Robust AI Project Team

Weekly Report

Heran Zhu

Electronic Information School, Wuhan University

Oct 9th, 2021

HVS(Human Visual System)产生对抗样本的相关论文

- 1. *The Human Visual System and Adversarial AI(2020)*: HVS对低频信息更敏感，对亮度的变化比色度的变化更敏感
- 2. *SSIMLayer: Towards Robust Deep Representation Learning via Nonlinear Structural Similarity(2018)*: 结构相似性度量 structural similarity metric, 一个提取结构信息的神经网络模块（模仿HVS）
- 3. *Demiguise Attack: Crafting Invisible Semantic Adversarial Perturbations with Perceptual Similarity*: 利用感知相似度(一种新的图像质量度量指标，可以模拟真实世界中光照和对比度变化)来产生扰动的黑盒攻击，使用面向hvs的图像度量来处理语义信息，以生成不可见的语义对抗扰动。可以作为一个部分融合到传统攻击方法中

HVS(Human Visual System)产生对抗样本的相关论文

- 4. *GreedyFool: Multi-Factor Imperceptibility and Its Application to Designing Black-box Adversarial Example Attack*: 根据影响人眼可感知性的因素(显著畸变(JND)、韦伯-费希纳定律、纹理掩蔽和信道调制)设计多因素度量损失产生对抗样本
- 5. *CDAE: Color decomposition-based adversarial examples for screen devices*: 为屏幕设备设计的基于颜色分解的对抗性示例方法DAE
- 6. *Semantic Adversarial Examples*: 语义对抗样本，约束优化问题，在HSV色彩空间上添加扰动(应该基于HVS对色度变化不敏感的特点)
- 7. *Feature Distillation: DNN-Oriented JPEG Compression Against Adversarial Examples*: 基于图像压缩技术的抗对抗实例攻击方法

影响HVS的因素

The Human Visual System and Adversarial AI

- **HVS对低频信息更敏感**
- **HVS对亮度的变化比色度的变化更敏感。**

GreedyFool: Multi-Factor Imperceptibility and Its Application to Designing Black-box Adversarial Example Attack

- **Just Noticeable Distortion**: JND。人眼无法感受像素周围的明显低于失真阈值以下的刺激
- **Weber-Fechner Law**: 一个心理物理学的观点，明显的刺激差异保持一个恒定的比率
- **Texture Masking**: 纹理掩膜。人眼对平滑区域像素的干扰比纹理区域的干扰更敏感（也就是对低频变化比高频变化更加敏感）
- **Channel Modulation**: 通道调制，人眼对颜色通道的敏感度是有差异的。对绿色最敏感，对蓝色最不敏感。

IQA(Image Quality Assessment)产生对抗样本的相关论文

- 1.*Feature Distillation: DNN-Oriented JPEG Compression Against Adversarial Examples*: 基于图像压缩技术的抗对抗实例攻击方法
- 2.*RAN4IQA: Restorative Adversarial Nets for No-Reference Image Quality Assessment*: 基于GAN的无参考IQA
- 3.*VR IQA NET: Deep Virtual Reality Image Quality Assessment using Adversarial Learning*: 将对抗学习应用到VR IQA中
- 4.*Generating Adversarial Examples with an Optimized Quality*: 直接利用IQA的指标来产生对抗样本
- 5.*A Novel Rank Learning Based No-Reference Image Quality Assessment Method*

Thanks!