

令和 6 年 6 月 18 日現在

機関番号：12612

研究種目：挑戦的研究（萌芽）

研究期間：2022～2023

課題番号：22K19808

研究課題名（和文）機能の重ね合わせを実現する深層学習におけるタスク融合学習

研究課題名（英文）Task Fusion Learning in Deep Learning

研究代表者

柳井 啓司（Yanai, Keiji）

電気通信大学・大学院情報理工学研究所・教授

研究者番号：20301179

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：本研究では、単一のニューラルネットワークに複数のタスクのための機能を同時に学習させ、学習した独立の機能の組み合わせ・重ね合わせによって、事前に学習した個別の単一機能とは異なる、新しい機能を実現することを目的として研究を行った。具体的には(1)条件信号による画像変換タスクの重ね合わせ、(2)Vision Transformerの継続学習、(3)Stable Diffusionによる学習なしでの任意の言葉に対する領域分割の実現、について成果を得ることができた。

研究成果の学術的意義や社会的意義

本研究によって、ニューラルネットワークには、異なる機能の同時学習に関してより柔軟に対応できる能力が備わっていることが示された。また、10億スケールの大規模画像言語ペアデータで学習されたテキストからの画像生成モデルには、テキストと視覚概念をピクセルレベルで対応付ける能力が備わっていることが示され、追加の学習なしで多様なタスクに活用できる可能性が大いに高まったと言える。今後は、この学習なしの能力を多様なタスクに対して実証し、それらの複合的な処理も学習無しで実現することを追求することで、大規模視覚言語モデルの応用可能性を大いに広げることが可能となる。

研究成果の概要（英文）：In this study, we started research to demonstrate that Neural Networks have general-purpose capabilities that are more similar to those of the human brain by having a single neural network learn functions for multiple tasks simultaneously, and by combining and superimposing the learned independent functions to realize new functions that are different from the individual single functions that were learned beforehand. The following three specific research projects have been studied. (1) Superimposition of image transformation tasks using conditional signals. (2) Continuous learning of Vision Transformer (ViT). (3) Stable Diffusion for region segmentation of arbitrary words without learning: Using a large-scale trained image generation model, we extract regions corresponding to words without additional learning.

研究分野：メディア情報学

キーワード：深層学習 継続学習 大規模モデル 視覚言語モデル

## 様式 C - 19、F - 19 - 1 (共通)

### 1. 研究開始当初の背景

深層学習技術の発展により、多種多様なタスクに対して高精度な画像認識・分類を実現することが可能となってきた。それに伴い、複数のタスクを同時学習するマルチタスク学習や、一度学習したモデルに新しいタスクや分類クラスを追加で学習する継続学習が広く研究されるようになってきている。マルチタスク学習では、それぞれのタスクを独立のモデルとして学習するよりもタスク間に相関がある場合、相乗効果によってお互いのタスクの性能向上が期待できる。また、追加学習では、工夫なしに追加の学習を行うと元々学習してあったタスクが破壊されてしまう破壊的忘却(catastrophic forgetting)が発生してしまうが、それを回避するための様々な方法が研究されている。

こうした学習の中には、条件信号を変更すると同じ出力から、異なるタスク処理の結果を出すタイプの学習があるが、我々はそれに着目し、条件信号を組み合わせることによって、多様なタスク、複合的なタスク、未学習のタスクに対応させることができることを発見し、その実証的な実験を行うことを計画した。

### 2. 研究の目的

本研究では、単一のニューラルネットワークに複数のタスクのための機能を同時に学習させ、学習した独立の機能の組み合わせ・重ね合わせによって、事前に学習した個別の単一機能とは異なる、新しい機能を実現し、Neural Network がより人間の脳に近い汎用的な能力を持っていることを実証することを目的とする。また、複数機能の組み合わせによる実用的な応用についても検討する。

### 3. 研究の方法

具体的には以下の3つの研究を実施した。

#### (1) 条件信号による画像変換タスクの重ね合わせ

深層学習ネットワークで可能な画像変換、例えば、ノイズ除去、領域分割、スタイル変換をそれぞれ、 $[1, 0, 0]$ 、 $[0, 1, 0]$ 、 $[0, 0, 1]$  という条件ベクトルとともに同一ネットワークに学習させ、 $[1, 1, 1]$  とするとそれぞれ単独の変換しか学習していないにもかかわらず、ノイズ除去をして、領域分割を行い、スタイル変換をすることが実現されるというのがこの条件信号による画像変換タスクの重ね合わせである。

#### (2) Vision Transformer の継続学習

当時、通常の CNN では継続学習の研究が多く行われていたが、Vision Transformer (ViT) では研究がなかったため、class incremental な継続学習を ViT で実現するアーキテクチャに関して、研究を実施した。

#### (3) Stable Diffusion による学習なしでの任意の言葉に対する領域分割の実現

Stable Diffusion は 10 億スケールの画像と文書のペアで学習された画像生成モデルであるが、単語の条件入力に cross-attention を利用している。この cross-attention の attention map は単語ごとに生成されるが、Stable Diffusion を画像編集に用いた場合の入力画像中の単語に対応する領域を map がおおよそ示していることが先行研究によって示されており、この性質を用いて任意の単語に相当する領域の推定を学習なしで実現できることを示す。

### 4. 研究成果

上記の(1)(2)(3)について研究成果を得た。

#### (1) 条件信号による画像変換タスクの重ね合わせ

3. で述べたように $[1, 1, 1]$  とするとそれぞれ単独の変換しか学習していないにもかかわらず、ノイズ除去をして、領域分割を行い、スタイル変換をすることが実現されることを期待したが、残念ながらそれは高品質には実現せず $[1, 1, 0]$ や $[1, 1, 1]$ などのすべての組み合わせについて少量の学習が必要であった。組み合わせの学習も行った結果、想定通りの変換が実現でき、期待する成果が得られた。なお、現在においては、大規模言語モデル(LLM)と画像生成・変換モデルを組み合わせることによって、プロンプトと呼ばれる文章の入力で画像変換を制御することができるようになっており、LLMを用いたプロンプトベースの入力で、それぞれ単独学習での画像変換を組み合わせることが可能になってきている。今後はその方向で、この研究を発展させる予定である。

#### (2) Vision Transformer の継続学習

この研究では、1つの Vision Transformer に複数の分類タスクを学習させるため、タスクに応じたごく少数のパラメータ(1タスクにつき全体の 0.5%程度)を導入することで、各タスクにネットワークを適応させるように学習させることを可能とした。

### (3) Stable Diffusion による学習なしでの任意の言葉に対する領域分割の実現

本研究の最終年の2年目は、画像変換の様な複数のタスクを融合可能なタスクを想定して、タスク融合学習の実現に向けて研究を進める予定であった。具体的には、領域分割、超解像、白黒画像のカラー化、インペインティング、スタイル変換、画像ドメイン変換、画像変換タスクなど、多様な画像変換タスクに関して、同時学習および組合せタスクの学習実験を行う予定であったが、最新の画像生成モデルである Stable Diffusion でのタスク融合学習での実現可能性について検討するうちに、大規模学習データで学習済の画像生成モデルである Stable Diffusion がゼロショットで任意の言葉に対応した領域分割が可能である性質を発見し、主にそれに関する研究に重点をおいて実施した。任意の単語に対応する領域の抽出を追加学習なしに学習済の Stable Diffusion Model をそのまま利用するだけで実現した。結果として大きな成果をあげることができ、日本最大の画像認識シンポジウムで優秀研究賞を受賞する成果を挙げる事ができた。

## 5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Higuchi Kotaro, Yanai Keiji	4. 巻 74
2. 論文標題 Patent image retrieval using transformer-based deep metric learning	5. 発行年 2023年
3. 雑誌名 World Patent Information	6. 最初と最後の頁 102217 ~ 102217
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.wpi.2023.102217	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 YUAN Honghui, YANAI Keiji	4. 巻 E107.D
2. 論文標題 Multi-Style Shape Matching GAN for Text Images	5. 発行年 2024年
3. 雑誌名 IEICE Transactions on Information and Systems	6. 最初と最後の頁 505 ~ 514
掲載論文のDOI (デジタルオブジェクト識別子) 10.1587/transinf.20231HP0010	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Benitez-Garcia Gibran, Takahashi Hiroki, Yanai Keiji	4. 巻 22
2. 論文標題 Material Translation Based on Neural Style Transfer with Ideal Style Image Retrieval	5. 発行年 2022年
3. 雑誌名 Sensors	6. 最初と最後の頁 7317 ~ 7317
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/s22197317	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Rosas-Arias Leonel, Benitez-Garcia Gibran, Portillo-Portillo Jose, Olivares-Mercado Jesus, Sanchez-Perez Gabriel, Yanai Keiji	4. 巻 23
2. 論文標題 FASSD-Net: Fast and Accurate Real-Time Semantic Segmentation for Embedded Systems	5. 発行年 2022年
3. 雑誌名 IEEE Transactions on Intelligent Transportation Systems	6. 最初と最後の頁 14349 ~ 14360
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TITS.2021.3127553	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

[学会発表] 計22件(うち招待講演 1件/うち国際学会 19件)

1. 発表者名 Zhongtao Chen, Yuma Hombu and Keiji Yanai
2. 発表標題 Mask-based Food Image Synthesis with Cross-Modal Recipe Embeddings
3. 学会等名 ACM Multimedia Asia (国際学会)
4. 発表年 2023年

1. 発表者名 Ryota Kaji and Keiji Yanai
2. 発表標題 VQ-VDM: Video Diffusion Models with 3D VQGAN
3. 学会等名 ACM Multimedia Asia (国際学会)
4. 発表年 2023年

1. 発表者名 Jingbin Xu, Junwen Chen and Keiji Yanai
2. 発表標題 Contextual Associated Triplet Queries for Panoptic Scene Graph Generation
3. 学会等名 ACM Multimedia Asia (国際学会)
4. 発表年 2023年

1. 発表者名 Junwen Chen and Keiji Yanai
2. 発表標題 QAHOI: Query-Based Anchors for Human-Object Interaction Detection
3. 学会等名 International Conference on Machine Vision Applications (MVA) (国際学会)
4. 発表年 2023年

1. 発表者名 Yuma Honbu and Keiji Yanai
2. 発表標題 Training-free Region Prediction with Stable Diffusion,
3. 学会等名 International Conference on MultimeMedia Modeling (MMM) (国際学会)
4. 発表年 2024年

1. 発表者名 Kento Terauchi and Keiji Yanai
2. 発表標題 CalorieCam360: Simultaneous Eating Action Recognition of Multiple People Using Omnidirectional Camera
3. 学会等名 ACM International Conference on Multimedia Retrieval (ICMR) (国際学会)
4. 発表年 2023年

1. 発表者名 Yingcheng Wang, Junwen Chen and Keiji Yanai
2. 発表標題 HowToEat: Exploring Human Object Interaction and Eating Action in Eating Scenarios
3. 学会等名 ACMMM Workshop on Multimedia Assisted Dietary Management (MADIMA) (国際学会)
4. 発表年 2023年

1. 発表者名 柳井 啓司
2. 発表標題 生成AI技術の基礎と展望
3. 学会等名 SID日本支部 講演会 「生成AIとメタバースが開く社会に向けたディスプレイ技術と課題」 (招待講演)
4. 発表年 2024年

1. 発表者名 本部 勇真, 山口 廉斗, 柳井 啓司
2. 発表標題 StableSeg: Stable Diffusionによるゼロショット領域分割
3. 学会等名 画像の認識・理解シンポジウム (MIRU)
4. 発表年 2023年

1. 発表者名 Kotaro Higuchi, Yuma Honbu, Keiji Yanai
2. 発表標題 Patent Image Retrieval Using Cross-entropy-based Metric Learning
3. 学会等名 Proc. of International Workshop on Frontiers of Computer Vision (IW-FCV), (国際学会)
4. 発表年 2023年

1. 発表者名 Daiki Shimizu, Keiji Yanai
2. 発表標題 Virtual Try-On Considering Temporal Consistency for Videoconferencing.
3. 学会等名 Proc. of the International Multimedia Modeling Conference (MMM) (国際学会)
4. 発表年 2023年

1. 発表者名 Jing Yang, Junwen Chen, Keiji Yanai
2. 発表標題 Transformer-Based Cross-Modal Recipe Embeddings with Large Batch Training.
3. 学会等名 Proc. of the International Multimedia Modeling Conference (MMM) (国際学会)
4. 発表年 2023年

1 . 発表者名 Kota Izumi,Keiji Yanai
2 . 発表標題 Zero-shot Font Style Transfer with a Differentiable Renderer
3 . 学会等名 Proc. of ACM Multimedia Asia ( 国際学会 )
4 . 発表年 2022年

1 . 発表者名 Junwen Chen,Keiji Yanai
2 . 発表標題 Parallel Queries for Human-Object Interaction Detection
3 . 学会等名 Proc. of ACM Multimedia Asia ( 国際学会 )
4 . 発表年 2022年

1 . 発表者名 Yuma Honbu,Keiji Yanai
2 . 発表標題 SetMealAsYouLike: Sketch-based Set Meal Image Synthesis with Plate Annotations
3 . 学会等名 Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management (MADIMA) ( 国際学会 )
4 . 発表年 2022年

1 . 発表者名 Kento Adachi,Keiji Yanai
2 . 発表標題 DepthGrillCam: A Mobile Application for Real-time Eating Action Recording Using RGB-D Images
3 . 学会等名 Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management (MADIMA) ( 国際学会 )
4 . 発表年 2022年



1. 発表者名 Kohei Yamamoto, Keiji Yanai
2. 発表標題 Text-based Image Editing for Food Images with CLIP
3. 学会等名 Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management (MADIMA) (国際学会)
4. 発表年 2022年

1. 発表者名 Shu Naritomi, Keiji Yanai
2. 発表標題 Real Scale 3D Reconstruction of a Dish and a Plate using Implicit Function and a Single RGB-D Image
3. 学会等名 Proc. of ACM MM Workshop on Multimedia Assisted Dietary Management (MADIMA) (国際学会)
4. 発表年 2022年

1. 発表者名 Mana Takeda, Keiji Yanai
2. 発表標題 Continual Learning in Vision Transformer
3. 学会等名 Proc. of IEEE International Conference on Image Processing (ICIP) (国際学会)
4. 発表年 2022年

1. 発表者名 Yuchen Qian, Kohei Yamamoto, Keiji Yanai
2. 発表標題 StyleGAN-based CLIP-guided Image Shape Manipulation
3. 学会等名 Proc. of International Conference on Content-based Multimedia Indexing (CBMI) (国際学会)
4. 発表年 2022年

1. 発表者名 Yuma Honbu, Keiji Yanai
2. 発表標題 Unseen Food Segmentation
3. 学会等名 Proc.of ACM International Conference on Multimedia Retrieval (ICMR) (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関