INFORMATION SECURITY Ew-401

Department of Informatics and Electronics

Demonstration on Show

MATSUURA LAB.

[Cryptography and Information Security]

Department of Informatics and Electronics

Information Security

Information and communication
engineering department

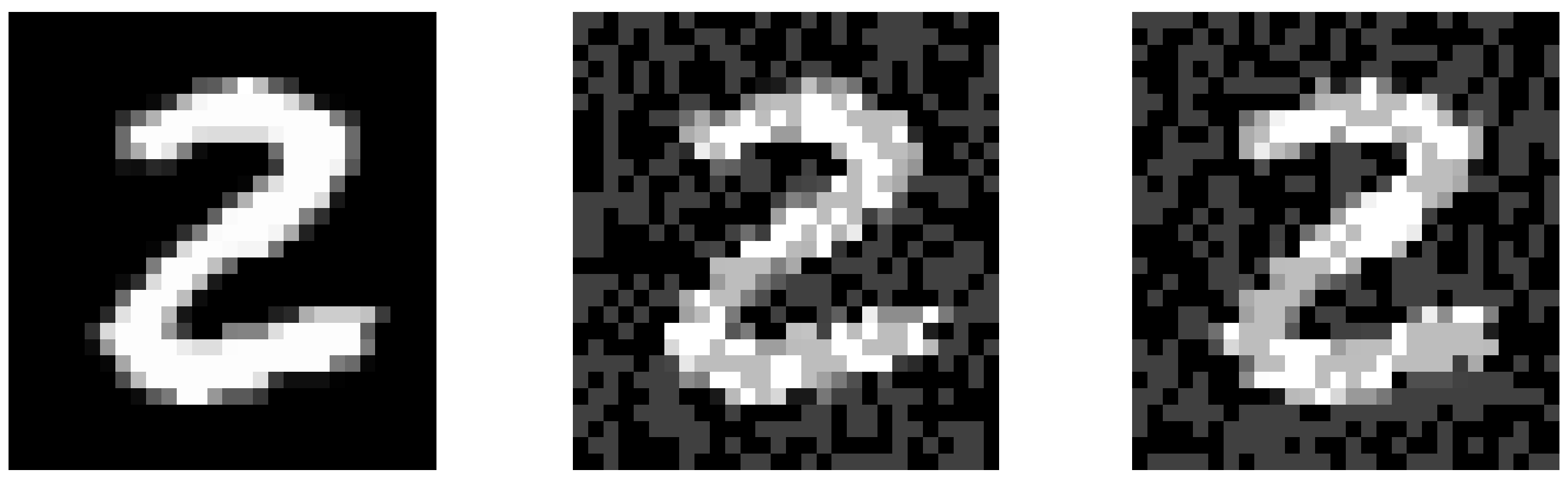
http://kmlab.iis.u-tokyo.ac.jp

- Image recognition by machine learning

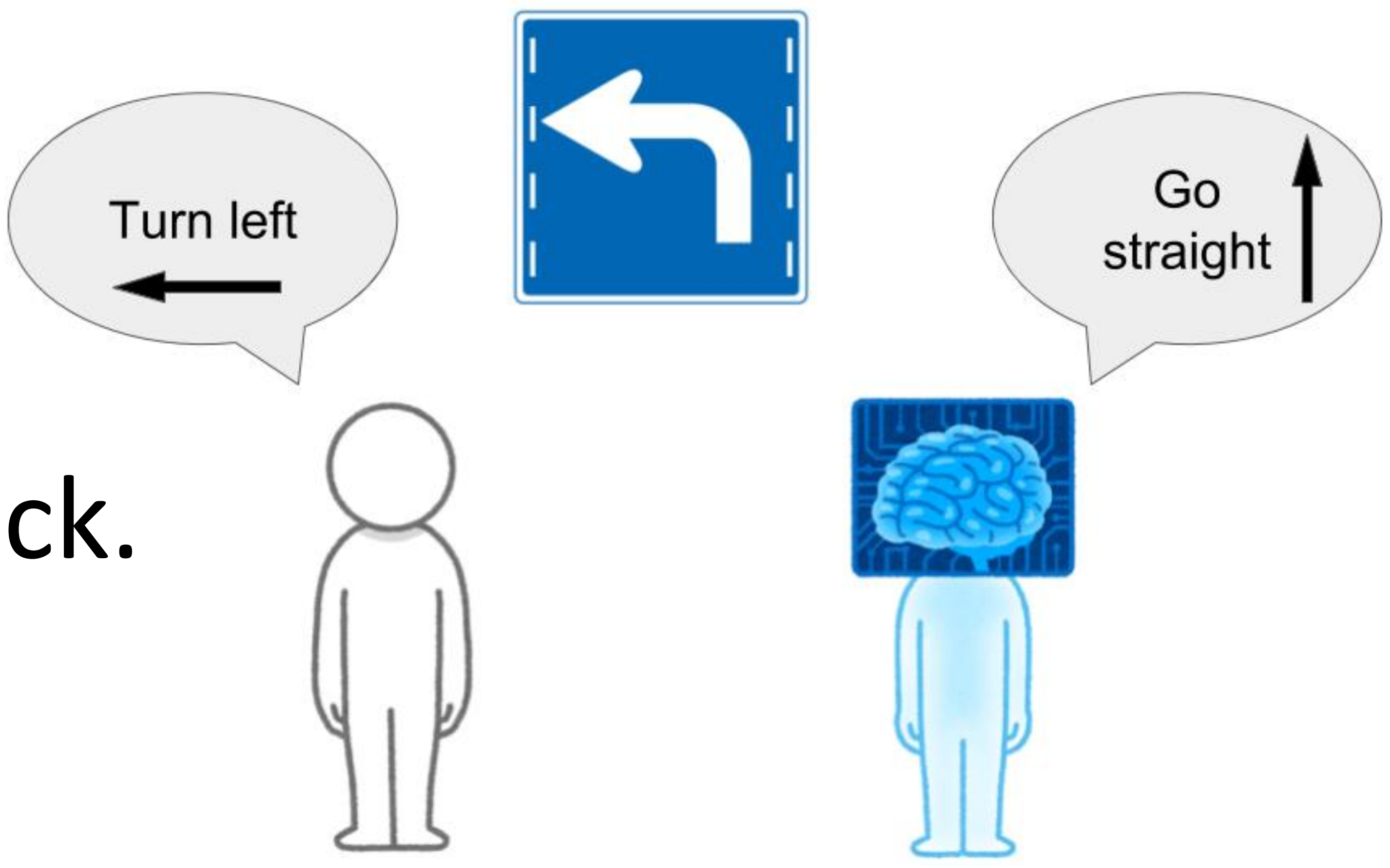
  The accuracy of image recognition is very increased by multilayer neural networks that learn from a lot of images. In the future, this technique is expected to apply in various fields for example autonomous driving car.

- Adversarial example

  It’s reported that machines, even high accuracy classifiers, have possibility of misrecognition that is far from the human sense.
  → It’s called adversarial example and sometimes regarded as an attack.



From left, original , noised and adversarial image.

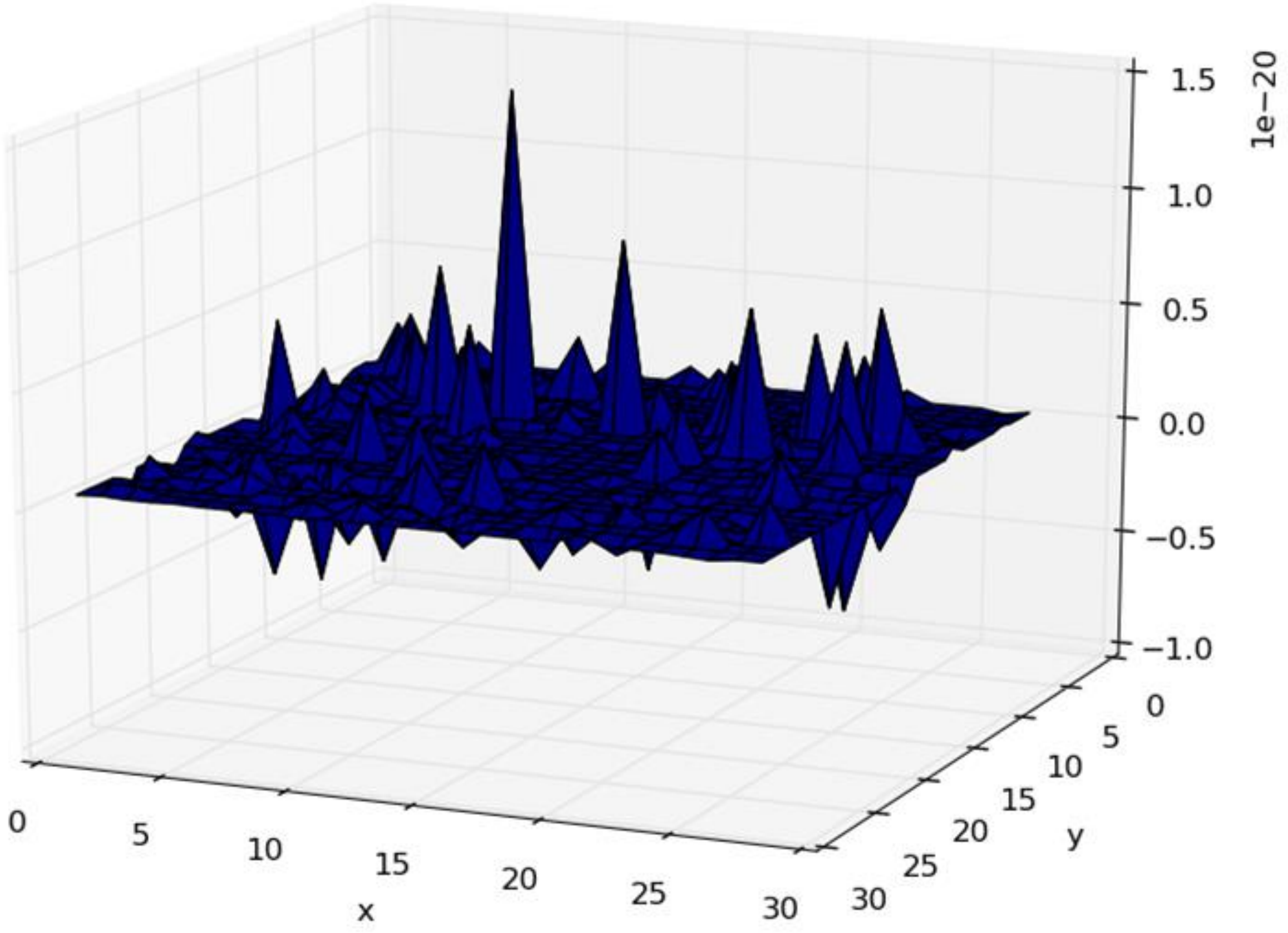


Concept image of machine’s misrecognition.

- Detection

  In this research, we tried to detect it by following approach as one of the countermeasures. As a result, we detected over 80%.

  1. Measure how easy to be classified as other class by calculating <u>saliency map</u>.
  2. Construct detector that is trained by values calculated by procedure 1 .



Saliency map