

# 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

高橋 悠太<sup>1\*</sup>・藤井 純一郎<sup>1</sup>・天方 匡純<sup>1</sup>・山下 隆義<sup>2</sup>

<sup>1</sup>正会員 八千代エンジニアリング株式会社 (〒111-8648 東京都台東区浅草橋 5-20-8 CS タワー)

\*E-mail: yt-takahashi@yachiyo-eng.co.jp (Corresponding Author)

<sup>2</sup>非会員 中部大学工学部情報工学科 (〒487-8501 愛知県春日井市松本町 1200)

我が国は、その国土に多くの河川を有し、管理作業の多くは人の手により行われている。ドローン (UAV) 技術の発展により、河川監視への適用事例が増えており、得られた画像に人工知能を適用し、不法投棄等を認識検知する研究も行われてきた。人による地上での撮像は、作業量も多く、場所の制約や、ばらつきに影響される。空撮画像は地上画像とは異なる画角を持ち、撮影条件も異なる。本研究では、河川維持管理データベースシステム RiMaDIS が有する大量の地上画像を、現在ではまだ数少ない空撮画像での検知精度向上に活用する。画像特徴量に基づき分類した不法投棄に関する画像を Faster R-CNN に入力し、推論結果を比較する。検証により、Bounding Box の画像占有率などが検知精度向上に寄与することを明らかにした。

**Key Words:** ドローン (UAV) 空撮, 河川巡視, 対象物検知 (Object Detection), Faster R-CNN

## 1. 背景

我が国は多くの河川を有し、治水利水のため、膨大な管理業務を要するが、多くは人の手により行われている。急変・甚大化する自然災害に対して、強靱で高効率なインフラを構築維持するためにも、AI (人工知能) による代替補助は効果的と考えられる。

河川管理業務のうち、AI 活用が容易な分野に、例えば河川巡視が挙げられる。河川巡視は、河道及び河川管理施設等の状況の把握、河川区域等における違法・違反行為の発見、河川空間の利用に関する情報収集、河川の自然環境に関する情報収集を対象として、概括的に行うものである<sup>1)</sup>。平常の河川巡視は、大別して一般巡視と目的別巡視に分けられる。あらかじめ設定した巡視項目全般について行う一般巡視に対し、巡視項目や目的・場所を絞り込んで行う目的別巡視は、水質状況や河川敷地の利用状況、ごみの不法投棄等を詳細に把握するため実施される。一般・目的別巡視に拘らず、写真を含む各情報は RiMaDIS (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) に記録される。RiMaDIS は、河川維持管理業務を効率化し、知見の効率的な集積を目的として構築され、今後の応用が期待されている<sup>2,3)</sup>。

RiMaDIS 内の画像は、人により撮影された地上画像である。車両等移動手段を使用するが、そのたびに撮影者が停止して撮影される「点」の画像であり、過去の撮影位置との紐づけ自動化や画角の一定保持は難しい。また、進入困難な場所の撮影もあり、労力を要する。

これに対し、ドローン (UAV) + AI を用いた河川巡視を提案する。ドローンによる不法投棄などの監視は 2012 年ごろから提案されていた<sup>4)</sup>が、解像度に応じてカメラが大きくなること、撮影範囲が広がるのに応じて画像枚数が増加することから、検討にとどまっていた。現在、高解像度カメラの小型化が進み、十分実用的な技術として、検証が行われている<sup>5)</sup>。また、人工知能による画像認識技術も、人の誤認率を凌駕<sup>6)</sup>してきている。

今、ドローン河川巡視 + AI を考えるとき、地上画像は豊富にあり、空撮画像はいまだ少ない状況と考える。現段階では、飛行開始から報告書まで、End to End の自動化は困難であるため、まずはドローンを飛行させて撮影し、画像から対象物を検知する工程について考える。空撮画像は、飛行中のカメラから、おおよそ一定の画角で得られる「面」の画像であり、GPS を基に設定される運行情報と撮影された対象位置情報との機械的な紐づけは、比較的容易であると考えられる。また、一部の河川

上空を除き、DID（人口集中区）等の進入困難な場所も少ない。ただし、ドローンによる空撮画像は、まだ数が少ないことから、AIに学習させるサンプル数を確保する必要がある。学習データを拡張（Augmentation）する方法も考えられるが、基となる画像データが少なすぎる状況での不適切な Augmentation は過学習を誘発するため、推論する画像群の持つ特徴量を考慮して行うのが効果的と考えられる。ここでは、既に RiMaDIS が持つ地上画像の活用を考える。地上画像と空撮画像には、対象物が同じでも、画角や距離といった条件に違いがあると考えられる。撮影条件に応じて、空撮画像の対象物検知精度が向上する地上画像サンプルを選択可能になれば、特定の検知手法に対して最適な地上画像を抽出し、空撮画像の少ない初期段階でも学習に活用可能と考えられる。そのため、検証は地上画像の中でもサンプル数が豊富な不法投棄に注目し、対象物検知を行う。

対象物検知（Object Detection）は画像内に特定の対象物が存在するかを検知し、対象物を囲む枠（Bounding Box）を座標で算出する画像分析技術である。機械学習、特に深層学習を用いた高精度な技術が特徴である<sup>7)</sup>。対象物領域の推定と、クラス分類や枠の座標推定を1ステップでまとめて行う手法と、2ステップに分けて行う手法がある。前者は後者と比較して、YOLO v4<sup>8)</sup>や SSD<sup>9)</sup>など、精度を維持しながら推論の高速化を志向した手法が多い。これに対し、後者は R-CNN（Regions with Convolutional Neural Networks）を代表とする、高精度な手法である。土木分野においては、コンクリート護岸などの構造物劣化領域検出手法として利用されている<sup>10)</sup>。本研究の想定するドローン河川巡視は、リアルタイムでの推論を基本的には要しないが、報告書へ検知結果を記載すると考えた場合、ある程度の適時性を要するため、R-CNNを高速化した Faster R-CNN<sup>7)</sup>を用いる。

よって本研究では、河川巡視項目のうち、地上画像が豊富な不法投棄に注目し、画像の持つ条件（特徴量）の違いがどのように Faster R-CNN による検知に影響を与えるか、いくつかの特徴量を比較して検証を行う。

## 2. 検証に用いるネットワークと特徴量

まず、検証に用いる Faster R-CNN の動作機序と層構成について説明する。次に、対象物検知の学習用データセット改善のため、空撮画像と地上画像の選り分けに用いる特徴量3つについて整理する。

### (1) Faster R-CNN

Faster R-CNN は深層学習を用いた物体検知アルゴリズムで、2015年、Microsoft により開発された。2012年の画

像内物体認識国際コンテストである ILSVRC（ImageNet Large Scale Visual Recognition Challenge）にて、深層学習を用いたチームが優秀な成績を残して<sup>5)</sup>から、画像認識技術として、急速に研究が進んだ。2015年にはいくつかの深層学習モデルが人間の認識能力を超えている<sup>11)</sup>。

深層学習にはいくつかのネットワークがあるが、本研究で用いる Faster R-CNN は特徴マップの出力に CNN（Convolutional Neural Networks）を使用している。単なる全結合 MLP（MultiLayer Perceptron）<sup>12)</sup>の組み合わせでは出力は次元になるが、畳み込み層を加えることで、入力次元を保った特徴量が抽出され、より高度な学習が可能になる。Faster R-CNN の処理は、RPN（Region Proposal Network）による物体範囲の特定という物体検知段階と、Faster R-CNN 本体による物体分類段階の2段階で行われる。RPNの損失関数は、参考文献<sup>7)</sup>より、以下に示す。

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

ここで、 $i$ は Anchor 点の番号を示し、 $p_i$ は Anchor 点 $i$ が物体である確率を示す。 $p_i^*$ は Ground-Truth との比較ラベルを示し、Anchor 点 $i$ が物体であった場合は1、それ以外は0となる。 $t_i$ は予想される Bounding Box の座標を、 $t_i^*$ は Ground-Truth での Bounding Box の座標を示す。参考文献<sup>7)</sup>では、 $N_{cls}$ をミニバッチサイズ、 $N_{reg}$ は特徴マップ内の Anchor 数としている。 $\lambda$ は右辺二項のバランスパラメータで、参考文献を基に、 $\lambda = 10$ とする。GPU を使用可能な環境では、 $N_{cls}$ はGPUの数に依存することがある。物体か否かに関する分類損失 $L_{cls}$ はクロスエントロピーであらわされ、Bounding Box の推定損失 $L_{reg}$ は矩形回帰smooth $L_1$ を用いて、以下に表す。

$$L_{reg}(t_i, t_i^*) = \text{smooth}_{L_1}(t_i - t_i^*), \quad (2)$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise.} \end{cases} \quad (3)$$

物体検知段階では、CNN 各 Anchor 点周りにおいて、アスペクト比の異なる $k$ 個の Anchor Box を適用して、物体であるか否かについて分類を行う。この時、Anchor Box と Ground-Truth 画像での Bounding Box の IoU（Intersection over Union）を算定し、閾値を設けて背景と物体を区別する。本実験では IoU < 0.3 は背景とし、IoU > 0.6 で物体とする。ここで、式(1)の右辺第二項は、物体が検知されなかった場合、考慮されない。すなわち IoU > 0.6 の時のみ計算される。この時、それぞれの Anchor Box と Ground-Truth 画像での Bounding Box とのずれである $L_{reg}$ を算出し、回帰させる。

物体分類段階では CNN によって出力された特徴マッ

プに RoI Pooling 層を適用し、固定長の特徴量ベクトルに変換する。これを2つの全結合層に接続し、物体の有無についての分類と、矩形回帰のための出力を得る。この2段階を交互に勾配更新することで、モデルを学習する。

## (2) 画像の選り分けに用いる特徴量

本研究では学習データに用いる画像を選り分けるため、t-SNE (T-distributed Stochastic Neighbor Embedding : t 分布型確率的近傍埋め込み法)<sup>13</sup>+k近傍法<sup>14</sup>, k平均法<sup>15</sup>, BB占有率 (画像全体における Bounding Box の占有率 : 本論文で提案), これら3つの特徴量を使用する。はじめの2つについては、ピクセルごとの類似度であり、画像が持つ背景情報の類似度を示しているといえる。3つ目も同様だが、特に画像撮影の条件 (画像内にどれほど撮影対象が捉えられているか) を示す指標として採用した。

### a) t-SNE + k 近傍法

t-SNE は非線形次元削減手法である。高次元データ  $X$  内の点  $x_i, x_j$  におけるデータ間距離  $d(x_i, x_j)$  を求め、類似度  $p_{ji}$  を以下のように定義する。

$$p_{ji} = \frac{\exp\left(-\frac{d(x_i, x_j)^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{d(x_i, x_k)^2}{2\sigma_i^2}\right)} \quad (4)$$

$$p_{ii} = 0$$

上式  $d$  がユークリッド距離の時、 $x_i$  に対し、 $x_j$  がガウス分布に比例して選ばれるならば、類似度は条件付確率で表現されることを示している<sup>13</sup>。次に、条件付き確率を対称化することにより結合確率  $p_{ij}$  を以下に定義する。ただし、 $N$  はデータ数を示す。

$$p_{ii} = \frac{p_{ij} + p_{ji}}{2N} \quad (5)$$

次に低次元空間データ  $Y$  内の点についても同様に類似度を求めると以下のように表される。

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_k \sum_{l \neq k} \left(1 + \|y_k - y_l\|^2\right)^{-1}} \quad (6)$$

$$q_{ii} = 0$$

この時、 $q_{ij}$  は t 分布を仮定しているのが t-SNE たる由縁になる。式(5)と式(6)から同時分布  $P$  および  $Q$  の間のカルバック・ライブラー情報量は以下のように表される。

$$KL(P||Q) = \sum_j \sum_{i=j} p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right) \quad (7)$$

二つの確率分布の差異を示す、この指標を最小化することで、t-SNE は次元削減を行う。

本研究では t-SNE によって得られた画像の2次元デー

タを k 近傍法を用いて、学習用の空撮画像周辺にある地上画像を選び出し、学習に用いる。t-SNE はカルバック・ライブラー情報量を最小化することから、画像全体で確率的に近い特徴を持つ画像が選ばれていると考えられる点に留意する。また、t-SNE は精度の良い非線形な次元削減を可能とするが、事後に得られたデータの分類に、再度全体の演算を必要とする点にも注意する。

### b) k 平均法

k 平均法は非階層型クラスタリング手法である。各データ  $x_i (i = 1, \dots, n)$  に対して、ランダムにクラスタを振り、各クラスタの中心 ( $V_j (j = 1, \dots, k)$ ) を計算する。各  $x_i, V_j$  間の距離を求め、 $x_i$  を最も近いクラスタに割り当てなおす操作を、閾値を満たすまで繰り返す (式(8))。

$$\arg \min_{V_1, \dots, V_k} \sum_{i=1}^n \min_j \|x_i - V_j\|^2 \quad (8)$$

k 平均法はこの通り、初期値に依存するため、本研究では改善された k-means++<sup>15</sup> を用いる。各データ間の距離を用いる点は t-SNE と同様であるが、次元圧縮を伴わないため、より画像データの特徴を残しつつ、空撮画像に類似する地上画像を選び出せると考えられる。

### c) BB 占有率

BB 占有率は本研究で新たに提案する手法である。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。これを利用した新しいデータセットも開発されている<sup>17</sup>。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。

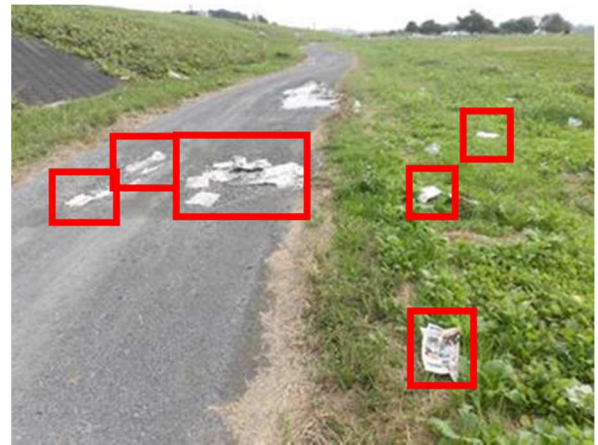
## 3. 3 種の特徴量を用いた比較実験と考察

### (1) ベンチマーク実験

地上画像を加えて、検知精度が向上するかを確認するため、空撮画像のみを使った場合をベンチマークとする。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比較して、画像サイズに対する不法投棄の割合がごくわずかであることがわかる。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  $640 \times 480$  [pixel] 程度 (図-1 内橙破線) に切り出し、学習に使用する。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。地上画像は最大 800 枚使用でき



(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

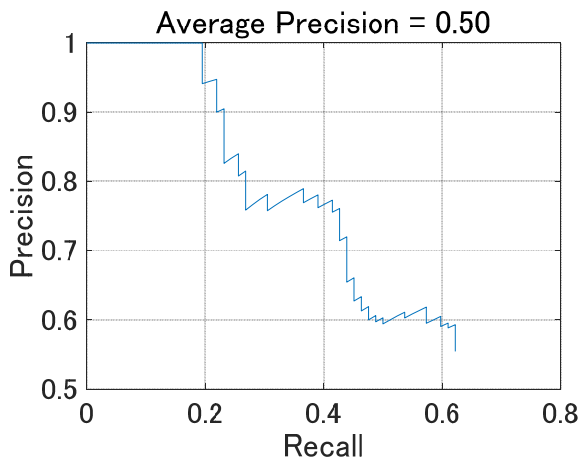


図-2 ベンチマークのPR曲線



図-3 ベンチマークで推論した画像

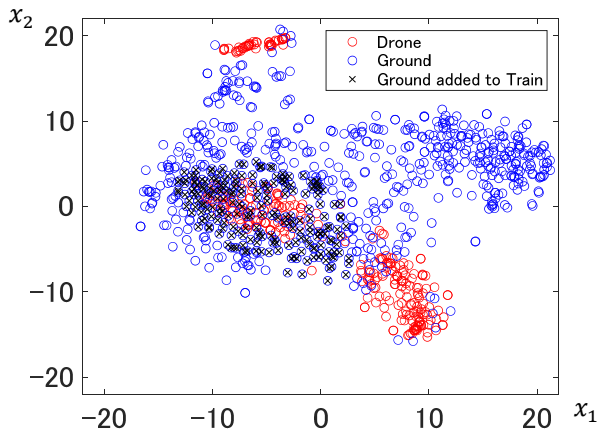
るとする。空撮画像は、疑似的に再現した不法投棄の上空から撮影している。そのため、空撮画像がほぼ直上から不法投棄を撮影しているのに対し、地上画像は横方向からになる。この条件の差異によって、影を特徴量としてどのように評価するか等の影響が予想される。特徴量の抽出を効率化するため、事前学習済み ResNet50<sup>11)</sup> を使用し、抜き出し層は 40 層の Relu 層とした。入力サイズが 224×224 の RGB 画像を用いた。学習には MATLAB 環境を利用し、勾配計算は SGDM (Stochastic Gradient Descent with Momentum : モーメンタム項付き確率的勾配降下法)<sup>16)</sup>、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 に固定して行った。以降の実験についても、これらの設定は変更しない。

図-2 にベンチマーク実験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られているが、精度・信頼度共に高くはないことがわかる。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

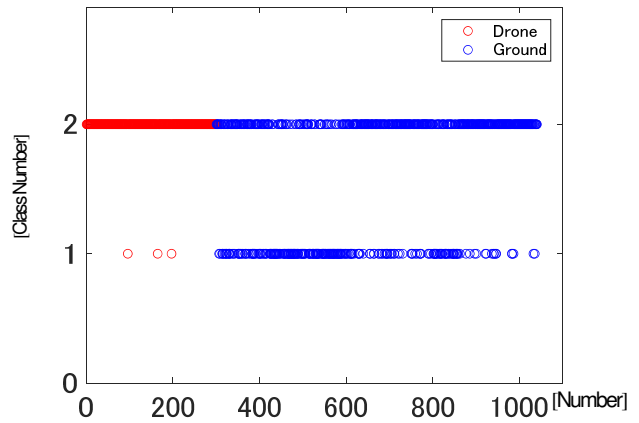
体は見切れているため、教師画像には不法投棄として設定されていないことから、今回は検知できていなくとも問題ない。また、Bounding Box の中心が不法投棄の影上にあることから、影の方により注目していると予想できる。原因の一つに、背景が灰色の強く、明るいことから、影への注目が強くなったと考えられる。これらの結果から、十分な学習が行われておらず、背景と不法投棄を判別できていないことが示唆される。ここから、地上画像を学習画像に追加し、精度が向上するかを検証する。

## (2) t-SNE + k 近傍法

t-SNE を用いて各画像データを 2 次元データにしたものを、図-5(a)に示す。横軸が第 1 成分  $x_1$ 、縦軸が第 2 成分  $x_2$  を表す。赤丸が空撮画像 (Drone) を、青丸が地上画像 (Ground) を示す。赤丸の内、青丸が周囲に多く存在する中央左の群 (103 枚) を空撮画像の学習に 83 枚、推論に 20 枚使用する。k 近傍法を用いて、周囲の 200 データ程度 (図-5(a)中の黒× : Ground added to Train) を目安として、地上画像 215 枚を選び、学習に追加する。データ数の配分は、考察の章において、表-1 に整理する。推論結果の PR 曲線を、図-6(a)、図-7(a)、に示す。ベンチ

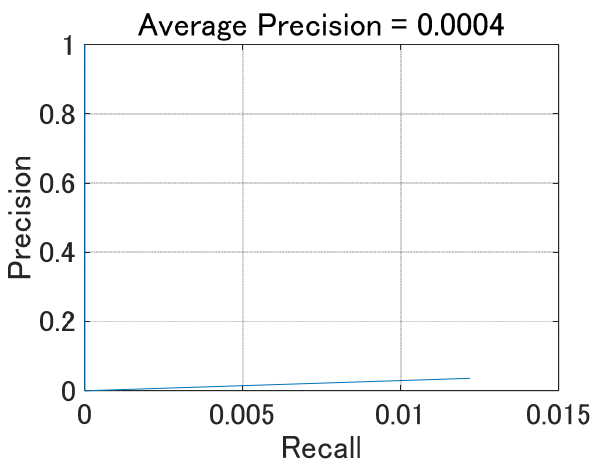


(a) t-SNE の 2次元プロットと k 近傍法の適用結果

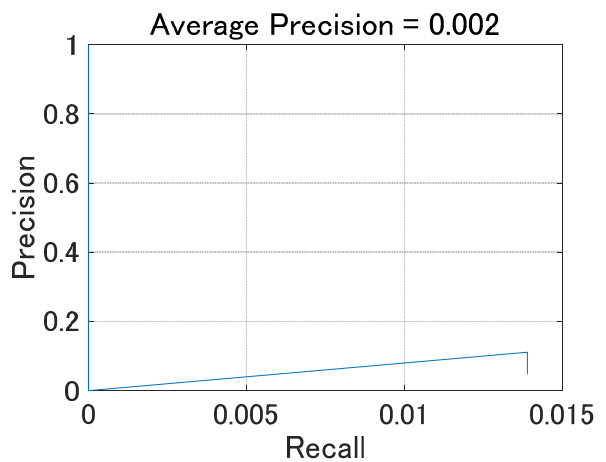


(b) k 平均法を用いた 2 クラス分類の結果

図-5 t-SNE + k 近傍法および k 平均法による分類結果



(a) t-SNE + k 近傍法



(b) k 平均法

図-6 t-SNE + k 近傍法および k 平均法を用いた学習・推論結果：PR 曲線



(a) t-SNE + k 近傍法



(b) k 平均法

図-7 t-SNE + k 近傍法および k 平均法を用いた学習・推論結果の一例

マークと比較して、全く検知できていないことがわかる。推論結果の一例を図-7(a)に示す。ベンチマークで半分程度の範囲は検知できていた不法投棄が、ほとんど検知できていない。背景のうち、足跡のような明暗のある地面を、不法投棄として検知しているように見える。結果から、地上画像の持つ特徴量が、空撮画像の持つ特徴量と混在し、検知精度が悪化していると予想できる。

### (3) k 平均法

次に、k 平均法を用いて 2 クラス分類を行う (図-5(b))。縦軸をクラス番号、横軸をデータ番号とする。空撮画像 (赤丸) 297 枚と、同じクラス (クラス 2) に多く分類された地上画像 (青丸) 482 枚を空撮画像に追加し、学習を行った。推論に空撮画像のうち 20 枚を使用した。t-SNE + k 近傍法の実験と比べ、全体の学習デー

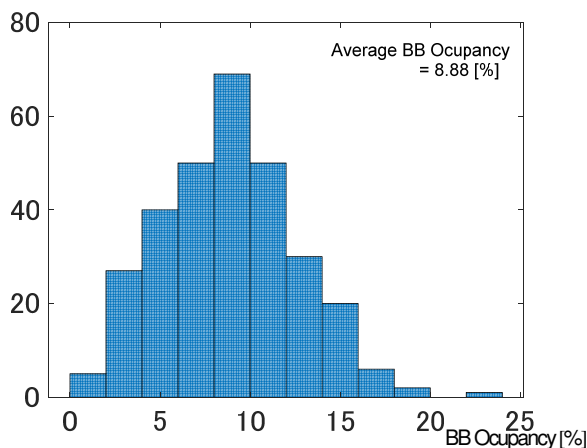


図-8 空撮画像の BB 占有率ヒストグラム

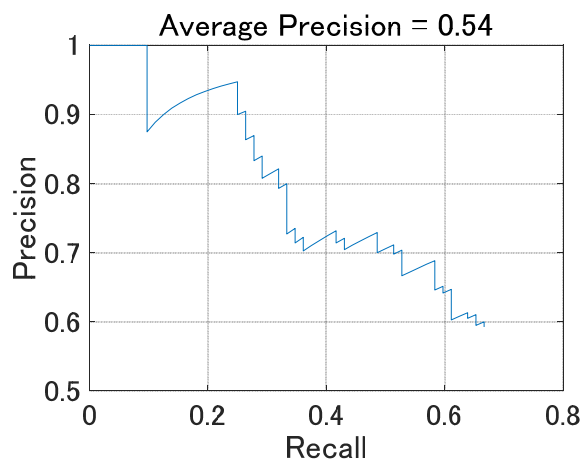


図-9 BB 占有率を用いた際の PR 曲線



図-10 BB 占有率を用いた際の推論結果の一例

表-1 各実験のデータ配分

実験番号	(1)	(2)	(3)	(4)
空撮画像	280	83	277	280
対(1)枚数	0	-197	-3	0
地上画像	0	215	482	95
合計	280	298	779	375

単位：枚

表-2 平均 Precision

実験番号	(1)	(2)	(3)	(4)
平均 Precision	0.50	0.0004	0.002	0.54
対(1)スコア	0	-0.4996	-0.498	+0.04

タは増加しており、特に地上画像は倍以上に増加している。PR 曲線と、推論結果の一例を図-6(b)、図-7(b)に示す。PR 曲線と平均 Precision は、t-SNE + k 近傍法の結果よりは改善しているが、十分でない。推論結果の一例を見ると、ベンチマークでも検知できていなかった中央付近のペットボトルが検知されていることがわかる。これは地上画像にペットボトルのゴミが多かったことで、空撮画像では捉えづらかったペットボトルを検知できるようになったことが想定される。このように、データセットが持ち合わせていない特徴量を、他のデータセットで補完することは転移学習などで行われており、本結果の妥当性を示唆している。

#### (4) BB 占有率

最後に、BB 占有率を用いて、学習用の地上画像を選び出す。BB 占有率は、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を行った。まず、図-8 に空撮画像の BB 占有率のヒストグラムを示す。横軸は BB 占有率になる。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 曲線を図-9 に示す。ベンチマーク (0.50) より向上して

いることがわかる。次に、推論結果の一例を図-10 に示す。他のすべての結果より、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、k 平均法を使用した場合と同様、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。

#### 4. 比較実験の考察

ベンチマークを含め、行った 4 つの実験について考察する。表-1、表-2 に各実験に使用したデータ配分と、得られた検知精度のスコアである平均 Precision を示す。実験番号は前章の項番号に対応する。使用した比較用の特徴量は、すなわち、(1)からベンチマーク、t-SNE + k 近傍法、k 平均法、BB 占有率の順となる。

選り分けにより、空撮画像の枚数が減少することも考えられる。これは、学習すべき特徴量を絞り込んでデータを洗浄 (Cleansing) しているとも考えられる。減少と洗浄のトレードオフを確認するため、(1)ベンチマークに対して、使用した空撮画像の増減を見る。最もスコア

の低い(2)は、地上画像の枚数は大きく増えているが、空撮画像の枚数が少なくなった。これは、(2)の特徴量を基にした場合、空撮画像の中でも地上画像より類似度が低い画像が多数となり、除外されることで、かえって学習データの確保が困難になることが予想される。次に、空撮画像の使用枚数がほとんどベンチマークと変わらない(3)においては、地上画像を多く追加して学習データを増やしているように見えるが、結果としてスコアは悪化している。これも、空撮画像とあまり共通していない特徴量を持つ地上画像を選んでしまい、かえって学習を困難にってしまったと考えられる。ただし、ベンチマークでは検知できなかったペットボトルを検知したことから、空撮画像の持つ特徴量のみでは検知困難な対象物を、地上画像により補完できていると考えられる。

ベンチマークと同数の空撮画像を用いて学習を行った(4)は、地上画像を追加することによって、スコアの上昇とともに、(3)と同様にペットボトルの検知に成功している。背景と対象物の比率が偏っていると、その背景と対象物に重要な相関性を想定してしまい、適切な学習が困難になるといった既往の研究<sup>18)</sup>等を考慮すると、ある程度整った比率で、異なる背景の画像を活用することが効果的であることを示唆している。したがって、データの Augmentation と同様に学習データを補完するだけでなく、適切な画像特徴量を選ぶことで、異なる状況（例えばペットボトルや違法行為である焚火そのものの画像）を学習データに追加し、空撮画像では未だ得られない状況の画像についても、検知の可能性が考えられる。

これら実験およびその考察から、ドローンによる空撮画像に、地上画像など撮影条件の異なる画像を、適切な特徴量を比較考慮しながら追加することで、学習結果を改善可能であると確認できた。

## 5. まとめ

ドローンによる空撮画像のみで得られた学習データに、人による地上画像を追加することで、学習を改善すること、その適切な指標・手法について探索検証することを目的とし、本研究で得られた知見を以下3つにまとめる。

- 1) データの選り分けに用いる特徴量を3つ (t-SNE + k近傍法, k平均法, Bounding Box 占有率) 提案し、比較実験を行った。提案した特徴量を基に、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近傍法やk平均法では、スコアが悪化した。このことから、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した。

- 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、地上画像を追加することで、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情報・状況を補完的に学習できる可能性が示唆された。

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、偏りが無いとは断言できず、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、ドローン河川巡視が一般化されていくに際し、空撮画像では得られていない状況についても、地上画像等で補っていく必要がある。RiMaDIS には不法投棄に限定しても数十万枚の画像が存在するため、これらを活用し、今回効果が確認された Bounding Box 占有率の洗練や、新たな特徴量指標の検討が可能であると考えられる。今後の検討予定として、今回特徴量抽出に用いた ResNet50 などのネットワークを用いて、データセットをよりよく整理する方法等を考えている。

謝辞：本研究は国土交通省の革新的河川技術開発の一環として実施したものです。国土交通省水管理・国土保全局、および中部地方整備局にはプロジェクト全般でのサポートとフィールドの提供で多大なご協力を頂きました。ここに感謝の意を表します。

## 参考文献

- 1) 国土交通省：河川砂防技術基準 維持管理編（河川編），pp.32-34，2015。
- 2) 工藤愛樹：河川管理データベース（RiMaDIS）について，月刊「建設」，一般社団法人全日本建設技術協会，pp.39-41，Vol.62，7月号，2018。
- 3) 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。

- 4) Lega, M., Ceglie, D., Persechino, G., Ferrara, C. and Napoli, R.M.A.: Illegal dumping investigation: a new challenge for forensic environmental engineering, *WIT Transactions on Ecology and The Environment*, Vol 163, 2012.
- 5) 清水隆博, 秋田麗子, 野間口芳希, 米倉瑠里子, 佐藤隆洋, 炭田英俊: 河川管理における UAV の活用方策に関する現地実証試験, こうえいフォーラム, 第 26 号, 2018.
- 6) Krizhevsky, A., Sutskever, I. and Hinton, G.: Imagenet classification with deep convolutional neural networks, *NIPS*, 2012.
- 7) 例えば, Ren, S. , He, K., Girshick, R. and Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks, *NIPS*, 2015.
- 8) Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv*, arXiv:2004.10934, 2020.
- 9) Liu, W., Anguelov, D. Erhan, D. Szegedy, C. and Reed, S.: SSD: Single shot multibox detector. *arXiv preprint*, arXiv:1512.02325, 2015.
- 10) 齋藤彰儀, 上総虎智, 平木悠太, 天方匡純, 吉田武司: 深層学習によるコンクリート護岸劣化領域検出システムの開発, デジタルプラクティス, 情報処理学会, Vol.10 No.2, 2019.
- 11) He, K., Zhang, X., Ren, S. and Sun., J.: Deep residual learning for image recognition, *Computer Vision and Pattern Recognition (CVPR)*, 2016.
- 12) Gardner, M.W. and Dorling, S.R.: Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, *Atmospheric Environment*, Vol.32, Issues 14–15, pp. 2627–2636, 1998.
- 13) van der Maaten, L. and Hinton, G. E.: Visualizing Data using t-SNE. *Journal of Machine Learning Research*, Vol.9, pp. 2579–2605, 2008.
- 14) Dasarathy, B. V.: Nearest-neighbor approaches, *Handbook of Data Mining and Knowledge Discovery*, pp. 88–298, Oxford University Press, 2002.
- 15) David, A. and Vassilvitskii. S.: K-means++: The Advantages of Careful Seeding., *SODA 2007: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035, 2007.
- 16) 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.
- 17) 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.
- 18) Ribeiro, M. T., Singh, S. and Guestrin, C.: “Why Should I Trust You?”: Explaining the Predictions of Any Classifier, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp.1135-1144, 2016.

(Received June 30, 2020)  
(Accepted July 31, 2020)

## APPLICATION OF GROUND PICTURE TO RIVER PATROL BY THE IMAGE RECOGNITION AI AND UAV – VERIFICATION OF THE FEATURES VALUE, AND THE USAGE –

Yuta TAKAHASHI, Junichiro FUJII, Masazumi AMAKATA and Takayoshi  
YAMASHITA

Japan has many rivers, and the almost management work is carried out manually. Due to the development of drone (UAV) technology, they are applied to river monitoring, and the research with AI, such as using the image taken by drone for detect illegal dumping, is also increasing. On the “ground”, taking a lot of picture for report by humans requires hard work and is also subject to the diversity of condition. The aerial image has a different angle of view from the ground image and the conditions are different. In this research, whether can utilize the ground image of river maintenance management database system RiMaDIS to improve the detection accuracy in less aerial image are verified. The image of illegal dumping taken from ground and drone are classified by three image feature value, and they were learned by Faster R-CNN. The verification result suggests that features value such as Bounding Box occupancy contribute to the improvement of detection accuracy.