Oral presentation | General Session | [General Session] 2. Machine Learning

[2A4][General Session] 2. Machine Learning

座長:椿 真史（産業技術総合研究所）

Wed. Jun 6, 2018 5:20 PM - 7:00 PM Room A (4F Emerald Hall)

6:40 PM - 7:00 PM

[2A4-05]Highlighting Non-contributing Pixels for Visual Explanation of CNNs

〇Koichi Ikeno[1], Satoshi Hara[1], Takashi Washio[1] (1. The Institute of Scientic and Industrial Research, Osaka University)

Keywords:Convolutional Neural Network, Interpretability

Explaining the output of Convolutional Neural Networks (CNNs) is a challenging topic.
A typical explanation is to identify which pixels are contributing to the output of CNN.
In this paper, we propose a new approach for explaining the output of CNNs by finding pixels that are \emph{not} contributing to the output.
To highlight non-contirbuting pixels, we propose optimizing a noise level so that additive noise to the input image does not change the CNN output.
The experimental results on MNIST show that the proposed method can idntify non-contributing pixels adequately.