

課題番号	Q20J-02
課題名（和文）	深層学習による劣悪品質データの認識と高精細データの生成に関する研究
課題名（英文）	Development of deep neural network models for low-quality data recognition and high-quality data generation
研究代表者	所属（学部、学科・学系・系列、職位） 理工学部 理工学科 理学系 准教授
	氏名 日高 章理
共同研究者	所属（学部、学科・学系・系列、職位）
	氏名
	所属（学部、学科・学系・系列、職位）
	氏名
	所属（学部、学科・学系・系列、職位）
	氏名
	所属（学部、学科・学系・系列、職位）
	氏名

研究成果の概要（和文）

本研究では大脳視覚野と AlexNet の画像刺激に対する応答特性を比較分析し、AlexNet の初期層が V1 野 の情報表現と類似する特性を獲得している可能性や、中高次層の情報表現特性が V4 野の特性と対応する可能性があることを確認した。また既存の物体検出手法である Faster R-CNN 法に Attention 機構を導入した手法を提案し、従来法の検出精度を向上し得ることを示した。また低画質顕微鏡画像の画質復元手法として SSIM U-net 法を提案し、Pix2pix 法と比べて微小粒子等の「ねつ造」を抑えた復元結果が得られることを示した。

研究成果の概要（英文）

First, we analyzed the response of the visual cortex in monkey's brain and AlexNet, and confirmed several similarities between the visual cortex and AlexNet; For example, the early layers of AlexNet may have acquired characteristics similar to those of the V1 cortex. Second, we proposed object detection methods that introduce several attention mechanisms into Faster R-CNN which have improve detection accuracy. Third, we proposed SSIM U-net for restoration of low-quality microscopic images, and showed that our method can restore images with less "false creation" of non-existing small particles, compared to Pix2pix.

1. 研究開始当初の背景

脳の処理過程を模した計算モデルであるニューラルネットワーク(以下 NN)の発展系である深層 NN が 2010 年代前半から目覚ましい発展と成功を示し続けている。特に 2010 年代後半からは、深層 NN によって画像や音声などのデータの意味内容を自動理解するだけでなく、そのようなデータを自動生成することも可能になってきていた。しかし複雑なネットワーク構造の中でどの計算過程が最終的な認識や生成の結果に最も寄与しているのか、あるいは悪影響を及ぼしているのかなど、内的情報処理の素性が不明であるという問題があった。またネットワーク内部のパラメータを最適化(学習学習)する際に用いるデータは、多くの場合ある一定の品質(画質・音質等)範囲内にあるため、その品質を大きく下回る入力データが学習後に与えられたときに十分な認識を行えないことや、逆にその品質を超えるレベルのデータ生成を行えない、などの問題があった。

2. 研究の目的

本研究の主な目的は、(1)深層学習モデルの視覚刺激への反応特性の分析、(2)劣悪品質データを高精度認識する深層学習手法の開発、(3)高品質データ(画像・音声等)を生成する深層学習手法の開発、の3点である。以下において(1)~(3)それぞれの研究方法およびその成果を示す。

3. 研究の方法

上記の目的(1)~(3)ごとに研究方法を示す。

(1) 深層学習モデルの視覚刺激への反応特性の分析

深層 NN の中でも最も基本的な物体識別用ネットワークである AlexNet は、5 つの畳み込み層、3 つの最大プーリング層、1 つの平均プーリング層、および3 つの全結合層より構成される畳み込みニューラルネットワーク(以下 CNN)の一種である。各畳み込み層および最終層を除く全結合層にはそれぞれ活性化層として ReLU 関数が付随する。本研究では機械学習ライブラリ PyTorch で実装されている AlexNet モデルを大規模な自然画像の教師ラベル付

きデータセットである ImageNet のサブセット(1000 クラス、約 120 万枚)により学習した。誤差関数には Cross entropy を用い、バッチサイズは 128、最大エポック数は 90、最適化手法には SGD 法を用い、誤差逆伝播学習を行った。学習係数は初期値を 0.01 とし、25 エポックごとにその値を 1/10 とした。

本研究ではこのようにして学習した AlexNet の内部応答とニホンザル(Macaca fuscata)の脳視覚野の生理的応答を比較した。ただし、ニホンザルの生理的応答のデータは本研究で取得したものではなく、他の研究グループが過去に取得したデータを提供されたものである。当該の生理的データはニホンザルへ石、樹皮、葉、花、果物、蝶翅、羽毛、皮の 8 カテゴリの自然物体表面画像(各 8 枚、合計 64 枚)を呈示し、初期視覚皮質(V1)、中次視覚皮質(V4)、下側頭皮質(IT)の神経活動電位を記録して得られたものである。本研究ではこの実験により計測された刺激呈示期間中の平均発火率と、当該実験で用いられた自然物体表面画像を AlexNet に入力して得られた各階層における中間応答のベクトル値の相関を算出することで、脳視覚野と AlexNet の類似性・相違性を定量的に検証した。

より具体的には、まず AlexNet に i 番目および j 番目の画像を入力した際の第 l 層における中間応答ベクトル $\mathbf{f}_i^l, \mathbf{f}_j^l$ との間で算出した非類似度 d_{ij}^l から成る非類似度行列を D^l とする。同様に、ニホンザルに i 番目および j 番目の画像を呈示した際の V1 野(または V4 野または IT 野)の神経活動電位データの間の非類似度 d_{ij}^{V1} (または d_{ij}^{V4} または d_{ij}^{IT}) から成る非類似度行列を D^{V1} (または D^{V4} または D^{IT}) とする。ここで、AlexNet から得られた非類似度行列 D^l とニホンザルの神経活動電位から得られた D^{V1} (または D^{V4} または D^{IT}) の間の相関値を求めると、その値の高低により、AlexNet の第 l 層とニホンザル脳視覚野の V1 野(または V4 野または IT 野)の反応特性の類似性・相違性を定量化することができる。本研究ではこのアプローチに基づき、AlexNet の低次層から高次層に至るまでの脳 V1 野(または V4 野または IT 野)

との相関値の高低や変化の様子を分析し、大脳視覚野と CNN の類似性・相違性について議論した。

(2) 劣悪品質データを高精度認識する深層学習手法の開発。

近年発達が著しい自動運転技術などで必須の要素となる路上物体検出問題において、夜間・大雨・降雪・濃霧などの視界劣悪な状況下では、自転車周辺の車両・歩行者などの画像中での映り方が著しく不鮮明となり、見落としや認識の遅れなどを招く恐れがある。しかし、そのような著しく不鮮明な対象の周辺には、同種の物体が比較的鮮明な状態で写っていることも多い。

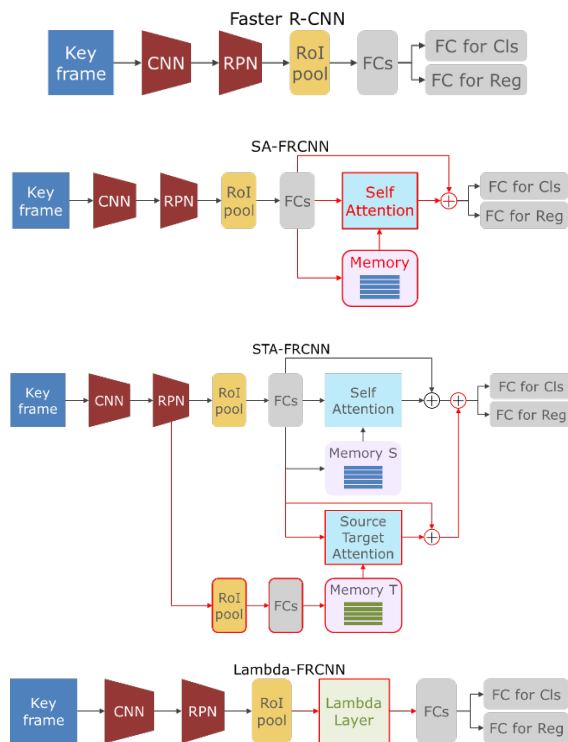
例えば下図左の画像のみを見たときに、一目でここに写っている物体が何かを確信を持って正しく視認することは難しいかもしれない。しかし「そのすぐ近くに下図右のような物体（＝歩行者）が写っている」という追加の情報が与えられると、我々は「下図左の物体も歩行者だろうか」という推測を自然に行うようになる（実際、下図左は街灯の脇にたずむ歩行者である）。



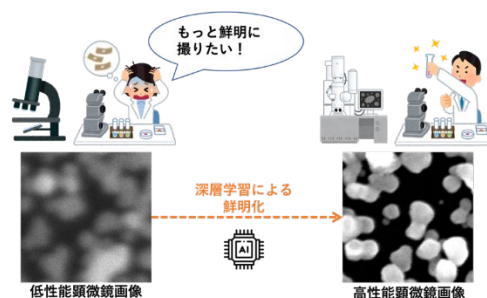
本研究では、既存の深層物体検出モデルである Faster R-CNN 法（以下 FRCNN）に対して上記のような認知プロセス、すなわち判別困難な物体の周辺にある比較的判別しやすい関連物体の「見え情報」を動的に検出プロセスに組み込む手法を提案した。

具体的には、FRCNN の内部モジュールである Region Proposal Network (RPN)が出力する物体候補領域群から得られる特徴ベクトル集合に Self-attention 機構、Source-target attention 機構、ならびに Lambda Layer 機構のいずれかを適用し、新たな特徴ベクトル集合に変換した上で最終的な物体種別の判定および正確な位置座標の推定を行う改良型 FRCNN 法を提案した。これにより、従来法では候補領域ごとにその領域内のみから得られる特徴

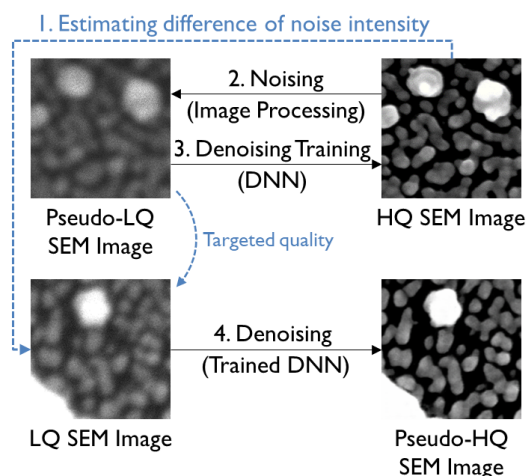
表現に基づいて最終的な種別判定や位置推定が行われていたものに対し、提案法では候補領域ごとにその周辺にある別の候補領域の「見え情報」を動的に取り込んでより高精度に種別判定・位置推定を行う機構を実現した。



(3) 高品質データ(画像・音声等)を生成する深層学習手法の開発。



本研究では比較的低価格（約 500 万円）の走査型電子顕微鏡（以下 SEM）で撮影した画像を高価格（約 5000 万円）の SEM で撮影された画像並の画質・解像度まで鮮明化する手法の開発を行った。そのためのアプローチとして、高価格 SEM で撮影した画像を低価格 SEM の画質並みに人工的に劣化させ、その「劣化写像」の逆写像を深層 NN の一種である画像翻訳変換モデルで学習する、というアプローチ（人工劣化復元法）を用いた。



画像翻訳変換モデルとし U-net 法と Pix2Pix 法を用いた。両手法の復元（鮮明化）結果を比較してそれぞれの長所・短所を分析したところ、まず Pix2Pix 法の方が U-net 法と比べてより鮮明な復元結果が得られるものの、Pix2pix 法の内部で用いられている敵対的生成ネットワーク（以下 GAN）の副作用により、劣化画像中のノイズ成分が復元画像においてニセの構造物（微小金粒子等）として「ねつ造」されやすい傾向があることがわかった。他方、U-net による復元結果は Pix2pix 法と比べて画像全体のボケ成分（Blur）を取り除ききれず、全体的にややぼやけた印象の結果となりやすいことがわかった。このため、U-net の復元結果の鮮明度をより向上させるため、損失関数を従来の平均二乗誤差（以下 MSE）から Structural Similarity（以下 SSIM）に置き換えた改良版 U-net を提案した。

4. 研究成果

上記の方法(1)~(3)ごとに研究成果を示す。

(1) 深層学習モデルの視覚刺激への反応特性の分析
ニホンザルの大脳視覚野の神経活動に基づく画像間非類似度行列と、1000 種類の物体から成る約 120 万枚の自然画像で学習した AlexNet モデルの各層の中間応答に基づく画像間非類似度行列の相関強度を分析したところ、AlexNet の初期層（第 1 ないし第 2 畳み込み層あたり）の中間応答が V1 野の神経活動と強く対応し、0.5~0.6 程度の正の相関を持つことが示された。しかしその相関強度は AlexNet の第 3

畳み込み層以降で 0.4 程度に減少していった。一方、V4 野の神経活動に基づく相関強度は、AlexNet の第 1 畳み込み層から第 3 畳み込み層にかけて 0.25 程度から 0.5 程度まで増加していった。さらに、AlexNet 第 3 畳み込み層以降の中高次層の応答に基づく画像間非類似度行列は、V1 野・V4 野・IT 野の中で V4 野と最も高い相関強度を示した。これらの結果は、AlexNet の初期中間層が V1 野の情報表現と類似する特性を獲得していることを示唆している。さらに、AlexNet の中高次層の情報表現特性が V4 野の特性と対応する可能性も示された。

当該研究成果については③の国内学会で口頭発表を行った。

(2) 劣悪品質データを高精度認識する深層学習手法の開発

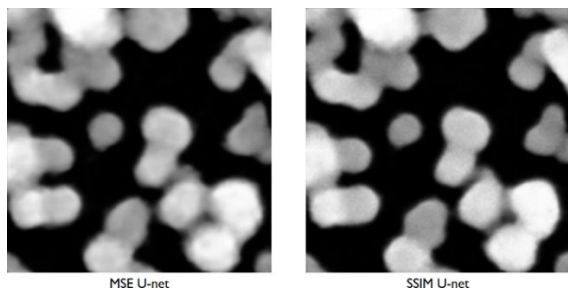
路上物体検出用の公開データセット BDD100K を用いて従来法（FRCNN）および 3 種類の提案手法（SA-FRCNN, STA-FRCNN, Lambda-FRCNN）を学習し、検出精度（Mean Average Precision: mAP）の比較を行った。提案法はいずれも従来法より mAP の値が高く、各種 Attention 機構の導入によって FRCNN の検出精度を向上し得ることが示された。提案法はいずれも画像中での面積が比較的大きな物体での検出精度（ AP_M , AP_L ）における精度向上が顕著であった一方、小物体に対する検出精度（ AP_S ）は向上しなかったため、この点の改善が今後の課題となる。

Model	mAP	AP_S	AP_M	AP_L
FRCNN	0.158	0.032	0.195	0.355
SA-FRCNN (t = 32)	0.162	0.028	0.196	0.380
SA-FRCNN (t = 0.32)	0.163	0.030	0.198	0.384
STA-FRCNN (t = 32)	0.165	0.029	0.201	0.392
STA-FRCNN (t = 0.32)	0.162	0.029	0.199	0.373
Lambda-FRCNN	0.170	0.032	0.203	0.397

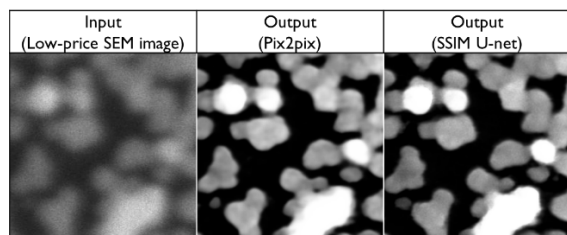
当該研究成果については④の国際学会で査読付き Proceedings 論文として発表した。また同論文の口頭発表においてその内容が高く評価され、④の Best presentation award を受賞した。

(3) 高品質データ(画像・音声等)を生成する深層学習手法の開発。

下図左が従来の MSE U-net, 右が提案法である SSIM U-net による人工劣化画像の復元結果である。SSIM の導入により, 従来法による復元結果で見られたボケの影響をより抑えた復元結果が得られるようになった。



下図は実際の低価格 SEM 画像を高価格 SEM 並の画質に復元した結果である。提案法 (SSIM U-net) の復元結果は Pix2pix 法と比べるとかすかにボケが残った復元結果となったものの, 復元の際にノイズ成分を微小粒子として描画してしまう「ねつ造」の発生件数は少なかった。今後は SSIM U-net の復元結果を Pix2pix 法と同等以上に鮮明化すること, ねつ造粒子の発生を 0 に近づけるための新たな機構の開発などが課題となる。



当該研究成果については②の国際学会で査読付き Proceedings 論文として発表した。

なお, 上記(1)~(3)に関する発展的研究課題が 2022 年度科研費基盤研究(C)「高難度データ認識と高精度データ生成のための「注意深い」深層学習モデルの開発」として採択されたため, 本研究における未解決問題や今後の課題は上記の科研費課題において引き続き解決を図っていく。

5. 主な発表論文等

[国際会議紀要論文] (計 2 件)

- ① Ryunosuke Ikeda and Akinori Hidaka, "Improvement of On-Road Object Detection using Inter-Region and Intra-Region Attention for Faster R-CNN", Proceedings of the 28th International Workshop on Frontiers of Computer Vision (IW-FCV2022), OS6-2 (15 pages), Feb. 2022. [査読有]
- ② Koshiro Nagano, Yoshiharu Mukouyama, Takashi Nishimura, Hiroyuki Fujioka, Kenji Watanabe, Takio Kurita, and Akinori Hidaka, "Noise Reduction of SEM Images using U-net with SSIM Loss Function," Proceedings of the 52nd ISCIE International Symposium on Stochastic Systems Theory and Its Applications (SSS'20), pp. 65-72, Osaka, Oct. 29-30, 2020. DOI:10.5687/sss.2021.65. [査読有]

[学会発表] (計 1 件)

- ③ 我妻伸彦, 日高章理, 田村弘, "サル視覚皮質と深層畳み込みニューラルネットワークに基づく物体識別モデルの画像情報表現", 電子情報通信学会ニューロコンピューティング研究会, 信学技報, vol. 121, no. 338, NC2021-36, pp. 30-33, 2022 年 1 月。

[受賞] (計 1 件)

- ④ IW-FCV2022 Best Presentation Award Presented to Ryunosuke Ikeda and Akinori Hidaka for their presentation of the paper entitled "Improvement of On Road Object Detection using Inter Region and Intra Region Attention for Faster R CNN," February 22, 2022.