1. Record Nr. UNISA996500066703316

Titolo	Computer vision - ECCV 2022 . Part V : 17th European Conference, Tel Aviv, Israel, October 23-27, 2022 : proceedings / / Shai Avidan [and four others]
Pubbl/distr/stampa	Cham, Switzerland : , : Springer, , [2022] ©2022
ISBN	3-031-20065-9
Descrizione fisica	1 online resource (804 pages)
Collana	Lecture Notes in Computer Science
Disciplina	006.37
Soggetti	Computer vision - Equipment and supplies Pattern recognition systems - Data processing
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Intro -- Foreword -- Preface -- Organization -- Contents - Part V -- Adaptive Image Transformations for Transfer-Based Adversarial Attack -- 1 Introduction -- 2 Related Work -- 2.1 Adversarial Attack -- 2.2 Adversarial Defense -- 3 Method -- 3.1 Notations -- 3.2 Overview of AITL -- 3.3 Training AITL -- 3.4 Generating Adversarial Examples with AITL -- 4 Experiments -- 4.1 Settings -- 4.2 Compared with the State-of-the-Art Methods -- 4.3 Analysis on Image Transformation Operations -- 5 Conclusion -- References -- Generative Multiplane Images: Making a 2D GAN 3D-Aware -- 1 Introduction -- 2 Related Work and Background -- 3 Generative Multiplane Images (GMPI) -- 3.1 Overview -- 3.2 GMPI: StyleGANv2 with an Alpha Branch -- 3.3 Differentiable Rendering in GMPI -- 3.4 Discriminator Pose Conditioning -- 3.5 Miscellaneous Adjustment: Shading-Guided Training -- 3.6 Training -- 4 Experiments -- 4.1 Datasets -- 4.2 Evaluation Metrics -- 4.3 Results -- 4.4 Ablation Studies -- 5 Conclusion -- References -- AdvDO: Realistic Adversarial Attacks for Trajectory Prediction -- 1 Introduction -- 2 Related Works -- 3 Problem Formulation and Challenges -- 4 AdvDO: Adversarial Dynamic Optimization -- 4.1 Dynamic Parameters Estimation -- 4.2 Adversarial Trajectory Generation -- 5 Experiments -- 5.1 Experimental Setting -- 5.2 Main Results -- 6 Conclusion -- References -- Adversarial

Contrastive Learning via Asymmetric InfoNCE -- 1 Introduction -- 2 Asymmetric InfoNCE -- 2.1 Notations -- 2.2 Asymmetric InfoNCE: A Generic Learning Objective -- 3 Adversarial Asymmetric Contrastive Learning -- 3.1 Adversarial Samples as Inferior Positives -- 3.2 Adversarial Samples as Hard Negatives -- 4 Experiments -- 4.1 Main Results -- 4.2 Transferring Robust Features -- 4.3 Ablation Studies -- 4.4 Qualitative Analysis -- 5 Related Work -- 6 Conclusions -- References.
One Size Does NOT Fit All: Data-Adaptive Adversarial Training -- 1 Introduction -- 2 Related Work -- 2.1 Adversarial Defense -- 2.2 Decline of the Natural Accuracy -- 3 Review of Standard Adversarial Training -- 4 Adversarial Perturbation Size Matters -- 5 Data-Adaptive Adversarial Training -- 6 Experiment -- 6.1 Investigation of Hyper-parameters -- 6.2 White-Box and Black-Box Attacks -- 6.3 Distribution of Adversaries -- 6.4 DAAT with Different Calibration Strategies -- 7 Conclusions -- References -- UniCR: Universally Approximated Certified Robustness via Randomized Smoothing -- 1 Introduction -- 2 Universally Approximated Certified Robustness -- 2.1 Universal Certified Robustness -- 2.2 Approximating Tight Certified Robustness -- 3 Deriving Certified Radius Within Robustness Boundary -- 3.1 Calculating Certified Radius in Practice -- 4 Optimizing Noise PDF for Certified Robustness -- 4.1 Noise PDF Optimization -- 4.2 C-OPT and I-OPT -- 5 Experiments -- 5.1 Experimental Setting -- 5.2 Universality Evaluation -- 5.3 Optimizing Certified Radius with C-OPT -- 5.4 Optimizing Certified Radius with I-OPT -- 5.5 Best Performance Comparison -- 6 Related Work -- 7 Conclusion -- References -- Hardly Perceptible Trojan Attack Against Neural Networks with Bit Flips -- 1 Introduction -- 2 Related Works and Preliminaries -- 3 Methodology -- 3.1 Hardly Perceptible Trigger -- 3.2 Problem Formulation -- 3.3 Optimization -- 4 Experiments -- 4.1 Setup -- 4.2 Attack Results -- 4.3 Human Perceptual Study -- 4.4 Discussions -- 4.5 Potential Defenses -- 4.6 Ablation Studies -- 5 Conclusion -- References -- Robust Network Architecture Search via Feature Distortion Restraining -- 1 Introduction -- 2 Related Works -- 3 Preliminary -- 4 Method -- 4.1 Network Vulnerability Constraint -- 4.2 Robust Network Architecture Search (RNAS) -- 5 Experiment -- 5.1 Experimental Setup.
5.2 Results on CIFAR-10 -- 5.3 Results on CIFAR-100, SVHN and Tiny-ImageNet -- 5.4 Upper Bound H of network vulnerability -- 6 Conclusion -- References -- SecretGen: Privacy Recovery on Pre-trained Models via Distribution Discrimination -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Problem Formulation -- 3.2 Method Overview -- 3.3 Generation Backbone of SecretGen -- 3.4 Pseudo Label Predictor -- 3.5 Latent Vector Selector -- 3.6 SecretGen Optimization -- 3.7 Discussion -- 4 Experiments -- 4.1 Experimental Setup -- 4.2 Evaluation Protocols -- 4.3 Attack Performance -- 4.4 Robustness Evaluation -- 4.5 Ablation Studies -- 5 Conclusion -- References -- Triangle Attack: A Query-Efficient Decision-Based Adversarial Attack -- 1 Introduction -- 2 Related Work -- 3 Methodology -- 3.1 Preliminaries -- 3.2 Motivation -- 3.3 Triangle Attack -- 4 Experiments -- 4.1 Experimental Setup -- 4.2 Evaluation on Standard Models -- 4.3 Evaluation on Real-world Applications -- 4.4 Ablation Study -- 4.5 Further Discussion -- 5 Conclusion -- References -- Data-Free Backdoor Removal Based on Channel Lipschitzness -- 1 Introduction -- 2 Related Work -- 2.1 Backdoor Attack -- 2.2 Backdoor Defense -- 3 Preliminaries -- 3.1 Notations -- 3.2 L-Lipschitz Function -- 3.3 Lipschitz Constant in Neural Networks -- 4 Methodology -- 4.1 Channel Lipschitz Constant -- 4.2 Trigger-

Activated Change -- 4.3 Correlation Between CLC and TAC -- 4.4 Special Case in CNN with Batch Normalization -- 4.5 Channel Lipschitzness Based Pruning -- 5 Experiments -- 5.1 Experimental Settings -- 5.2 Experimental Results -- 5.3 Ablation Studies -- 6 Conclusions -- References -- Black-Box Dissector: Towards Erasing-Based Hard-Label Model Stealing Attack -- 1 Introduction -- 2 Background and Notions -- 3 Method -- 3.1 A CAM-driven Erasing Strategy -- 3.2 A Random-Erasing-Based Self-KD Module.
4 Experiments -- 4.1 Experiment Settings -- 4.2 Experiment Results -- 5 Conclusion -- References -- Learning Energy-Based Models with Adversarial Training -- 1 Introduction -- 2 Related Work -- 3 Background -- 3.1 Energy-Based Models -- 3.2 Binary Adversarial Training -- 4 Binary at Generative Model -- 4.1 Optimal Solution to the Binary at Problem -- 4.2 Learning Mechanism -- 4.3 Maximum Likelihood Learning Interpretation -- 4.4 Improved Training of Binary at Generative Model -- 5 Experiments -- 5.1 Image Generation -- 5.2 Applications -- 5.3 Training Stability Analysis -- 6 Conclusion -- References -- Adversarial Label Poisoning Attack on Graph Neural Networks via Label Propagation -- 1 Introduction -- 2 Preliminaries and Related Work -- 2.1 Graph Convolution Network and Its Decoupled Variants -- 2.2 Label Propagation and Its Role in GCN -- 2.3 Adversarial Poisoning Attacks on Graphs -- 3 Label Poisoning Attack Model -- 3.1 Label Propagation -- 3.2 Maximum Gradient Attack -- 3.3 GCN Training with Poisoned Labels -- 4 Experiments -- 5 Attack Performance Evaluation -- 6 Conclusion -- References -- Revisiting Outer Optimization in Adversarial Training -- 1 Introduction -- 2 Analyzing Outer Optimization in AT -- 2.1 Notations -- 2.2 Comparison of Gradient Properties -- 2.3 Revisiting Stochastic Gradient Descent -- 2.4 Example-Normalized Gradient Descent with Momentum -- 2.5 Accelerating ENGM via Gradient Norm Approximation -- 3 Experiments and Analysis -- 3.1 Comparison of Optimization Methods -- 3.2 Combination with Benchmark at Methods -- 3.3 Comparison with SOTA -- 3.4 Ablation Studies -- 4 Conclusion -- References -- Zero-Shot Attribute Attacks on Fine-Grained Recognition Models -- 1 Introduction -- 2 Related Works -- 3 Fine-Grained Compositional Adversarial Attacks -- 3.1 Problem Setting.
3.2 Compositional Attribute-Based Universal Perturbations (CAUPs) -- 3.3 Learning AUPs -- 4 Experiments -- 4.1 Experimental Setup -- 4.2 Experimental Results -- 5 Conclusions -- References -- Towards Effective and Robust Neural Trojan Defenses via Input Filtering -- 1 Introduction -- 2 Standard Trojan Attack -- 3 Difficulty in Finding Input-Specific Triggers -- 4 Proposed Trojan Defenses -- 4.1 Variational Input Filtering -- 4.2 Adversarial Input Filtering -- 4.3 Filtering Then Contrasting -- 5 Experiments -- 5.1 Experimental Setup -- 5.2 Results of Baseline Defenses -- 5.3 Results of Proposed Defenses -- 5.4 Ablation Studies -- 6 Related Work -- 7 Conclusion -- References -- Scaling Adversarial Training to Large Perturbation Bounds -- 1 Introduction -- 2 Related Works -- 3 Preliminaries and Threat Model -- 3.1 Notation -- 3.2 Nomenclature of Adversarial Attacks -- 3.3 Objectives of the Proposed Defense -- 4 Proposed Method -- 5 Analysing Oracle Alignment of Adversarial Attacks -- 6 Role of Image Contrast in Robust Evaluation -- 7 Experiments and Results -- 8 Conclusions -- References -- Exploiting the Local Parabolic Landscapes of Adversarial Losses to Accelerate Black-Box Adversarial Attack -- 1 Introduction -- 1.1 Related Works -- 2 Background -- 3 Theoretical and Empirical Study on the Landscape of the Adversarial Loss -- 4 The BABIES Algorithm -- 5 Experimental Evaluation -- 5.1 Results of BABIES on MNIST, CIFAR-10 and ImageNet