

Vision&Language 研究のこれまで・これからと、 日本でのコミュニティ形成の試み

品川 政太郎^{1,a)}

概要：Vision&Language は、自然言語処理と画像処理の融合領域を指す研究分野である。この分野は、深層学習による両分野の進展と共に、image-captioning や visual question answering など代表的な研究課題が注目され、今では一つの研究分野として広く認知されるようになった。Vision&Language 研究の面白さの一つは、実世界に紐づいた課題を、自然言語を介した情報伝達により解決できる点にある。例えば、目の見えない人の代わりに機械が周りに何があるかを説明する、人間の指示に従って機械が物を運ぶといった課題が挙げられる。これらの課題を解くには、画像情報と言語情報をいかに統合して紐づけるかが鍵となる。また、自然言語は多様な表現が可能のために、人間同士ですら正確な情報伝達を行うことは簡単ではない。情報伝達にミスが発生した時に、外界との相互作用を考慮しながらどのように意図を擦り合わせるかという対話戦略まで考えることが、実用性を追求する上で重要であり、実際に分野内でもそのような方向を目指す流れが始まっている。では、Vision&Language 研究によってこれまで何が可能になり、何がこれからの課題となるのだろうか？本講演では、Vision&Language の原点といえる SHRDLU に始まる過去の議論や、最近の研究動向から、この議論のたたき台を提供したい。また、Vision&Language を主軸とする研究機関は国内でも数が少ないことから、より高度な課題解決に向けて、自然言語・画像の両分野の密な連携が必要であることを述べ、そのためのコミュニティ形成の試みについて紹介する。

¹ 奈良先端科学技術大学院大学

^{a)} shinagawa.seitaro.si8@is.naist.jp