画像・映像をより深く理解する人工知能
「視覚」と「言語」による創発

中島 悠太
NAKASHIMA Yuta
大阪大学データビリティフロンティア機構
准教授

画像・映像を入力とする人工知能（AI）で、バイアス（例えば、特定の社会的グループの人が不利になるような識別結果を出力するなど）が大きな問題となっています。結果をどのように導き出したかのプロセスを人間が理解可能な「説明可能なAI」は一つの対策となりますが、抜本的な解決には、「データの本質を見る」モデルを創り出す必要があります。しかし、それは人には簡単でも、現在のAIには難しい作業です。

AIと人の大きな違いとして、人には言語があります。人は推論プロセスを説明して、間違いがあれば他の人から指摘してもらえます。我々の研究では、現在のAIで広く利用されている単一のベクトルによる意味の記述（図1左）ではなく、独自の言語を学習によってAIに獲得させることで、人と同じように画像・映像を理解できるのではないかという仮説の下、新しいAIの可能性を模索しています（図1右）。この研究は、これからの画像・映像に関わる多くの研究や、その応用の基盤となる技術としての展開が期待されます。

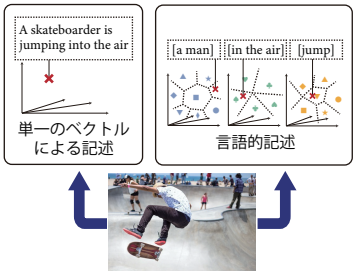

図1 従来のAIによる画像のベクトル表現（左）と今回の研究で提案する言語的表現（右）

キーワード
コンピュータビジョン、パターン認識、深層学習、画像・映像の理解、言語、表現学習

応用分野
画像・映像の利活用を加速するコンテンツに基づく検索

［研究の先に見据えるビジョン］フレーム問題を超えて知識を集約する社会基盤へ

画像・映像に関わる多くのアプリケーションはもちろん、ロボットやエージェントが活躍する未来の社会では、視覚から様々なもの・ことを認識するAIが社会基盤としての役割を担うと予想します。この研究では、学習で獲得した言語によって、フレーム問題（現実に起こりうる問題すべてに対処することができないこと）を超えて映像・画像などから知識を自動的に集約し、その知識を広く活用する新しいAIの創出を目指します（図2）。

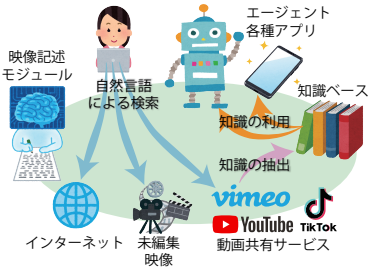

図2 社会基盤としての視覚AI