Vol.2018-CVIM-211 No.15
2018/3/2

Deep Learningによる視覚・言語融合の最前線

牛久 祥孝[1,a)]

概要：Deep Learning の恩恵は，音声認識や画像認識，機械翻訳といった種々のタスクの精緻化だけではない．それぞれのタスクにおける手法が畳込みニューラルネットワークや再帰型ニューラルネットワークによって表されるようになり，異分野の研究者にとっても理解しやすくなった．結果として，かつては独立に扱われていたようなモダリティのデータを融合し，理解・生成するような研究が大きな広がりを見せつつある．本講演では，画像や動画といった視覚的なデータと，自然言語とを融合した挑戦的かつ萌芽的なタスクとして，「画像/動画からのキャプション生成」「画像も含めた質問応答システム」「キャプションからの画像生成」「画像を伴う言語横断検索・キャプション翻訳」を，歴史的な経緯も踏まえながら紹介する．

Frontiers of Vision and Language: Bridging Images and Texts by Deep Learning

Ushiku Yoshitaka[1,a)]

[1] 東京大学
[a)] ushiku@mi.t.u-tokyo.ac.jp