

# 深層学習と領域認識を組み合わせた画像変換技術

## Image Conversion Technology Utilizing Deep Learning & Semantic Segmentation



西森 稔晃\*<sup>1</sup>  
Toshiaki Nishimori

平出 昂\*<sup>2</sup>  
Ko Hirade

高桑 義直\*<sup>3</sup>  
Yoshinao Takakuwa

清水 裕量\*<sup>4</sup>  
Hirokazu Shimizu

中尾 健太\*<sup>5</sup>  
Kenta Nakao

自動運転車両の検証として、実車による走行試験が行われているが、自然現象や交通流の組み合わせは無数に存在し、検証データ収集の時間・費用の削減、その効率化が求められている。そこで当社では、東京大学との共同研究により、物体認識において重要な役割を占めているカメラ画像を対象に、深層学習を用いて、ある環境で撮影された実画像を異なる環境の画像に変換するAI(Artificial Intelligence)アルゴリズム(関連特許出願済)を開発した。本アルゴリズムでは、細部の変換精度を向上させるため、従来の画像変換技術に領域認識技術を組み合わせた。この結果、自動運転車両の検証に必要な大量のデータが短時間で生成でき、検証データ収集の時間・費用の削減が可能となる。

## 1. はじめに

自動運転車両には、車両周辺の環境を認識するために、Radar (Radio Detection And Ranging)やLiDAR (Light Detection And Ranging)など様々なセンサが用いられているが、中でも画像センサ(カメラ)は物体認識において重要な役割を果たしている。

カメラ画像の認識技術の検証には、実車による走行試験が行われているが、自然現象や交通流の組み合わせは無数に存在し、検証データ収集の時間・費用の削減、その効率化が求められている。

そこで当社は、最新のAI技術を活用しモビリティ開発に貢献するために、東京大学大学院・情報学環 上條研究室との共同研究を行った。研究の結果、深層学習を用いて、ある環境で撮影された実画像を夜間や雨天などの異なる自然環境の画像に変換するAIアルゴリズムを開発した。これにより、自動運転車両の検証に必要な、大量のデータを短時間で生成でき、検証データ収集の時間・費用の削減が可能となる。

本報では、開発した画像変換技術、実験結果について報告する。

## 2. 画像変換

### 2.1 敵対的生成ネットワーク

敵対的生成ネットワーク(GAN:Generative Adversarial Networks<sup>(1)</sup>)とは、生成ネットワーク(Generator)と識別ネットワーク(Discriminator)を競わせながら入力データの特徴を学習し、実在しない疑似データを生成する人工知能アルゴリズムである。そのアーキテクチャの柔軟性から様々な分野で研究されており、本報の目的である入力画像を別の画像に変換する手法も多数提

\*1 三菱重工機械システム株式会社 設備インフラ事業本部 制御システム技術部 主席

\*2 三菱重工機械システム株式会社 設備インフラ事業本部 制御システム技術部

\*3 三菱重工機械システム株式会社 設備インフラ事業本部 プロジェクト推進部 部長 技術士(電気電子部門)

\*4 三菱重工機械システム株式会社 設備インフラ事業本部 プロジェクト推進部 主席 技術士(電気電子部門)

\*5 デジタルイノベーション本部 CIS部 主席技師 技術士(情報工学部門)

案されている。

CycleGAN<sup>(2)</sup>は、異なる2種類の画像群(A と B)を与えるだけで両群の根底にある特徴的な違いを自動的に学習し、入力した A に属する画像を B に属する画像へと変換可能な手法である。A と B の画像群を厳密に対応付ける必要がない“教師なし学習”のため、学習データを容易に準備可能な点が特徴である。本手法を応用し、日中に撮影された車載カメラの画像を夜間に撮影された画像のように変換する研究などが行われている<sup>(3)</sup>。

## 2.2 画像変換の課題

先に示した CycleGAN による日中から夜間への画像変換例を図1に示す。(a)が入力の日中画像、(b)が変換された夜間画像であり、(b)は(a)と比較して全体的に輝度が低下し、夜間らしい画像となっている。しかし、細部を確認すると何もない空中に高輝度な光が発生している点や前方走行車両のテールランプが点灯していない等、偽の情報や夜間らしくない画像であることが確認できる。

このような事象が発生する最も大きな原因は、CycleGANが画像特徴量空間における画像群同士の違いを学習するネットワークであり、画像中の各領域において、それがどんな物体かを表すラベル情報やその特性を加味した学習を行っていないことにある。つまり、“この領域は空であり夜間の場合は光らない”、“この領域はテールランプであり夜間の場合は光る”といった、画像に映っている対象物ごとの特性について考慮されていないことを意味する。

そこで、本報では各領域のラベル情報を考慮した画像変換技術の開発に取り組んだ。



図1 CycleGAN による日中から夜間への画像変換

## 2.3 領域のラベル情報を保持する画像変換

### (1) セマンティックセグメンテーション(Semantic Segmentation)の活用

画像中の各領域におけるラベル情報を活用するため、セマンティックセグメンテーション用に学習されたネットワークを CycleGAN の生成ネットワークに転移学習させる手法を開発した。

セマンティックセグメンテーションとは、画像内の全ての画素に対してラベルやカテゴリを関連付ける深層学習アルゴリズムであり、自動運転/運転支援システムでも交通シーンの理解に応用されている。図2に示す通り、入力画像と正解ラベルをペアとして学習することにより、任意の入力画像に対して自動的に画素単位のラベル付けが可能となる<sup>(4)</sup>。

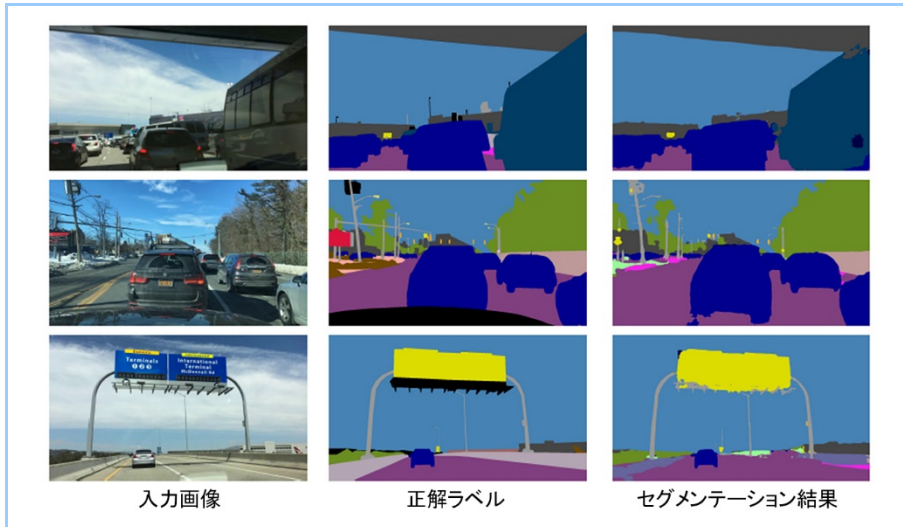


図2 セマンティックセグメンテーションの例

(2) ネットワークの学習方法

開発した画像変換技術のアーキテクチャを図3に示す。Step1 では、日中画像  $X_S$  とその正解ラベル  $Y_S$  を用いてセマンティックセグメンテーションのネットワーク  $S$  を学習する。学習は  $S$  によって生成されたセグメンテーション結果  $S(X_S)$  と正解ラベル  $Y_S$  の差を表す損失関数  $loss1$  を最小化することで実現される。これにより、図2で示したように画素単位のラベル付けが可能なネットワークを得ることができる。

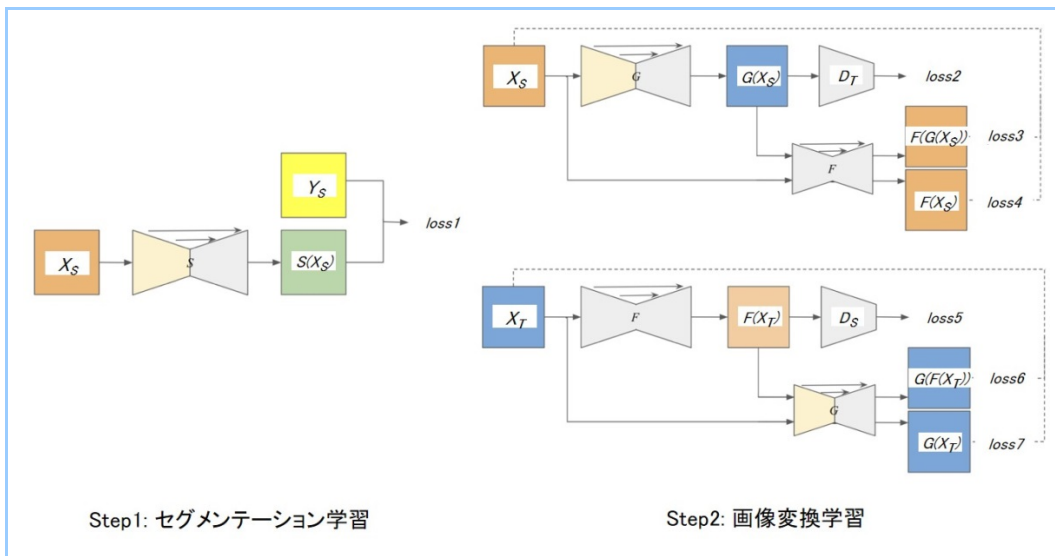


図3 開発した画像変換技術の学習アーキテクチャ

次に、Step2 では Step1 で学習したネットワーク  $S$  をベースに、CycleGAN のネットワーク  $G$  と  $F$  を学習する。 $G$  は日中画像を夜間画像に変換する生成器、 $F$  は夜間画像を日中画像に変換する生成器を表す。 $X_T$  は夜間画像、 $G(X_S)$  は  $X_S$  から変換された疑似的な夜間画像である。 $D_T$  は  $G(X_S)$  に対する識別器であり、 $G(X_S)$  が変換された夜間画像か、実際の夜間画像かを判定する。 $loss2$  は識別器の確からしさを表す目的関数であり、 $G(X_S)$  を変換された夜間画像と判定できれば値が大きく、実際の夜間画像と誤判定すれば値が小さくなる。したがって、 $G$  は  $loss2$  を最小化しようとし、識別器は  $loss2$  を最大化しようとして互いに学習を行う。 $F(G(X_S))$  は  $F$  を通して  $G(X_S)$  を更に日中画像に戻した画像であり、 $loss3$  が変換した画像を更に変換することで元の画像に戻らなければならないという制約を表す。 $F(X_S)$  は日中画像  $X_S$  を  $F$  に通して変換される画像であり、 $loss4$  が夜間画像から日中画像へ変換する  $F$  に日中画像を入力してもそのまま出力されなければならないという制約を表している。 $loss5 \sim loss7$  は入力画像を  $X_T$  にした場合の値

であり、考え方は  $loss2 \sim loss4$  と同じである。

セマンティックセグメンテーションのネットワークを転移学習しているため、一般的な CycleGAN では考慮されない、各領域のラベル情報を保持した変換が可能となり、細部を含めた変換精度の向上が期待できる。

### 3. 実験結果

#### 3.1 データセットと学習

本報では、日中及び夜間撮影画像として、自動運転の研究で広く用いられている Berkeley Deep Drive (BDD) データセット<sup>(4)</sup>と、独自に収集した東京都内の車載カメラ画像を使用した。

Step1 のセマンティックセグメンテーション用ネットワークは、画素単位の正解ラベルが付与されている BDD データセットを用いて学習し、Step2 の CycleGAN 用ネットワークは、BDD データセットを用いた場合、東京都内データセットを用いた場合の2ケースで学習を行った。

#### 3.2 変換結果と評価

BDD 及び東京都内のデータセットから学習に使用しなかった日中画像を夜間画像に変換した結果をそれぞれ図4に示す。従来技術とは、セマンティックセグメンテーションを組み合わせしていない、一般的な CycleGAN による変換を表している。赤丸で囲った部分が、従来技術と比較して優れた変換結果となっている箇所を示す。従来技術では、街灯ではなく樹木が光る誤変換を生じさせているが、開発技術では正しく街灯のみが光っている。また、前方を走行する車両のテールランプもより強く光る結果が得られている。このように街灯やテールランプが夜間特有の画像に変換できていることは、開発した技術がセグメンテーションによって得られる画像中のラベル情報を CycleGAN に正しく転移学習できていることを示している。また、従来技術では道路面にも元の日中画像では見られない不自然な明るい領域が発生しているが、開発技術ではそのような偽信号も発生しておらず、“道路”というラベル情報が誤変換を抑制していると考えられる。

以上の結果から、画素単位のラベル情報を推定するセマンティックセグメンテーションを組み込む開発技術の有効性を確認できた。



図4 BDD データ(上)及び東京都内データ(下)に対する変換結果

また、GAN による画像生成の評価に使用される Fréchet Inception Distance (FID)<sup>(5)</sup>を用いて、変換結果を定量的に評価した結果を表1に示す。FID はある画像の集合Aと、別の画像の集合Bにおいて、特徴量空間における距離を算出したものであり、その距離が短い(値が小さい)ほど両集合の画像が似ていることを表す。今回のデータセットにおいて、変換前の日中画像と学習に使



用した夜間画像の FID は 3000 であるが、変換された夜間画像と学習に使用した夜間画像では 200 まで減少しており、日中画像から夜間画像に近付ける変換が行えていることを確認した。同様に、変換前の日中画像と学習に使用した日中画像の FID は 80 であるが、変換された夜間画像と学習に使用した日中画像は 2400 であり、日中画像から夜間画像に近付ける変換が行えていることを確認した。

表1 FID による変換画像の類似度比較

	学習画像 (夜間)	学習画像 (日中)
変換前画像 (日中)	3000	80
変換後画像 (夜間)	200	2400

(値が小さいほど画像が似ていることを表す)

## 4. まとめ

画像中の各領域におけるラベル情報を活用することで、細部の変換精度を向上させる画像変換手法を開発した。本手法を用いて日中画像を夜間画像に変換した結果を、FID により定量的に評価し、夜間画像に近付ける変換が行えていることを示した。

本手法に大雨、霧、雪などの学習画像を用いることで、自動運転車両の検証に利用可能な異なる天候の画像を短時間で生成することが可能となる。今後の取組みとして、時間帯や天候などを別々に学習し、複数の変換の組み合わせることで、複合的な外乱画像の生成を行う予定である。また、Radar や LiDAR についても、深層学習を用いて悪天候のセンサデータを生成する手法を開発中であり、検証データ収集の時間・費用の更なる削減をし、モビリティ開発に貢献していきたい。

最後に、本研究にご協力いただいた東京大学大学院・情報学環 上條俊介准教授、李珍鎬氏、塩塚大気氏に心より感謝申し上げます。

## 参考文献

- (1) I. Goodfellow, et al., Generative Adversarial Nets, In NIPS, 2014
- (2) J. Zhu, et al., Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, In ICCV, 2017
- (3) 益田慎太ほか, 多様な仮想空間を構築するための画像モダリティ変換, 第 32 回人工知能学会全国大会, 2018
- (4) F. Yu, et al., BDD100K: A diverse driving video database with scalable annotation tooling, arXiv preprint arXiv, 1805.04687, 2018
- (5) M. Heusel, et al., GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, NIPS, 2017