计算机顶会投稿分享与交流

中国科学院信工程研究所

加小俊

目 录

CONTENTS

01 个人简介

02 如何撰写好一篇论文

03 个人心得

04 研究领域推荐

个人简介

Biography

I received my B.S. degree in Software Engineering from China University of Geosciences, China. I am now a Ph.D student in University of Chinese Academy of Sciences. My research interests include computer vision, adversarial attack, adversarial training and reinforcement learning etc. I am the author of referred journals and conferences in CVPR, AAAI, ACM Multimedia etc. I will graduate in June 2023 (expected). Currently, I am looking for both academic and industrial job opportunities (especially post-doc positions regarding AI security) starting from Summer 2023. Feel free to connect if there are suitable opportunities!

Publications

- Comdefend: An efficient image compression model to defend adversarial examples, **Xiaojun Jia**, Xingxing Wei, Xiaochun Cao, Hassan Foroosh**CVPR, 2019** | **Project** | Github
- Adv-watermark: A novel watermark perturbation for adversarial examples, **Xiaojun Jia**, Xingxing Wei, Xiaochun Cao, Xiaoguang Han**ACM MM, 2020** | **Project** | Github
- Defending against Model Stealing via Verifying Embedded External Features, Yiming Li, Linghui Zhu, **Xiaojun Jia**, Yong Jiang, Shu-Tao Xia, Xiaochun Cao **AAAI, 2021** | **Project** | Github
- LAS-AT: Adversarial Training with Learnable Attack Strategy, **Xiaojun Jia**, Yong Zhang, Baoyuan Wu, Ke Ma, Jue Wang, Xiaochun Cao **CVPR, 2022** | **Project** | Github

个人网站: https://jiaxiaojunqaq.github.io/

如何撰写好一篇论文

1. Abstract (***)
2. Introduction (***)
3. Related Work
4. Motivation/Observation Section [selected]
5. Method (***)
6. Experiments (***)
7. Conclusion

Abstract部分

1-2句话背景（是什么，为什么重要）+motivation，因此我们做了什么+具体怎么做+这么做取得了什么效果/意义

非常识性短语需要在后面给解释(e.g./i.e., …)

自己方法涉及到的关键性短语可以斜体进行强化

时态记得一般现在时

关键词3-5个，相关性从高到低排序

Adversarial training (AT) is always formulated as a minimax problem, of which the performance depends on the inner optimization that involves the generation of adversarial examples (AEs). Most previous methods adopt Projected Gradient Decent (PGD) with manually specifying attack parameters for AE generation. A combination of the attack parameters can be referred to as an attack strategy. Several works have revealed that using a fixed attack strategy to generate AEs during the whole training phase limits the model robustness and propose to exploit different attack strategies at different training stages to improve robustness. But those multi-stage hand-crafted attack strategies need much domain expertise, and the robustness improvement is limited. In this paper, we propose a novel framework for adversarial training by introducing the concept of "learnable attack strategy", dubbed LAS-AT, which learns to automatically produce attack strategies to improve the model robustness. Our framework is composed of a target network that uses AEs for training to improve robustness, and a strategy network that produces attack strategies to control the AE generation. Experimental evaluations on three benchmark databases demonstrate the superiority of the proposed method, and the proposed method outperforms state-of-the-art adversarial training methods.

Introduction部分

第一段：背景

第二段：详细展开研究的任务/意义

第三段：自己方法，包括motivation、method详细（以及为什么这么做）、优势（但是不要单纯写性能好）

第四段：贡献，3-4条

注：行文能让读者觉得你这么做是合适且理所应当的为佳；时态一般现在时；依旧把读者当“小白”（该解释的解释，不要假设对方很懂你的研究领域）；可以加一张题图来帮助理解/分析。

Related Work部分

- **详略得当**（e.g.，一篇做防御的文章一般需要同时review攻击和防御，防御部分的related work篇幅应当大于攻击部分)
- **最好能有自己的总结，而不是简单的按照时间顺序划分**(e.g., 给现有的工作进行分类总结)
- **别人的工作用一般过去时**
- **引用格式保持一致**
- **et al 等符号记得斜体**

Motivation/Observation部分

用来合理的引出自己的方法，可以是某个特别的实验现象+分析，也可以是现有方法的不足+分析。通过这些分析来引入你的method的合理性，让读者觉得你的做法是有insight而不是incremental的。最后一句一定是一个引子，用来引出下一个section的method。

Method部分

- **最好能有一张好看的示意图总结方法的大致流程**
- **先总后分** （每个subsection开头先讲这个部分做什么，以及为什么，再讲具体的做法）
- **如果方法有很多部分（≥ 3），在第一/二个subsection先总结general pipeline，再每个subsection讲每个部分具体怎么做**
- **可以考虑分第一个subsection讲prelim**

Experiments部分

- 一般是main experiments + discussion experiments (e.g., ablation study)
- 画图记得美观
- 表格和图分配合理（不要全是表或者全是图）
- 实验分析要有更多的细节，尤其是分析自己的方法相比baseline为什么好，好在哪里
- baseline和数据集的选择要合理（至少包括<2年内的baseline和业内公认的数据集）。

Conclusion部分

- 是abstract的改写版本，一般可以remove一些背景的细节和方法的Motivation等
- 记得用一般过去时

Check List

- 写的时候要注意呈现出来的连贯性，不要打断读者思维，大段的证明等可以放附录。
- 把审稿人都当小白，尤其是不是很common sense的东西一定要在出现的时候解释清楚
- 注意、主被动的分布得当，尽量多用主动语态
- 尽量不要写很长或很复杂的句子，拆成多个简单句
- 尽量避免口语化的用词和没有意义的冗余
- 英文中一般没有三个名词加在一起的，一般最多两个名词
- 写的时候不要拉仇恨，比如避免general，special case这种
- 尽量不要写没有被证明或者实验验证的claim，实在要写也要加presumably

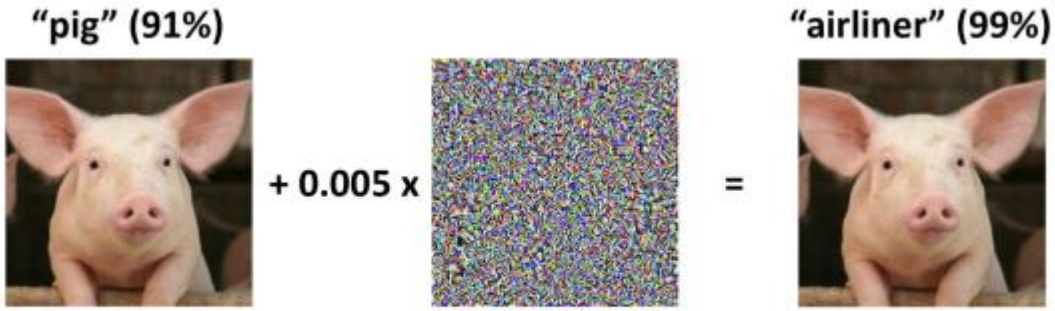
Check List

- 证明不要跳步，一步一步都写清楚
- 检查时态，别人的方法过去时，自己的方法现在时
- 图表要self-consistent，caption加上必要的说明，让读者光看图表就能知道在干什么
- 检查参考文献，包括所有的方法、数据集是不是都合理引用了及格式是否统一
- 检查图片，包括自己的方法在颜色上区分是否明显，label的大小是不是能看得清
- Grammarly查语法，Linggle查短语/介词搭配
- e.g., i.e., 后面要跟逗号
- 检查reference的格式是一致的（例如会议全部用缩写好了）

个人心得

- 多看文章，最重要的还是要做实验，在coding中会收获更多
- 要跟进前沿的研究，最好是跨领域的研究，比如对抗，后门这些安全研究领域和GIS的结合
- 多尝试多投稿，尽早加入相关实验室或团队，接触科研。
- 多关注一些学术类型的公众号，参加里面科研分享

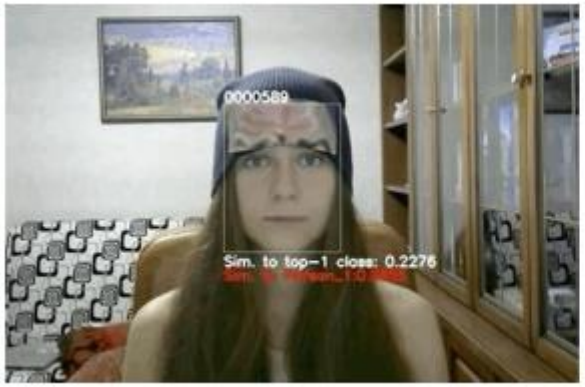
研究领域推荐-对抗

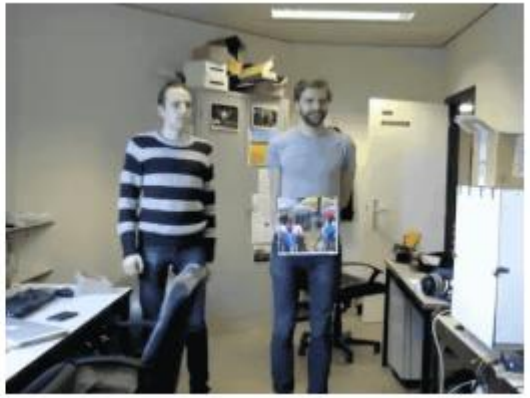


[Szegedy et al. 2014]: Imperceptible noise (adversarial examples) can fool state-of-the-art classifiers

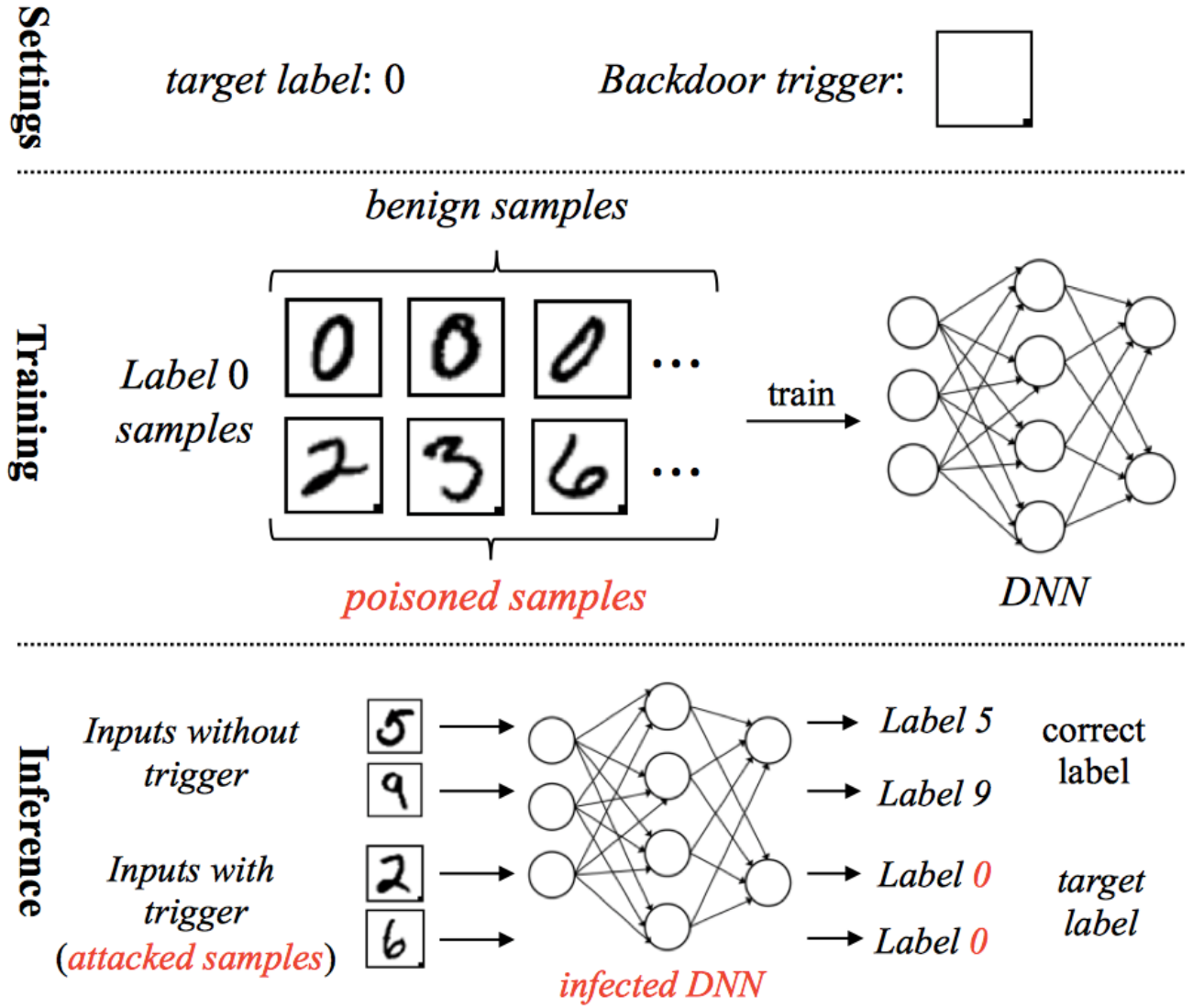
[Sharif et al. 2016]:
Glasses that fool face recognition

Eykholt et al., Robust Physical-World Attacks on Deep Learning Visual Classification, CVPR 2018

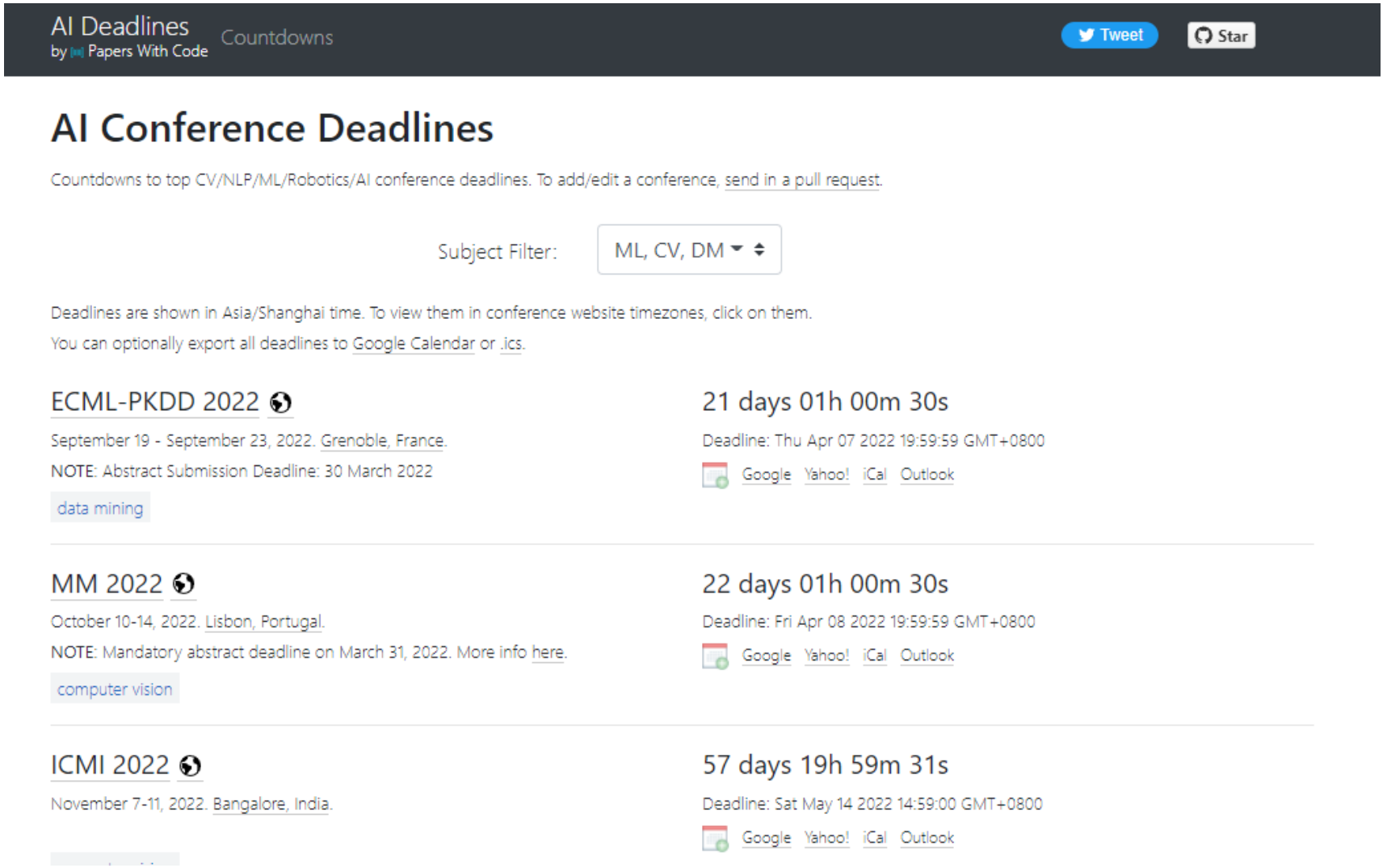




研究领域推荐-后门


Settings
target label: 0
Backdoor trigger:
benign samples
Training
Label 0 samples
train
poisoned samples
DNN
Inference
Inputs without trigger
Inputs with trigger (attacked samples)
Label 5
Label 9
Label 0
Label 0
correct label
target label
infected DNN

研究领域推荐



https://aideadlin.es/?sub=ML,CV,DM

研究领域推荐

写作：https://www.overleaf.com/project
PDF剪切：https://pdfresizer.com/crop
投稿网站:https://cmt3.research.microsoft.com
https://openreview.net/group?id=ICLR.cc
对抗前沿：https://nicholas.carlini.com/writing/2019/all-adversarial-example-papers.html
后门前沿：https://github.com/THUYimingLi/backdoor-learning-resources
作图：PPT or Visio

Thanks