

青山学院大学 物理科学科 コロキウム

2023年度 第2回

下記の通りコロキウムを企画致しました。学生や分野の違う方にもわかるレベルから始めて下さるようお願いしてあります。どなた様もご自由に是非ご聴講ください（事前参加登録なし）。

（世話人：田中 周太、連絡先：sjtanaka@phys.aoyama.ac.jp）

「主催：物理科学科、基礎科学コース、機能物質創成コース」

講演者 瀧 雅人氏 (立教大学人工知能科学研究科)

日時 8月29日（火）15:05 – 16:35から

場所 青山学院大学 理工学部 L棟6階 L 603室

講演題目 コンピュータビジョンの現在：Transformerによる変革

この10年の深層学習の発展によって、画像処理をはじめとする様々なAI分野で飛躍的な性能向上が達成された。このような成功の秘密はネットワーク構造のデザインを通じて、我々の持つ様々な事前知識（帰納バイアス）をネットワークの中に柔軟に実現できる点にある。その典型的な例は、画像タスクにおける畳み込みニューラルネットワーク (CNN) である。この10年の画像タスクにおける深層学習の大きな成功も、CNNの持つ画像処理に対する帰納バイアスをベースにしていると言って良い。

ところが2020年になり、畳み込み構造に一切依存しない深層学習 Transformer によって、最先端のCNNに匹敵する性能を実現できることが判明した。この新しいモデルは Visual Transformer (ViT) と呼ばれる。また翌年には MLP-Mixer と呼ばれる手法で、Transformer を使わずとも、多層パーセプトロン (MLP) をベースにしたモデルだけでも同等の性能の画像認識が可能であることが判明した。

このように、現在 CNN パラダイムとは全く別の画像認識の方向性が明らかになりつつある。この ViT、及び MLP-Mixer の発見は、データセットサイズがスケールアップした深層学習では、必ずしも CNN のような強い帰納バイアスは必要でなくなることを実証している。また訓練データセットがスケールしていくとやがて ViT が CNN を凌駕することから、場合によっては強い帰納バイアスがモデルの性能に限界をもたらすことも具体的にわかってきた。

これらのアプローチは現在、MetaFormer という統一的な枠組みで研究されている。このセミナーでは、MetaFormer の現状とこれまで判明している興味深い事実についてする。また、MetaFormer に関する我々のグループの成果についても紹介したい。