CH-002

少量のラベル付きデータを用いた蚊の分類学習方法の検討
A Study on a Mosquito Classification Learning Method Using a few Labeled Data

大城 慶知[1)]　遠藤 聡志[2)]　斎藤 美加[3)]
Yoshitomo Oshiro　Satoshi Endo　Mika Saitou

1　はじめに

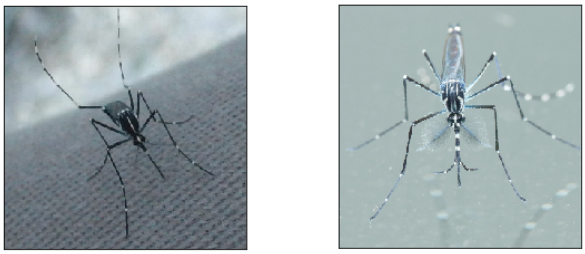

図 1　ヒトスジシマカ (左) とネッタイシマカ (右)

蚊媒介性感染症のデング熱は分布の拡大、罹患率、重篤度から公衆衛生上の脅威である。沖縄や日本には、デング熱を媒介するヒトスジシマカが広範囲で生息しているが、より媒介能の高いネッタイシマカの移入に関して監視する必要があり、蚊の簡便で正しい分類は喫緊の課題である。本研究では、調査の補助のために高精度な蚊の分類モデルを作成することを目的とする。用いるデータセットはヒトスジシマカとネッタイシマカの合計して 375 枚しかなく、2 クラスともハエ目カ科に分類されており似た特徴を持っている。一般的に少量のラベル付きデータセットを用いた分類問題では、転移学習を用いることが多い。しかし、蚊の分類問題では蚊のみを捉え、かつ重要な特徴量が微小な領域に含まれることもあり適切な重みを学習することが難しい。よって、ラベルのない蚊に近いデータセットを用いて教師なし事前学習を行い, 小さい特徴量を分類するための手法の検証を行う。近年、教師なし事前学習で教師あり事前学習を超えたと報告される研究がいくつかあり、そのうちの一つが対照学習を用いた He, Kaiming らの Momentum contrast for unsupervised visual representation learning(MoCo)[1] である。Kaiming ら [1] は画像間の得られた特徴マップを効率的に比較することで、様々なタスクやデータセットで教師あり学習を凌駕し、質の高い特徴量を獲得することに成功している。また、教師なし事前学習の分野では ImageNet を用いて事前学習を行い、PASCAL VOC などに転移学習を行うのが主流であり、蚊のような小さなドメインに対してあまり行われていない。最も望ましいのはデータのドメインに合わせた特徴量を学習することができる学習タスクを構築することである。以上を踏まえ、本論文では先行研究 2.1 の精度向上を目的として MoCo[1] を用いた教師なし事前学習を行い ImageNet 重みと比較し考察を行う。また、得られた特徴量の可視化を行い新たな改善案を模索する。

1）琉球大学大学院理工学研究科情報工学専攻,Graduate School of Engineering and Science, University of the Ryukyus
2）琉球大学工学部工学科知能情報コース,Computer Science and Intelligent Systems, University of the Ryukyus
3）琉球大学医学研究科 ,Medical Research Unit, University of the Ryukyus

2　先行研究

2.1　西銘らによる蚊の分類モデルに関する研究

西銘らは、iNaturalist[3] と呼ばれる生物学者を対象にした SNS からネッタイシマカとヒトスジシマカの画像を取得し、専門家によりラベル付けされた 426 枚のデータセットを作成した。図 2 のように、前処理として iNaturalist[3] から取得したデータに、Van Horn らの The inaturalist species classification and detection dataset[4] で学習された物体検出モデルで蚊が写る領域の検出を行った。

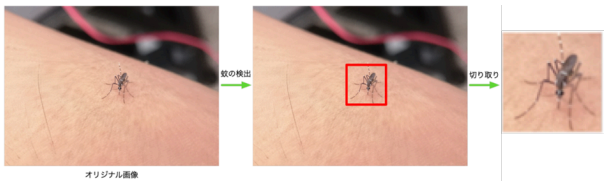

図 2　蚊の部分を検出して切り出し処理

図 3　データかさ増し

次に図 3 のように回転・明るさを変更し 1 枚から 10 枚のデータかさ増しを行い、300x300x3 にリサイズした。426 枚のデータセットを学習データ 340 枚、テストデータ 86 枚に分割した。分類モデルには、ResNet50 の ImageNet 学習済みモデルに転移学習を適応し、2 層の隠れ層と 1 層の出力の全結合層 3 層を追加して学習を行った。その結果、テストデータに対して表.1 に示した精度が得ている。

Class	Precision	Recall	F-measure
ネッタイシマカ	0.66	0.88	0.75
ヒトスジシマカ	0.93	0.77	0.84

表 1　先行研究精度

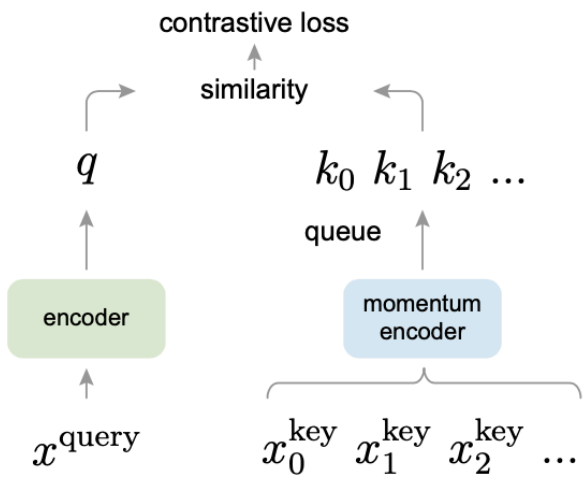

図 4　Momentum contrast for unsupervised visual representation learning [1] より引用

2.2　Momentum contrast for unsupervised visual representation learning

He, Kaiming らの Momentum contrast for unsupervised visual representation learning(MoCo)[1] は画像分野で用いられる教師なし表現学習の一種である。学習の概念は NLP の分野で教師なし表現学習として成功した BERT[5] や対照学習で類似した先行研究の CMC[6] を参考に作られている。具体的には、同じ画像同士は似たような特徴量ベクトルを生成し、違う画像とは離れた特徴量ベクトルを生成するタスクを構築している。Encoder や Momentum Encoder は CNN の特徴ベクトルを出力するように設計しており、Query となる画像を入力することで特徴量ベクトルを取得し、key となる画像の特徴量ベクトルとの類似度を比較する。一連の流れは MoCo[1] より引用した図を図 4 に示す。key の 0 番目には query と同じ画像に別の Augmentation をかけたものを正解データとして扱う。さらにより多くの key と比較するために step ごとに query の特徴量ベクトルを Queue に保存していくことで、メモリを抑えることに成功し大量の key との比較で良い表現を得ることに成功した。また、query と key の特徴量ベクトルを取得する際に、同じモデルを利用すると特徴量ベクトルが類似して適切に学習できないことを検証で示された。そこで query を出力するモデルのみを学習し、key を出力するモデルは学習せずに query を出力するモデルの重みに momentum を用いて徐々に近づくように設計することで学習の安定化が図られた。PASCAL VOC と COCO データセットを用いた評価実験では、ImageNet で学習した教師あり事前学習モデルと比較した場合 MoCo の方が優れた結果を示した。

3　アプローチ

西銘らの先行研究では、少量のラベル付きデータセットで学習させるために公開されている ImageNet 学習済みモデルを用いて転移学習を行いベースラインを作成した。しかしながら ImageNet では蚊の分類に必要な特徴量を学習をできていない可能性がある。その理由として、ImageNet[2] は学習データが 120 万枚の 1000 クラスで構成されており、蚊のクラスは 1000 クラスのうち 1 クラスであることが挙げられる。そこで蚊に近いデータセットのみで事前学習を行うことで蚊の分類に必要な特徴量を学習しようと考え、ラベル付きデータセットが少ないケースに最適な手法である教師なし事前学習に着目した。教師なし事前学習には、Jigsaw[8] や DeepCluster[7] など様々なタスクが研究されている。その中でも、蚊という小さなドメインに絞った場合でも同画像間の特徴ベクトルを近付けるというタスクでは学習が行えることから教師なし事前学習に MoCo[1] を選択した。蚊の分類に適した重みを事前学習するために、ネッタイシマカとヒトスジシマカに分類されていないハエ目カ科とハエ目ユスリカ科を蚊に近い構造を持つデータとして iNaturalist[3] から収集した。そのデータを用いて西銘らの先行研究と同様に物体検出を行いデータセットを作成した。蚊に近い構造を持つデータセットと MoCo を用いて事前学習を行い、事前学習で得た重みを蚊の分類問題に転移学習する。

4　実験

4.1　先行研究の再現実験

西銘らのデータセットを以下のように変更したため、比較対象となる教師あり事前学習での再現実験を行なった

1　データセットのサイズが 300x300 を満たさないデータは、300x300 にリサイズした際に足りない箇所が近いピクセルの値から補完されてしまい、ノイズになってしまうためデータセットから除外した。表 2 に示したように全体のデータセット数は 327 枚になり、261 枚を学習データとし、66 枚をテストデータとして分割した。
2　データかさ増しを学習の step ごとに回転、明るさ、上下反転、crop and pad を指定した値の範囲でそれぞれランダムで行うように変更した。

表 2 に示したように、ネッタイシマカとヒトスジシマカでは生息範囲の違いによりネッタイシマカの画像が少いため偏りが生まれていることがわかる。Optimizer はデータセットが少ないこともあり Adam の learning-rate=0.001 を用いて、ResNet50 の ImageNet 学習済みモデルを用いて Fine-tuning を行った結果、表 3 に示した精度が得られた。表 1 と表 3 を比較すると、ヒトスジシマカの F 値が高くなり、ネッタイシマカの F 値が低下していることから、ヒトスジシマカのデータが多いことで分類が偏っている考えられる。

Class	学習データ	テストデータ	合計
ネッタイシマカ	88	23	111
ヒトスジシマカ	173	43	216
合計	261	66	327

表 2　データセット

4.2　MoCo を用いた事前学習

iNaturalist[3] からネッタイシマカとヒトスジシマカに分類されていないハエ目カ科約 10,000 枚とハエ目ユス

Class	Precision	Recall	F1-score
ネッタイシマカ	0.684	0.565	0.619
ヒトスジシマカ	0.787	0.86	0.822
合計	0.757	0.757	0.757

表 3　先行研究の再現実験の精度

リカ科約 10,000 枚の合計 20,000 枚を取得しデータセットを作成した。

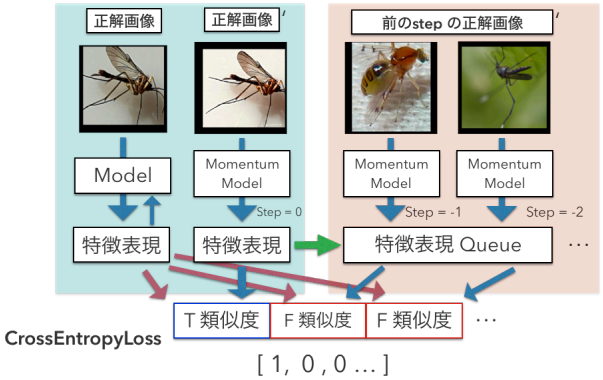

図 5　MoCo の全体図

蚊の類似画像を用いた MoCo の学習の流れを図 5 に示す。ハイパーパラメータでは Optimizer や temperature は MoCo[1] と同様の値を使用し、Encoder の momentum と Queue size に関しては変更した。Encoder の Momentum は step ごとに学習していく Encoder の重みにどれだけ近づけるかを決めるパラメータであり、MoCo[1] の実験で一番良い精度を得ていた 0.999 を用いた。また、Queue size はデータセットが 20,000 であることから 300 と設定した。

GPU には RTX 2080Ti を用いており、epoch100 を学習するのに一日要し、Loss は最終的に 0.331 で学習が終了した。

4.3　実験結果

モデル重み	ネッタイ	ヒトスジ	全 2 クラス
ランダム	0.0	0.79	0.65
ImageNet	0.62	0.82	0.75
MoCo	0.62	0.82	0.75

表 4　MoCo で学習した重みによる蚊の分類精度。数値は F1-score

先行研究の再現実験と同様に MoCo で学習した重みを用いて蚊の分類問題への転移学習を行った。初期の重みをそれぞれランダム、ImageNet、MoCo の 3 種類で蚊の分類問題を学習させ、テストデータに対する F1-score の結果を比較した (表 4)。ランダムな重みから学習を行った場合、データが多いヒトスジシマカに分類が偏ってしまうケースが見られたが、MoCo や ImageNet では全てヒトスジシマカと分類することを回避することができている。このことから MoCo を用いて事前学習を行うことは、少なからず表現は獲得できているといえる。しかしながら、MoCo と ImageNet の精度を比較するとほとんど変化がないことから、それぞれのモデルを可視化することで捉えている表現の違いを考察する。

4.4　GradCAM による可視化

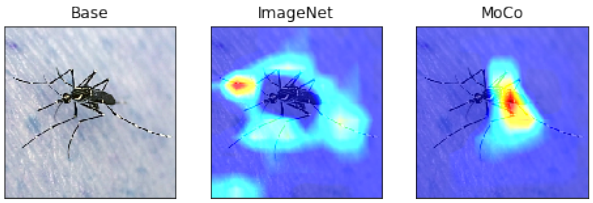

図 6　GradCAM による可視化 元画像 (左)・ImageNet(中央)・MoCo(右)

MoCo で得られた特徴量が蚊の分類に寄与しているのかを確かめるために、図 6 のように GradCAM[9] を用いてテストデータに対して可視化を行った。GradCAM による可視化で蚊を捉えている数を調査するために、全てのテストデータ 66 枚に対して手動で計数した。図 6 より、ImageNet(中央) では蚊の部分を避けているため蚊を捉えていないと判断し、MoCo(右) の場合には蚊がいる部分が赤く活性化されているため蚊を捉えていると判断する。

4.5　考察

予測クラス	ヒトスジシマカ	ネッタイシマカ	合計
ヒトスジシマカ	**20/34**	3/9	23/48
ネッタイシマカ	9/9	12/14	21/23
合計	29/43	15/23	43/66

表 5　ImageNet ベースモデル 蚊を捉えている枚数

予測クラス	ヒトスジシマカ	ネッタイシマカ	合計
ヒトスジシマカ	36/39	**13/14**	49/53
ネッタイシマカ	3/4	3/9	6/13
合計	39/43	16/23	55/66

表 6　MoCo ベースモデル 蚊を見ている枚数

表 5 及び表 6 には、蚊を捉えている枚数をカウントした結果を示した。表 5 及び表 6 は列にはクラス、行にはモデルが予測したクラスで分けられ、(蚊を捉えている枚数)/(母数) の形式で表記している。表 5 の太字で記述されている箇所はヒトスジシマカと正解したデータにもかかわらず蚊を注視していない枚数が多いことが見て取れる。また、同様に表 6 ではヒトスジシマカと誤回答したデータに蚊を注視していない枚数が多くみられた。このことからそれぞれ ImageNet はネッタイシマカにおいては蚊を捉えるモデル、MoCo はヒトスジシマカの場合に蚊をよく捉えるモデルと考えられる。

4.6　フィルタの可視化

GradCAM の可視化とその集計により、ImageNet ではネッタイシマカの体や胴体を捉えることができ、MoCo ではヒトスジシマカを捉えることが可能なのではないかと考えた。そこでヒトスジシマカに着目して Conv15 層目、Conv30 層目、Conv46 層目で最も活性化されたフィ

ルターを可視化した。ここで最も活性化されたフィルターはモデルに画像を入力してあるレイヤーの全てのフィルターの平均値をとり最大値をとるフィルターのことを指す。フィルターは一般的に用いられる可視化手法を使用している。具体的にはランダムに生成した画像を入力にして、フィルターの活性化を平均したものを損失関数として扱い勾配上昇法を用いて入力画像に変更加えていくことで、入力画像をフィルターの可視化として取得する方法である。

4.7 考察

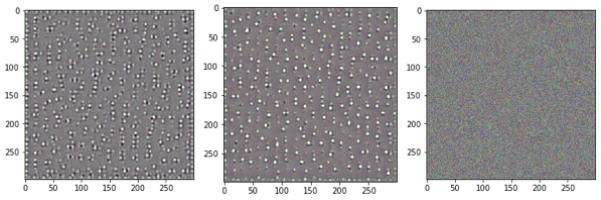

図 7　MoCo ヒトスジシマカに関連するフィルター
Conv15(左)・Conv30(中央)・Conv46(右)

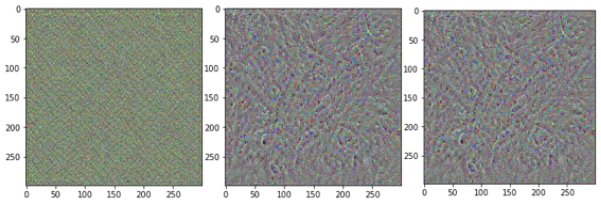

図 8　ImageNet ヒトスジシマカに関連するフィルター
Conv15(左)・Conv30(中央)・Conv46(右)

図 8 の結果から、MoCo は Conv30 層 (中央) において白黒の斑点のフィルターがよく活性化していることが見て取れる。これは蚊の画像において白黒の斑点模様は胴体に多く見られる特徴を捉えていると思われる。図 7 の結果は、多様な色を含んだ模様になっており、蚊の輪郭を捉えたような形状をしている。蚊は白黒で構成されているため、白黒の模様が現れていないことから蚊と背景との境界線を含めて注視していると思われる。上記の結果より MoCo は ImageNet と比較して、蚊のデータに対して白黒を注視した特徴が得られている、そのため、データ量の多いヒトスジシマカを捉えることで分類できたのではないかと考えられる。

5　終わりに

西銘らの ImageNet 学習済みモデルによる蚊の分類モデルの再現実験を行い、MoCo を用いた事前学習によって作成した重みとの比較を行った。ImageNet の学習データ数 120 万に比べて、MoCo を用いた 2 万枚のラベルなし学習データで同精度の結果が得られたが、フィルターを可視化した結果 MoCo が蚊の模様に近いフィルター、ImageNet が蚊の境界線を表したようなフィルターを重視していた。それぞれ違う特徴量を重視していることから ImageNet の重みを生かしつつ蚊の模様を分類を捉える方法を模索する必要がある。また、今後の課題として小さい特徴量を捉えるというタスクにおいて、蚊のデータでは、偏りや少量のサイズであることと問題を分割することが難しい問題であったため、小さい特徴を生成したダミーデータを作成して実験・考察する必要がある。

参考文献

[1] He, Kaiming, et al. ”Momentum contrast for unsupervised visual representation learning.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[2] Deng, Jia, et al. ”Imagenet: A large-scale hierarchical image database.” 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.

[3] iNaturalist https://www.inaturalist.org

[4] Van Horn, Grant, et al. ”The inaturalist species classification and detection dataset.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[5] Devlin, Jacob, et al. ”Bert: Pre-training of deep bidirectional transformers for language understanding.” arXiv preprint arXiv:1810.04805 (2018).

[6] Tian, Yonglong, Dilip Krishnan, and Phillip Isola. ”Contrastive multiview coding.” arXiv preprint arXiv:1906.05849 (2019).

[7] Caron, Mathilde, et al. ”Deep clustering for unsupervised learning of visual features.” Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[8] Noroozi, Mehdi, and Paolo Favaro. ”Unsupervised learning of visual representations by solving jigsaw puzzles.” European Conference on Computer Vision. Springer, Cham, 2016.

[9] Selvaraju, Ramprasaath R., et al. ”Grad-cam: Visual explanations from deep networks via gradient-based localization.” Proceedings of the IEEE international conference on computer vision. 2017.