-DIM and Ens-Admix-TI-DIM to generate adversarial examples, respectively, on the ensemble of Inc-v3, Inc-v4, IncRes-v2 and Res-152, and test on three adversarially trained models, while the Admix-TIM-DIM base method crafts adversarial examples on the Inc-v3 model.

The results are summarized in Table 4. We can see that the ensemble attack of Ens-Admix-TI-DIM has considerably higher transferability than Admix-TI-DIM, and our SVRE-Admix-TI-DIM further promotes the performance.

C. Visualization on Crafted Examples

In Figure 6, we visualize six randomly selected raw images and their corresponding adversarial examples crafted by Ens-MI-FGSM and SVRE-MI-FGSM, respectively. The adversarial examples are crafted on the ensemble of Inc-v3, Inc-v4, IncRes-v2 and Res-152 models. It shows that the adversarial examples crafted by our SVRE method are imperceptible to human eyes.

Table 4. The black-box attack success rates (%) against three adversarially trained models using Admix-TI-DIM as the base method.

Attack method	Inc-v3$_{ens3}$	Inc-v3$_{ens4}$	IncRes-v2$_{ens}$	Average
Admix-TI-DIM	80.8	78.9	63.2	74.3
Ens-Admix-TI-DIM	96.5	96.4	93.6	95.5
SVRE-Admix-TI-DIM	**98.6**	**98.3**	**96.8**	**97.9**

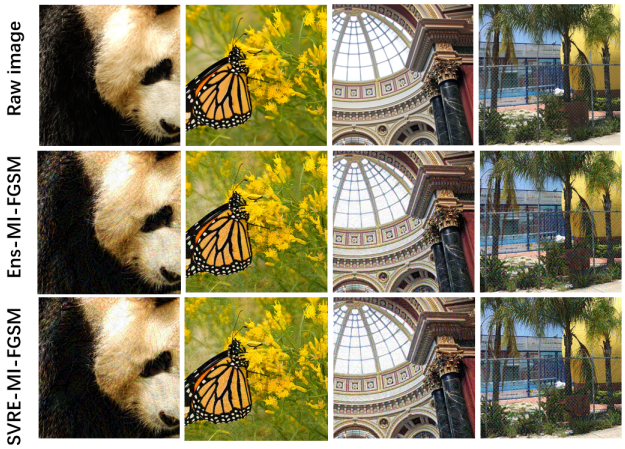

Figure 6. Adversarial examples generated by Ens-MI-FGSM and SVRE-MI-FGSM, respectively. The adversarial examples are crafted on an ensemble of Inc-v3, Inc-v4, IncRes-v2 and Res-152 models.