StableDiffusionなどの画像生成AIを用いたドローン河川巡視用AIのデータ増強

The data augmentation for UAV river patrol AI using generated image by generative models such as Stable Diffusion

高橋 悠太[*1] 藤井 純一郎[*1] 天方 匡純[*1]
Yuta Takahashi Junichiro Fujii Masazumi Amakata

[*1] 八千代エンジニヤリング株式会社 技術創発研究所

The data in the civil engineering field is less data with much variety. Drone river patrols fly vast river areas and detect illegal dumping, including general garbage, using AI. The patrol drones are not constantly flying, they are rarely captured by aerial images, and it is even more difficult to detect temporary illegal occupation. In previous studies, it has been confirmed that adding images taken on the ground with different angles of view to the learning data improves learning, but the number of images is required for training even if the images are taken on the ground. In order to improve the learning of the detection model, this study verified whether the image for data augmentation can be generated and learned by image generation AI such as Stable Diffusion.

1. はじめに

土木分野における特異な物体の検知はその多様さに比べてデータが少ない. ドローン河川巡視はドローンにより広大な河川領域を撮影し, 一般ごみを含む不法投棄などを AI により検知することになる. 巡視用ドローンが常時飛行でない場合, 空撮で捉えられることは少なく, 一時的な不法占用等の検知はさらに困難となる. 既往の研究において, 画角は異なるが地上で撮影された画像を学習データに加えることで学習を改善する効果が確認されているが, 地上撮影であっても画像数が要求される. 本研究では, 特に検知モデルの学習を改善するため, データ増強用の画像を Stable Diffusion[Rombach 2021]などの画像生成 AI により生成・学習が改善するか検証を行った.

2. 方法論

本研究では既往の研究を参考に, Faster R-CNN[Ren 2015]を用いる. 特に生成画像した画像に対し, 既往研究で学習された学習済みモデル(既往モデル[高橋 2020]を使用し, 生成画像の有効性を確認する. 既往モデルは特定の河川で得られた疑似不法投棄画像を学習したモデルで, 同様の疑似不法投棄に対しては未学習であっても十分な精度が得られている. 次に, 生成に用いた Stable Diffusion のパラメータについて整理する. 本モデルはローカル環境で使用可能な日本語化ツール[AUTOMATIC1111]が存在し, 本論文では 2023 年 1 月段階での Git の訳を基準とする. Stable Diffusion は Latent Diffusion モデルを利用した生成モデルで, Transformer[Vaswani 2017]を用いてテキストを画像との空間に埋め込み, プロンプトと呼ばれる入力テキストによって求める画像を, 崩壊なく得られるように開発された. 使用する日本語ツールは Stable Diffusion 以外のモデルも使用可能であるが, 今回は Stable Diffusion v-21-768-pruned モデル[Stability AI]を safetensor 形式で使用した.

3. 実験方法

Stable Diffusion の txt2img 機能を用い, プロンプトに入力して画像を生成し, 既往モデルに入力して検知結果を検証する. 生成する画像は不法投棄として多い, ペットボトル(plastic bottle), ゴミ袋(plastic bag), 三角コーン(safe corn)としてプロンプトに入力した. 空撮風の画像を得たいため, 空撮画像(aerial image)を入力した. 空撮画像であるため, 対象は小さく写る必要があることから, Small を入力し, 定冠詞 a をつけて対象が大量に生成されることを防止した. 加えて Negative ワードに many, garbage を追加することで, 同じく大量のゴミが生成され, 実際の環境とも異なる上検知結果の検証が困難な画像が生成することを防止した. 背景について, 通常河川敷の状況は各河川で異なるため, 今回草原(grass)岩場(rock)土(cray)の三種類を入れ替えて生成した. また河川敷であることが強調されるよう, river side をポジティブに入れた. すなわちペットボトルが岩場にある画像を生成したい場合, プロンプトは a small plastic bottle, on rock field, in river side, aerial image.となる. 画像サイズは 768*768 pixel, サンプリング回数は 30, CFG スケールは 14 とした.

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

4. 生成結果と既往モデルによる検知結果

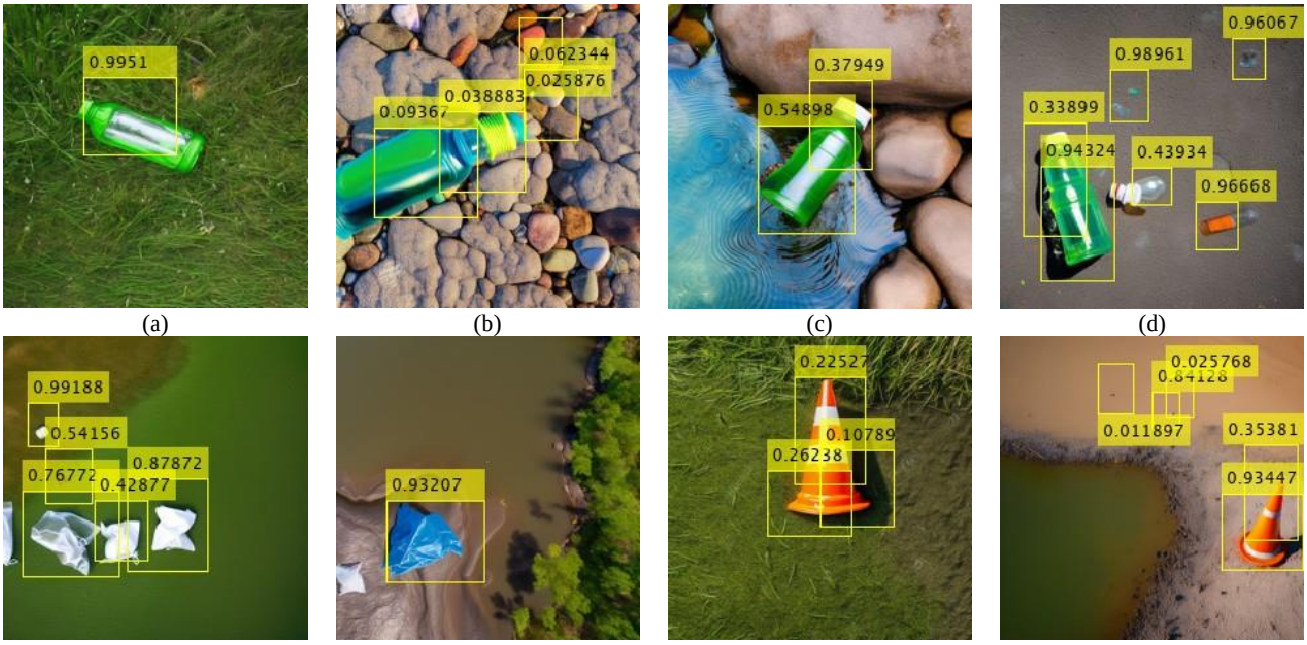
4.1 生成結果

図-1 に各生成結果について示す. 草原の上のペットボトル:(a), 岩場のペットボトル:(b)-(c), 土の上のペットボトル:(d), 草原の上のゴミ袋:(e), 岩場のゴミ袋:(f), 草原の上の三角コーン:(g), 土の上の三角コーン:(h)となった. 土の上のゴミ袋および岩場の三角コーンは大量に生成される, 対象の輪郭が崩壊するなどしたため, 記載していない. 学習データの影響か, ペットボトルは緑色になることが多かった. また, ネガティブに many を入れていたとしても, (d)のように破片のような形で複数のペットボトルが生成されている. これは海外の河川敷や河口部で得られた河川海洋ゴミの画像を学習しているためと考えられる. また, 本来ドローンは高さ 30-50m 程度から撮影することが望ましいため, 今回対象ピクセル数は 1 辺 200 ピクセル前後のため, 地上解像度 1cm の場合, ペットボトルが 2m の大きさとなってしまう. そのため, 一般的な解像度より高い解像度のカメラを使用したあるいは低高度で撮影したと考える. 2L ペットボトルの高さがおおよそ 40cm とすると今回の生成結果を現実の解像度に直した場合, 地上解像度は 0.2cm であったといえる. 既往モデル

(a) (b) (c) (d)

(e) (f) (g) (h)

図-1 生成画像



(a) (b) (c) (d)

(e) (f) (g) (h)

図-2 生成画像に対する既往モデルでの検知結果

に入力するため，224*224 pixel に圧縮して入力するため，形状情報は担保されると考え，検知を行う．

4.2 検知結果

既往モデルに生成画像を入力し，不法投棄を検知した結果を図-2 に示す．既往モデルは ResNet50 をバックボーンとしているため，生成画像を224*224 pixelに圧縮して入力している．黄色枠に付した数字は Confidence を示し，閾値は 0.01 とした．ペットボトル：図-2(a)-(d)は既往モデルにおいてもデータ数が十分あり，学習できたためか，高い Confidence で検知できていることが分かる．ただし，草原のような場所においては高く，岩場のような場所では低くなっていることが分かる．これは既往モデルでも岩場のデータが少なかったことが考えられる．次に，ごみ袋：図-2 (e)(f)について，一部アノマリーの検知漏れ・過剰検知はあったが，検知すべき対象については 0.5 以上の Confidence で検知できていることが分かる．0.5 以下の場合であっても，Negative Maximum Suppression (NMS)などで補完可能であること，また実際の河川巡視では一部検知できていれば周辺を同様の現象と定義するため，十分検知できているといえる．加えて，三角コーン：図-2(h)の中でいくつか小さなアノマリーを検知したが，図-2(g)(h)についてもある程度検知できていることが分かる．

5. 考察

5.1 プロンプトについて

プロンプトは生成画像の変化に直結するため，微調整が望ましい．本実験では背景や対象の変化に伴うテキスト変化の影響を確認するため，一部テキストを入れ替えることで，目的の画像生成を試みた．土の上のゴミ袋および岩場の三角コーンは大量に生成される，対象の輪郭が崩壊するなどしたが，これらはプロンプトを微調整することで目的の画像に近づけることができたと考えられるが，学習用の画像として用いる場合，大量に必要となり，ある程度の画像の安定性を担保したいことから，これらの微調整による改善がどれほど見込めるかは検証を要する．

5.2 生成結果と検知結果について

生成画像に対する検知結果は，検知対象をある程度の Confidence で検知できていることが分かった．これは既往モデルが生成に用いたプロンプトに含まれる対象(plastic bag, plastic bottle, safe corn)を十分学習できていたこと，生成結果が既往モデルで学習したような空撮された模擬不法投棄に似た画像を生成できたことを示す．すなわち，生成した画像を既往モデルに入力し，検知できなかったあるいは過剰に検知したアノマリーを有する画像を生成・学習することで，より汎用的な検知モデルの構築が可能と言える．

6. まとめ

ドローン河川巡視 AI のための学習用空撮画像が対象の多様性に対し不足する課題に対し，Stable Diffusion などの生成モデルを利用して学習データが補完可能か，検証を行った．生成した画像を既往研究で学習したモデルに対し入力し，検知結果を分析することで以下の知見を得た．

1) 不法投棄としてゴミ袋，ペットボトル，三角コーンをプロンプトとしてそれぞれ plastic bag, plastic bottle, safe corn として入力し，背景を草原(grass)，岩場(rock)，土(cray)，ネガティブワードとして many，garbage を入力し Stable Diffusion で画像生成した．結果，対象が大量に写ることなく，安定した画像を取得できた．

2) 生成した画像を既往モデルに入力し，不法投棄として入力したプロンプトの対象を検知可能か，検証を行った．一部既往モデルでは学習データの少なかった背景に対して Confidence が下がり，他のアノマリーを検知することあったが，概ね対象を高い Confidence で検知できた．

3) 以上の結果から，既往モデルが生成に用いたプロンプトに含まれる対象(plastic bag, plastic bottle, safe corn)を十分学習できていたこと，生成結果が既往モデルで学習したような空撮された模擬不法投棄に似た画像を生成できたことを確認し，生成した画像を既往モデルに入力し，検知できなかったあるいは過剰に検知したアノマリーを有する画像を生成・学習することで，より汎用的な検知モデルの構築が可能であることを確認した．

本研究では生成した画像を学習に用いる場合アノテーションが必要となる．現在，プロンプトに含める対象の画像上での位置を規定して生成するモデルも研究されている[Li 2023]ことから，これらを利用してアノテーション作業も不要に学習データを生成可能になると考えられる．今後はこれらモデルを活用する方法，アノテーションされた画像を学習に用いて，モデルが改善可能か検証を行う．

参考文献（論文誌と同じスタイルを推奨）

[Rombach 2021] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer: High-Resolution Image Synthesis with Latent Diffusion Models, arXiv:2112.10752, 2021.

[Ren 2015] S. Ren, K. He, R. Girshick, and J. Sun: Faster r-cnn: Towards real-time object detection with region proposal networks, arXiv preprint arXiv:1506.01497, 2015..

[AUTOMATIC1111] https://github.com/AUTOMATIC1111/stable-diffusion-webui

[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.

[Vaswani 2017] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin: Attention Is All You Need, arXiv:1706.03762, 2017.

[Stability AI] https://huggingface.co/stabilityai/stable-diffusion-2-1

例えば[Li 2023] Y. Li, H. Liu, Q. Wu, F. Mu, J. Yang, J. Gao, C. Li, Y. J. Lee :GLIGEN: Open-Set Grounded Text-to-Image Generation, arXiv:2301.07093, 2023.