様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 １５ 日現在

機関番号：１３５０１
研究種目：基盤研究(C)（一般）
研究期間：2019～2022
課題番号：１９Ｋ１２１５５

研究課題名（和文）海馬と大脳皮質における宣言的記憶の形成モデル

研究課題名（英文）Declarative memory model between hippocampus and neocortex

研究代表者
服部　元信（Hattori, Motonobu）

山梨大学・大学院総合研究部・教授

研究者番号：４０２９３４３５

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：本研究では，脳を模倣した工学的に有用な知的システムの構築を目指し，脳における宣言的記憶の形成過程のモデル化を行い，以下の成果を得た．(1)神経細胞を精緻にモデル化したスパイキングニューラルネットワークの新しい教師あり学習法を考案し，脳のように破局的忘却を抑制した記憶形成が可能であることを示した．(2)海馬CA2を導入した新しい海馬モデルによって複雑なエピソードの記憶及び想起が可能となることを示した．(3)情報の覚えやすさと忘却のしやすさについて調査し，忘却の抑制に効果的な学習法を考案した．(4)長期記憶を担うモデルのロバスト性及び汎化性能を向上させる手法を考案した．

研究成果の学術的意義や社会的意義
意識的に思い出すことのできる記憶(宣言的記憶)は，思考や推論といった脳の高次情報処理に利用されているため，人間のように知的で柔軟な情報処理システムを構築するにあたっては，如何にして宣言的記憶の形成過程を模倣するかが重要である．特に人間特有の記憶については，脳の一部を損傷させて機能を解明するような研究は倫理的に不可能なため，計算モデルによる研究の果たす役割が大きい．本研究では，主に脳の海馬に関する知見に基づき，宣言的記憶を形成するモデルの構築，忘却を抑制しつつ記憶を形成するための学習方法の構築を行い，人間の記憶の形成過程を工学的に模倣するモデルの基礎としての成果を得た．

研究成果の概要（英文）：In human memory, memories related to facts and events are called declarative memories and are the basis of our intelligent information processing in the brain. In this study, we aimed to construct an intelligent system that mimics the brain and is useful from an engineering perspective. We modeled the formation process of declarative memory in the brain and obtained the following results: (1) We devised a new supervised learning method for spiking neural networks, which use elaborate models of neurons, and demonstrated the possibility of memory formation with suppression of catastrophic forgetting. (2) We showed that a new hippocampal model introducing the CA2 region enables the storage and retrieval of complex episodes. (3) We investigated the ease of memorization and forgetfulness of information and devised an effective learning method to suppress forgetting. (4) We developed learning methods to improve long-term memory robustness and generalization performance.

研究分野：情報工学

キーワード：海馬　大脳皮質　エピソード記憶　宣言的記憶　ニューラルネットワーク　破局的忘却　スパイキングニューラルネットワーク　生成モデル

１．研究開始当初の背景

脳の機能において記憶はとりわけ重要であり，我々人間の知的な活動は記憶を抜きにしては語れない．記憶は，まず毎日の個人的な出来事の経験，すなわち，エピソード記憶として獲得され，そこから時間的文脈が除去された事実に関する記憶（意味記憶）が形成されると考えられている．これら意識的に思い出すことのできる記憶（宣言的記憶）は，思考や推論といった高次な情報処理に利用されているため，人間のように知的で柔軟な情報処理システムを構築するにあたっては，如何にして宣言的記憶の形成過程を模倣するかが大変重要である．近年の神経心理学的な臨床研究や動物実験による神経科学的研究により，宣言的記憶には，脳の海馬と嗅周囲皮質（大脳皮質の一部）が重要な役割を果たしていることが明らかになっている．また，脳に与えられた宣言的記憶に関する情報は，初めに海馬に蓄えられた後，徐々に大脳皮質へと転写されていくことが強く示唆されている．しかし，(1) 宣言的記憶に関する情報が海馬にどのように蓄えられ，それがどのような仕組みで取り出され大脳皮質へ転写されるのか，(2) 大脳皮質に既に固定されている記憶を破壊することなく新しい情報をどのように追加するのか，(3) エピソード記憶からどのように時間的文脈が除去されて意味記憶が形成されていくのか，さらには，(4) 個々の意味記憶からどのようにして体系的な知識が形成されていくのか，などそのメカニズムの全容は未解明である．

本研究で対象とする宣言的記憶において，海馬が重要な役割を果たしていることは，てんかんの治療のために，海馬とその周辺を切除した患者の臨床的研究から明らかになった．その後，現在までの約半世紀に渡って，マウスなど動物の海馬を用いた神経科学的な研究や，海馬を損傷した患者を対象とした記憶課題による神経心理学的な研究が盛んに行われてきた．しかし，海馬のスライスによる実験や脳の一部の機能を遺伝的に制限する実験などは，実験動物では可能であっても，人間を対象とした研究では倫理的な問題が大きく行うことはできない．そのため，高次情報処理の基盤となる，人間に特有の宣言的記憶の解明には，計算モデルによる研究が極めて重要な役割を果たす．これまでにも，計算機科学の分野において，海馬の計算モデルや工学モデルがいくつか提案されてきた．しかし，従来の研究では，海馬やその一部のみをモデル化の対象とすることがほとんどであった．それに対し我々は，海馬のみならず，長期記憶の座である大脳皮質までを考慮し，記憶の形成過程を巨視的にシステムとして捉えた研究を行ってきている．

２．研究の目的

本研究では，我々のこれまでの研究を発展させ，宣言的記憶を形成する，より生物学的に妥当な海馬―大脳皮質モデルを構築することを目的とし，以下の課題に取り組むことを計画した．

(1) 宣言的記憶を可能とする海馬モデルとその特性

初期的な記憶の座である海馬では，次々と新しい記憶を蓄えていく必要があるため，高速な学習が不可欠である．一方，比較的小さい部位であるため，効率的な学習によって大記憶容量を実現していると考えられるが，一般に，高速な学習で大記憶容量を実現するのは非常に難しい．同時に，海馬においても，過去の記憶を破壊することなく新規な情報を記憶する機構が必要である．本研究では，神経科学的な知見を，生物学的に妥当なニューロンモデルと学習法を採用した海馬モデルに導入し，海馬に必要とされる機能が発現するかどうか，この他にどのような仕組みが必要となるのかを明らかにする．

(2) 海馬モデルからの記憶の抽出の仕組み

脳に入力された新規な情報は，大脳皮質における破局的忘却を回避するため，一旦，海馬に記憶され，その後，徐々に大脳皮質へと転写され長期記憶として固定されると考えられている．しかし，情報がニューロン間の重みに分散されて記憶される人工神経回路網において，そこに蓄えられている記憶項目を取り出すことは極めて困難である．一方，近年の研究によると，睡眠中に海馬からの記憶の抽出，及び大脳皮質への記憶の転写，固定が行われている可能性がある．ここでは，より生物学的に妥当なスパイキングニューロンモデルからなる海馬モデルに，睡眠時の脳活動に関する知見を採り入れ，海馬モデルに蓄えられた情報を自律的に想起し，長期記憶を担う大脳皮質モデルへ転写する仕組みを計算機実験により明らかにする．

(3) 長期記憶形成のための大脳皮質モデルと破局的忘却抑制の仕組み

海馬から抽出された情報は，最終的な記憶の貯蔵庫である大脳皮質に蓄えられる．その際，既に大脳皮質に蓄えられている情報を破壊することなく，新規な情報を追加的に学習する必要がある．通常の人工神経回路網では，新規な情報のみで追加学習を行うと，新規情報を記憶するために重みが書き換えられてしまうため，既に蓄えられていた情報の記憶が破壊されてしまう(破局的忘却)．そのため，従来，この問題は，記憶すべき情報を予め全て用意することなどで回避してきた．しかし，予め全ての学習データを用意することは困難なことが多い．何より，我々生物はこのような学習を行ってはいない．我々人間の脳がこの忘却をどのように回避しているかは未解明の問題である．人工的なニューラルネットワークは，学習によって自らの性能を改善していく手法であるため，脳のように新規な情報の

みで継続的に学習し，経験とともに性能が改善されていくことが望ましい．本研究では，スパイキングニューロンモデルからなる大脳皮質モデルを中心に，長期記憶を担うネットワークにおいて破局的忘却を抑制する方法の構築を行う．

(4) モデルの妥当性評価と工学的な有用性の向上

本研究で構築した宣言的記憶を形成する海馬―大脳皮質モデルの特性について，神経心理学的実験によって得られている人間の記憶特性との整合性について調べ，本モデルの妥当性を検証する．同時に，将来の工学的な応用を見据え，長期記憶を担うネットワークにおけるロバスト性，並びに汎化能力を向上させる方法について検討し，評価を行う．

３．研究の方法

本研究課題は以下のように実施した。

(1) スパイキングニューラルネットワークにおける破局的忘却の抑制

我々のこれまでの研究では，抽象度の高い形式ニューロンモデルで破局的忘却を抑制する方法を考案してきたが，より生物学的に妥当なスパイキングニューロンモデルにおいては，まだその方法が明らかになっていない．そこで，生物学的に妥当な教師あり学習法として提案されている，Dopamine-modulated STDP (DA-STDP)学習法に，適応的シナプス可塑性(Adaptive Synaptic Plasticity: ASP)を導入した学習法を考案した．これを入力層，興奮性・抑制性層，出力層からなるネットワークに適用し，破局的忘却の抑制性能の調査を行った．

(2) スパイキングニューラルネットワークのための教師あり STDP 学習法

スパイキングニューラルネットワーク(SNN)では，情報処理の最小単位である神経細胞を精緻にモデル化しており，学習には，生体の神経細胞間の学習則を模倣した，スパイクタイミング依存性シナプス可塑性則(STDP: Spike Timing Dependent synaptic Plasticity) が用いられている．しかし，学習の効率面，性能面において，単純なニューロンモデルからなるネットワークを誤差逆伝播学習法で学習させる従来手法のような優れた性能はまだ得られていない．ここでは，SNN の学習の効率性を重視した新たな学習法を構築することを目的とし，重みの減衰機構を導入した新しい教師あり STDP 学習法を考案した．従来の STDP 学習法では，教師なし学習が採用されていたため，識別問題に適用するには，学習後，各ニューロンにラベルの割り付けを行う必要があったが，本手法は教師あり学習であるためこのコストが不要になった．また，学習効率を改善するために，重みの減衰機構の導入も行った．

(3) CA2 の構造と機能を取り入れた海馬エピソード記憶モデル

脳の海馬は宣言的記憶の形成に重要な役割を果たしていると考えられており，各部位の解剖学的特徴や機能的な特徴が調べられて来た．こうした研究は，従来，主に海馬 CA1,CA3,DG といった部位に対して行われてきており，モデル化に際してもこれらの部位のみを考慮することが多かった．これに対し，近年，海馬 CA2 の研究が盛んに行われるようになってきた．CA2 は CA3 と CA1 の間にある小さな部位であるが，海馬の各部位と結合を有しており，海馬ネットワークにおいて中心的な役割を果たしているとも考えられている．また，CA2 は社会性記憶の形成に重要な役割を果たしていることがわかっているが，近年では時間的な情報の記憶にも深く関与していると考えられている．そこで本研究では，海馬モデルに CA2 を導入し，その役割を調査することを目的とする研究を行った．モデル化に際しては，CA2 に関する生物学的な知見に基づくとともに，θリズムに基づいた学習方法を取り入れた．

(4) データの覚えやすさと忘却しやすさの関連性を用いた破局的忘却の抑制

人工的なニューラルネットワークにおける破局的忘却を防ぐ最も簡単な方法は，新規な情報を学習する際に，それまでに記憶した情報も一緒に学習させることであるが，過去に学習した情報をすべて保存しておくことは現実的ではない．そもそも人間はこのような学習をせずとも，破局的忘却を生じることはない．ここでは，破局的忘却を抑制する一つの方法として，生成モデルを用いる手法の研究を行った．この手法では，生成モデルの学習によって，過去に学習したデータが生成され，それを新規な学習データと一緒に学習させることで破局的忘却を抑制する．過去のデータは生成モデルによって生成されるため，保存しておく必要はない．また，このようなモデルにおいて，データの覚えやすさと忘却のしやすさとの関係を調査した．

(5) ロバスト性向上のための学習法

近年，大規模なニューラルネットワークを大量のデータで学習させることにより，画像分類，物体認識などのタスクにおいて人間に匹敵するあるいは凌駕する性能が得られるようになってきている．その一方で，人間には知覚できない微小なノイズによって容易に誤認識が引き起こされるという脆弱性も明らかになっている．このような入力は，Adversarial Examples (AE) と呼ばれ，セキュリティが重視されるようなシステムで致命的な結果をもたらす可能性があるが，そもそも人間にはあり得ない誤認識を生じる点が人間の視覚モデルとして重大な欠陥といえる．こうした入力に対する耐性を向上させることは，海馬―大脳皮質モデルを工学的に応用する際のロバスト性を改善するのに有用である．このような背景から，ここでは，AE への耐性の向上を目的とし，まず，人工的なデータに

よる詳細な解析を行い，その結果を基に耐性を改善する手法を考案した．また，脳の視覚野における情報処理のモデル化を取り入れた手法の構築も行った．

(6) 汎化能力向上のための学習法

海馬―大脳皮質モデルの工学的な応用に際しては，ロバスト性だけでなく汎化能力もできるだけ向上させることが望ましい．ここでは，将来の応用を見据え，アンサンブル学習を用いた知識蒸留によって汎化能力を向上させることを目的とし，教師となるネットワークの多様性を強化する方法，並びに学習の進度に応じて徐々にアンサンブルを実施していく方向を考案し，その効果について調査を行った．

４．研究成果

(1) スパイキングニューラルネットワークにおける破局的忘却の抑制

従来の ASP では，ある程度破局的忘却を抑制した学習が可能であったが，教師なし学習法であるため，興奮性層のニューロンにラベルを割り付ける際には，過去に学習した全ての情報が必要になり，継続的な学習に適していなかった．これに対し，本研究では，興奮性・抑制性層と出力層の学習に DA-STDP 法を用いた手法を考案した．これにより，新規な情報のみでの効率的な学習が可能となった．さらに，興奮性・抑制性層と出力層の重みの正規化を双方向に 2 段階で行うことによって，優れた破局的忘却抑制性能が得られることがわかった．

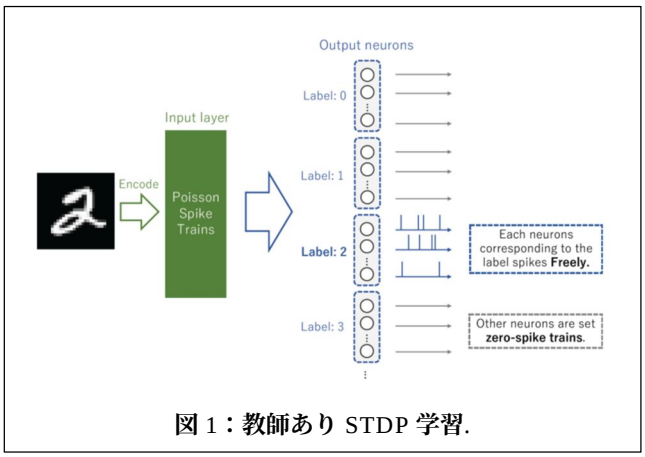

図 1：教師あり STDP 学習．

(2) スパイキングニューラルネットワークのための教師あり STDP 学習法

従来の SNN の学習では，STDP 学習則を教師なし学習として用いてきた．しかし，識別モデルとして利用するには，学習外で興奮性ニューロンへのラベル割り当て作業が必要であり，そのコストが大きかった．ここでは，STDP 学習則がローカルな学習であることに着目し，図 1 に示す新たな教師あり学習の方法を考案した．また，識別能力向上のために，あまり意味のない重みに抑制性の役割を持たせる重みの減衰機構を考案して導入した．計算機実験の結果，表 1 に示すように，従来法と比較してより小規模なネットワークにおいて，より少ない学習回数で優れた性能が得られることを明らかにし，本手法が学習効率の面で優れていることを明らかにした．さらに，新規な学習データのみによる追加学習が可能であり，破局的忘却を回避できることも明らかにした．

表 1：手書き数字データに対する精度(%)の比較．

モデル	出力層ニューロン数 100	400	1600
Diehl&Cook	77.22 ±0.467	87.89 ±0.411	88.99 ±0.480
This work	89.99 ±0.201	93.05 ±0.156	93.99 ±0.139

(3) CA2 の構造と機能を取り入れた海馬エピソード記憶モデル

本研究では，図 2 のように海馬モデルに CA2 を導入し，その役割を調査することを目的とする研究を行った．モデル化に際しては，CA2 に関する生物学的な知見に基づくとともに，θ リズムに基づいた学習方法として，TSP(Trisynaptic Pathway)，DSP(Disynaptic Pathway)，MSP(Monosynaptic Pathway)の順で重みの更新を行う手法を用いた．計算機実験の結果，CA2 を導入した海馬モデルでは，CA2 と CA3 の相互作用により，文脈情報を考慮したエピソードの記憶及び想起が可能となることがわかった．CA2 がなくても単純なエピソードの記憶及び想起は可能であったが，過去の情報に依存した複雑なエピソードの記憶及び想起には CA2 が必要であった．また，CA2 の規模が大きくなるとエピソードの記憶性能が向上することもわかった．

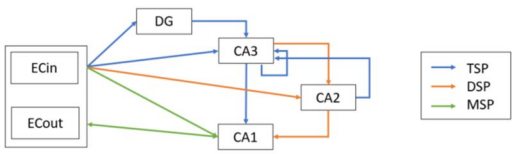

図 2：CA2 を取り入れた海馬モデル．

(4) データの覚えやすさと忘却しやすさの関連性を用いた破局的忘却の抑制

本研究では，破局的忘却を抑制する一つの方法として，生成モデルを用いる手法の研究を行った．この手法では，生成モデルの学習によって，過去に学習したデータが生成され，それを新規な学習データと一緒に長期記憶を担うネットワークに学習させることで破局的忘却を抑制する．過去のデータは生成モデルによって生成されるため，保存しておく必要はない．また，このようなモデルにおいて，データの覚えやすさと忘却のしやすさとの関係を調査した．図 3 では，Fashion-MNIST データの 8 番目と 9 番目のデータを初めに学習した後，4 番目と 5 番目のデータのみを追加的に学習したときの再現率の変化を表している．この図から，覚えやすいデータ(8 番目)ほど，後の学習によって忘却されやすいという傾向があることがわかった．これを利用して，学習の時点で予測した忘却しやすいデータを生成モデルに重点的に学習させる方法を考案したところ，図 4 に示すように忘却の抑制に効果があることがわかった．

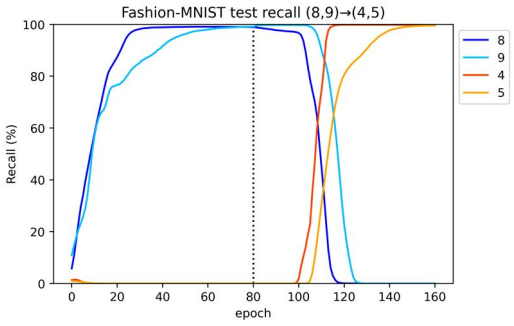

図 3：再現率の推移. (8,9)→(4,5)の順で学習.

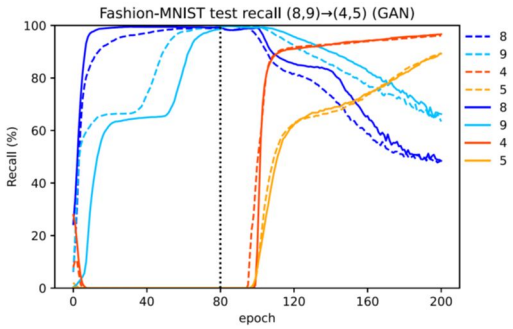

図 4：生成モデルによる忘却抑制．追加データ構成比：”8”:”9”=1:1(点線)，”8”:”9”==3:1(実線).

(5) ロバスト性向上のための学習法

AE への耐性を獲得する手法として，Adversarial Training という学習法が知られている．ここでは，まず，小規模な人工データを用いて，どのような原理で耐性が得られるのかを解析した．その結果，AE を通常の入力と区別する中間層ニューロンが学習によって獲得されることが耐性向上に寄与している可能性を示した．また，この結果を基に，当該ニューロンがより入力を区別しやすくなるように，AE の検出器を用いた新たな学習法を考案した．一方，脳の 1 次視覚野における情報処理を模倣することでも AE への耐性が得られることが明らかになっているが，ここに局所応答正規化を導入した方法を考案した．一般的な局所応答正規化では，近傍の特徴マップに存在するニューロンから抑制を受けることになるが，隣り合う特徴マップが似た特徴を抽出しているとは限らない．そこで，本手法では，フィルタの類似度を定義し，似た特徴を抽出している特徴マップから抑制を受ける手法を考案した．この抑制方法により，特徴マップに抽出される特徴がより明瞭になり，AE に対する耐性が向上することが期待できる．実際，手書き数字データに対するロバスト性の比較を行ったところ．図 5 に示すように提案手法によって優れたロバスト性が得られることがわかった．

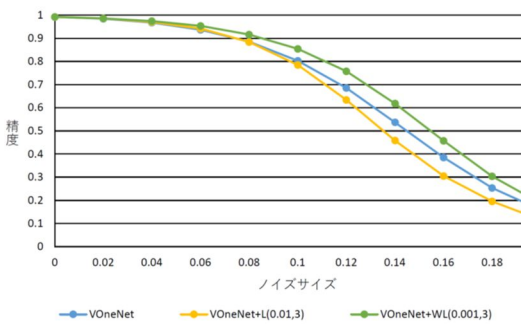

図 5：MNIST データに対するロバスト性.
青：従来法，緑：提案法.

(6) 汎化能力向上のための学習法

長期記憶を担うモデルの汎化能力を向上させる方法として，アンサンブル学習を用いたオンラインの知識蒸留法を考案した．この手法では，個々のネットワークの多様性を強化するために，ネットワーク毎に異なる学習データを与える方法を考案した．また，学習の初期において，アンサンブルから得られる教師信号は信頼性が低いため，学習の進度に応じて徐々にアンサンブルからの教師信号の影響が強くなっていく学習法を考案した．表 2 は，Tiny ImageNet データを用いた際の従来法との精度比較を表している．この結果から，アンサンブルに用いるネットワークが最小の 2 個の場合であっても，提案手法によって優れた汎化性能が得られることがわかった．

表 2：TinyImageNet データに対する汎化性能.

Model	ResNet-44	ResNet-32
KDCL-MinLogit ($\lambda = 0.1$)	49.78±0.12	47.83±0.29
KDCL-MinLogit ($\lambda = 0.3$)	49.28±0.46	47.98±0.19
KDCL-MinLogit ($\lambda = 0.5$)	48.68±0.24	47.19±0.63
KDCL-MinLogit ($\lambda = 0.7$)	46.55±0.15	45.44±0.11
KDCL-EDGE (proposed)	**50.84**±0.14	**49.06**±0.34

５．主な発表論文等

〔雑誌論文〕　計3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　0件）

１．著者名 小宮山亮太，服部元信	４．巻 Vol.J104-D, No.4
２．論文標題 Adversarial Examples の考察に基づく Adversarial Robustness の向上	５．発行年 2021年
３．雑誌名 電子情報通信学会論文誌D	６．最初と最後の頁 406-414
掲載論文のDOI（デジタルオブジェクト識別子） 10.14923/transinfj.2020PDP0004	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 服部元信	４．巻 51
２．論文標題 海馬から大脳皮質への宣言的記憶の転写モデル	５．発行年 2019年
３．雑誌名 細胞	６．最初と最後の頁 43-46
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 高木純平，　服部元信	４．巻 139
２．論文標題 自己蒸留によるDNNの蒸留の効率化	５．発行年 2019年
３．雑誌名 電気学会論文誌C	６．最初と最後の頁 1509-1516
掲載論文のDOI（デジタルオブジェクト識別子） 10.1541/ieejeiss.139.1509	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 －

〔学会発表〕　計6件（うち招待講演　0件／うち国際学会　1件）

１．発表者名 宮澤隆太，服部元信
２．発表標題 局所応答正規化を導入した視覚野モデルによるAdversarial Examplesへのロバスト性の向上
３．学会等名 情報処理学会第85回全国大会
４．発表年 2023年

１．発表者名
Qiuyue Li, Motonobu Hattori, Wei Zhang and Zhigang Gao

２．発表標題
Online Knowledge Distillation via Collaborative Learning with Enhanced Diversity and Gradual Ensemble

３．学会等名
IEEE International Workshop on Computational Intelligence and Applications (IWCIA)（国際学会）

４．発表年
2021年

１．発表者名
宮澤隆太，服部元信

２．発表標題
特徴の明示的学習と検出器を利用した Adversarial Examples の防御手法

３．学会等名
令和3年度電気学会東京支部学生発表会

４．発表年
2021年

１．発表者名
荒木裕史，服部元信

２．発表標題
スパイキングニューラルネットワークのための適応型重み減衰を取り入れた教師ありSTDP学習

３．学会等名
情報処理学会第83回全国大会

４．発表年
2021年

１．発表者名
小宮山亮太，服部元信

２．発表標題
Adversarial Training の考察に基づくAdversarial Examples への耐性の向上

３．学会等名
電子情報通信学会ニューロコンピューティング研究会

４．発表年
2020年

１．発表者名 小宮山亮太，服部元信
２．発表標題 Adversarial Training の考察に基づくAdversarial Examples への耐性の向上
３．学会等名 電子情報通信学会東京支部学生会第25回研究発表会講演論文集
４．発表年 2020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
研究協力者	小宮山　亮太 (Komiyama Rryota)		
研究協力者	高木　純平 (Takagi Jumpei)		
研究協力者	荒木　裕史 (Araki Hiroshi)		
研究協力者	兪　雪蕾 (Yu Xuelei)		
研究協力者	李　秋月 (Li Qiuyue)		

6．研究組織（つづき）

	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
研究協力者	西牟田　航平 (Nishimuta Kohei)		
研究協力者	宮澤　隆太 (Miyazawa Ryuta)		
研究協力者	佐直　祐弥 (Sajiki Yuya)		

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