情報・エレクトロニクス系部門

デモ展示中

松浦研究室

[暗号と情報セキュリティ]

生産技術研究所 情報・エレクトロニクス系部門

Department of Informatics and Electronics

情報理工学系研究科

電子情報学専攻

情報セキュリティ

http://kmlab.iis.u-tokyo.ac.jp

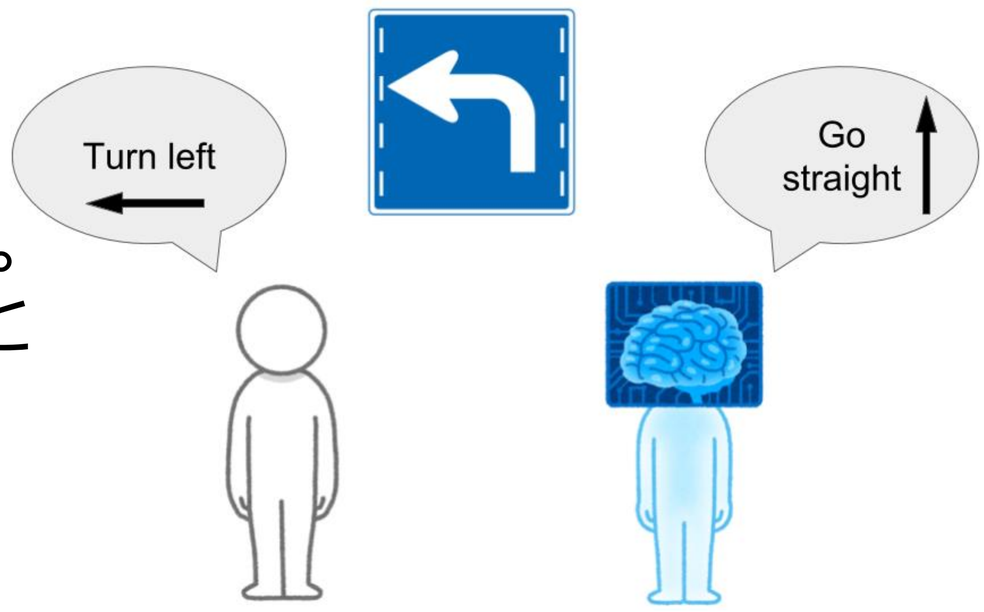
- 機械学習による画像認識

  多層のニューラルネットワークを用いた、これまででは考えられないほど大量の画像を教師として学習する機械学習の手法により、画像認識の精度が大きく向上してきている。将来的には自動車の自動運転などへの応用も期待される。

- Adversarial Example

  高い精度を誇る画像認識器でも、人間の感覚からは考えられない間違いをすることがわかっている。
  → これは adversarial example と呼ばれ、意図的に用いれば機械への攻撃となりうる。

左から元画像、ランダムノイズを加えた画像、adversarial example の例

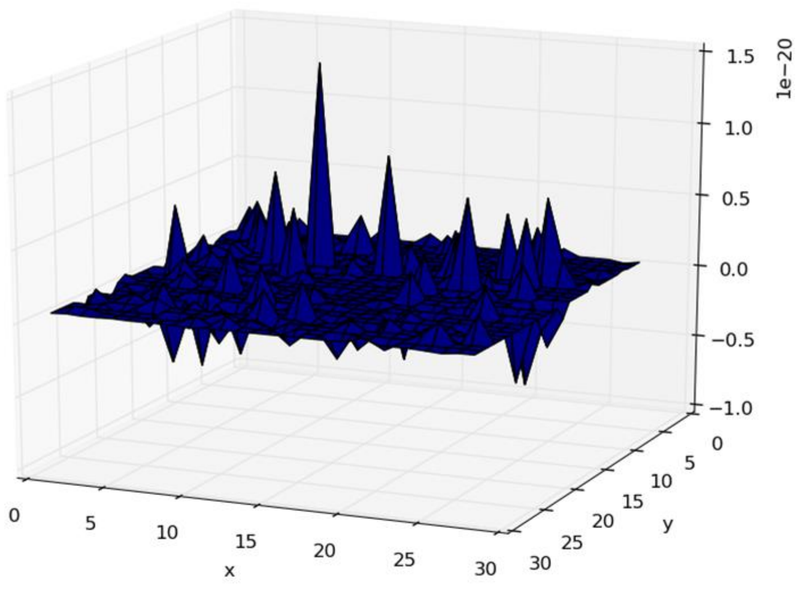


機械が認識を誤るイメージ図

- 検知

  本研究では、対策の一環として以下のアプローチで検知を試みた。この結果、8割以上の精度で検知できることを確認した。

  1. 「他クラスへの移りやすさ」を saliency map を計算することによって計測。
  2. あらかじめ用意した正例・負例に対し上記 1 を行った結果を教師として利用し、検知器を構築。



Saliency map