地上画像を用いたドローン河川巡視で不足する空撮画像特徴量補完の可能性検証

The verification of complement for UAVs image feature with ground image in RiMaDIS

高橋 悠太[*1] Yuta Takahashi
藤井 純一郎[*1] Junichiro Fujii
天方 匡純[*1] Masazumi Amakata
山下 隆義[*2] Takayoshi Yamashita

[*1] 八千代エンジニヤリング株式会社技術創発研究所
Research Institute for Infrastructure Paradigm Shift, Yachiyo Engineering Co., Ltd.

[*2] 中部大学工学部情報工学科
Department of Information Engineering, Chubu University.

Despite the development of deep learning, the number of application cases in the civil engineering field has not so increased. The definition of the boundary conditions is difficult, and there are various abnormalities to be detected. Additionally, the anomaly data may be less or not exist. For example, illegal dumping in river patrols can take various forms depending on the context. At the start stage using UAV / AI, a situation where there are few aerial images can be assumed. This study verified whether the learning data could be complemented by learning together with images obtained on the ground registered in river maintenance database RiMaDIS. Model uses Faster R-CNN. Images close to the feature space of aerial ones were selected based on several criteria and methods. The ground images selected by the occupancy rate of the Bounding Box and the Deep Network (ShuffleNet, Inception v3) improved the average Precision.

1. 背景

深層学習の発展がありながら，土木分野での適用事例があまり増えていない原因に，解くべき問題の境界条件を規定することが困難であり，また検知すべき異常が多様で，それぞれのデータ自体は僅か，あるいは存在しないことが挙げられる[全 2020]．例えば，河川巡視における不法投棄は，文脈によって多様な形態を取りうる．この状況で AI を設計すると，河川巡視の文脈で検知すべき対象を網羅するように学習する必要がある．

今ドローン/AI を用いた河川巡視を考えると，開始段階では，空撮画像が少ない状況を想定できる．少ない学習データで効果的に学習する方法として，転移学習や Augmentation が考えられる．ImageNet で学習したネットワークを抽出層に使用する転移学習が有効だが，追加する検知層の学習には多様・大量のデータが必要になる．ドローンがある程度の高さから撮影した画像は広範囲が撮影され，背景は多様になる．対象は小さく写るため，切り出して推論する必要があり，Augmentation（反転・回転等）の適用性は高い．ただし，学習データ量を増やす方法であり，多様性を高めるわけではない点に注意する．そのため，空撮画像に不足する対象の特徴量を補完する必要がある．

ここで，河川維持管理データベース RiMaDIS に登録されている地上で得られた画像(地上画像)に注目する．RiMaDIS は河川維持管理情報に関するデータベースで，限りある人員のもと効率的，効果的に河川維持管理を行うことを目的に平成25年度より試行・運用を開始した[桃田 2019]．地上画像は空撮画像が必要とする対象を捉えていると考えられるが，撮影角度や対象が異なる．この違いによる影響を考慮し，適切な指標・手法により地上画像を選別・追加することで，検知 AI の性能向上が既往の研究で確認されている[高橋 2020]．

連絡先：高橋悠太，〒111-8648 東京都台東区浅草橋 5-20-8 CS タワー3F，yt-takahashi@yachiyo-eng.co.jp，03-5822-2903

本研究では空撮画像に地上画像を併せて学習し，空撮画像を補完可能か，Faster R-CNN を用いて追加的検証を行った．地上画像は，空撮画像の持つ特徴量空間に近しい画像を，既往の研究に 2 つの深層ネットワークを加え，5 つの基準・方法(t-SNE [van der Maaten 2008]& k-NN [Dasarathy 2002]，kmeans++ [David 2007]，Boundary Box 占有率 [高橋 2020]，ShuffleNet [Zhang 2017]，Inception v3 [Szegedy 2015]) により選出した．

2. 実験概要

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 t-SNE&k-NN，Case.3 は kmeans++，Case.4 は Boundary Box 占有率，Case.5 は ShuffleNet，Case.6 は Inception v3 を用いて選出した地上画像を追加する．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，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しているのに対し，地上画像は横方向から写る．したがって，影を特徴量と評価するか等，影響が予想される．特徴量抽出を効率化するため，事前学習済み ResNet50 [He 2016]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）[Bottou 2010]，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．以降の実験にも，これらの設定は変更しない．

3. 結果と考察

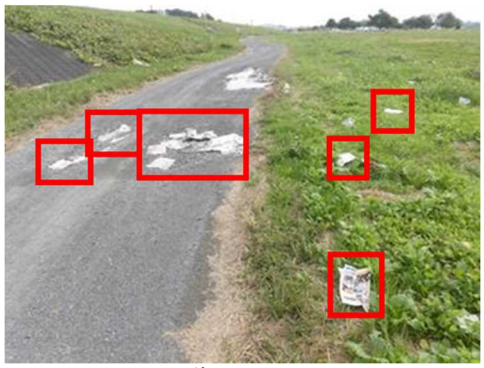
3.1 ベンチマーク(Case.1)

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．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．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，今回は検知できていなくとも問題ない．また，Bounding Box の中心が不法投棄の影上にあることから，影の方により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る．これらの結果から，十分な学習が行われておらず，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，地上画像を学習画像に追加し，精度向上を検証する．

3.2 改善が見られた指標・手法（Case.4，5，6）

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の学習パラメーターと学習結果については，**表-3** に示す．改善が見られた手法（Case.4，5，6）のみ，特にまとめる．

(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

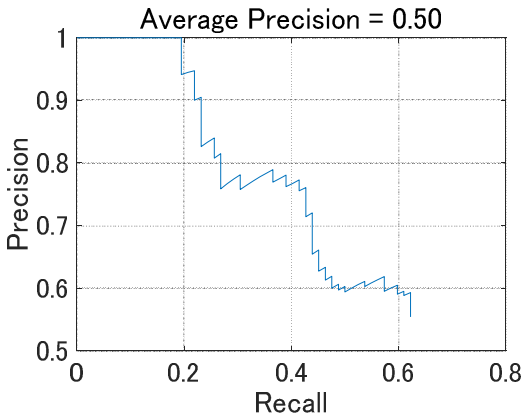


図-2 ベンチマーク（Case.1）の PR 曲線

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	6
空撮画像	280	83	277	280	280	280
対 Case.1 増減	0	-197	-3	0	0	0
追加地上画像	0	215	482	95	161	109
合計	280	298	779	375	441	389

Unit: image.

表-2 各Caseでの平均 Precision

Case	1	2	3	4	5	6
平均 Precision	0.50	0.0004	0.002	0.54	0.65	0.61
対 Case.1 スコア	0	-0.4996	-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

(1) Bounding Box 占有率:BBO(Case.4)

BBO は本研究に関連し，新たに提案している手法である[高橋ら 2020]. 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，画像の特徴量を示す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. このような考えを利用した新しいデータセットも開発されている[Shao 2019]. 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できると考える. 本研究では Bounding Box の面積と，画像サイズの pixel 比を算出し，空撮画像と同程度になるように地上画像を選出して学習を行った.

BBO を用いて，学習用の地上画像を選び出す. 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 になる. 平均は 8.88[%]だった. 平均値の 25%上下間を持つ地上画像(95 枚)を，学習用として選び出す. PR 曲線と推論結果の一例を**図-5** に示す. ベンチマーク(0.50)より向上しており，信頼度が高く，Box が重複することも，見落としもないことが確認できる. また，ベンチマークでは検知できていなかったペットボトルを高い信頼度で検知しており，地上画像を増やすことで，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と確認できた.

(2) ShuffleNet(Case.5)

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.

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，対象が写っている画像と同程度の枚数となるように選出した. 第 1 クラスは対象の写った空撮画像，第 2 クラスは背景画像とする. ネットワークは ImageNet に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. すべての対象に対して信頼度が高くなっており，左側の対象はボックスの中心に検出されている. この結果は，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) Inception v3(Case.6)

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

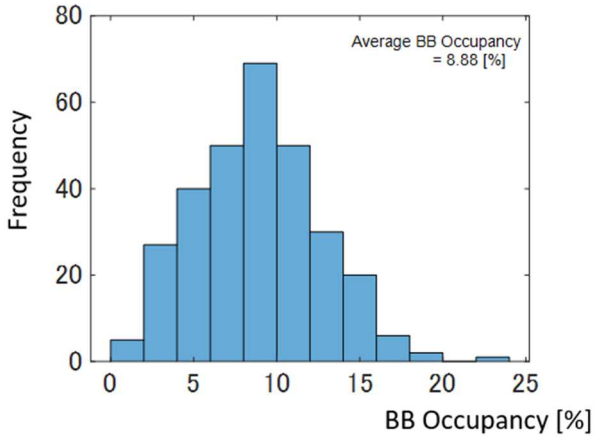
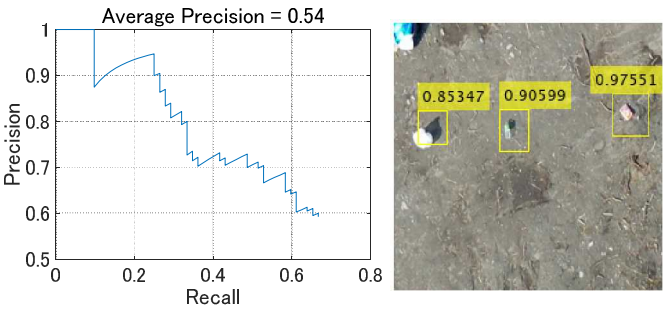
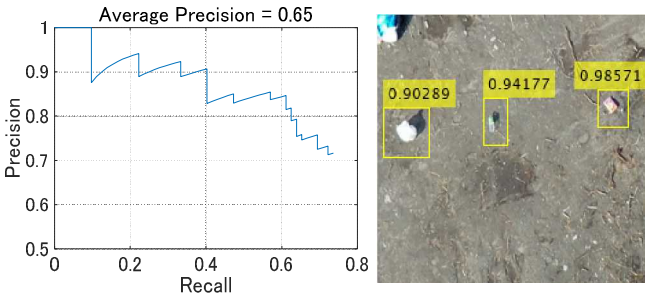


図-4 空撮画像のBBOヒストグラムと平均BBO



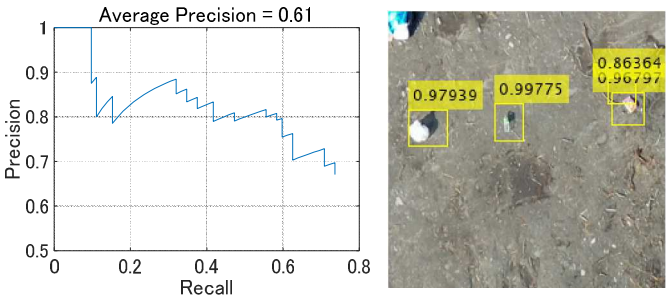
(a) The PR 曲線　　(b) 推論例

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

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 BBO の結果より高くなったが, ShuffleNet の結果より低くなった. 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. この結果は, 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.

4. データ数とスコアの関係に対する考察

表-1, **表-2** に示す結果から, データクレンジングより, データを減らさずに追加した方が, 高い改善効果を得られることが確認できる. ただし, kmeans++の結果(Case.3)では, ほとんど空撮画像が減っていないにもかかわらず, スコアは大きく低下している. 学習回数を固定しているため, 学習が不十分であったことも考えられる. Case.4 でスコアが上昇したことは, 今回のような特徴量補完の試みにおいて, BBO が有効である可能性を示している. Case.5, 6 で, パラメーターの少ない ShuffleNet が高スコアであった理由に, 少ないデータでの学習により適していたと考えられる. Inception v3 の学習結果は, 過学習である可能性があり, 分類結果について, 個々の画像について 2 つのネットワークの違いを分析する必要があると考えられる.

5. まとめ

本研究は, ドローンが撮影した少量の空撮画像からのみで構成される学習データセットに, 地上画像を追加することで空撮画像に不足する特徴量の補完が可能か, 検証を行った. 本研究で得られた知見を以下 4 つにまとめる.

1) 空撮画像に類似する地上画像の選出に用いる 5 つの指標・手法(t-SNE & k-NN, kmeans++, Boundary Box 占有率, ShuffleNet, Inception v3)を提案し, 比較実験を行った. 提案した指標・手法を基に, それぞれ画像を分類し, 類似度の高い空撮・地上画像を学習データとして, 実験を行い, 特徴量の違いによる影響を確認した.

2) BB 占有率, ShuffleNet, Inception v3 を用いて分類した地上画像を追加したとき, 空撮画像のみで学習したベンチマークからスコアである平均 Precision が改善した. ただし, t-SNE & k-NN や kmeans++では, スコアが悪化した. 適切な特徴量を用いて画像を事前比較することで, 学習データを改善することができることを確認できた.

3) 空撮画像だけでは十分に検知できない対象物(ペットボトル等)について, 地上画像の追加で検知が可能になった. これは単に学習データが増えたことによる学習の安定化に加えて, ペットボトル等のゴミが多い地上画像を学習したことで, 検知が可能になったとも考えられる. よって, 学習データに空撮画像とは異なる撮影条件の画像であっても, 適切な特徴量選択の上で追加することで, 空撮画像では得づらい対象物の特徴量を補完的に学習できる可能性が示唆された.

4) プレフィルターとして, 深層ネットワーク ShuffleNet および Inceptionv3 を提案した. Inception v3 は ShuffleNet より大きなパラメーターと入力サイズを有するが, ShuffleNet を用いた場合が最も好スコアであった。この結果は, ShuffleNet の少ないデータでの学習に対する堅牢性が優位であったと考えられる. したがって, プレフィルター自体が肥大化しないようにすると, データセットを改善するためのプレフィルターとしては, ShuffleNet などの小規模なネットワークでも十分であることを確認できた.

今後の課題について整理する. 今回データで用いた画像は, 空撮・地上ともに, 最低限学習が可能な数は得られていたが, 偏りがないとは断言できず, また物体検知に利用した Faster R-CNN のアルゴリズム特性が, Bounding Box 占有率のそれと合致したとも考えられる. 今後の検討予定として, 各指標・手法により選出された画像にどのような傾向があるか比較すること, 途中層の特徴量を分析することで, データセットの改善要因について整理する. また, RiMaDIS には莫大な地上画像が蓄積されているため, 空撮画像の増加に伴い, 地上画像についても, 空撮画像と同数以上追加した巨大学習データセットでの検証実験をすることで, より大きなデータでの再現性についても確認する.

参考文献

[全 2020] 全邦釘: 土木工学分野における人工知能技術活用のために解決すべき課題と進めるべき研究開発, AI・データサイエンス論文集, 第 1 巻, J1 号, pp.9-16, 2020.

[桃田 2019] 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.

[van der Maaten 2008] van der Maaten, L. and Hinton, G. E.: Visualizing Data using t-SNE. Journal of Machine Learning Research, Vol.9, pp. 2579–2605, 2008.

[Dasarathy 2002] Dasarathy, B. V.: Nearest-neighbor approaches, Hand-book of Data Mining and Knowledge Discovery, pp. 88–298, Oxford University Press, 2002.

[David 2007] David, A. and Vassilvitskii. S.: K-means++: The Ad-vantages of Careful Seeding., SODA 2007: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035, 2007.

[高橋 2020] 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.

[Zhang 2017] Zhang, X., Zhou, X., Lin, M. and Sun, J.: *ShuffleNet: An extremely efficient convolutional neural network for mobile device*, arXiv:1707.01083, 2017.

[Szegedy 2015] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna. Z.: *Rethinking the inception architecture for computer vision*, CoRR, abs/1512.00567, 2015.

[He 2016] He, K., Zhang, X., Ren, S. and Sun., J.: Deep residual learning for image recognition, Computer Vision and Pattern Recognition (CVPR), 2016.

[Bottou 2010] Bottou, L.: Large-Scale Machine Learning with Stochastic Gradient Descent, Proceedings of the 19th International Conference on Computational Statistics (COMPSTAT' 2010), pp. 177–187, 2010.

[Shao 2019] Shao, S., Li, Z., Zhang, T., Peng, C., Yu, G., Zhang, X., Li, J. and Sun, J.: Objects365: A large-scale, high-quality dataset for object detection, Proceedings of the IEEE international conference on computer vision, pp. 8430–8439, 2019.