Audio Adversarial Examples

Preliminary Talk
Karla Markert, 07 July 2020
Outline

About
  About Me
  About My Department at AISEC

Introduction to Adversarial Examples
  Neural Networks
  Adversarial Images
  Adversarial Audio

Organizational Stuff
Table of Contents

About

Introduction to Adversarial Examples

Organizational Stuff
About

About Me

Name  Karla Markert
Department  Cognitive Security Technology
Role  Research assistant
Background  Mathematics, political science and computer science
About My Department at AISEC

Cognitive Security Technologies:
Intersection of artificial intelligence and IT security.
About

About My Department at AISEC

Applications:
- CAN traces,
- malware,
- wireless networks

Anomaly Detection

unsupervised

semi-supervised

supervised
About My Department at AISEC

Applications:
- encryption,
- privacy attacks on memory networks,
- architectures for data analysis
About My Department at AISEC

Applications:
- GDPR,
- source code,
- textual descriptions
About My Department at AISEC

Applications:

- face recognition,
- speech recognition,
- deep fake detection
Table of Contents

About

Introduction to Adversarial Examples

Organizational Stuff
Deep learning “is an approach to AI. Specifically, it is a type of machine learning, a technique that enables computer systems to improve with experience and data. [...] Deep learning is a particular kind of machine learning that achieves great power and flexibility by representing the world as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones.” [2]
Introduction to Adversarial Examples

Neural Networks

Visualization of a neural network with one hidden layer.

Image taken from Wikipedia\(^1\).

Introduction to Adversarial Examples

Adversarial Images

Images taken from [5, 4].
Introduction to Adversarial Examples

Adversarial Images

Images taken from [7]. Ostrich means Strauß in German.
Introduction to Adversarial Examples

Adversarial Images

Stop sign recognized as stop sign.

Stop sign recognized as bottles.

Images taken from [6].
Introduction to Adversarial Examples

Adversarial Images

Images taken from [1].
Introduction to Adversarial Examples

Adversarial Audio

Original
Transcription: without the dataset the article is useless

Adversarial
Transcription: okay google browse to evil dot com

Examples taken from [3].
Table of Contents

About

Introduction to Adversarial Examples

Organizational Stuff
In this seminar, we take a look at different audio adversarial attacks and possible mitigations.

- **Level:** Bachelor and Master
- **Number of Participants:** 8
- **Language:** English
- **Requirements:** Basic knowledge in machine learning (especially deep neural networks) and IT security.
Organizational Stuff

**Time:** This course will be held as a block seminar.

- **July 10** (Friday), 14:00 - 14:45 *Preliminary talk*
- **August 11** (Tuesday) 14:00-15:00 *Kick Off*
- **November 26** (Thursday) and **November 27** (Friday), 9:00-17:00 *Presentations*
- **December 4** (Friday), 9:00-10:00 and **January 15** (Friday), 9:00-10:00 *Debriefing*
- **December 7** (Monday), 23:59 *Deadline for paper*
Organizational Stuff

Goals:

- familiarization with scientific paper reading and scientific presentations;
- better understanding of attacks against machine learning algorithms;
- active participation and insights into topics of current research. For more information, see module description IN0014 and IN2107.
Organizational Stuff

**Method:** The seminar is organized as follows.

- Every participant gives a *presentation on a scientific paper*, which is assigned in the kick off session.
- Every student is required to write a *four page hand out* summarizing the main points of the paper (LaTeX template will be provided).

We attach great importance to all students profiting from the others’ presentations.
Organizational Stuff

The grade is composed up of:

- 10% active participation,
- 25% presentation (structure of the talk, introduction to the topic, clear problem definition and motivation, sound style of delivery...),
- 25% hand out (language, structure of the hand out...),
- 40% quality of the content (main points of the paper, good discussion and outlook...)
Everyone interested in participating is required to send an e-mail to karla.markert@aisec.fraunhofer.de until July 16 indicating her/his interest. You may include a letter of motivation, a CV or a transcript of records.

...any questions so far?


Contact Information

Karla Markert

Department
Cognitive Security Technologies

Fraunhofer-Institute for
Applied and Integrated Security (AISEC)

Address: Lichtenbergstr. 11
85748 Garching (near Munich)
Germany

Internet: www.aisec.fraunhofer.de

Phone: +49 89 3229986-136
E-Mail: karla.markert@aisec.fraunhofer.de