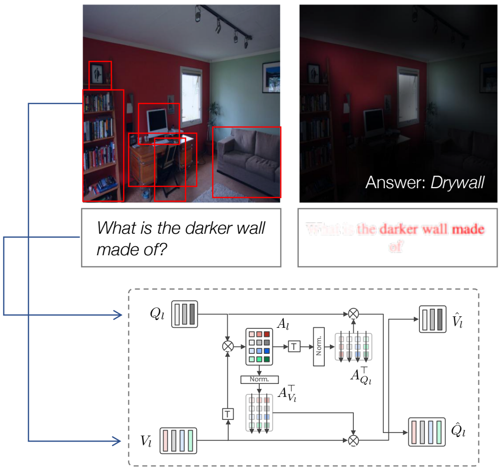
Toward visual recognition of Shitsukan concepts by multi-modal representation learning

Takayuki Okatani
Tohoku University / RIKEN Center for Advanced Intelligence Project
okatani@vision.is.tohoku.ac.jp



For the past several years, we have been conducting research on the development of a computer vision system that can recognize *Shitsukan* as humans do, with a particular focus on applications of deep learning methods. We first studied supervised learning of Shitsukan concepts, for which it is hard to provide true labels unlike object category classification. To mitigate the scalability issue with the approach, we next studied if it is possible and how to learn Shitsukan concepts from public data on the Web. Conducting a preliminary study of generating natural language text representing Shitsukan concepts, we now believe that the most important remaining problem is how to build a representation space of Shitsukan concepts bridging the two modalities, vision and language. Towards a solution to this, we have proposed a neural architecture for fusing vision and language representations, named dense co-attention networks, applying it to visual question answering (i.e., the task of answering a given question about the contents of a given scene image). We further extended this to enable to conduct multi-task learning of different vision-language tasks with a single network, aiming at coping with the issue of dataset bias, which has recently been recognized to be a major issue with applications of deep learning to high-level AI tasks.

Reference:

Duy-Kien Nguyen and Takayuki Okatani, "Multi-Task Learning of Hierarchical Vision-Language Representation", Proc. CVPR 2019: 10492-10501, 2019.

Duy-Kien Nguyen and Takayuki Okatani, "Improved Fusion of Visual and Language Representations by Dense Symmetric Co-Attention for Visual Question Answering", Proc. CVPR 2018: 6087-6096, 2018.

Sirion Vittayakorn, Takayuki Umeda, Kazuhiko Murasaki, Kyoko Sudo, Takayuki Okatani, Kota Yamaguchi, "Automatic Attribute Discovery with Neural Activations", Proc. ECCV (4) 2016: 252-268, 2016

Biography:

Takayuki Okatani received his B.Sc. degree as well as his M.Sc. and Ph.D degrees in Mathematical Engineering and Information Physics from Graduate School of Engineering at Tokyo University, 1994, 1996, and 1999, respectively. Currently, he is directing the Computer Vision Laboratory at Tohoku University. He also serves as a Leader of Infrastructure Management Robotics Team at RIKEN Center for Advanced Intelligence Project (AIP) from 2016. His research interests are in the field of computer vision and machine learning.

Homepage: http://www.vision.is.tohoku.ac.jp