公開

（事後評価・研究者用）

研　究　報　告　書

「マルチメディアデータから新しい概念を発見する高階モデル学習」

研究期間： 2018 年 4 月～2020 年 3 月
研究者番号： 50203
研 究 者： 井上　中順

1.　研究のねらい

本研究の狙いは，マルチメディアデータから新しい概念を発見・創出する手法の確立である．コンピュータによる画像の認識・理解に関する従来の研究では，学習データ作成のために，大規模なデータに人手でアノテーション（タグ付け）を行うことが前提とされていた．しかし，近い将来，さらなるデータの大規模多様化が進むと，このようなアノテーションのコストが膨大化する．そのため，新しい認識対象となる物体や人物動作について，これまでよりも少量のアノテーションで，より高精度な学習を実現する手法が切望されている．

本加速フェーズ研究の基となった，1年4ヶ月の ACT-I 研究期間では，「学習用サンプルが与えられていない物体・動作・シーンを発見すること」を目標として掲げ，Zero-Shot Learning の研究を実施してきた．ここで，Zero-Shot Learning とは，学習とテストの段階で検出対象が異なることを想定したモデル学習の枠組みである．例えば，飛行機のモデルを，車・鳥など他の物体の画像と，物体間の関係を表す外部データから学習する．ここで，飛行機の画像は学習に用いず，辞書などの外部データを利用する点が，従来の教師あり学習とは異なる部分となる．

加速フェーズ研究期間では，この方式と従来の教師あり学習を統一し，数枚の画像から高精度にモデル学習を行う Few-Shot 学習の実現を目指す．さらに，Few-Shot 学習向けの評価用のデータセット（画像 URL 集）を作成し，これまでよりも多角的な評価方針を提案，それに関して国際ワークショップの開催を目標とする．

2.　研究成果

（1）概要

加速フェーズ期間における主な成果と実績は，以下の2点である．

（A）大規模データに対する Few-shot 適応学習手法の提案

（B）Few-Shot Verb Image データセットの作成

前者については，マルチメディア情報処理分野のトップカンファレンスである ACM Multimedia に論文(フルペーパー)が採択されている．これは Zero-Shot 学習と Many-shot 学習（大量のデータを用いた教師あり学習)を統一的に定式化することで，少量の学習データに適応するモデルの構築を実現したものである．評価は大規模な画像・映像データで実施している．

後者のデータセットは，人物動作を表す画像の URL 集であり，Few-Shot 学習のベンチマーキングに適している．この研究に関しては，コンピュータービジョン分野のトップカンファレンスである IEEE/CVF International Conference on Computer Vision (ICCV)のワークショップを主催する形で発表を行った．本分野におけるワークショップは大規模なものであり，採択率が低く，国際的には顕著な実績として評価されるものである．

（2）詳細

（A）大規模データに対する Few-shot 適応学習手法の提案

本研究項目では，少数の画像データから物体，動作，シーンといった意味的コンセプトに関するモデルの学習を行うFew-Shot Adaptation 法を提案した．これは従来の教師あり学習(Supervised Learning)および Zero-Shot 学習を統一的に扱う枠組みである．

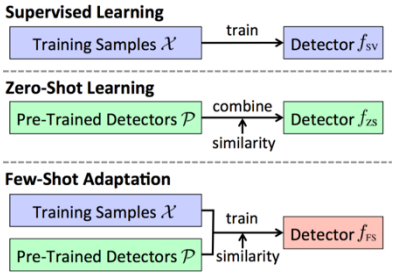

図 1. 従来法と提案法の入出力

図1は従来手法と本手法の仕組みを比較したものである．まず，従来の教師あり学習では学習データ X を入力とする．ここには，人手でタグ付けが行われた画像が含まれている．次に，Zero-Shot 学習では，事前学習済み識別器の集合 P を入力とし，複数の検出器を組み合わせることで，新たな対象に関する識別器を構成する．提案手法は，これら2つの方式を統一したものである．具体的には，教師あり学習と Zero-Shot 学習の最適化問題を，目的関数の重み付け和を取ることで統合し，サンプル数が少ない場合は Zero-Shot 学習に近い結果を，サンプル数が多くなるにつれて教師あり学習に近い結果を出力するように設計されているものである．

評価実験では，大規模映像データセットとして知られる，TRECVID Semantic Indexing データセットを用いて提案手法の効果を示した．主な実験結果は図2の通りである．ここでグラフの横軸は学習サンプル数，縦軸は Mean Average Precision (%)である．提案手法は Few-Shot Adaptation（赤)のもので，線形モデルに基づいたもの(Linear)とカーネル化を行ったもの(Kernelized)の2通りを，3種類の従来手法と比較している．Zero-Shot 学習用の事前学習モデルは ImageNET および Places データセットのものを導入している．本実験より，従来手法よりも提案手法の方がどの場合においても高い精度を示していることが分かった．これは，サンプル数が増加するに連れて，Zero-Shot 学習から教師あり学習への滑らかな遷移が行われたためである．

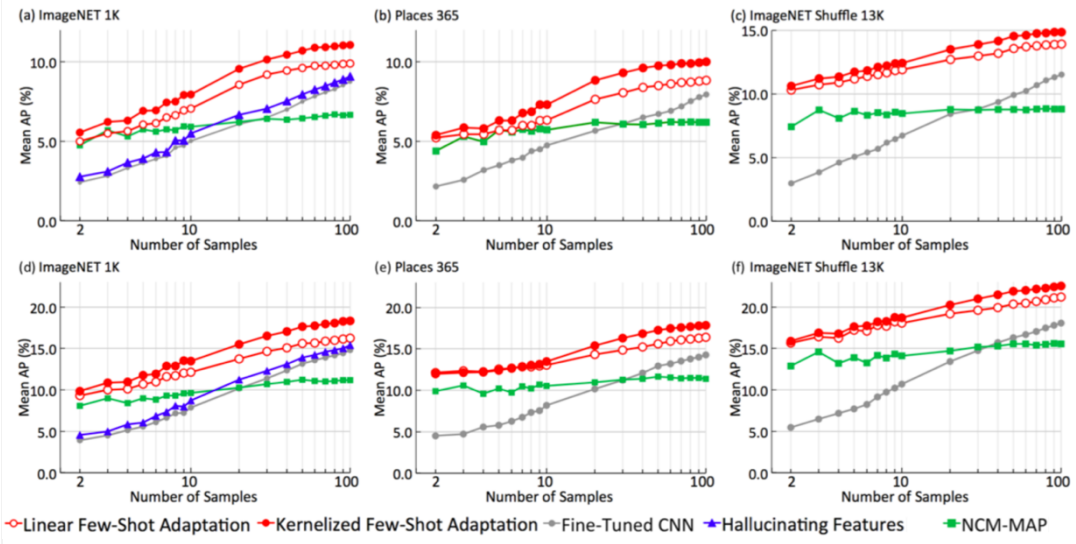

図 2. TRECVID データセットにおける Few-Shot 学習の実験結果

(B) Few-Shot Verb Image データセットの作成

Zero-Shot 学習および Few-Shot 学習の研究をさらに発展させるには，カテゴリ数の多いデータセットが必要であることが，これまでの評価実験で明らかとなったため，ベンチマーキングデータセットの作成に着手した．本データセットは，人物動作の決定的瞬間を捉えら画像に関する URL 集であり，1,000 カテゴリを有するものを作成し，国際会議 ICCV でワークショップを主催して公表した．

本データセットから得られる画像の例を図3に示す．動作を表すカテゴリは，WordNet の動詞概念(verb synset)単位であり，物体画像を有する ImageNet データセットの拡張として利用することができるものとなっている．当初は映像データへのタグ付けを検討していたが，動作が明確である瞬間を捉えることを重視して画像 URL 集を作成することとした．また，それによりクラウドソーシングによるアノテーションコストを削減することもできた．

本データに関する画像認識精度は，ResNet50 を用いた場合で表1の通りである．ImageNet データセットで事前学習を行うことで精度が向上するものの，依然として高精度な認識が難しい問題を設定できている．

本データセットは，今後，大規模なプロジェクトとして，100 倍程度のカテゴリ数を実現できると，大規模 Few-Shot 学習が実現し，意味的コンセプト間のモデル化に大きな発展を齎すことが期待できる．しかし，大規模化にはリンク切れ等の問題が生じてしまい，その解決には Instagram などの画像共有サービスを有する機関の協力が必要である．

表 1. 認識精度

Pre-Training	Train Accuracy		Val Accuracy	
	Top-1	Top-5	Top-1	Top-5
Scratch	73.1	92.2	11.6	24.2
ImageNet	75.0	90.6	14.9	30.4

また，本項目に関する国際ワークショップ提案は，2019 年度のコンピュータビジョンに関する国際会議 ICCV に採択され，2019 年 10 月にワークショップ開催が実現した．ワークショップでは，データセット公表とともに，IEEE/CVF の査読付きフルペーパーを 14 本，アブストラクトを 12 本採択し，当日の参加者は 300 名と盛況であった．現在のところ，国内機関からのトップカンファレンスにおけるワークショップ提案採択は稀であり，今後はさらに活動の枠を広げる必要がある．

図 3. 動作画像のサンプル

3.　今後の展開

本研究の成果は，少数サンプル環境下での学習を実現するものであり，大規模データ収集が困難な分野への波及効果が期待できる．特に，最近では医療分野など，収集可能なデータの量が限られている応用が注目を浴びており，巨大IT企業に限らず，深層学習技術の導入が望まれているため，研究成果をそれぞれの応用分野に適用することで，幅広く社会貢献ができる可能性がある．

4.　自己評価

・研究目的の達成状況

当初の第一目標であった，データセット作成に関するトップカンファレンス国際ワークショップの開催が実現した．また，Few-Shot 学習法もマルチメディア分野の国際会議フルペーパーで採択されたため，概ね目標は達成できている．

・研究の進め方

東京工業大学の有するスーパーコンピュータ TSUBAME3.0 を活用して研究実施を行った．大規模な評価実験が出来た点は評価できる．データセットおよび実験の規模は Google など他の研究機関が公表しているものに比べると小規模であるが，最大限に研究費を活用できている．

・研究成果の科学技術及び学術・産業・社会・文化への波及効果

Few-Shot 学習に関する成果はデータ収集が困難な分野への波及効果が期待できる．また，その原理は情報分野の発展に貢献するものである．

・研究課題の独創性・挑戦性

研究提案時，Few-Shot 学習という課題自体は独創性の高いものであった．データセットおよび実験の大規模化は高い挑戦性を有していたが，個人型研究では実施スピードが限られており，他機関と比較すると実施体制を見直す必要がある課題であった．

5.　主な研究成果リスト

（1）論文（原著論文）発表：0件

（2）特許出願

研究期間累積件数：0 件

（2）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

1. Nakamasa Inoue and Koichi Shinoda, Few-Shot Adaptation for Multimedia Semantic Indexing, ACM Multimedia, 2018.
2. Nakamasa Inoue, Chihiro Shiraishi, Aleksandr Drozd, Koichi Shinoda, Shi-wook Lee, Alex Chichung Kot, Activity Detection in Extended Video using Action Tubelets (VANT at TRECVID 2018), NIST TRECVID workshop, 2018.
3. Nakamasa Inoue, Multi-Discipline Approach for Learning Concepts Workshop, Organizer's talk, IEEE/CVF International Conference on Computer Vision, 2019.