異なる画角画像の混合学習と特徴量補完
～空撮画像は地上画像で補完できるのか？～

高橋 悠太 [1a)]　藤井 純一郎 [1b)]　天方 匡純 [1c)]　山下 隆義 [2d)]

概要

ドローンの社会実装に向けた法整備が進んでいる．ドローンで長大・広域的な対象を撮影する場合，特定対象の検知は，画像認識技術が得意とする．深層学習による検出 AI の適用が考えられるが，飛行しなくては得られない空撮画像を，本格的な飛行前に大量に得ることは現実的でない．本研究では，小規模空撮画像学習データに，空撮画像とは画角等が異なるはずの地上画像を追加し，検出 AI の性能を向上可能か，どの画像が影響するか，また特徴量は何か，いくつかの指標・手法の比較検証により，明らかにする．

1. はじめに

ドローンによって長大，広域的な範囲を撮影し，特定の対象を検知するとき，画像認識技術，特に深層学習による AI が効果的と考えられる．土木分野において，ドローンによる河川巡視実現に向けて検討が進んでおり，巡視で得られた画像を AI で学習・推論し，不法投棄や違法行為を発見することが目的となる．ドローン河川巡視が実施される初期段階では，ドローンによる空撮画像は十分でないと考えられるが，多様かつ大量なデータを本格的な飛行前に得ることは現実的でない．一方，既に RiMaDIS[1)]（River Management Data Intelligent System：全国統一版河川維持管理業務支援データベースシステム）には地上で撮影された，河川巡視項目に関連する画像が保存されている．画角や対象の写る範囲は異なるが，現実の巡視項目に応じた画像が網羅されていると考えられる．本研究では，このデータに着目し，空撮画像の量的・質的不足を，地上画像を学習に加えることで補完(Domain Adaptation)可能か，検証を行う．追加する地上画像は不法投棄の画像を用い， k 平均法(kmeans++)，Boundary Box 占有率(BBO)[2)]，ShuffleNet，Inception v3 を画像選定用の指標ないしプレフィルターとする．空撮画像と同クラスとされた地上画像を学習に追加し，比較試験を行う．スコアが改善されたデータセットに共通となる画像と，その他の画像を比較することで，どのような画像が補完に貢献したか否か，検証を行う．

[1] 八千代エンジニヤリング株式会社 技術創発研究所
[2] 中部大学 工学部 情報工学科
[a)] yt-takahashi@yachiyo-eng.co.jp
[b)] jn-fujii@yachiyo-eng.co.jp
[c)] amakata@yachiyo-eng.co.jp
[d)] takayoshi@isc.chubu.ac.jp

2. 実験概要

2.1 学習条件と地上画像の分類手法

地上画像は最大 800 枚使用できるとする．空撮画像と地上画像を**図-1** に示す．赤枠は河川巡視記録に多い，不法投棄を示す．空撮画像は地上画像と対し，画像サイズに対する不法投棄の割合はごく僅かなことがわかる．学習効率の低下を避けるため，使用する地上画像に最も多い 640×480[pixel]程度（**図-1(a)**内橙破線）に切り出して使用する．空撮画像は，実際の河川に用意した疑似不法投棄の上空から撮影している．そのため，空撮画像がほぼ直上から不法投棄を撮影するのに対し，地上画像は横方向から写る．そのため，影を特徴量と評価するか等，影響が予想される．検知モデルは Faster R-CNN[3)]を用いた．特徴量抽出を効率化するため，事前学習済み ResNet50 を使用し，抽出層は 40 層の Relu とした．入力サイズが 224×224 の RGB 画像を用いた．学習には MATLAB 2020a 環境を利用し，勾配計算は SGDM（Stochastic Gradient Descent with Momentum：モーメンタム項付き確率的勾配降下法），ミニバッチサイズは 2，学習率は 0.0001，学習エポック数を 10 に固定した．以降の実験でも，これらの設定は変更しない．

実験は，少数の空撮画像のみで構成された学習データを用いたベンチマークとする．空撮画像の学習データは 280 枚，推論データに 20 枚使用する．5 つの基準・方法で選出した地上画像をそれぞれ加え，平均 Precision の変化を比較する．Case.1 はベンチマーク， Case.2 は kmeans++により地上画像と空撮画像を 2 クラス分類し，空撮画像がより多く選ばれたクラスと同クラスの地上画像を，Case.3 は空撮画像での BBO 平均値に近いを地上画像を加える．Case.4 は ShuffleNet，Case.5 は Inception v3 が対象あり・なし空撮画像をそれぞれ学習し，対象ありと分類された地上画像を追加する．各 Case の空撮画像と地上画像枚数を**表-1** に，スコアの比較を**表-2** に示す．改善があった Case の結果および共通して選ばれた地上画像について比較し，考察する．

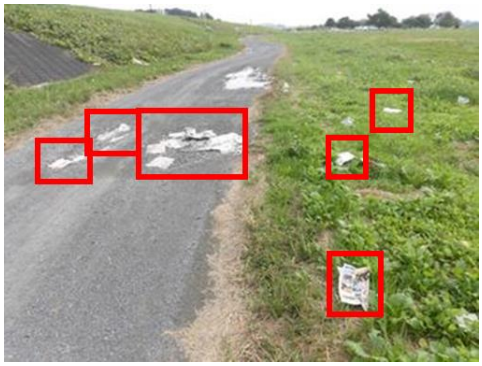
2.2 空撮画像のみでの学習・推論結果 BM

図-2, 図-3 にベンチマーク実験(Case.1)の推論結果についての，PR（Precision-Recall）曲線と，推論結果の一例を示す．平均 Precision は曲線下の面積を示す．推論結果の一例において，黄色の枠が推論された Bounding Box を，数字はその信頼度（Confidence）を示す．左上の青い物体は見切れているため，教師画像には不法投棄としてアノテーションされておらず，今回は未検知でも問題ない．半分程度，Bounding Box 内に左側の不法投棄を捉えているが，精度・信頼度共に高くないことが分かる．ペットボトル（中央付近）や雑誌（右上）はうまく検知できていない．また，Bounding Box の中心が不法投棄の影上にあることから，影に注目していると予想できる．原因の一つに，背景が灰色の強く，明るいことから，影への注目が強くなったと考えられ，背景と不法投棄を判別できていないことが示唆される．以降，地上画像を追加し，精度向上を検証する．

3. 地上画像追加結果と考察

データセットに用いた空撮画像と地上画像の枚数と，各 Case の平均 Precision を表-1，表-2 に示す．またプレフィルター用ネットワーク(ShuffleNet，Inception v3)の学習パラメーターと学習結果については，表-3 に示す．

(1-a) 空撮画像（3840×2160 [pixel]）



(1-b) 地上画像（640×480 [pixel]）

図-1 学習に用いる空撮画像と地上画像の例

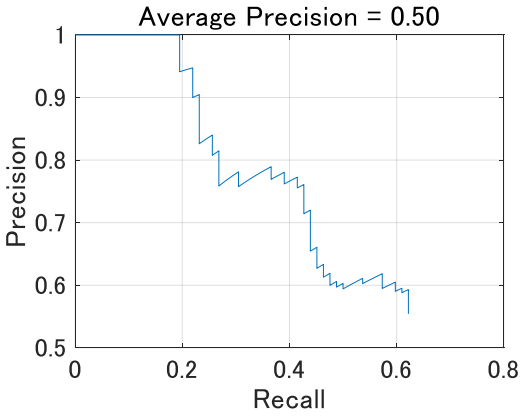


図-2 ベンチマーク（Case.1）の PR 曲線

図-3 ベンチマーク（Case.1）での推論例

表-1 ベンチマーク（Case.1）に対する空撮・地上画像の増減

Case	1	2	3	4	5
空撮画像	280	277	280	280	280
対 Case.1 空撮画像増減	0	-3	0	0	0
追加地上画像	0	482	95	161	109
合計	280	779	375	441	389

Unit: image.

表-2 各 Case での平均 Precision

Case	1	2	3	4	5
平均 Precision	0.50	0.002	0.54	0.65	0.61
対 Case.1 スコア	0	-0.498	+0.04	+0.15	+0.11

表-3 プレフィルター用ネットワーク学習パラメーターと学習結果

	ShuffleNet	Inception v3
全体学習率	0.0001	
MBS	128	32
FVL [%]	0.15	0.14
FVA [%]	0.93	0.98

3.1 Bounding Box 占有率：BBO（Case.3）

BBO は本研究に関連し，新たに提案している手法である[2]．物体検知は背景と対象物を分ける作業を含むことから，背景画像の範囲は少ないデータほど，中間層のテンソルはスパースになりづらく，対象物の特徴量が畳み込み層で抽出されやすい．同様の考えを利用した新しいデータセットも開発されている[4]．ここで，入力画像サイズに対し，同程度の Bounding Box 面積を持つならば，同程度特徴量が抽出可能と考える．本研究では Bounding Box の面積と，画像サイズの pixel 比を算出し，空撮画像の平均値と同程度になるような地上画像を選出，学習した．

BBO は，画像の Bounding Box 面積の合計値を画像サイズで除した値になる．この時，Bounding Box の重複を許して計算した．まず，**図-4** に空撮画像の BBO のヒストグラムを示す．横軸は BBO を示し，縦軸はその頻度を示す．BBO 平均値は 8.88[%]だった．平均値の 25%上下間を持つ地上画像（95 枚）を，学習用とする．PR 曲線と推論結果の一例を**図-5** に示す．Case.1（0.50）より向上しており，信頼度も高い．**図-5** 内左の対象を Box 中央で捉えられておらず，影に注目しているように見られるが，Box の重複や，見落としもないことが確認できる．また，Case.1 では未検知だったペットボトルを高い信頼度で検知しており，地上画像の追加により，空撮画像データセットに不足した特徴量を補完可能であることが確認できた．

3.2 プレフィルターとしての深層ネットワーク

3.2.1 プレフィルターの学習データ

プレフィルターとして，データセットの改善に深層ネットワークを適用するには，例えば，2 クラス分類などの学習が必要となる．空撮画像のみの場合は教師なし学習となり，不安定になる可能性がある．一方，空撮画像と一部の地上画像を利用した学習では，選んだ地上画像が持つ特徴の影響を強く受けると考えられる．本研究では，学習対象を捉えていない背景画像を第 2 クラスとして学習した．背景画像は容易かつ大量に入手できるだけでなく，対象を捉えていた空撮画像の敵対データとして直接学習していると考えることができる．すなわち，背景として分類された地上画像は学習に適していないことを直接示唆している．

3.2.2 ShuffleNet（Case.4）

学習データは検出モデルと同じ 280 枚の空撮画像を使用する．640×480 ピクセルサイズの背景画像（309 枚）は，切り出し処理の残りから，対象が写る画像と同程度の枚数となるように選出した．第 1 クラスは対象の写る空撮画像，第 2 クラスは背景画像とする．ネットワークは ImageNet で事前学習済みのものを用いる。検証率は 0.3，全体学習率は 0.0001，学習回数は 8，ミニバッチサイズ（MBS）は 128，最終畳み込み層以降を学習し，畳み込み層の重み学習率は 10 倍とした．最終的な検証の損失と精度（FVL と FVA）は 0.15，0.93 であった（**表-3**）．推論により第 1 クラスと分類

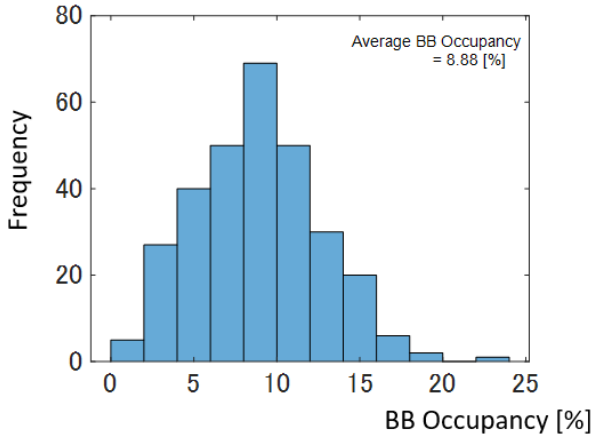
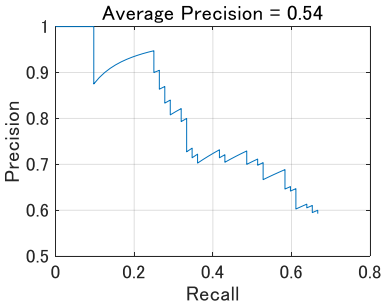


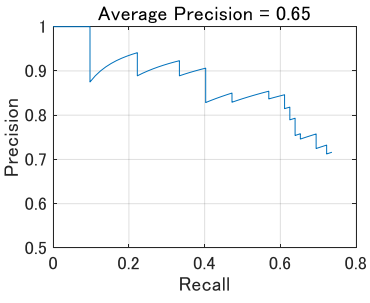
図-4 空撮画像の BBO ヒストグラムと平均 BBO



(a) The PR 曲線

(b) 推論例

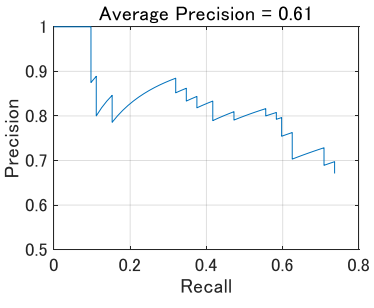
図-5 BBO:推論結果



(a) The PR 曲線

(b) 推論例

図-6 ShuffleNet:推論結果



(a) The PR 曲線

(b) 推論例

図-7 Inception v3:推論結果


(a) Case. 2,3,4,5 で選出
BBO: 7.75


(b) Case. 2,3,4,5 で選出
BBO: 7.33


(c) Case. 3,4,5 で選出
BBO: 8.82


(d) Case. 3,4,5 で選出
BBO: 10.11

図-8 各ケースで共通した学習データ

された地上画像は 161，閾値は 0.85 とした．PR 曲線を図-6(a)に示す. 平均 Precision は BBO の結果より高くなった．推論例を図-6(b)に示す．全対象に対して信頼度が高く，左の対象は Box 中心に検出されている．本結果はプレフィルターとしての深層ネットワークの有効性を示唆している．

3.2.3 Inception v3（Case.5）

ShuffleNet と同様，Inceptionv3 は背景画像を用いて学習している．Inception v3 では学習回数と MBS を変更した．損失が ShuffleNet と同様にとなるよう，また条件を可能な限り一致させ，メモリリークを回避するよう，それぞれ 9 と 32 とした．FVL と FVA は 0.14，0.98 だった．パラメーターと学習結果は表-3 に示す．推論では、第 1 クラスに分類された地上画像は 109 枚，閾値は 0.85 だった．PR 曲線を図-7(a)に示す．平均 Precision は Case.3 より高くなったが，Case.4 より低くなった．推論例を図-7(b)に示す．右側のオブジェクトで検出が重複していたが，全体として最も高い自信度だった．プレフィルターとして深層ネットワークを安易に使用するだけでは，改善の限界があることを示唆している．Inception v3 はパラメーター数も多く，本来多くの学習データを要するため，過学習であったとも考えられるが，プレフィルターはより少ないデータ・時間で学習可能な方が良いと考えられることから，実用のためにデータを増やす必要性は高くないと考えられる．

4. 共通して選出された画像についての考察

各 Case での学習データセットの内, 共通した地上画像について比較する Case.2 以降で共通した地上画像を確認した．それぞれ図-8 に示す. BBO は図-8(a)から 7.75, 7.33, 8.82, 10.11 だった．スコアが低下した Case.2 においても選出された図-8(a),(b)は比較的まとまった状態で写っている．一方，図-8(c),(d)では比較的小さな対象が散乱した状態で写っていた．これらの差異は Case.2 に用いた手法の数学的特徴に影響すると考えられる．確率空間を仮定し，次元圧縮後の距離によって類似度を分類するため，対象が散在するとノイズのようになり，有意な距離に満たないと考えられる．そのため，Case.3,4,5 において，小さな対象（ペットボトルなど）も含め，空撮画像との類似度が高い地上画像を選出・学習したことで，空撮画像に不足していた特徴量を補完し，スコアを改善したことが示唆される．

5. まとめ

ドローンで撮影した空撮画像に検知 AI を適用するための学習に用いる際，不足すると考えられる空撮画像を地上で撮影した画像で補完可能か，比較検証を行った．検証により，Boundary Box 占有率・ShuffleNet，Inception v3 を用いて，空撮画像に類似する地上画像を選出・学習に用いた場合，ベンチマークからスコア改善が見られた．この結果から, 適切な指標･手法により選出した地上画像を用いて，検知 AI の性能を高めることが分かった．ただし，空撮・地上に共通してアスペクト比が近く，Faster R-CNN の特性に合致した可能性について，今後検証を要する．

今回最も改善効果の高かった深層ネットワークを用いる場合，どのような画像の貢献が最も高いか，明示的に解釈することは困難である．データを小バッチに分けて追加しながら学習・評価する方法も考えられるが，データ個々の特性に影響を受けることから，今後は検知モデルの学習中の中間層を分析し，学習に影響を与える画像群を追跡検証する. あるいは BBO に加え, SIFT などの各指標を基に，プレフィルター用深層ネットワークに分類された画像特性の把握により, 客観的指標に基づく選定が可能か検討する．

References

[1] 桃田美雪，河川維持管理データベースシステム「RiMaDIS(リマディス)」の活用について，中部地方整備局管内事業研究発表会，2019.
[2] 高橋悠太，藤井純一郎，天方匡純，山下隆義： UAV と画像認識 AI による河川巡視を補う地上画像の特徴量とその利用法検討，AI･データサイエンス論文集，第 1 巻，J1 号，pp.580-587，2020.
[3] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, NIPS, 2015.
[4] S. Shao, Z. Li, T. Zhang, C. Peng, G. Yu, X. Zhang, J. Li, and J. Sun, Objects365: A large-scale, high-quality dataset for object detection, Proceedings of the IEEE international conference on computer vision, pp. 8430–8439, 2019.