Figure 13. Adversarial examples generated by gradient stabilization attacks, along with original images and the baseline PGD attack.

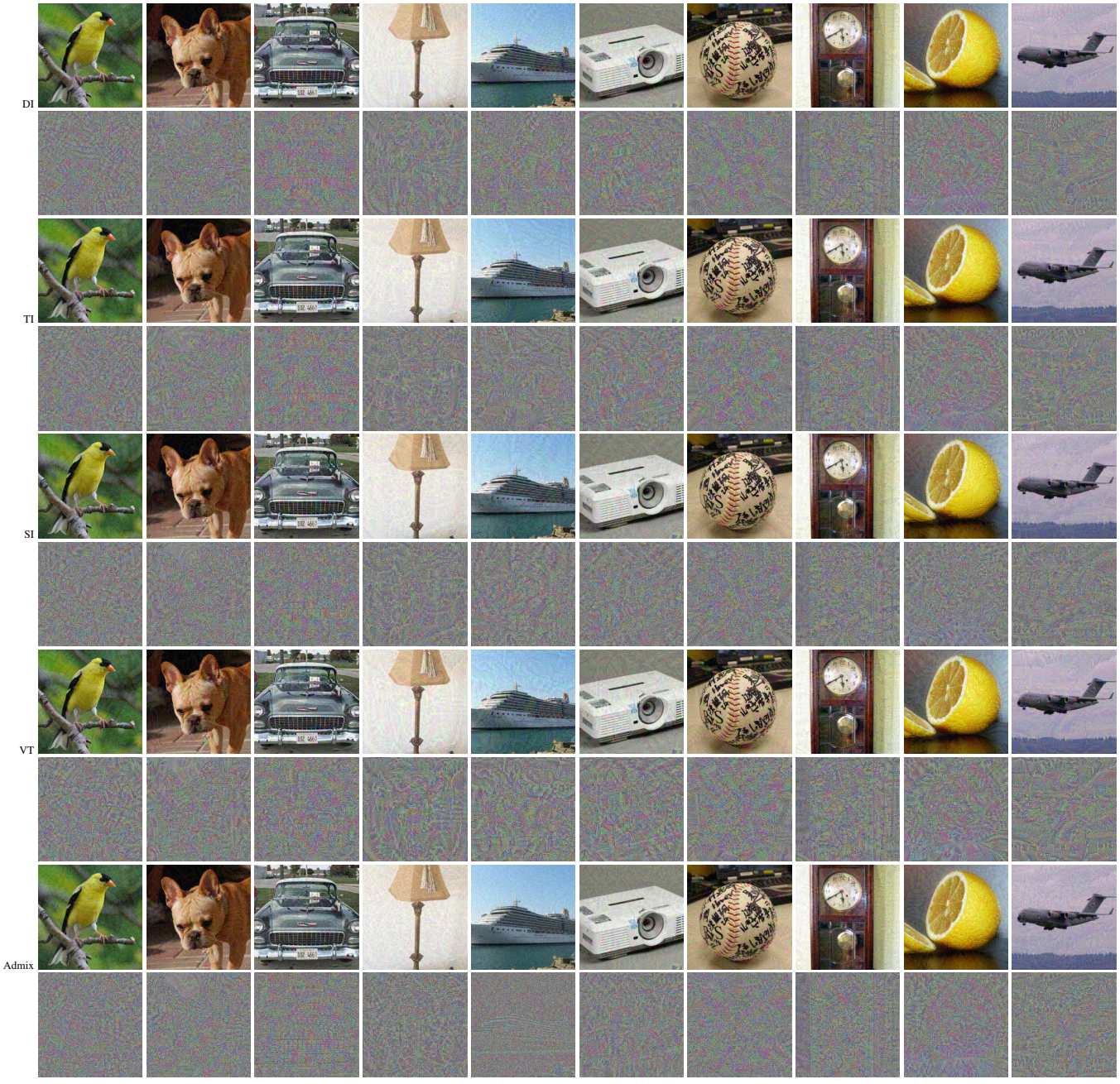

Figure 14. Adversarial examples generated by input augmentation attacks.

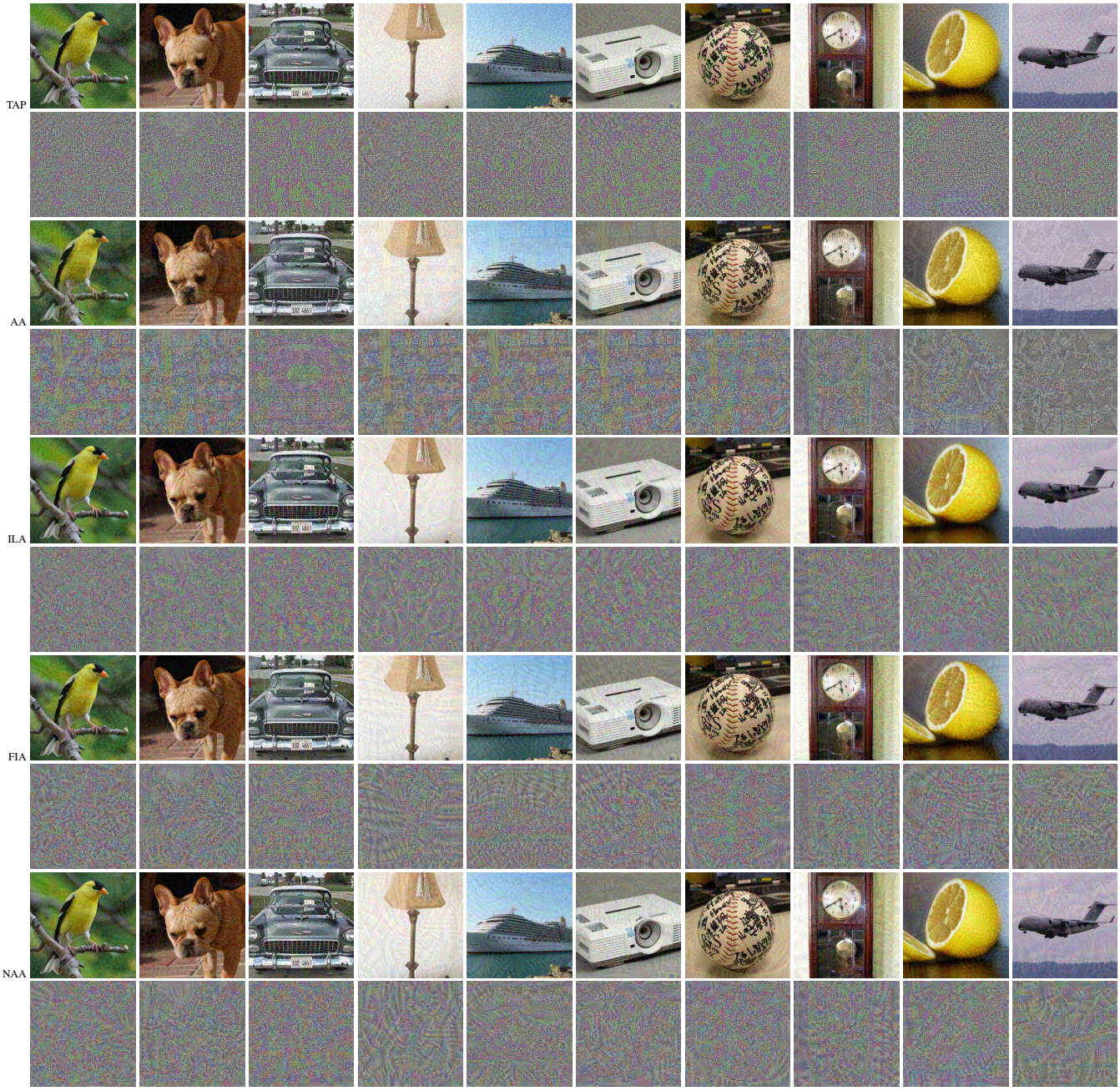

Figure 15. Adversarial examples generated by feature disruption attacks.

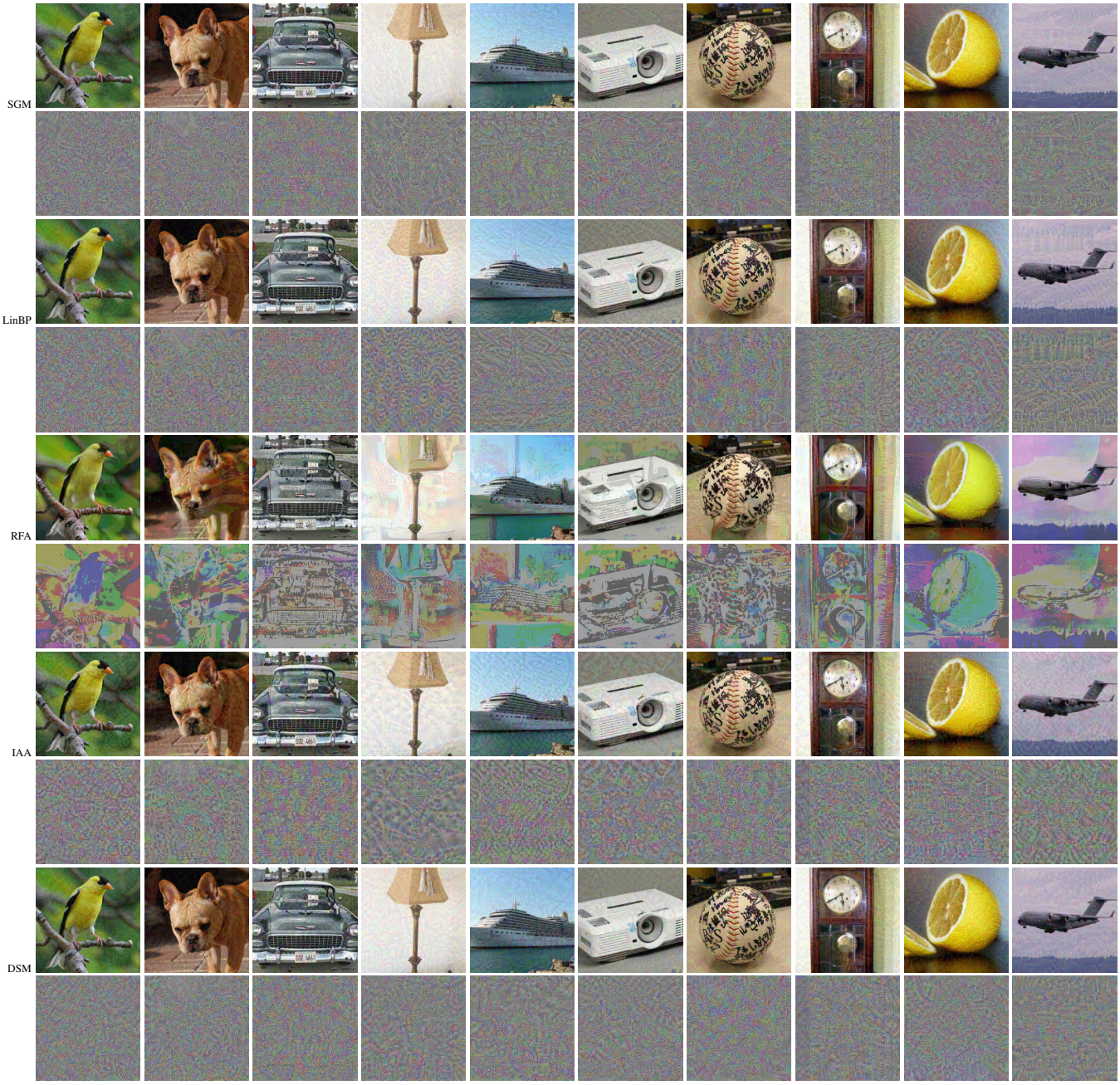

Figure 16. Adversarial examples generated by surrogate refinement attacks.

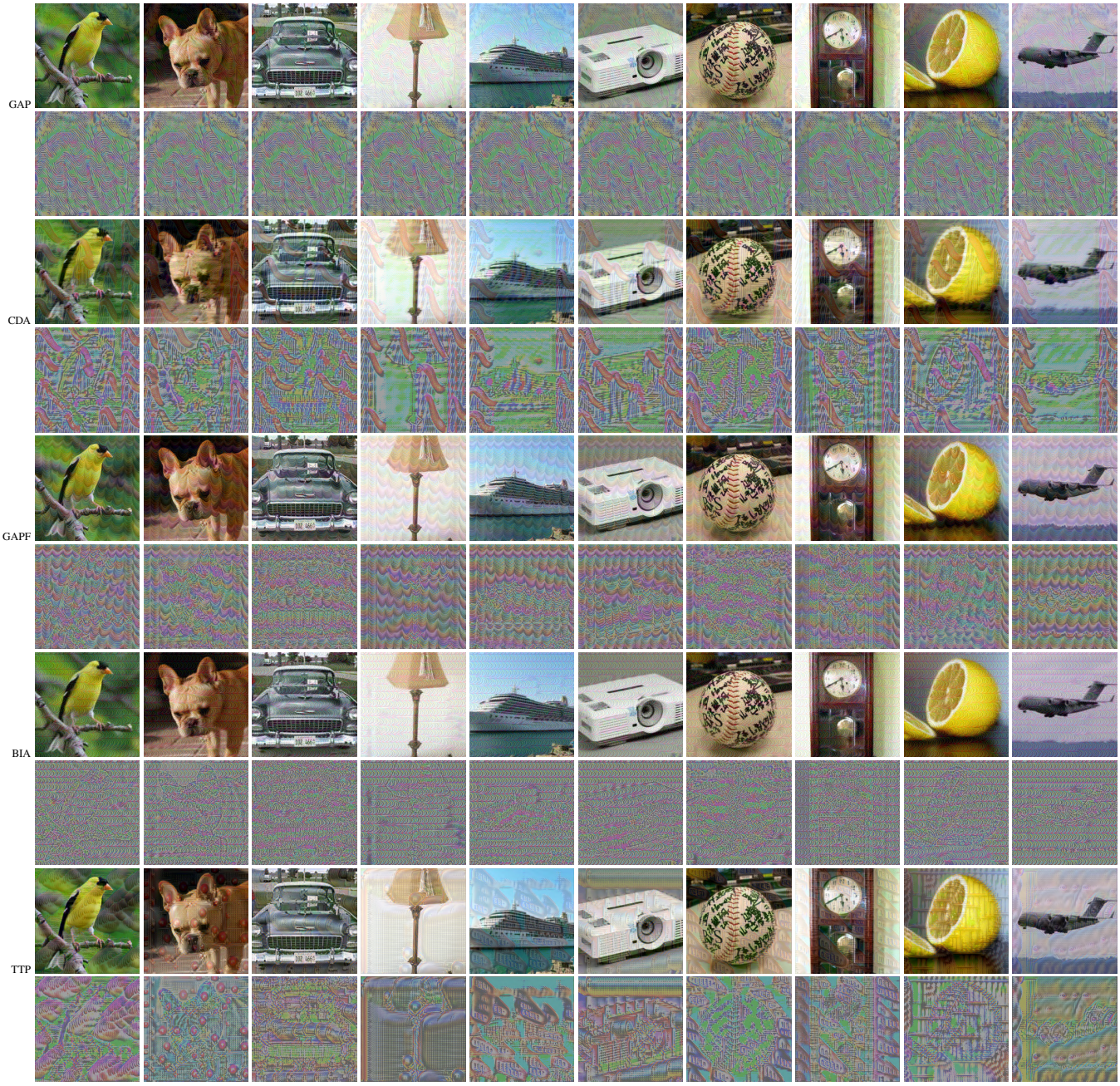

Figure 17. Adversarial examples generated by generative modeling attacks.