

高照度差シーンにおける深層ニューラルネットワークを用いた 車載カメラの露光時間制御

— 高速道路トンネル出口における物体検出性能の向上 —

福島 光瑠*・高井 勇

株式会社豊田中央研究所 〒480-1192 愛知県長久手市横道 41-1

Exposure Time Control of Vehicle-Mounted Camera Using Deep Neural Network in High Contrast Scenes: Improvement of Object Detection Performance at Highway Tunnel Exits

Hikaru FUKUSHIMA* and Isamu TAKAI

Toyota Central R&D Labs., Inc., 41-1 Yokomichi, Nagakute-shi, Aichi, 480-1192

For more advanced autonomous driving systems in next generation, this paper introduces a scene-adaptive exposure time control method for visible light cameras using a deep neural network (DNN) controller. Images captured on actual roads frequently contain overexposed areas that make object detection impossible. The proposed DNN controller understands current scenes from captured images and provides the camera with an appropriate exposure time to enable object detection. A designed lightweight DNN controller prototype predicts the next exposure time in 2.4 ms from the current exposure time and image. Object detection performance was evaluated at actual tunnel exits, as one of the most difficult types of scenes. Due to the scene-adaptive exposure by the DNN controller significantly reducing overexposed areas without losing vehicles in a tunnel, the number of detected objects with the proposed method was increased by more than 80% compared with the conventional method.

Key words: autonomous driving (AD), vehicle-mounted (in-car or on-board) camera, exposure time control, deep neural network, object detection, high contrast scenes

はじめに

近年、自動運転 (AD: autonomous driving または automated driving) システムの市販化に向けた開発競争が世界中で激化している。現在の車載システムにおける車両周辺環境をセンシングするための代表的なセンサーは、カメラ (可視光, 近赤外など), ミリ波レーダー, LiDAR (light detection and ranging), ソナーなどである。その中でも可視光カメラは最も搭載実績があり, 高性能化と低コスト化が進んだことから, 一台の車に複数搭載され, 主要センサーとして重要な役割を担っている。また, このカメラの画像から深層ニューラルネットワーク (deep neural network; DNN) を用いて物体を高速かつ高精度に検出する物体検出器¹⁻⁴⁾の急速な進化が, 自動運転システムの実現可能性をさらに高めている。

しかしながら, その高い検出性能も, カメラから提供さ

れる画像の質によって大きく左右される。例えば, トンネルの出入口といった照度差 (明暗差) の大きいシーンでは, カメラの露光時間を適切に制御することは非常に困難であり, 多くの場合, 取得される画像には“白飛び”や“黒つぶれ”の領域が発生する。そして, いかに優れた物体検出器であっても, そのような領域に存在する物体は検出できない。自動車用のシステムにおいて物体の未検出はシステムの信頼性に大きく影響するため, 重要な問題である。このような問題に対し, 高ダイナミックレンジカメラ⁵⁻⁹⁾の使用は解決策のひとつであるが, 複数のカメラ搭載が前提と考えれば, ハードウェアコストの増加の可能性がある。また解像度などの他の基本性能とトレードオフの関係となる可能性もあるため, 実績のある通常のカメラで解決することが第一に期待される。また, 複数の露光時間画像を合成して高ダイナミックレンジ画像を生成する画像

*E-mail: fukushima@mosk.tytlabs.co.jp

処理技術¹⁰⁻¹³⁾も古くからあるが、車載カメラで撮影されるシーンは移り変わりが激しいため、生成画像にモーションアーチファクトが生じるという問題がある。また、カメラの露光時間をシーンに合わせて調整、制御することでこのような問題の解決を試みる技術¹⁴⁻¹⁶⁾もある。しかしながら実際の走行シーンは多種多様であり、それらすべてに柔軟かつ短時間に対応させることは容易ではない。さらに、不適切な露光条件により露光過多や露光不足が部分的に生じている画像をDNNで補正する技術¹⁷⁻¹⁸⁾も提案されているが、トンネル出口のように信号飽和に達する白飛びが広範囲に生じるようなシーンでは、補正に必要な情報が極度に不足するため、適正露光画像への復元は事実上困難である。また、一番の懸念は、そのような制御や補正技術が、すでに搭載されている物体検出器と機能的に独立しており、それらによって得られた画像が物体検出器にとって必ずしも適切なものとは限らないことである。

本論文では、照度差の大きな照明環境における課題の解決により次世代のさらに高度な自動運転システム実現への貢献を狙った、通常の可視光カメラとDNNを用いたシーン適応型の露光時間制御法を提案する。本提案法の技術的注力点は大きく2つある。1つは、搭載制約が厳しく、またシーンの移り変わりが激しい自動車用システムであることを考慮し、露光時間制御器として用いるDNNのモデルサイズを可能な限り小型化して、制御時間の短縮と実装負担の軽減を図ったことである。2つ目は、莫大な時間とコストをかけて開発された高性能な物体検出器の性能が、本提案法の採用によって棄損されず、発揮されるように制御することである。特に後者が本技術の要となる。そして、提案法を実装した車載システムを構築し、照度差の大きい代表シーンとして、トンネル出口における実道評価によって有効性を実証する。

1. 提案法の概要

提案法を含む可視光カメラ（以降、カメラ）を用いた自動運転システムの概要をFig. 1に示す。システムの標準的な動作は、カメラで撮影された画像 $Img(t)$ から物体検出器で物体を検出し、その検出結果（物体の位置、分類、信頼度スコアなど）を自動運転の中核システムに送る。この中核システムは、統合されたさまざまな情報をもとに、走行ルート決定や、最終的には車両制御などを行う。

通常、カメラの露光時間 T_{exp} はカメラに内蔵された専用チップなどで決定され、次回以降の撮影時に適用される。最も一般的なアルゴリズムは、撮影された画像内の任意の領域の平均輝度が設定された値に近づくように段階的に制

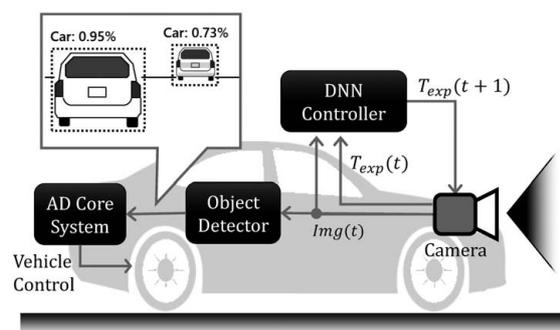


Fig. 1 Overview of autonomous driving (AD) system using proposed method.

御する方法などである。しかしながらこのような方法では、移り変わりの激しい走行シーンにおいて制御が常に遅れをとる形となり、それらが複数フレームにわたる白飛びなどを引き起こす。

提案法では、カメラ搭載の専用チップの代わりに、モデルサイズが小型のDNNを用いた露光時間制御器（以降、DNN制御器とする）を適用する。DNN制御器は、カメラから出力された画像 $Img(t)$ とその画像が撮影されたときの露光時間 $T_{exp}(t)$ が入力されると、現在のシーンにおいて次に設定すべき適切な露光時間 $T_{exp}(t+1)$ を出力するように学習されている。段階的ではなく一足飛びに露光時間が決定されることから、高速な走行シーンにおいても対応可能となる。ここで重要なことは、“適切な露光時間”の定義である。DNN制御器を学習する際は、これが教師（正解ラベル）となる。詳細は次章で述べるが、提案法では、搭載された物体検出器の“検出性能が最もよくなる露光時間”である。これは、DNN制御器の採用によりすでに搭載されている物体検出器の性能を悪化させることなく、引き出すようにする、という考えに基づく。このシステム目線の考えが、従来の独立した制御技術との大きな差異である。

2. 制御器

DNN制御器を学習するためのデータの収集方法をFig. 2に示す。ここではトンネル出口から±5 sを出口区間と定義し、この区間のデータを学習用データとして収集する。これは、高速道路での一般的な制限速度である100 km/hで走行中、トンネル出口直後に停止車両があると仮定すると制動のために100 m以上先から検出する必要があるため、最低限必要な時間である。データ収集効率の向上を目的としてフロントガラスに隣り合わせで2台のカメラ（非同期）を設置し、ほぼ同時刻のシーンを異なる露光時間で撮影する。なお、2台のカメラにそれぞれ設定される露光

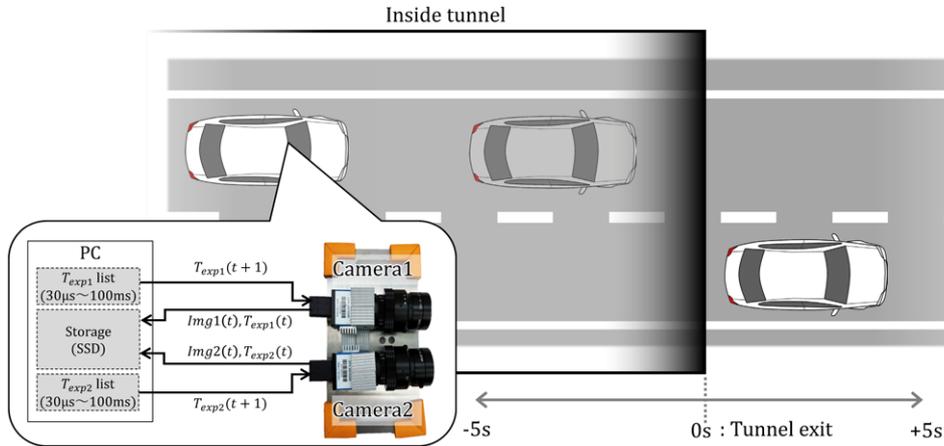


Fig. 2 Block diagram of collecting training data.

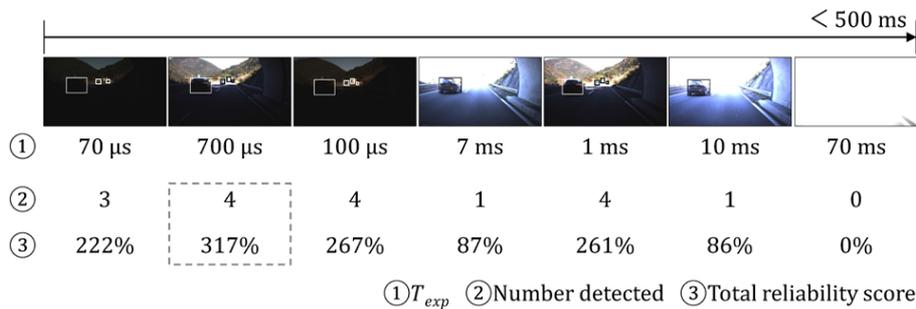


Fig. 3 Method for generating correct labels.

時間は、リスト (T_{exp1} list, T_{exp2} list) に従い、お互いに重複のない 30μ s から 100 ms の間の値である。この値が順次それぞれのカメラに設定され、さまざまな露光時間の画像とその露光時間がデータセットとして収集される。使用カメラは Imaging Source 社の DFK33UX249 (SONY 製 IMX249 搭載)、取得画像の解像度は 1280×720 (カラー) である。走行ルートは東海北陸自動車道の一部の区間 (晴れ、区間往復時のトンネル数 44) である。

収集画像から正解ラベルを生成する方法を Fig. 3 に示す。数多く存在する高性能な物体検出器の一例として、YOLOv5¹⁹⁾ を本論文では採用する。したがって、YOLOv5 の物体検出性能が最もよくなる露光時間が“適切な露光時間”と定義され、これが正解ラベルとなる。まず、画像から YOLOv5 (信頼度スコアの閾値: 0.25) で物体を検出し、検出数 (Fig. 3 の ②) と検出された物体の信頼度スコアの合計 (Fig. 3 の ③) を算出する。なお、検出対象は車両 (トラック、バス、普通乗用車) であり、誤検出は除外する。500 ms 以内に撮影された 6 枚以上の画像でグループを作成し、物体検出数が最大の画像を探索する。最大検出数が同じ画像が複数あれば、次に信頼度スコアの合計が最大の画像を探索する。そして、その画像の露光時間 (Fig.

3 の ①) がそのシーン (グループ) における正解ラベルとなり、それをグループ内の全画像に一律に付与する。この図では 700μ s である。他のグループでも同様の処理を繰り返して、グループ単位で正解ラベルを付与する。なお、白飛びまたは黒つぶれ領域が 80% 以上を占める画像は学習用データから除外する。この正解ラベルを生成する過程では、物体検出数を評価指標とし、誤検出は考慮に入れない。誤検出は物体検出器の進化やさらに後段の処理過程で排除できる可能性があるが、露光時間のミスマッチによって生じる未検出はいかに高度な信号処理技術であっても対応は困難である。したがって、システムの初段で機能する本提案手法では、物体の未検出が最小となることを優先している。また、グループ内の画像は撮影時刻が同一ではないため完全な同一シーンではないが、ここでは撮影時刻範囲を 500 ms 以内に限定することで同一シーンであると見なしている。

Fig. 4 に DNN 制御器の構造を示す。入力は 160×90 (8bit, グレー) の $Img_{PM}(t)$ と $T_{exp_PM}(t)$ 、出力は $T_{exp_PM}(t+1)$ である。なお、カメラから出力される画像を 8 分の 1 の解像度にし、グレースケール化することで DNN のモデルサイズを小型化している。中間層の前段は画像の特徴量を抽

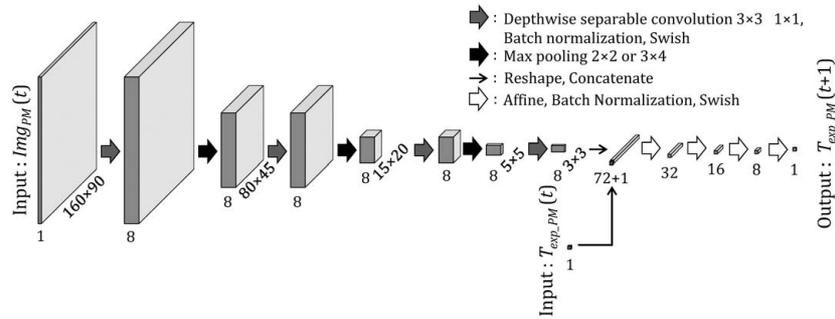


Fig. 4 Structure of deep neural network (DNN) controller.

出するための畳み込み層²⁰⁾、後段は連続値である露光時間を予測するための全結合層で構成される。画像から得られる情報のみで絶対値である $T_{exp_PM}(t+1)$ を予測することは困難であるため、全結合層に画像の特徴量と $T_{exp_PM}(t)$ の両方を入力することで、次に設定すべき露光時間の予測を可能にしている。積和演算器の数は約 780 k、パラメータ数は 3875 個である。ロス関数を式 (1) に示す。なお、 T_{exp_PM} は正解の露光時間、 T_{exp_PM}' は予測露光時間である。

$$\text{Loss} = \left| 10 \log_{10} \left(\frac{T_{exp_PM}'}{T_{exp_PM}} \right) \right| \quad (1)$$

学習に用いたトレーニングデータは 59564 枚、バリデーションデータは 7111 枚である。また、エポック数は 2000、最適化アルゴリズムは AdaGrad (学習率: 0.01)、バッチサイズは 512 である。DNN 制御器を実装した PC (CPU: Intel core i7-8700, メモリー: 16 GB, GPU: なし) 上で 100 回推論した際の推論時間の平均は 2.4 ms であった。また、露光時間を設定する時間なども含めた制御時間の平均は 135 ms であった。これは時速 100 km において 3.75 m に 1 回の制御性能であり、シーン変化に対する制御遅延の懸念は小さいと考えられる。

3. 実道実験

3.1 実験条件

提案法の性能評価を行う。車載実験システムの構成を Fig. 5 に示す。フロントガラスに取り付けられた 2 台のカメラ (非同期) で同じシーンの画像を取得する。一方のカメラの露光時間は提案法によって制御され、もう一方のカメラは従来の最も一般的な自動露光時間制御アルゴリズム (従来法) によってカメラ内で制御される。これにより同一シーンにおける比較評価を行う。従来法は画像内の任意領域の平均輝度を 128 (8 bit, 256 階調の半分) に近づくよう制御する。さらに比較のため、平均輝度の計算領域が画像全体である従来法 A と、中心部 (縦横 50% 幅) である従

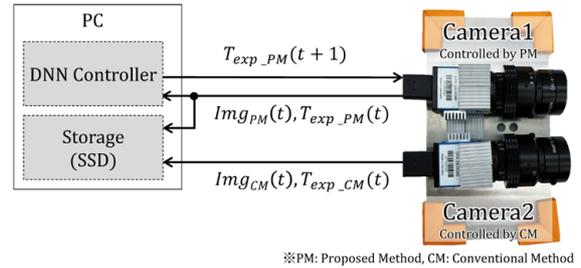


Fig. 5 Block diagram of in-vehicle test system.

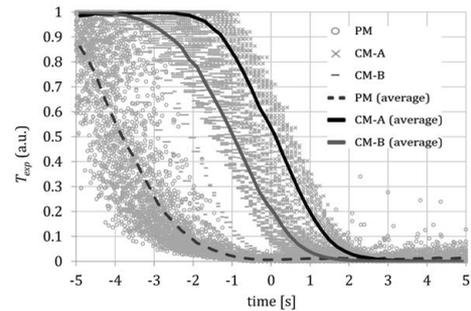


Fig. 6 Exposure time control results for each control method. PM: proposed method, CM-A: conventional method A, CM-B: conventional method B.

来法 B の 2 種類を用いる。走行ルートは学習データ収集時と同様に東海北陸自動車道の一部の区間 (晴れ、区間往復時のトンネル数 44) である。照度はトンネル外で数千から数万ルクス、トンネル内で数十ルクスのオーダーであり、2 桁から 3 桁の照度差がある。提案法と従来法 A、提案法と従来法 B の組み合わせでそれぞれ一往復し、評価用データを収集する。そして、 ± 5 s のトンネル出口区間における物体検出数で比較評価する。

3.2 露光時間制御結果

各制御法におけるトンネル出口区間の露光時間の制御結果を調べる。制御法ごとの全体的な制御傾向を把握するため、各制御法における各トンネル出口区間の最大露光時間を 1、最小露光時間を 0 として正規化した結果を Fig. 6 に

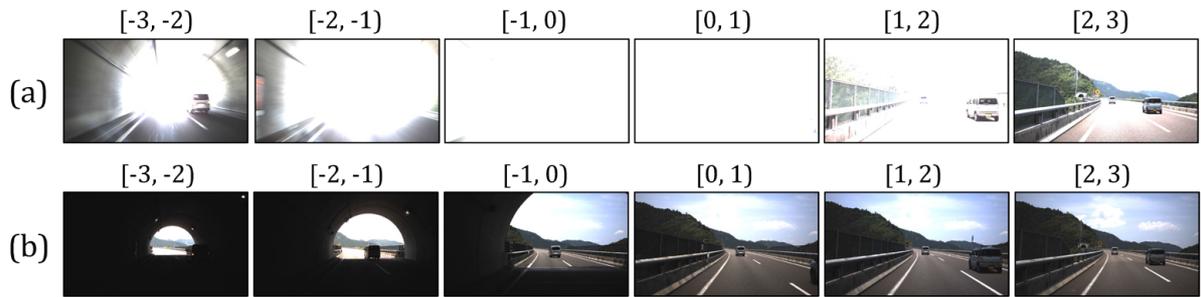


Fig. 7 Examples of images obtained by each control method. (a) Conventional method A, (b) proposed method.

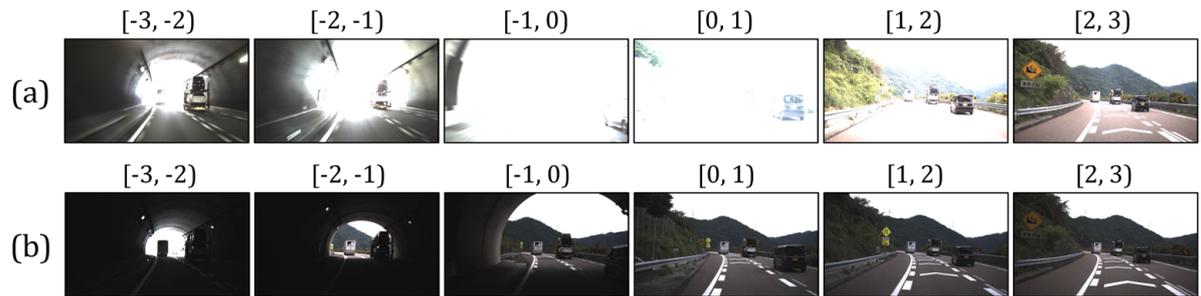


Fig. 8 Examples of images obtained by each control method. (a) Conventional method B, (b) proposed method.



Fig. 9 Examples of object detection by each control method. (a) Conventional method A, (b) proposed method.

示す。また、正規化後のデータに対し、200 ms ごとに露光時間の時間平均をとった結果も合わせて示す。なお、トンネル出口直後に再度トンネルが現れるシーンは、傾向が異なるため除かれている。従来法に比べて提案法はトンネル外の照明環境に対応した露光時間の制御が早期に開始されており、出口 (0 s) 到達前にすでに完了している。それと比較して従来法は、制御の開始が数秒遅い。従来法 B は従来法 A と比較して画像中央部に注目しているためトンネル外環境への対応が早いですが、それでも出口時点で制御が完了していない。

Fig. 7 および Fig. 8 に、従来法と提案法の制御によって得られたトンネル出口区間 (± 3 s から抽出) の画像の例を示す。従来法 A、従来法 B とともに白飛びしており、前方車両の検出が困難である。提案法では白飛びが低減され、

トンネル内部を走行している時点でトンネル出口の先が鮮明に見えるように露光時間が調整されている。

3.3 物体検出結果

各制御法により収集した画像から YOLOv5 (信頼度スコアの閾値: 0.25) で物体検出を行い、検出数 (true positive) を算出した²¹⁾。なお、検出対象は車両 (バス、トラック、普通乗用車など) である。Fig. 9 および Fig. 10 に、従来法と提案法のトンネル出口区間における物体検出結果画像の例を示す。従来法では白とびが発生し、トンネル外の車両の検出は困難である。これに対し提案法では、トンネル外の車両が検出されるようになり、さらにトンネル内の車両も同時に検出されている。つまり、提案法はトンネル外への対応を単に優先しているわけではなく、そのシーンを理解し、トンネル内部も含めてそのシーンにおい



Fig. 10 Examples of object detection by each control method. (a) Conventional method B, (b) proposed method.

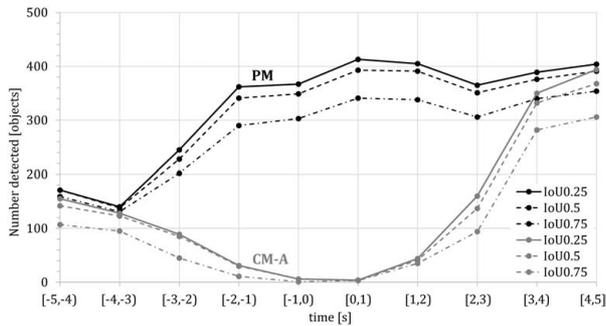


Fig. 11 Number of objects detected by proposed method (PM) and conventional method A (CM-A). IoU: Intersection over Union.

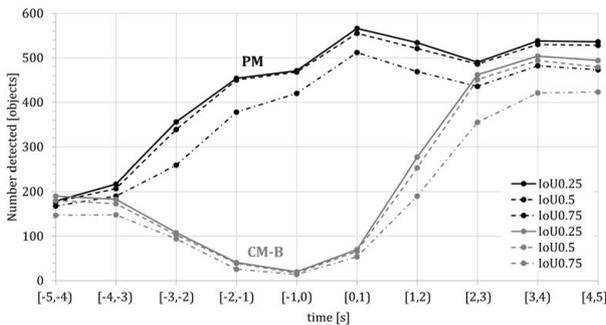


Fig. 12 Number of objects detected by proposed method (PM) and conventional method B (CM-B). IoU: Intersection over Union.

て未検出を最小化するように露光時間を制御していることが確認できる。Fig. 11 および Fig. 12 に、提案法と従来法のトンネル出口区間における 1 秒ごとの物体検出数を示す。パラメータは物体の正解位置（バウンディングボックス）と予測位置の重なり度合いを示す物体検出評価指標 IoU (intersection over union²²⁾) である。いずれの結果においても、従来法では出口付近で白飛びが発生し、出口の先の車両が検出できないため検出数が大きく減少している。一方、提案法では変化するシーンに適応した露光時間が設定され、-4 s 程度から出口の先が鮮明に撮影される

Table 1 Total number of objects detected by proposed method (PM) and conventional method A (CM-A). IoU: Intersection over Union.

IoU	CM-A	PM	PM/CM-A
0.25	1361	3261	240%
0.5	1267	3129	247%
0.75	979	2764	282%

Table 2 Total number of objects detected by proposed method (PM) and conventional method B (CM-B). IoU: Intersection over Union.

IoU	CM-B	PM	PM/CM-B
0.25	2349	4341	185%
0.5	2255	4261	189%
0.75	1872	3787	202%

ようになったことで検出数が大幅に増加している。IoU の違いによる傾向の差異は見受けられない。Table 1 および Table 2 に、提案法と従来法のトンネル出口区間における合計物体検出数を示す。提案法の検出性能は従来法と比較して 1.8 倍以上に向上している。

4. 結 論

トンネル出口といった照度差の非常に大きい照明環境における白飛びなどの課題を解決するために、モデルサイズが小型の DNN 制御器を用いたシーン適応型の露光時間制御法を提案した。さらに、提案法が高性能な物体検出器の性能を棄損しないように、露光時間を制御する仕組みを導入した。提案法を実装した車載システムを構築し、トンネル出口において物体検出性能を評価した。その結果、白飛びが低減されるよう露光時間が適応的に制御され、従来法と比べて車両の検出数が 1.8 倍以上と大幅に向上した。実道実験により提案法の有効性が実証された。この技術は、学習過程に歩行者や他の物体を加えることで、車両以外の対象であっても有効に機能すると考えられる。今後は対象物体拡大の検討、トンネル入口や逆光なども含めたさらに

多様な照明環境での有効性を確かめていく。また、DNN制御器のモデルサイズのさらなる小型化、動作速度の高速化にも取り組んでいく。

文 献

- 1) J. Redmon and A. Farhadi: arXiv (2018) (doi: 10.48550/arXiv.1804.02767).
- 2) A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao: arXiv (2020) (doi: 10.48550/arXiv.2004.10934).
- 3) N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov and S. Zagoruyko: "End-to-end object detection with transformers," *European Conference on Computer Vision (ECCV)*, 12346 (Glasgow, 2020) pp. 213–229.
- 4) W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. C. Berg: "SSD: Single shot multibox detector," *European Conference on Computer Vision (ECCV)*, 9905 (Amsterdam, 2016) pp. 21–37.
- 5) 太田 淳 (監修) : CMOS イメージセンサの最新動向—高性能化, 高機能化から応用展開まで— (シーエムシー出版, 2007).
- 6) Y. Ota, K. Morimoto, T. Sasago, M. Shinohara, Y. Kuroda, W. Endo, Y. Maehashi, S. Maekawa, H. Tsuchiya, A. Abdelghafar, S. Hikosaka, M. Motoyama, K. Tojima, K. Uehira, J. Iwata, F. Inui, Y. Matsuno, K. Sakurai and T. Ichikawa: "A 0.37W 143dB-dynamic-range 1Mpixel backside-illuminated charge-focusing SPAD image sensor with pixel-wise exposure control and adaptive clocked recharging," *International Solid-State Circuits Conference (ISSCC)* (2022) pp. 94–96.
- 7) Y. Sakano, T. Toyoshima, R. Nakamura, T. Asatsuma, Y. Hattori, T. Yamanaka, R. Yoshikawa, N. Kawazu, T. Matsuura, T. Iinuma, T. Toya, T. Watanabe, A. Suzuki, Y. Motohashi, J. Azami, Y. Tateshita and T. Haruta: "5.7 A 132dB single-exposure-dynamic-range CMOS image sensor with high temperature tolerance," *International Solid-State Circuits Conference (ISSCC)* (San Francisco, 2020) pp. 106–108.
- 8) Y. Fujihara, M. Murata, S. Nakayama, R. Kuroda and S. Sugawa: *IEEE Trans. Electron Devices*, **68** (2021) 152–157.
- 9) 映像情報メディア学会 (編) : CMOS イメージセンサ (コロナ社, 2012).
- 10) デジタル画像処理編集委員会 (編) : デジタル画像処理改訂第二版 (公益財団法人 画像情報教育振興協会, 2020).
- 11) 米本和也 : CCD/CMOS イメージ・センサの基礎と活用 (CQ出版社, 2003).
- 12) トランジスタ技術編集部 (編) : CCD/CMOS イメージ・センサ活用ハンドブック (CQ出版社, 2010).
- 13) フランクリン, クレイグアール : 特開 2022-71177.
- 14) 服部 理, 下浦 弘 : 特開 2006-109165.
- 15) 佐光靖史 : 特開 2011-9863.
- 16) 廣岡慎一郎, 吉田大輔, 野中雄一 : 電子情報通信学会技術研究報告, **114**, No. 233 (2014) 45–50.
- 17) M. Afifi, K. G. Derpanis, B. Ommer and M. S. Brown: "Learning multi-scale photo exposure correction," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021) pp. 9157–9167.
- 18) G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk and J. Unger: *ACM Trans. Graphics*, **36** Article 178 (2017) 1–15.
- 19) J. Glenn: YOLOv5 release v6.1. (2022) (<https://github.com/ultralytics/yolov5/releases/tag/v6.1>)
- 20) F. Chollet: "Xception: Deep learning with depthwise separable convolutions," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, 2017) pp. 1800–1807.
- 21) A. Dutta and A. Zisserman: "The VIA annotation software for images, audio and video," *Proc. 27th ACM International Conference on Multimedia* (Nice, 2019) pp. 2276–2279.
- 22) H. Rezatofghi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid and S. Savarese: "Generalized intersection over union: A metric and a loss for bounding box regression," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (California, 2019) pp. 658–666.