

早稲田大学大学院国際情報通信研究科

博士論文審査報告書

論 文 題 目

Semantic Image Recognition Methods
by Using Deep Learning

深層学習を用いた意味的画像認識手法

申 請 者

漢字名	ダン	ホアン アン
英字名	Hoang Anh	DANG

2020年8月

従来、人工知能は、主として知的エージェントに関する研究領域と認識されていた。しかし、2012年に提案された AlexNet による深層学習のブレークスルーによって、各種のメディア、市場関係者及び研究者の間では、人工知能はより広い意味で語られるようになってきた。例えば、コンピュータビジョン、各種のパターン認識、コンピュータによる自然言語処理、ロボット工学等も、人工知能に関する研究分野と認識されるようになってきている。つまり、現在、人工知能は、多岐に渡る人間の知能や理解に関する能力を、様々な応用分野で機械によって実現する研究課題として理解されている。

深層学習実現の先駆けとなった AlexNet を契機に、深層学習は理学及び工学を含んだ様々な分野でその応用が研究されている。深層学習が最もその有効性を示しているのは、主に、コンピュータビジョン、自然言語処理、知的エージェントの分野である。特にコンピュータビジョンの分野においては、畳み込みニューラルネットワークの登場によって、人間のレベルに達する画像分類性能が 2016 年に示された。これにより、世界のトップ研究者が集り、10 年弱に渡って画像認識及び画像分類性能を競った国際会議である ILSVRC (ImageNet Large Scale Visual Recognition Challenge) は、2017 年にその役目を終えた。このような深層学習研究の進展により、より困難な研究課題へと研究分野がシフトしている。例えば、セマンティックセグメンテーション、生成モデル、行動認識、画像質問応答等の研究課題が現在精力的に研究されており、より正確な、また、より深い意味的画像認識の研究が様々な角度から行われている。

以上を背景として、本論文では、深層学習を用いた意味的画像認識手法に関して論じている。具体的には、(1) ラスタ画像中の図形構成要素を意味的に理解し、ベクトル画像の一種である SVG (Scalable Vector Graphics) への変換を強化学習によって行う手法、(2) 提案するラベルプーリング手法等によってストリートファッション画像中の服飾アイテムをより正確に、かつ、意味的に理解するセマンティックセグメンテーション手法を提案し、従来手法と比較して、提案手法の優位性及び有効性について明らかにしている。

以下、各章ごとに概要を述べ、評価を加える。なお、本論文は英語で執筆されている。

第 1 章「Introduction」では、研究の背景、目的及び論文の構成について述べている。

第 2 章「Deep Learning」では、本論文に関連する研究として、ニューラルネットワーク、深層学習及び知的エージェントにおける種々のマイルストーンを整理している。本章は、それらの技術を系統的に分かりやすく整理した点を評価できる。

第 3 章「SvgAI」では、強化学習を利用した、ラスタ画像をベクトル画像である SVG へ変換する手法について述べている。まず、ラスタ画像からベクトル画像への変換に関する従来手法の問題点を整理し、それらの分析を行っ

ている。本章で提案する手法は、SVG エディタと名付けられた SVG 画像編集モジュールによって、与えられたラスタ画像と同等の SVG 画像を生成する手法であり、生成過程には Q-Learning と Policy Gradient を利用する強化学習を用いている。ここでは、より正確な SVG 画像を生成するため、SVG 画像生成プリミティブを基本図形の付加とそれらの要素（位置、線幅、線の色、領域の色等）の変更の二種類に分類して効率よく SVG 画像を生成する過程を学習する Dual ϵ Greedy 手法、並びに、必要最小限の探索によって得られる報酬を推定する Policy Gradient 手法を提案している。実験結果から、従来の学習手法と比べ、提案手法が優れていることを示している。

本章では、深層学習を利用したラスタ画像からベクトル画像への変換フレームワークを新しく提案した点、ラスタ画像から SVG 画像への変換を効率よく行う新しい学習方法を提案した点、並びに、これらの提案手法によって SVG 画像が従来手法よりも正確に生成できる点が評価できる。

第 4 章「Semantic Segmentation for Street Fashion Photos」では、セマンティックセグメンテーションによるストリートファッション画像の意味的分割手法について述べている。セマンティックセグメンテーションに関しては、従来、SegNet、DeepLabv3+、PSPNet、U-Net 等の深層学習による手法が提案されている。これらに対して、本章では、スケールリングを伴ったエンコーダ・デコーダネットワーク構成中で各層間の情報を相互的に利用する新しい深層学習手法、複数の補助的損失を利用した新しい学習手法、並びに、二次元のラベルプーリングに基づくスケールリングを利用した新しい学習手法を提案している。これらの提案手法を、ストリートファッション画像を 5 万枚強含む ModaNet データセットに対して適用し、従来手法と比較して評価している。従来の評価指標である mIoU (mean Intersection over Union) に加え、個々のクラスの分類精度とノイズを考慮した新しい評価指標 mIoU+ を提案し、両方の評価指標で評価を行っている。実験の結果、従来手法に比べ、提案手法は分類性能が高く、また、学習と推定に必要な時間も削減できることを示している。

本章では、より正確なセマンティックセグメンテーションのための新しいネットワーク構成を提案した点、そのネットワークを効率的に学習する手法を提案した点、並びに、ストリートファッション画像によく見られる小さな服飾アイテムを精度よく認識できるラベルプーリング手法を新たに提案した点が評価できる。また、性能及び計算量の点でも、提案手法は従来手法と比較して優れている点も評価できる。

第 5 章「Conclusion」では、本論文で得られた成果をまとめ、考察し、今後の課題について述べている。

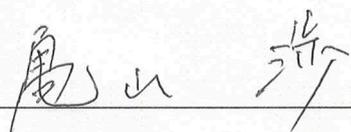
以上、要するに本論文は、意味的画像認識に関して、ラスタ画像から SVG 画像への変換、並びに、ストリートファッション画像のセマンティックセグメンテーションを具体的課題として取り上げ、それらに関する新しい手法を

提案し、その有効性を明らかにしたものである。これらの研究成果は、意味的画像認識の研究分野に新しい貢献をもたらすものであり、国際情報通信学の発展に寄与するところ極めて大きい。よって、本論文は博士（国際情報通信学）の学位を授与するに値するものと認める。

2020年8月20日

審査員：

主任 早稲田大学教授（専門：情報通信工学）
工学博士（早稲田大学） 亀山 渉



早稲田大学教授（専門：情報通信工学）
工学博士（北海道大学） 渡辺 裕



早稲田大学教授（専門：情報通信工学）
博士（工学）（東京大学） 甲藤 二郎

