Some papers may refer to, but notice that these papers are not classical or excellent ones, just pick some papers from different perspectives.

Unsupervised adversarial training

1. Learning from Simulated and Unsupervised Images through Adversarial Training

<https://openaccess.thecvf.com/content_cvpr_2017/papers/Shrivastava_Learning_From_Simulated_CVPR_2017_paper.pdf>

1. Pairwise Adversarial Training for Unsupervised Class-imbalanced Domain Adaptation

<https://dl.acm.org/doi/abs/10.1145/3534678.3539243>

Adversarial learning on Multimodal

1. On the Robustness of Large Multimodal Models Against Image Adversarial Attacks

<https://arxiv.org/abs/2312.03777>

Summary on adversarial training attacks and defense

1. Recent Advances in Adversarial Training for Adversarial Robustness

<https://arxiv.org/abs/2102.01356>

1. Adversarial Attacks and Defences: A Survey

<https://arxiv.org/abs/1810.00069>

Some theoretical analysis on adversarial training

1. Theoretical Understanding of Learning from Adversarial Perturbations

<https://www.semanticscholar.org/paper/Theoretical-Understanding-of-Learning-from-Kumano-Kera/723bf5b549cfe8195db7ac45f248bb91d19a3065>

1. Adversarial Training from Mean Field Perspective

<https://proceedings.neurips.cc/paper_files/paper/2023/hash/edcd1aa172dceda2ea9d45a48f25d3e3-Abstract-Conference.html>

1. Theoretical Analysis of Adversarial Learning: A Minimax Approach

<https://arxiv.org/abs/1811.05232>