単眼深度推定器の脆弱性検証を目的とした実環境評価型敵対的攻撃の試み

日下部尊[1,a)]　小野智司[1,b)]

概要：近年, 深層ニューラルネットワーク（Deep Neural Network: DNN）の進歩により単眼深度推定の性能は大幅に改善されている．一方で DNN は, 入力画像に微小な摂動が加わることで誤分類を引き起こす敵対的攻撃の危険性が明らかにされており, 単眼深度推定用の DNN にも同様の脆弱性が懸念されている．DNN を用いたシステムの実社会への応用が進んでおり，これらのシステムの頑健性の強化が急務となっている．このため，本研究では，対象シーンにプロジェクタを用いて摂動光を投影することで，単眼深度推定器の誤認識を引き起こす投光型敵対的攻撃方式を提案する．特に，摂動の設計を行う最適化において，実環境を用いて解候補の評価を行う実環境評価型進化計算を利用する．実験により，本来のロッカーの位置を後ろへ誤推定させる結果を確認した．

キーワード：単眼深度推定器，敵対的攻撃，物理攻撃，実環境最適化

Real-world assessment-based adversarial attacks to verify the vulnerability of monocular depth estimators.

Abstract: Recent advances in Deep Neural Networks (DNNs) have significantly improved the performance of monocular depth estimation. On the other hand, DNNs have been shown to be vulnerable to adversarial attacks that cause misclassification when small perturbations are applied to the input image, raising concerns about the vulnerability of DNNs for monocular depth estimation. As DNN-based systems are increasingly being applied to real-world applications, there is an urgent need to enhance the robustness of these systems. Therefore, this study proposes a projection-based physical adversarial attack method in which perturbed light is projected onto a target scene using a video projector to cause the misrecognition of a monocular depth estimator. Particularly, the proposed method introduces a physics-in-the-loop optimization method for perturbation design, which evaluates solution candidates using actual devices rather than the simulation. Experimental results demonstrated that the monocular depth estimator misrecognized a rocker's position in the scene as being deeper than it actually was.

Keywords: monocular depth estimation，adversarial attacks，physical attack，physics in the loop

1. はじめに

単眼深度推定とは, 単眼カメラで撮影されたシーンの 3 次元情報を推定する技術である．近年，深層ニューラルネットワーク（Deep Neural Network:DNN）の発展により深度推定の精度は飛躍的に向上し，工場や倉庫における物資の自動搬送装置，自動車の自動運転などへの活用が期待されている [1–4].

一方で，入力画像に微小な摂動を加えることで，意図的に画像分類器モデルの誤認識を誘発させる敵対的事例（Adversarial Example: AE）と呼ばれる脆弱性が存在することが明らかにされている [5]．単眼深度推定用の DNN も，畳込み層など画像分類用の DNN に共通する構造を多く含むため，同様の危険性が懸念される．単眼深度推定器を自律移動ロボット等の自動運転に用いる場合，DNN の

[1] 鹿児島大学
Korimoto, Kagoshima, Kagoshima 8900065, Japan
[a)] k9140536@kadai.jp
[b)] ono@ibe.kagoshima-u.ac.jp

誤推定が事故に繋がる可能性がある．このため，単眼深度推定器を含むコンピュータビジョンシステムを実世界に応用するためには，DNN の脆弱性の調査を行うことが必要である．

DNN の物理攻撃手法では，生成した敵対的事例を T シャツの表面に張り付け物体検出器を回避する攻撃が提案されている [6]．また，深度推定器に対して，敵対的なパッチを車の後方に張り付け車の一部が消えたかのように誤推定を引き起こす攻撃も提案されている [7]．近年では，赤外線レーザーを用いて道路標識にスポット光を照射し，自動運転車の認識システムを誤認識させる攻撃も報告されている [8]．物理攻撃の性能を向上させるために広く利用されている技術も存在し，期待値変換（Expectation over Transformation:EOT），非印刷性スコア（Non-printability score:NPS）損失，全変動（Total Variant:TV）損失，デジタル - 物理（Digital-to-Physical:D2P）が提案されている．しかしながら，Daimo らの手法 [9] を除くと DNN をブラックボックス条件下かつプロジェクタを用いた物理攻撃の研究は，著書らが調査した限りでは行われていない．

本研究では，対象シーンにプロジェクタを用いて摂動光を投影することで，単眼深度推定器の誤認識を引き起こす敵対的攻撃方式を提案する．また，摂動の設計を行う最適化において，実環境を用いて解候補の評価を行う実環境評価型進化計算を利用する．提案手法は，人間の視覚で判断した奥行と深度推定器が出力した奥行に相違が生じるような AE 生成を可能にする．評価実験により，提案手法が単眼深度推定用の DNN において，対象シーンとして設定したロッカーの本来の位置を後ろへ誤推定させることを確認した．

2. 関連研究

2.1 単眼深度推定

近年，深度推定技術に関する多くの研究が行われ，推定精度と推定計算速度は従来のアプローチと比較し大幅に向上している．これは，CNN を初めとする DNN の活用によるもので，特徴抽出機能が補強されているためである [10]．Laina らは，完全畳み込みアーキテクチャを用いることで，より少ないデータかつ，より短時間で学習できるだけでなく推定精度の高い結果を達成した [11]．深度推定の研究は，自律走行だけでなく，生物医学，ロボット工学，様々な産業用移動ロボットに応用されている．また，深度推定モデルの発展に伴い，推定に使用される情報の解析の研究も行われている [12]．

2.2 DNN に対する敵対的攻撃

Goodfellow らにより，入力画像に対して微小な摂動を加えることで DNN に誤認識をさせることが可能と明らかにされて以来，DNN に対する敵対的攻撃の研究が広く行われている [5]．脆弱性を発見する敵対的攻撃は，参照可能な情報によって一般的にホワイトボックス攻撃とブラックボックス攻撃に大別される．ホワイトボックス攻撃は，DNN モデルのパラメータ，勾配情報などを含む内部情報を利用する手法である．一方，ブラックボックス攻撃は，内部情報を用いない攻撃手法である．商用システムでは内部構造及びパラメータへのアクセスが禁止されていることも多いため，内部情報を利用しないブラックボックス攻撃による DNN の脆弱性を検証する技術の重要性が高まっている．

2.2.1 電子攻撃

電子攻撃の設定では，対象モデルの入力画像をピクセルレベルで任意に変更する柔軟性を持っている．したがって，これらの攻撃はカメラ等の入力システムを制御していることを前提としている [13]．デジタル環境下での敵対的攻撃の研究は以前から行われており，多くの手法が提案されている [5] [14]．

2.2.2 物理攻撃

物理攻撃の設定では，実世界からの入力の操作をするもので，カメラで撮影された画像に依存する．そのため，正確で微小な摂動を反映できる電子攻撃とは異なるため，カメラによって正確に捉えることが難しい．また，環境光の変化による外乱や歪みなどの様々な要因から，デジタル攻撃よりも困難な条件下での攻撃となっている．

物理攻撃手法は，パッチベース，カモフラージュベース，投光ベースに大別される [15]．また，パッチベース，カモフラージュベースは侵略的攻撃，投影ベースは非侵略的攻撃に分類される．投影ベースの攻撃は，対象物体に対して接触をせずに摂動を付加できる点，自然現象によって発生する可能性が高い点から脅威性が高い [8]．

2.3 先行研究: 単眼深度推定器に対するブラックボックス攻撃

Damimo らはブラックボックス条件下で単眼深度推定器に対する AE を生成する手法を提案した [9]．進化計算を用いて分類器に対する AE を生成する方式 [16] を応用し，人間の視覚で評価された奥行と深度推定器が算出した奥行との間に差異が生じるような AE 生成を可能にした．

評価実験では，実環境を再現した CG シミュレーション条件下での敵対的投光パターンの生成を行った．CG シミュレーション内では，対象物体であるロッカーが原画像と比較し全体的に奥に移動したような誤推定を引き起こす結果が得られた．しかし，CG シミュレーション内で生成された摂動パターンを実環境にて投影した際では，CG シミュレーションと同等の誤推定結果を得ることはできなかった．これは，物体の反射特性・環境光による外乱・カメラノイズによる歪みなど，様々な要因によって引き起こされる CG シミュレーションと実環境との間の大きな差により攻撃精度が低下したと考えられる，

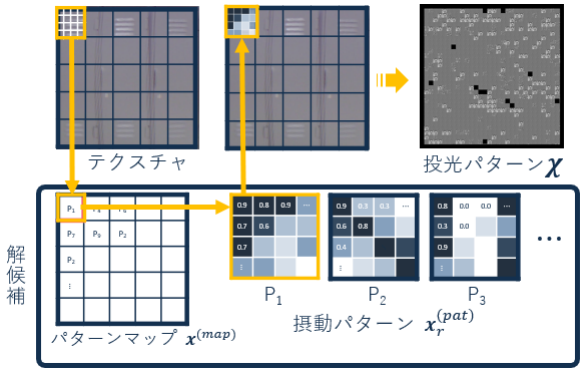


図 1 摂動設計手法

3. 提案手法

3.1 基本アイデア

本研究では，対象シーンに摂動光を投影することで，単眼深度推定器の誤認識を引き起こす敵対的攻撃方式を提案する．また，摂動の設計を行う最適化において，実環境を用いて解候補の評価を行う実環境評価型進化計算を利用する．本方式の基本アイデアを以下に示す．

実環境評価型最適化: 実環境を再現した CG シミュレーションでは影・室内光・物体の材質などの再現が不十分だったため，実環境ではシミュレーション上で得られた結果程の誤推定を引き起こせなかったと考えられている [9]．そのため，解候補の評価を，実環境化で実際に投光を行って対象 DNN の出力を得て行う．このアイデアは，Minamata らの提案手法である実環境評価型最適化の枠組みを適用している [17]．実環境評価型最適化を行うことで，室内照明の反射光・物体の材質・カメラノイズなどの実環境要因を考慮した条件下での摂動の生成を行える．

進化型多目的最適化（Evolutionary Multi-criterion. Optimization: EMO）[18] アルゴリズムの適用: AE を設計する際，深度推定誤差と摂動量の 2 つの目的関数はトレードオフの関係にある．多目的最適化において，異なる目的間のトレードオフを考慮した複数の最適解が得られるため有効な最適化手法と考える．また，進化計算は目的関数の構造や微分可能性に依存せず，ブラックボックスとなる最適化問題に有用な手法と言える．

物体表面への摂動の付加: 単眼深度推定器は画像認識器と同様に，対象画像において顕著性が高い領域に摂動が加わることによって，敵対的攻撃の影響が強まることが知られている [12]．すなわち，画像内の重要な特徴やエッジ，物体の境界の特徴が，深度推定の手がかりになっていると考えられている．このため，提案手法は対象物体の表面に摂動光を投影する．

3.2 定式化

提案手法は，ブロック単位の摂動設計手法 [16] を採用する．すなわち，摂動を $N_{pat} \times N_{pat}$ 画素のブロックに分割して構成することとし，最適化により，局所摂動パターンをいくつか設計すると同時に，摂動付与範囲内の各ブロックにどの局所摂動パターンを付与するかを決定する．これにより，高解像度のテクスチャにおいても設計変数の削減を可能とする．図 1 は，ブロックごとの摂動パターンを示す．解候補 $\boldsymbol{\chi}$ を構成する設計変数には，以下の 2 種類ある．

テクスチャ画像 $\boldsymbol{I}$ は $x_{u,v}^{(map)} \in \{0, 1, 2, \ldots, N_{AP}\}$ のように．(u, v) で区切られたパターン割り当てマップ $x_{u,v}^{(map)}$ によって表す．$x_{u,v}^{(map)} > 0$ の場合に，ブロック (u, v) に対応するブロック単位の摂動パターンが適用され，それ以外の場合は，摂動はそのブロックに追加しない．変数 $x_{p,q,r}^{(pat)}$ は r 番目の摂動パターン，すなわち，パターン r の局所的な座標 (p, q) における画素値の変動量を表す $(r \in \{1, \ldots, N_{AP}\})$.

$$\boldsymbol{\chi} = \boldsymbol{x}^{(map)} \cup \left\{ \boldsymbol{x}_r^{(pat)} \right\}_{r \in \{1,\ldots,N_{AP}\}} \tag{1}$$

$$\boldsymbol{x}^{(map)} = \{x_{u,v}^{(map)}\}_{(u,v) \in \boldsymbol{I}} \tag{2}$$

$$\boldsymbol{x}_r^{(pat)} = \{x_{p,q,r}^{(pat)}\}_{p,q \in \{1,\ldots,N_{pat}\}} \tag{3}$$

目的関数は深度推定誤差と摂動量の 2 つを設定し最小化する．目的関数 f_1 は，対象物体のテクスチャを含む周辺をマスクした領域（$i \times j$ 画素）における，推定された深度マップの深度値 $d_{(i,j)}^{est}$ とターゲット深度マップの深度値 $d_{(i,j)}^{target}$ との絶対誤差の総和とする．画像サイズを $W \times H$ 画素とする．また，目的関数 f_2 を摂動量 ρ の L2 ノルムとする．

$$\text{minimize } f_1(\boldsymbol{\chi}) = \sum_{(i,j)} \left| d_{i,j}^{(est)}(\boldsymbol{\chi}) - d_{i,j}^{(target)} \right|_{i \in \{1,\ldots,W\}, j \in \{1,\ldots,H\}} \tag{4}$$

$$\text{minimize } f_2(\boldsymbol{\chi}) = ||\rho||_2 \tag{5}$$

3.3 処理手順

提案手法の処理手順を図 2 に示す．対象シーンであるロッカー表面のテクスチャに対して，プロジェクタを用いて摂動光を投影する．投影後に撮影し，この画像を深度推定器に入力として与え，出力である深度マップから目的関数の計算を行う．この処理を繰り返すことで，目的関数を最小化し AE の生成を行う．

3.4 実機システム

本実験システムのハードウェア構成を図 3 に示す．カメラの撮影素子は Basler acA1300-30gc（1/3 インチ，1294 × 912 ピクセル，30fps），レンズは C-MOUNT VARI-FOCAL LENS DV3.4X3.8SA-1（焦点距離 3.8mm - 13mm，絞り範囲 F1.4 - CLOSE）を使用した．本実験で使用する GigE カメラへの電源供給は PoE 電源でのオプションを行った．

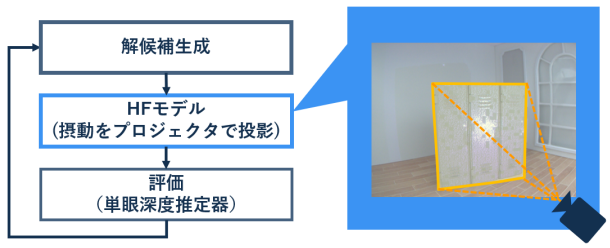


図 2 提案手法の処理手順

図 3 実機システムの構成図

プロジェクタは EPSON EB-E01（明るさ 3300lm）を使用した．プロジェクタのカラーモードは sRGB モード（明るさ 80，コントラスト 100，色の濃さ 79，色合い 23，シャープネス 20）に設定した．本実験環境では 3300lm での投影は光が強く，対象物体のテクスチャ以外の領域に対しての投影が際立ってしまった．そのため，プロジェクタのレンズに K & F Concept 58mm 可変 ND フィルターを装着し光量を抑えた．ND 濃度は ND128 に選択した．また，適切な実験環境下で実験をする際に，実験外からの外乱を防ぐため暗幕内で行った．

4. 評価実験

4.1 実験設定

本実験では，屋内シーンのデータセット NYU Depth v2 を訓練した Laina ら [11] の単眼深度推定器を攻撃対象とした．実験で生成した AE は，対象物体がシーンから消えるかのような誤推定を引き起こすように設計した．

提案手法の有効性を検証するため，屋内シーンを実際の 1/12 のスケールで再現した模型を利用して実験を行った．対象物体としてはロッカーを選択した．より現実環境のシーンを再現するために，ロッカーの材質は鉄とステンレスで制作されたものを使用している．

最適化アルゴリズムは MOEA/D [19] を使用した．スカラー化関数として Chebyshev 法を選択し，近傍サイズ $N_n = 10$，$\delta = 0.8$，$n_r = 1$，個体数 N_p を 40，世代数は 500 で行った．また，ブロック単位の摂動パターン数 N_{AP} を 10，サイズ N_{pat} を 8×8 に設定した．

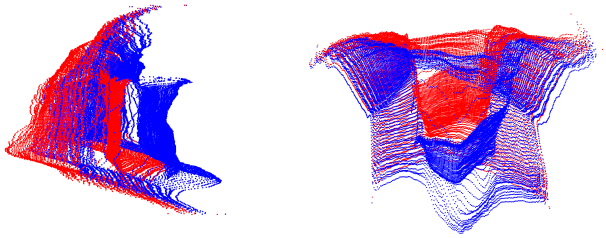
(a) 元画像　(b) 生成された AE

図 4 生成された AE の一例

(a) 元画像の深度推定結果

(b) AE の深度推定結果

図 5 元画像および AE の深度推定結果



(a) 横からみた点群データ　(b) 上からみた点群データ

図 6 元画像と AE の深度推定結果を 1 つの点群に位置合わせ

4.2 実験結果

提案手法により生成した AE を図 4(b) に示す．AE は元画像と比較して，投光により物体表面の明るさが上昇していることがわかる．

図 5 は深度推定結果を 3 次元点群で示している．図 5 の点群は，濃い青が前方を表し赤になるほど奥を表している．AE は元画像と比較すると，点群の色は青から黄緑に変化しており後ろへ誤推定していることがわかる．

図 6 は元画像と AE の 3 次元点群を位置合わせした結果を示している．図 6 の点群は，元画像の点群の色を青で統一，AE の点群の色を赤で統一している．図 6 より，AE は元画像と比較して，ロッカー全体が本来の位置より壁付近まで誤推定していることがわかる．

図 7 は，$f_1(\boldsymbol{\chi})$ スコアの推移を示している．最適化が進むにつれて，$f_1(\boldsymbol{\chi})$ スコアは減少していることがわかる．しかし，100 世代以降はなだらかな減少となっている．これは，100 世代目付近で局所解に陥った可能性が考えられる．

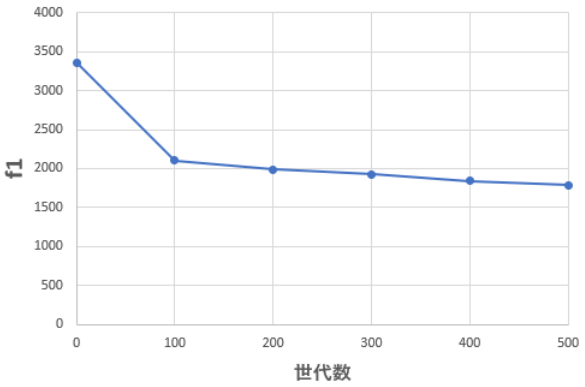


図 7 $f_1(\boldsymbol{\chi})$ スコアの推移

5. 結論

本研究では，対象シーンにプロジェクタを用いて摂動光を投影することで，単眼深度推定器の誤認識を引き起こす敵対的攻撃方式を提案した．提案手法は摂動の設計を行う最適化において，実環境を用いて解候補の評価を行う実環境評価型進化計算を利用することで，実環境でも頑健な AE を生成できる点に特徴がある．実験により，対象シーンとしての本来のロッカーの位置を壁近くまで誤推定させる結果を確認した．今後，最適化手法の見直しや別タスクへの有効性について検討する．

謝辞

本研究の一部は JSPS 科研費 JP22K12196 の助成による．

参考文献

[1] Keisuke Tateno, Federico Tombari, Iro Laina, and Nassir Navab. Cnn-slam: Real-time dense monocular slam with learned depth prediction, 2017.

[2] Xin Yang, Jingyu Chen, Yuanjie Dang, Hongcheng Luo, Yuesheng Tang, Chunyuan Liao, Peng Chen, and Kwang-Ting Cheng. Fast depth prediction and obstacle avoidance on a monocular drone using probabilistic convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 22, No. 1, pp. 156–167, 2021.

[3] Huaizu Jiang, Erik Learned-Miller, Gustav Larsson, Michael Maire, and Greg Shakhnarovich. Self-supervised relative depth learning for urban scene understanding, 2018.

[4] Gabriel Coll-Ribes, Iván J Torres-Rodríguez, Antoni Grau, Edmundo Guerra, and Alberto Sanfeliu. Accurate detection and depth estimation of table grapes and peduncles for robot harvesting, combining monocular depth estimation and cnn methods. *Computers and Electronics in Agriculture*, Vol. 215, p. 108362, 2023.

[5] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[6] Kaidi Xu, Gaoyuan Zhang, Sijia Liu, Quanfu Fan, Mengshu Sun, Hongge Chen, Pin-Yu Chen, Yanzhi Wang, and Xue Lin. Evading real-time person detectors by adversarial t-shirt. *CoRR*, Vol. abs/1910.11099, , 2019.

[7] Zhiyuan Cheng, James Liang, Hongjun Choi, Guanhong Tao, Zhiwen Cao, Dongfang Liu, and Xiangyu Zhang. Physical attack on monocular depth estimation with optimal adversarial patches, 2022.

[8] Takami Sato, Sri Hrushikesh Varma Bhupathiraju, Michael Clifford, Takeshi Sugawara, Qi Alfred Chen, and Sara Rampazzi. Invisible reflections: Leveraging infrared laser reflections to target traffic sign perception. *arXiv preprint arXiv:2401.03582*, 2024.

[9] Renya DAIMO and Satoshi ONO. Projection-based physical adversarial attack for monocular depth estimation. *IEICE Transactions on Information and Systems*, Vol. E106.D, No. 1, pp. 31–35, 2023.

[10] Somnath Lahiri, Jing Ren, and Xianke Lin. Deep learning-based stereopsis and monocular depth estimation techniques: A review. *Vehicles*, Vol. 6, No. 1, pp. 305–351, 2024.

[11] Iro Laina, Christian Rupprecht, Vasileios Belagiannis, Federico Tombari, and Nassir Navab. Deeper depth prediction with fully convolutional residual networks. In *2016 Fourth international conference on 3D vision (3DV)*, pp. 239–248. IEEE, 2016.

[12] Junjie Hu and Takayuki Okatani. Analysis of deep networks for monocular depth estimation through adversarial attacks with proposal of a defense method. *arXiv preprint arXiv:1911.08790*, 2019.

[13] Amira Guesmi, Muhammad Abdullah Hanif, Bassem Ouni, and Muhammad Shafique. Physical adversarial attacks for camera-based smart systems: Current trends, categorization, applications, research challenges, and future outlook. *IEEE Access*, Vol. 11, pp. 109617–109668, 2023.

[14] Nicholas Carlini and David Wagner. Towards evaluating the robustness of neural networks, 2017.

[15] Donghua Wang, Wen Yao, Tingsong Jiang, Guijiang Tang, and Xiaoqian Chen. A survey on physical adversarial attack in computer vision. *arXiv preprint arXiv:2209.14262*, 2022.

[16] Takahiro Suzuki, Shingo Takeshita, and Satoshi Ono. Adversarial example generation using evolutionary multi-objective optimization. In *2019 IEEE Congress on Evolutionary Computation (CEC)*, pp. 2136–2144. IEEE, 2019.

[17] Tomoki MINAMATA, Hiroki HAMASAKI, Hiroshi KAWASAKI, Hajime NAGAHARA, and Satoshi ONO. A coded aperture as a key for information hiding designed by physics-in-the-loop optimization. *IEICE TRANSACTIONS on Information and Systems*, Vol. 107, No. 1, pp. 29–38, 2024.

[18] Carlos M Fonseca, Peter J Fleming, Eckart Zitzler, K Deb, and L Thiele. Evolutionary multi-criterion optimization. In *Second International Conference, EMO 2003*. Springer, 2003.

[19] Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on evolutionary computation*, Vol. 11, No. 6, pp. 712–731, 2007.