

博士論文

夜間走行シーンにおける歩行者認識の為の  
夜間画像の視認性向上

公立はこだて未来大学大学院 システム情報科学研究科  
システム情報科学専攻

小倉 亮太  
2020年 8月

**Doctoral Thesis**

**Improvement of Nighttime Image Visibility for  
Pedestrian Recognition at Nighttime Driving Scene**

by

Ryota Ogura

Graduate School of Systems Information Science  
Future University Hakodate  
August, 2020

## 概要

本論文では、夜間歩行者認識を行う為に単眼 RGB 車載カメラから撮影した夜間走行画像に対し DeepLearning を用いて画像変換を行い、夜間画像の視認性向上を目的とする。

近年、自動車メーカー(自動車部品メーカー)、巨大 IT 企業を中心に先進運転支援システム(Advanced Driving Assistant System :ADAS)や自動運転に関する技術開発が加速している。これらのシステム開発には検知用のセンサが必要であり、主にカメラやミリ波レーダー、近年では LiDAR(Light Detection and Ranging)と呼ばれるセンサが登場、活用されている。これらのセンサで昼夜を問わず安定した物体検知が可能なものとしてミリ波レーダーと LiDAR があり、夜間での測距や物体の有無の判断を行っている。しかしこれらのセンサは物体の種類の特が困難であることや、高価であるという課題がある。LiDAR にいたっては数十万～数百万円するものがあり、安価な LiDAR では照射する赤外線レーザーの本数が少なく近傍しか検知できないことがある。一方、カメラを用いた物体認識の研究は構造がシンプルであり低価格であることから古くから行われ、近年では DeepLearning を利用することで認識性能を大きく向上させている。しかし、DeepLearning を利用するには学習データが大量に必要であり、且つ収集した画像データに対し認識対象のアノテーション作業(正解情報の付与)に膨大なコストが必要である。その為、夜間の歩行者認識を実現する為に、夜間歩行者画像の収集及び、アノテーション作業を実施することは困難とされている。その為、既に収集した昼間の歩行者画像(アノテーションデータ含む)を活用して夜間歩行者認識を行うことが望まれる。

本研究では、既に収集した昼間の歩行者画像を使用した物体認識モデルを活用する為に、夜間の画像を昼間の画像のように変換するアプローチを検討した。まず夜間直線道路における画像変換を行う。画像の明るさを改善する手法として複数の画像を足し合わせる処理がある。その処理をニューラルネットワークの中間層に応用した。その結果、夜間の入力画像に対して歩行者のエッジを保持したまま明るさを改善することができた。変換後の画像に対し、DeepLearning を利用した既存の物体認識手法を適用した結果、歩行者を認識することができ、原画像の夜間画像による物体認識と比較して提案手法が優位であることを確認した。

又、上記で述べた複数画像の足し合わせ処理についてフレーム間の移動量が多いシーンとしてカーブ道路がある。移動量が多い場合足し合わせをした時に対象物のずれにより画質の改善が困難である。そこでニューラルネットワークに複数画像を入力する際、過去フレームにおける変換画像を合わせて入力することで過去フレームからカーブの形状を徐々に推測するニューラルネットワークを検討した。

これらの結果から、提案手法による夜間直線道路とカーブ道路における画像変換手法を検討し、効果を確認した。

## Abstract

In this paper, the purpose of this study is to improve the visibility of nighttime images by performing nighttime running images captured by a monocular RGB on-board camera using deep learning to recognize nighttime pedestrians.

In recent years, the development of advanced driving assistance system (ADAS) and autonomous driving technology has been accelerated mainly by automobile manufacturers (automobile parts manufacturers) and large IT companies. Sensors for detection are required for the development of these systems, and mainly cameras and millimeter-wave radars, and in recent years, sensors called LiDARs (Light Detection and Ranging: LiDAR) have appeared and are being used. Millimeter-wave radars and LiDARs can detect objects stably both day and night with these sensors. They measure distance at night and judge the presence or absence of objects. However, these sensors have a problem that it is difficult to identify the type of the object and it is expensive. Some LiDARs cost hundreds of thousands to millions of yen, and inexpensive LiDARs can detect only a small number of infrared lasers and only nearby ones. On the other hand, research on object recognition using cameras has been conducted for a long time because of its simple structure and low cost. In recent years, the recognition performance has been greatly improved by using deep learning. However, to use deep learning, a large amount of training data is required, and enormous cost is required for the annotation work (addition of correct answer information) to the recognition target for the collected image data. Therefore, it is difficult to collect nighttime pedestrian images and perform annotation work to realize nighttime pedestrian recognition. Therefore, it is desirable to perform nighttime pedestrian recognition by using already collected daytime pedestrian images (including annotation data).

In this study, we considered an approach to transform nighttime images like daytime images in order to utilize an object recognition model using already collected daytime pedestrian images. First, image conversion on a straight road at night is performed. As a method to improve the brightness of an image, there is a process of adding multiple images. The processing was applied to the hidden layer of the neural network. As a result, it was possible to improve the brightness of the input image at night while retaining the pedestrian edge. As a result of applying the existing object recognition method using deep learning to the converted image, the pedestrian can be recognized, and the proposed method is superior to the object recognition based on the night image of the original image. It was confirmed.

In addition, a curved road is a scene in which the amount of movement between frames is large in the process of adding a plurality of images described above. When the amount of movement is large, it is difficult to improve the image quality due to the displacement of the object when adding. Therefore, when inputting multiple images to a neural network, we studied a neural network that gradually estimates the shape of a curve from a past frame by inputting the transformed images in the past frame together.

From these results, we examined the image conversion method on the night straight road and the curved road by the proposed method, and confirmed the effect.

# 目次

第1章	研究の背景.....	9
1.1	自動運転車の開発背景と市場動向.....	9
1.2	自動運転車の定義.....	12
1.3	ADAS/自動運転システムの市場動向.....	14
1.4	自動運転システムの変遷.....	16
1.5	自動運転車に必要なセンサの種類.....	21
第2章	関連研究.....	27
2.1	昼間環境での物体認識(画像処理).....	27
2.2	昼間環境での物体認識(DeepLearning).....	29
2.3	赤外線を利用した夜間物体認識.....	35
2.4	DeepLearning を利用した画像変換手法.....	37
第3章	画像変換を使用した夜間歩行者認識の概要.....	42
3.1	一般的な歩行者認識手法の処理の流れ.....	42
3.2	提案手法による歩行者認識手法の流れ.....	43
3.3	本研究の目的.....	44
第4章	画像データ, 物体認識手法の準備.....	46
4.1	シミュレーション画像の収集.....	46
4.2	カメラ画像の収集.....	49
4.3	歩行者認識手法の準備.....	52
4.3.1	学習データ(PascalVOC).....	52
4.3.2	評価指標.....	53
4.4	認識限界距離.....	54
第5章	夜間直線道路における画像変換手法の検討.....	60
5.1	提案手法の概要.....	60
5.2	提案手法の詳細.....	61
5.3	実験.....	64
5.3.1	学習データテストデータで同等の車速の場合.....	64
5.3.2	自車速の変化に対する検討.....	74
5.3.3	過去フレームの枚数の違いによる検討.....	77
第6章	夜間カーブ道路における画像変換手法の検討.....	79
6.1	提案手法の概要.....	79
6.2	提案手法の詳細.....	80
6.3	実験.....	82

6.3.1	曲率の違いによる画像変換結果(シミュレーション画像).....	82
6.3.2	実車走行画像(カメラ画像).....	86
6.3.3	十字路走行時の対応(限界性能).....	89
第7章	結論.....	92
謝辞	.....	94
参考文献	.....	95
図目次	.....	98
表目次	.....	101
研究業績	.....	103

## 各章の概要

### 第1章 研究背景

近年、政府、カーメーカーが ADAS/自動運転車の研究開発を加速している。その研究開発の背景を、近年の交通死亡事故、自動運転システムの定義づけ、ADAS/自動運転の開発の歴史、自車周辺の物体認識に必要なセンサ(カメラ、ミリ波レーダー、LiDAR、超音波センサ)について述べる。

### 第2章 関連研究

関連研究について、物体認識に関する研究と画像変換に関する研究で分けて述べる。まず物体認識に関する研究について従来の画像処理による特徴量計算、識別器に関する研究を述べ、次に近年、高精度な成果が報告されている DeepLearning を利用した物体認識について述べる。その後、夜間歩行者認識に有効とされる赤外線を利用した研究について述べる。その後、2014年に報告された画像変換技術(GAN)について述べ、それぞれの問題を述べる。

### 第3章 画像変換を使用した夜間歩行者認識の概要

本研究の夜間歩行者認識のアプローチの概要を述べる。従来の DeepLearning を利用した認識手法では、夜間歩行者の学習データ収集及びアノテーション作業に膨大なコストが必要であるが、本研究では、昼夜のシミュレーション画像を用いた画像変換手法を提案し、歩行者認識には既存の物体認識モデルを使用することでコスト低減を図る。

### 第4章 画像データ、物体認識手法の準備

本節では、次節第5章、6章の実験で使用するシミュレーション画像や実車カメラ画像について本研究における収集手段及び物体認識で使用する学習データの概要、物体認識可能な距離、認識率の指標である Precision, Recall について述べる。

### 第5章 夜間直線道路における画像変換手法の検討

5章では交通事故の多いシーンである直線道路において、既存の物体認識手法で歩行者認識できるよう夜間画像の変換を行う。暗い画像の画質改善手法として複数フレームの画像を使用し、ニューラルネットワークの中間層で値を畳み込むことで画質改善する手法を提案する。

### 第6章 夜間カーブ道路における画像変換手法の検討

5章では直線道路における画質改善を行った。しかし、カーブにおいては物体の移動量が大きく、改善には至らなかった。そこで、過去フレームに対しても画像変換を行い、

その変換結果を現在フレームと一緒に入力することでニューラルネットワークにカーブへの変換を学習する手法を提案する.

## 第7章 結論

本研究のまとめを記述する. 第1章から第6章までの内容から得られたことを記述し, 本研究の今後の展望を述べる.



# 第1章 研究の背景

## 1.1 自動運転車の開発背景と市場動向

近年、ドライバーの運転支援操作 Advanced Driver Assistance Systems(ADAS)や自動運転車 Auto Drive(AD)の開発がトヨタ、メルセデスベンツ等の国内外のカーメーカーや Google 等の巨大 IT 企業を中心に行われている。各社技術開発が進み、価格が下がることで、昔は高級車しか搭載されていなかった上記システムが大衆車にも普及している。更に、センサ等のハードウェアの開発に加えて、ソフトウェアの開発も盛んに行われている。2014 年には Graphics Processing Unit(GPU)の発達により Deep Learning を活用した人工知能(AI)技術開発が活発に行われている。

ADAS や AD の開発目的は、人手不足による輸送力の低減の抑制や、新たな交通サービスの創出や安全性の向上があり、特に安全性の向上に各社努めている。その理由として新たな交通サービスの創出や輸送力として使用するには、大前提として事故を起こさずに目的地までたどり着くことが必須であるからである。ここで国内の事故統計に目を向ける[1]。歩行中、自動車乗車中、二輪車乗車中、自転車乗用中と 4 パターンに分けると、それぞれ年々減少しているものの、歩行者による交通事故による死者が最も多いことがわかる。

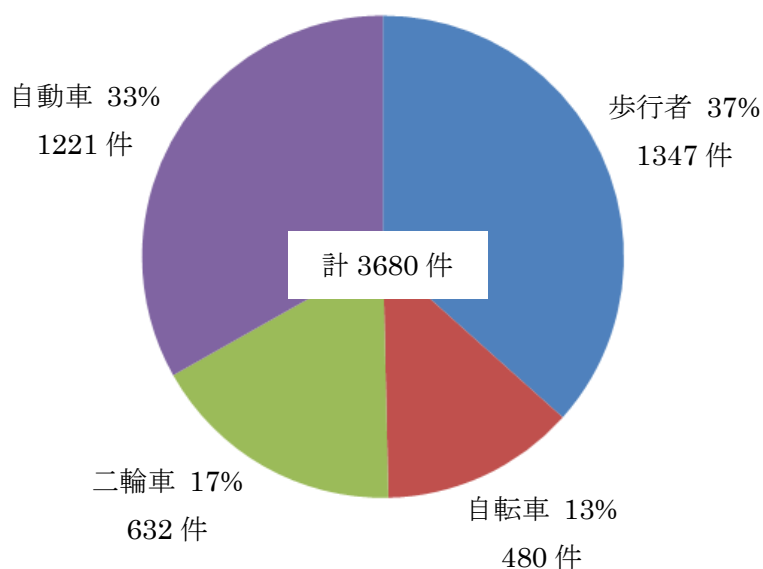


図 1.1 : 状態別交通死亡事故発生状況

次に時間帯別の負傷者事故件数及び、死亡事故件数を調査する。負傷者事故発生件数では、歩行者、自転車、原動機付自転車、二輪車の4つで夜間より昼間の方が、負傷者事故件数が多いことがわかる。一方、死亡事故件数を見ると、夜間の歩行者事故が昼間の事故件数の約2倍となっていることがわかる。このことから歩行者は夜間の交通事故発生時に死亡する可能性が高いことがわかる。夜間の交通死亡事故が多い理由として以下の理由が考えられる。

■スピードを出しやすい

夜間は昼間と比較して交通量が少ないことや、周囲が見にくい為、実際の車速よりスピードを遅く感じやすい為、スピードを出しやすい傾向がある。

■歩行者の発見が遅れる

夜間走行時は道路の視認性が悪く、歩行者が道路を横断していてもドライバーが気付かず急ブレーキによる減速無しで衝突してしまうことが考えられる。

■漫然運転や居眠り運転になりやすい

夜間は仕事帰りによる疲労から漫然運転になりやすい。また眠気による居眠り運転にもなりやすく歩行者に気付かず衝突してしまうことが多い。

■飲酒運転

近年、取り締まりが厳罰化しており件数は減っているものの、飲酒運転時は正常な注意力や判断力が失われ操作を誤ることが多い。

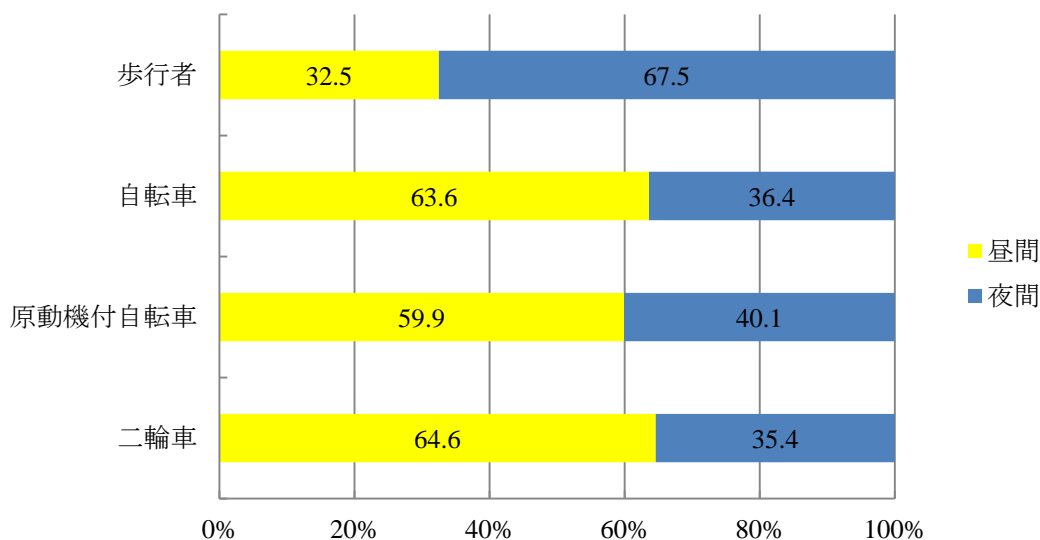


図 1.2 : 2018 年 昼夜別交通事故負傷者数

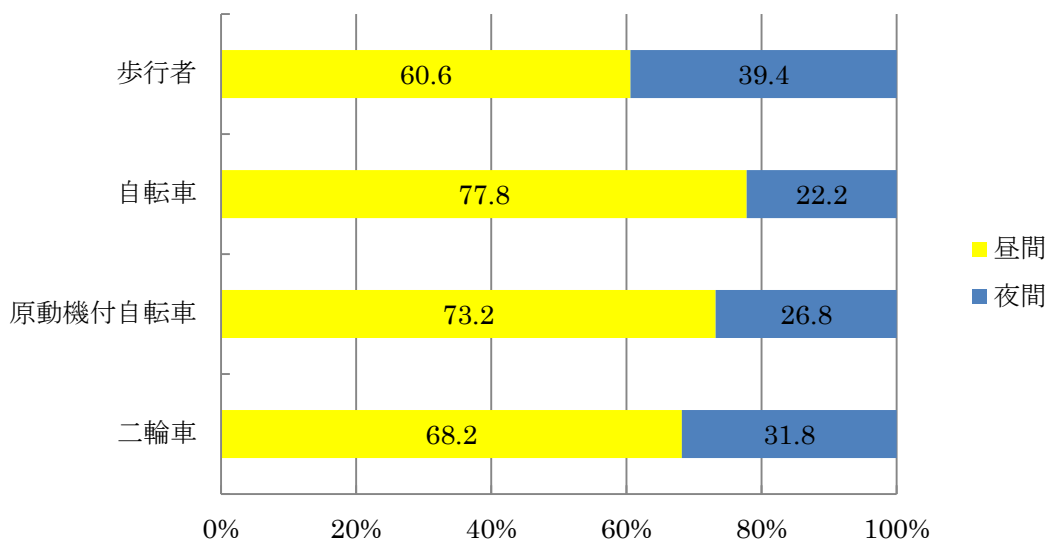


図 1.3 : 2018 年昼夜別交通事故死者数

又, 歩行者の状態別事故件数を調査[2]した結果, 図に示すように道路横断中が 57.9%, 自動車に対し背中を向けて歩行する背面通行中が 9.3%, 自動車に向かい合って歩行する対面通行中が 5.6%, 道路工事や交通整理をする為の移動である路上作業中が 1.8%, 道路工事を行う路上停止中での事故が 2.8%, それ以外の行動による事故が 22.6%あることがわかり, 道路横断中の事故を防ぐことが歩行者保護, 自動運転社会の発達にとって重要であることがわかる.

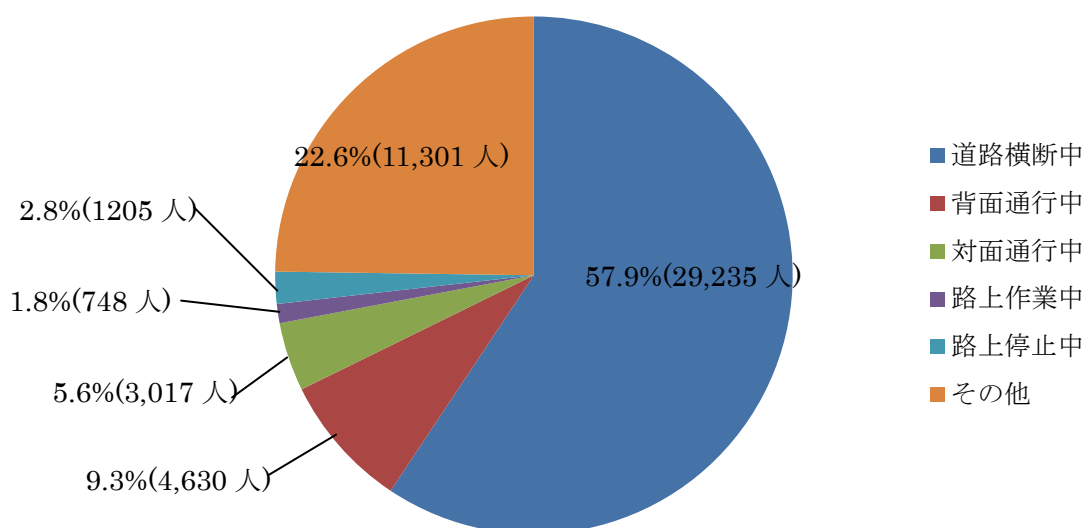


図 1.4 : 歩行者の状態別事故件数

## 1.2 自動運転車の定義

自動運転とはドライバーに代わり機械やコンピュータが自動車を制御することを言う。この自動運転には搭載されている機能や責任の所在毎にレベルがある。このレベルを定義したのは、米国運輸省道路交通安全局（NHTSA：National Highway Traffic Safety Administration）の一部で乗り物に関する標準化機構である SAE：Society of Automotive Engineers と呼ばれる(1905年に設立、設立時は別の名称)組織によって定義された。日本は2017年からSAEに基づき定義を行っており、表1.2に示すレベル0～レベル5までの6段階で定義されている。このレベルでは表1.1に示す4つの項目に基づいてレベルの住み分けを行っている。以下に各レベルにおける対応表を記載する[3]。

表 1.1：自動運転レベルの住み分け定義

番号	項目名	概要
①	運転操作	自動車のアクセルやブレーキ、ハンドルを操作する主担当者。
②	周囲認識	自動車の周辺環境を知覚して前進や後退の判断をする主担当者。
③	システム障害発生時の対応	事故や燃料切れ等の通常走行以外のイベントが発生した時に状況の判断及び自動車の操作をする主担当者。
④	運転環境	高速道路や一般道路等の自動運転を想定する走行領域。

表 1.2：自動運転レベル別定義

レベル	名前	運転操作	周囲認識	システム障害発生時対応	運転環境
0	手動運転	ドライバー	ドライバー	ドライバー	限定無し
1	運転支援	ドライバー (一部システム)	ドライバー	ドライバー	限定無し
2	部分自動運転	ドライバー (一部システム)	ドライバー	ドライバー	限定無し
3	条件付き自動運転	システム (一部ドライバー)	システム (一部ドライバー)	ドライバー	限定領域
4	高度自動運転	システム	システム	システム	限定領域
5	完全自動運転	システム	システム	システム	限定無し

以下にレベル0からレベル5までの概要を記す。

レベル0：運転操作においてシステムが介入することなく、ドライバーが周辺認識、アクセル、ブレーキの制御等のすべてのタスクを行う。

レベル1：レベル1は「運転支援技術」と呼ばれ、「自動運転」に必要な要素技術のことを言う。システムがハンドルの操作又は加速(減速)のどちらかを支援する。例えば、車線の逸脱を検知するとシステムが補正して車線の中央に戻したり、先行車との距離を一定に保つ為に自動でスピードを加減速するシステム(ACC：Adaptive cruise control)がある。これらのようにシステムがハンドル操作やスピード調整のどちらかを支援し、ドライバーがもう一方を制御するシステムをレベル1と言う。

レベル2：システムがハンドル操作又は加速(減速)のどちらも支援する。例えば、上記で挙げた車線逸脱防止とACCの両方を行うシステムである。現在、多くのカーメーカーに搭載されている。レベル2も「運転支援技術」と呼ばれる。

レベル3：限定された領域においてシステムが全てを操作する。緊急時はドライバーが対応する。高速道路等の特定の場所に限り、周辺の交通状況を認識して、ハンドル操作や加減速を行う。限定された領域を出た又は、システムが自動運転できないと判断した場合、ドライバーが運転を変える必要がある。その為、ドライバーはハンドル操作する必要は無いが、前方を注視する等の意識は必要である。海外では、レベル3の自動運転車の販売が行われており、一般人がレベル3の自動運転車を運転することができる。しかし、日本ではレベル3以上の自動運転車への法整備が進んでいない為、実験車両以外の自動運転車が走行することはできない。

レベル4：緊急時を含めて、特定の場所での運転をシステムが操作する。高速道路等の特定の領域において、システムが自車両の状態、周辺環境を認識して次のアクションを判断・実行する。まだ一般販売はされておらずコンセプトカーやテスト走行の段階である。ラストマイル(公共交通の最終地点と自宅等の目的地を結ぶ為の移動システム)等がレベル4の自動運転車に該当する。

レベル5：場所の限定なくシステムが操作する自動運転の理想的な状態。システムが全道路での交通状況を認識して、運転操作、緊急時での対応を行う。運転に関わることを自動車が全て行う為、車内にハンドルやアクセルといった機器が不必要になる。その為、車内は乗員の会議スペースやエンターテインメントをする場所となる。しかしレベル5の実現には自車周辺環境の詳細な把握やレベル4以下の自動車との混走時のやり取りなどが必要になる為、今日までのセンサーや認識技術より一段上の技術が必要である。

また事故時の責任がドライバーにあるのかシステムにあるのかの議論が必要になる。

自動運転は、運転操作はすべて機械が行うが、人が監視して何かあったら人が運転しなければならぬレベル2が始まりといわれる。レベル3では、機械が運転操作も監視も行い、人はセカンドタスクが可能になり、ここから本格的な自動運転になる。すべての運転操作を機械が担当するのはレベル4からで、レベル4と5は運行設計領域(環境制約)の有無の違いがあり、制約のないレベル5は完全なる自動運転を意味する。

### 1.3 ADAS/自動運転システムの市場動向

ADAS/自動運転システムの世界市場の調査結果によると、2030年までの新車におけるADAS/自動運転システムの世界搭載台数は図の通りとなる。レベル別に見ると、日米欧の新車に自動ブレーキや衝突警報などで標準化が進んでいるADASのレベル1が2,114万8,000台で世界市場全体の88.7%を占め、ステアリング操舵とブレーキ/アクセルを同時に自動化するレベル2の運転支援システムは270万4,000台となり、2018年から日欧の自動車メーカーを中心に高級車から中級車まで搭載車種が広がっている。現状レベル2はLKS(車線維持支援)とACC(車間距離制御)を組み合わせることで車線中央を自動走行する機能の搭載が中心であるが、高級車においてはドライバーの指示器操作によるオートレーンチェンジ(自動車線変更)や、ドライバーが降車後にキーやスマートフォンで遠隔操作して自動駐車することのできるリモートパーキングなどが実用化されている。

レベル1は2025年以降日米欧中からASEAN諸国、インドなどの新興国に需要の中心が移り、2025年の2,060万台から縮小して2030年の搭載台数は1,274万5,000台を予測する。レベル2では2020年に595万8,000台、2023年にはレベル1の搭載台数を上回り3,295万3,190台に増加すると予測する。2025年に4,357万4,000台、2030年は5,213万台に達し、最も増加する。2025年以降はV2Xの普及が日米欧中で進むことから、大部分の車両がレベル2以上の運転支援システムを搭載する。

レベル3の自動運転システムについては、2025年以降レベル3とレベル4のシステムコスト差が縮小することから、乗用車(自家用車)でも高級車を中心にレベル3からレベル4(高速道路限定)への切り替えが進み、2030年は373万台の横這いとどまるものと考える。

レベル4以上の自動運転システムについては、日米欧中において2020年からカーシェア/ライドシェア、公共交通、物流などにおいて自動運転車の試験的利用が行われ、2025年以降に普及が拡大する予測である。特に中国においてはICV(Intelligent Connected Vehicle)の技術開発と普及を政府が後押ししており、V2Xを利用した自動運転車のテスト走行がスマートシティ実証試験区で始まっている。このため、中国におけるレベル4の自動運転システムの需要は2025年以降に伸びると予測する。レベル4/5の世界搭載台数は、2025年には179万5,600台であるが、2030年は商用車に加えて乗用車(自家用車)での搭載が予測できることから、1,530万台に成長する見込みである。そして

2030年以降では、レベル1からレベル5までのADAS/自動運転システムの世界搭載台数は8,390万5,000台に達すると予測する。

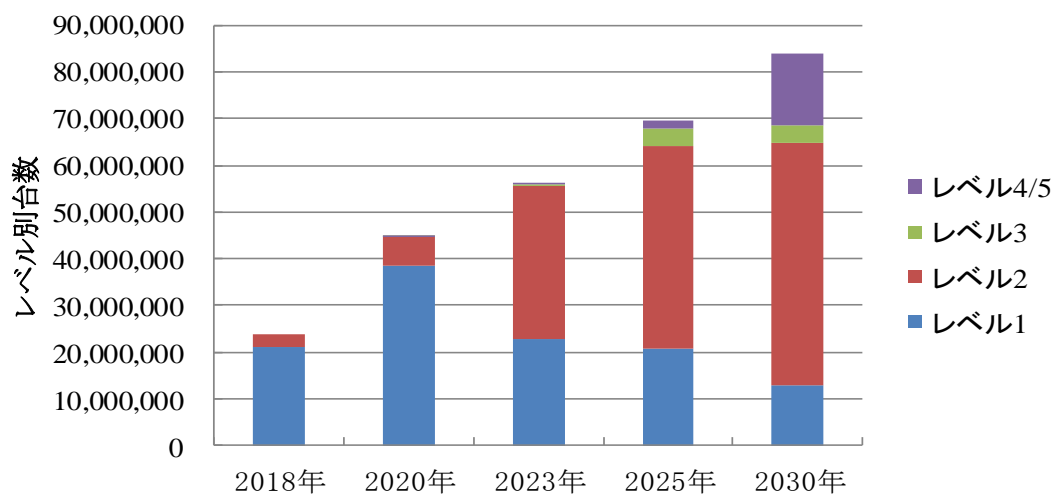


図1.5 : ADAS/自動運転システムの世界市場規模予測(概要)

表 1.3 : ADAS/自動運転システムの世界市場規模予測(レベル別詳細)

(単位：台)

レベル別	2018年	2020年	2023年	2025年	2030年
レベル1	21,148,000	38,666,000	22,689,000	20,600,000	12,745,000
レベル2	2,706,000	5,958,000	32,953,190	43,574,000	52,130,000
レベル3	0	800	220,000	3,701,000	3,730,000
レベル4/5	0	7100	207,900	1,795,600	15,300,000
合計台数	23,854,000	44,631,900	56,070,090	69,670,600	83,905,000

## 1.4 自動運転システムの変遷

本節では自動運転システムの開発状況について記載する[4]。自動運転の移り変わり時期は大きく分けて表 1.4 に示すように 4 つに分けられる。

### 第 1 期：路車協調型

第 1 期の自動運転システムでは、道路側に車両を誘導する誘導ケーブルを設置してラテラル制御を行う路車協調システムである。1950 年代後半から 60 年代まで米国の RCAGM, R. フェントン教授らオハイオ州立大学、英国の道路交通研究所、ドイツのジューメンズ社等が開発を進めていた。日本では、1960 年代前半に通商産業省機械技術研究所(現産業技術研究所)で開発が行われ、1967 年に当時の試作機が 100km/h で走行した。

誘導ケーブルを使用したラテラル制御は 1950 年代に工場内の無人運搬車を制御するシステムであり、路面に設置した誘導ケーブルに交流電流を流すとケーブル周囲に磁界が発生する。又、車両前部のバンパーには前後一組のコイルが装着されており、ケーブル側の磁界内を移動する際に発生する電流を基にコース外れを検出することができる。このような誘導ケーブルを設置するシステムは自動運転車の走行経路を明確に示すことができるが、ケーブルの設置や送電という作業やコストが運用上必須である為、テストコース等の限定された場所であり、タイヤ等の車両部品の耐久試験での使用に限定される。

誘導ケーブルが実際の公道で設置され、自動運転した例として、2 箇所ある。1 つ目はスウェーデンのハルムスタード、2 つ目はドイツのフルトである。これらの利用方法として路線バスを運行する際、停留所に正確に停止する必要がある(この正確な停止作業をプレジジョンドッキングと言う)。この正確な停止作業を実現する為に誘導ケーブルを設置した。その為、誘導ケーブルの設置は路線バスの停留所付近に限定され、大幅な設置コスト増にはならなかった。

表 1.4：自動運転システムの歴史概要

No	年代	特徴
第 1 期	1950 年代～1960 年代	路車協調型
第 2 期	1970 年代～1980 年代	自律型
第 3 期	1980 年代後半～1990 年代後半	ITS プロジェクトにおける各技術の開発
第 4 期	2000 年以降	実用化を目指した開発





図 1.6：路面下に設置された誘導ケーブルとパンパーに装着されたコイル[4]

## 第 2 期：自律型自動運転

1970 年代～1980 年代ではカメラセンサとコンピュータを使用した自動運転車が開発された。自律型自動運転の目的はカメラセンサとコンピュータの使用により、道路側に特別な装置を設置しなくても走行可能にすることである。カメラセンサとコンピュータによる自動運転システムをマシンビジョンシステムと呼ばれた。日本ではマシンビジョンシステムをいち早く開発し、1977 年に試作機を完成させた。この試作機は自車速 30km/h で走行し、道路上に設置されたガードレールを検知して衝突せずにテストコースを走行することができた。また 1980 年代にはオドメーターと自車位置を測位する装置により自分が走行した経路を記録するナビゲーションシステムを開発した。このナビゲーションシステムでは、マシンビジョンシステムでガードレール等の障害物を検知して避け、過去に走行した経路図を用いて約 300m のテストコースを車載機器だけで走破することができた。

1980 年代後半には、PVS(Personal Vehicle System)が試作され、マシンビジョンシステムによって道路上のレーンマーカーを検知しレーンマーカーに追従して走行する自動運転実験を実施した。PVS にはナビゲーションシステムも搭載されており、過去に走行した経路が保存され、この経路を用いて次回の走行時には目的地まで最短経路で走行できるように最適化アルゴリズムも搭載されていた。しかし、各情報を処理するコンピュータが大掛かりなものであったため、実験車両はマイクロバスのサイズを必要とした。

海外に目を向けると、アメリカではメリーランド大学やマーチンマリエッタ社が軍用の自動運転車(Autonomous Land Vehicle)を開発しており、主にオフロードでの使用を想定したものであった。ドイツでは、1980 年代中盤からディックスマンズ教授の所属するミュンヘン連邦国防大学にて自動運転車両 VaMoRa(Versuchsfahrzeug fuer autonome Mobilitaet und Rechnersehen)の開発が行われた。この VaMoRa も車両サイズはマイクロバス程度あり、1980 年代後半には約 90km/h で走行可能であった。

### 第3期：ITS プロジェクトにおける各技術の開発

1980年代後半から各国が国を挙げてITSプロジェクトを立ち上げ、自動運転システムの開発を行っている。これまで車両単独での自動運転であったが、複数車両が連なった隊列走行であったり、普通乗用車に加えてバスやトラック等のサイズ違いも実験対象となった。

ヨーロッパでは1986年から8年間行われたITSプロジェクトである

PROMETHEUS(Programme for a European Traffic with Highest Efficiency Unprecedented Safety)がある。このプロジェクトでダイムラーベンツのVITAII(Vision Technology Application)はカメラセンサ18台と60台のマイクロプロセッサからなるマシンビジョンシステムを搭載し、100km/hの速度で先行車認識や障害物回避、車線追従、車線変更を行うことができた。このシステムにはVaMoRaを改良したVaMPが搭載されており、近傍領域を撮影するカメラと遠方領域を撮影するカメラの2種類が搭載され、撮影した連続道路画像にカルマンフィルターを適応することで車線や先行車を高精度に認識している。VaMPは1995年の実車実験でドイツからデンマークまで約1700kmの道のりで400回以上車線変更に成功し、120km/hでの自動運転に成功した。

アメリカでは、ISTEA(Intermodal Surface Transportation Efficiency Act:総合陸上交通効率化法)という法律があり、ISTEAに基づき、ITSプロジェクトの1つAHS(Automated Highway Systems)を進め、1997年に実車実験を行った。この実験ではサンディエゴ市内の約12kmのテストコースを以下の7チームがそれぞれの自動運転車で実験するものである。

表 1.5 : ITS プロジェクトにおける各機関の取り組み

研究機関名	自動運転車の実験内容
カリフォルニア PATH	8台の乗用車が、車間距離6.3m、車速96km/hで隊列走行し、先頭から2台目の乗用車が車線変更した後、最後尾に合流する。路車協調システムで磁気マーカーの間隔は1.2mである。
カーネギーメロン大学	マシンビジョンシステムによる自律型自動運転。乗用車2台、ミニバン1台、路線バス2台をマシンビジョンシステムにより自動運転させた。
オハイオ州立大学	自律型自動運転と路車協調の組み合わせ。マシンビジョンシステムに加えて、路面にはレーダー波反射テープを貼り付け、反射波を利用して車両のコースずれを検知することができる。このテープの欠点として反射テープにほこりやゴミが付着すると反射が減衰して検知性能が低下する。

トヨタ	ACC(Adaptive Cruise Control)による車間，速度の制御を実施.
ホンダ	マシンビジョンシステムとカリフォルニア PATH の磁気マーカーを利用した自動運転. マシンビジョンシステムではインフラが整備されてないへき地での運用，路車協調はインフラが整備されている都市部での運用を想定.
イトン・ボラド社	大型トラックでの ACC の運用.
カリフォルニア運輸省	カリフォルニア PATH の磁気マーカーのメンテナンスを行う作業車両の実験. 磁気マーカーの有無をマシンビジョンシステムで検知してメンテナンスを行う.

上記，実車実験後，1998 年アメリカ運輸省は自動運転の将来的な導入が困難であり，産業への寄与が限定的であることから AHS による ITS プロジェクトを中止した.

日本では，建設省(現国土交通省)が 1995 年にテストコース，1996 年に上信越高速道路で磁気マーカーによる路車協調型の自動運転システムの実験を行った. しかし，自動運転の難易度が高く，導入が難しいことから自動運転道路システム AHS(Automated Highway Systems)の開発を中止し，走行支援道路システム AHS(Advanced Cruise-Assist Highway Systems)の開発をメインに行うことにした. その一環として 2000 年に通商産業省機械技術研究所と自動車走行電子技術協会(現日本自動車研究所)は車車間通信が可能な 5 台の乗用車を用意し，隊列走行する実験を行った. 各車両には GPS による自車位置計測と高精度地図データ，車車間通信による互いの位置と速度の情報をやり取りすることで隊列の合流と車線変更した.

#### 第 4 期：実用化を目指した開発

第 4 期の ITS プロジェクトの開発は国家機関によるプロジェクトとは異なり，民間企業による開発がメインとなる. また対象車両は乗用車から路線バスやトラック等の用途が限定された車両となる. 路線バスやトラックなどが対象となった理由として，これらの車両は決まった道路，高速道路を走行し，想定される走行シーンが絞られる為と考えられる.

ドイツのアーヘン工科大学ではトラックの隊列走行時にドライバーの人数を減らす為にマシンビジョンシステムを使用する実験を行っている. 実験は 2005 年から 2009 年まで行われ，4 台のトラックが車間距離 10m，速度 80 km/h で隊列走行する実験を行った. 先頭のトラックはドライバーが運転し，後続のトラックはマシンビジョンで検知したレーンマーカーに沿って自動運転する. マシンビジョンシステム以外にも車間距離計測にはレーダーを使用し，無線 LAN による車車間通信する機能を備えている.

又，従来の自動運転のようにシンプルに自動運転車が道路を自律して走行する状況だ

と想定される状況が多く技術的なハードルが高いが、シーンを限定した自動運転、運転支援システムが開発された。ヨーロッパでは HAVEit(Highly Automated Vehicle for Intelligent Transport)と呼ばれるコンセプトである。このコンセプト例として低速走行時の運転支援をするもので、渋滞中や市街地での低速時(0 km/h～30 km/h)の衝突回避ブレーキ、白線追従による簡易的な操舵アシストである。

アメリカでは DARPA(国防高等研究計画局)は2つのコンペティション Grand Challenge と Urban Challenge を開催した。これらのコンペティションの参加団体はカーネギーメロン大学、スタンフォード大学、マサチューセッツ工科大学、オハイオ州立大学等の大学が参加している。2つのコンペティションの概要を以下に記載する。

表1.6：Grand Challenge と Urban Challenge の概要

コンペティション名	概要
Grand Challenge2005	走行コースが砂漠のオフロードである為、GPS と地図を組み合わせたナビゲーションを行い、車両の屋根には複数のレーザースキャナーを設置して自車周辺の障害物を検知する。優勝したのはスタンフォード大学。
Urban Challenge2007	走行コースがカリフォルニア州の一般道路であり、カリフォルニア州の交通規制、制限速度、交通標識を守るという指示や、他車の運転する車両が入り混じった道路を走行するなど実際の走行シーンを想定したルールとなる。優勝したのはカーネギーメロン大学の車両。

上記コンペティションの結果を基に Google のトルン博士が自動運転車を試作した。この自動運転車のセンサの構成はレーザースキャナー、マシンビジョンシステム、レーダー等のセンサが搭載されており、信号機のある市街地を走行することができる。この実験は2009年からカリフォルニア州の公道で実験が開始され、2015年までに160万kmの走行距離を無事故で達成した。

また、2013年以降、Googleの自動運転開発に触発されて自動車メーカーや自動車部品メーカーによる自動運転車の開発が加速されている。ダイムラーはドイツのマンハイムからブファルツハイムまでの市街地、田舎道の約100kmを自動運転で走行した。この時のセンサの構成はマシンビジョンシステムとミリ波レーダーのみである。

自動運転開発の第2期から第4期までと自動運転車が路車協調せずに単独で走行するには多数のセンサが必要であり、カメラを使用したマシンビジョンシステムは古くから使用されている。

## 1.5 自動運転車に必要なセンサの種類

自動運転を実現するには様々なセンサが必要である。例えば、カメラや LiDAR、ミリ波、加速度センサ、GPS など物体認識用途から自車両の状態を計測する為に使用するセンサと多種に存在する。本節では、以下の自車両の周辺環境認識に必要なセンサについて着目し、1.4.1 節よりそれぞれのセンサについて詳細を述べる。

- カメラ(単眼・ステレオ・IR)
- LiDAR
- ミリ波レーダ
- 超音波センサ

### 1.4.1 カメラ(単眼・ステレオ・IR)

#### ■単眼カメラ

単眼カメラとは、今最も自動車に普及しているセンサであり、車両の前方監視及び後方監視やアラウンドビューモニターなど幅広い用途がある。また構成部品としてイメージセンサとレンズから成る。

表 1.7: 単眼カメラのメリット及びデメリット

メリット	低コスト、サイズが小さい為、設置場所の自由度が高い。物体認識が比較的容易である。カラー画像を利用する場合、信号機の色から交差点の状況を把握することができる。キャリブレーションが容易である。
デメリット	距離計測において、認識した画像座標の縦方向のピクセル位置から算出することは可能だが、誤差は一般的に大きい。



図 1.7: 単眼カメラ例 : STC-MCS241U3V(オムロンセンテック製)



図 1.8：単眼カメラ例：デンソー製 Toyota Safety Sense

■ステレオカメラ

二つのレンズ，イメージセンサから成り，視差を利用して3次元認識が可能である。その為，撮影シーンの奥行情報を正確に把握することが可能である。

表 1.8：ステレオカメラのメリットとデメリット

メリット	単眼カメラと同様に物体の認識が可能であり，物体の正確な距離が測定可能である。
デメリット	二つのカメラのキャリブレーションが難しい。



図 1.9：ステレオカメラ例：デンソー製ステレオカメラ

## ■IR カメラ

近赤外線カメラ，遠赤外線カメラが存在するが，用途が限られることから車載用に普及はしていない．近赤外線カメラは近赤外線を照射する，遠赤外線カメラは物体から放射される遠赤外線を受光する為，夜間での物体認識に強みを持つ．

表 1.9: IR カメラのメリットとデメリット

メリット	夜間での物体認識が容易である．
デメリット	近赤外線の照射距離により撮像距離が変化する．撮像範囲は約30m となり，可視光カメラと比較して短い． 遠赤外線カメラの価格が高価である． カメラサイズが可視光カメラより大きい．



図 1.10: IR カメラの例 : JVC ケンウッド製遠赤外線カメラ

#### 1.4.2 LiDAR(Light Detection and Ranging)

LiDAR は、レーザー光を走査しながら対象物に照射してその散乱や反射光を観測することで、対象物までの距離を計測したり対象物の性質を特定したりする、光センサー技術のことである。豊田中央研究所所属の城殿ら[5]は下記 Velodyne 製 LiDAR を使用している。

表 1.10 : LiDAR のメリットとデメリット

メリット	物体までの距離測定が可能である。 位置や形状の把握が可能である。 天候による明るさの影響を受けにくい。
デメリット	高価格(数万円～数百万円) サイズが大きい



図 1.11 : LiDAR 例 : Velodyne 製 VLS-128-AP

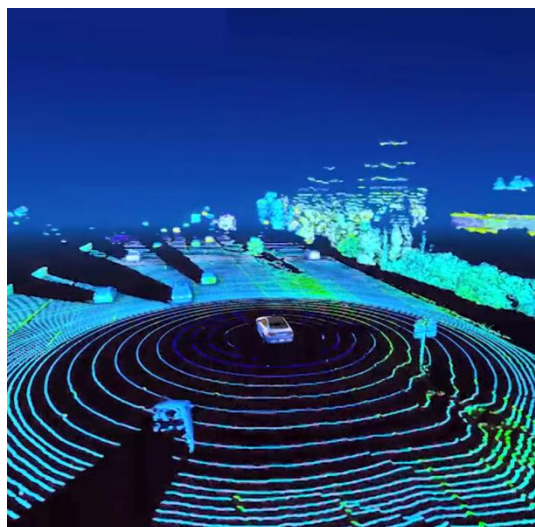


図 1.12 : LiDAR による 3 次元点群データの取得結果(引用元 Velodyne HP より)



### 1.4.3 ミリ波レーダー

周波数 30GHz～300GHz 帯の電波を飛ばし、対象物による反射波を利用して測距を行う

表 1.11：ミリ波レーダーのメリットとデメリット

メリット	短波長である為、対象物を高精度で検出することができる。 分解能が高く最小で 0.1 mmの動きを検出可能である。 電波は赤外線や超音波と比較すると直進性が高い為、雨天や霧における悪影響が小さい。
デメリット	カメラと比較してコストがかかる。



図 1.13：ミリ波レーダー例：富士通テン性ミリ波レーダー

### 1.4.4 超音波センサ

超音波センサは、送波器にから超音波を対象物に向け発信し、その反射波を受波器で受信することにより、対象物の有無や対象物までの距離を検出する。対象の反射率に依存しない為、駐車時のバックソナーとして利用される。

表 1.12：超音波センサのメリットとデメリット

メリット	ガラス等の透明な物体の検出が可能。 ホコリや汚れに強い。 金網等の複雑な物体でも検出可能。
デメリット	検出距離が約 10m までと短い。



図 1.14：超音波センサ例：タムラ製作所製超音波センサ

それぞれのセンサの特徴を表に記載する。下表よりカメラセンサ，特に単眼カメラは周囲環境の変化に弱いデメリットがあるものの，このデメリットをソフトウェア上で克服することで，応用が広がることがわかる。その為，本研究では単眼カメラを使用して研究を行う。

表 1.13：自動運転に必要なセンサ機能

機能 センサ		道路上の物体認識の可否			センサの特性・性能		
		区画線等	信号灯火	一般物体	照明変化	天候変化	距離精度
カメラ	単眼	○	○	○	△	△	△
	ステレオ	○	○	○	△	△	△
	IR	△	×	△	○	△	○
LiDAR		×	×	△	○	○	○
ミリ波レーダー		×	×	△	○	○	○
超音波センサ		×	×	×※	○	○	△

※超音波センサでは物体の種類認識はできず，物体の存在の有無のみがわかる為，今回は×とした。

## 第2章 関連研究

本章では、物体認識の手法について従来の画像処理を利用した手法から、現在利用されている DeepLearning を利用した手法について述べる。

### 2.1 昼間環境での物体認識(画像処理)

単眼カメラを利用した物体認識に関する研究は、昼間の環境では盛んに行われている [6-9]。Dalal ら[6]は歩行者の輪郭の勾配に着目した Histogram of Oriented Gradient(HOG) 特徴量を提案し、Support Vector Machine (SVM) と組み合わせることで歩行者の認識性能を向上させている。HOG 特徴量は、対象領域を複数のセルに分割し、各領域内で方向ヒストグラムをまとめて多次元ベクトルを構成し、特徴量とする。これにより、画像内部の物体の形状を捉えるに適した画像特徴量となり、歩行者検出に利用される。渡辺ら[7]は、HOG 特徴量を改良した Co-occurrence Histogram of Oriented Gradient(CoHOG)特徴量を提案し、詳細な形状抽出を行っている。画像は輝度値の異なる画素が集まったものであり、この各画素について輝度値が変化する方向(輝度勾配)を算出することで、物体の各パーツのエッジの局所的な形状を求めることが出来る。求めた各パーツの画素毎のエッジは一定の範囲における出現頻度を集計すると安定した特徴量となる。この集計処理で CoHOG では、離れた 2 箇所形状を 1 組として、その出現頻度を特徴量とすることで得られる特徴量の種類を増やし、より多くの形状を表現できるようにした。その結果、物体の分類性能を HoG より向上することができる。

Dollar らの研究[8,9]では色情報や勾配情報等を組み合わせて複数の特徴量を生成し、分類器として AdaBoost を使用した Integral Channel Features(ICF)がある。ICF の基本的な考え方は、顔検出で有名な Viola-Jones(VJ)手法を拡張したものである。VJ は以下の 3 つの要素から成り立っており、Haar-like 特徴量、Adaboost による特徴の選択と学習、cascade 型分類器による高速化を実現する。この VJ では輝度だけを使用して特徴を計算しているが、ICF では、入力画像のカラーチャンネルに対して、線形・非線形の変形処理を行い、複数のチャンネル画像を作成する(図 2.1)。作成した複数のチャンネルに対して Haar-like 特徴量をより簡素化した特徴を計算する。得られた特徴量は VJ の輝度だけを使用した特徴量と比較してより次元の多い特徴となる。また ICF は検出対象に応じて抽出する特徴量を追加することが容易な構造となる。特徴の選択と学習は VJ と同様の Adaboost、分類器には cascade 型分類器である soft cascade を利用する。抽出する特徴量として、Dollar らは(a)入力カラー画像を輝度画像にする、(b)表色系を変換する、(c)ガボールフィルタをたたみ込んで特定方向の勾配画像を求める、(d)DoG フィルタをたたみこむ、(e)勾配強度を求める、(f)エッジ検出する、(g)直線のフィルタをたたみ込んで特定方向の勾配画像を求める (ガボールフィルタとほとんど同じ)、(h)閾値処理で二値化する等がある。得られた特徴量のどの領域を見るかで特徴量の数は膨大となるが、

Adaboost を使用することで、有効な特徴を選択し組み合わせで分類をすることが可能である。Dollar らの実験では、人認識には勾配方向(ガボールフィルタか直線のフィルタをかけた画像)と勾配強度、LUV 表色系の各チャンネル、勾配方向を6方向に量子化した場合には合計 10 チャンネルを使うことが多い。画像を各チャンネルに変換した結果と、AdaBoost により選択された特徴がどのチャンネルのどの領域をみるものが多かったかを示すヒートマップとなる(図 2.2)。このヒートマップ図から人の肩周辺のエッジや、顔の肌色に反応する特徴が選択されている。ICF の処理速度について  $640 \times 480$  の画像から異なるスケールの人を認識するのに約 2s/枚の計算時間が必要だとされており、更なる高速化するための手法[9]が提案されたり、GPU を用いたりリアルタイム検出が可能になったりしている。さらにステレオカメラを利用して画像の 3 次元情報を取得、人が存在する可能性が高い領域を推定して検出範囲を小さくすることで 100fps を達成したという研究がある。

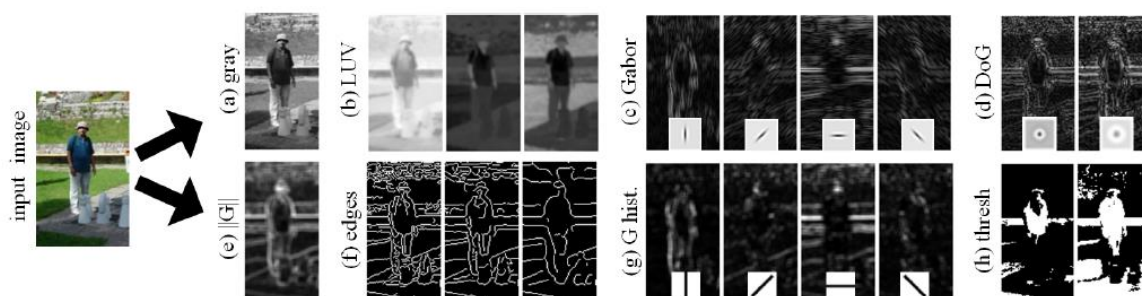


図 2.1 : ICF による複数の特徴抽出器による特徴画像[8]

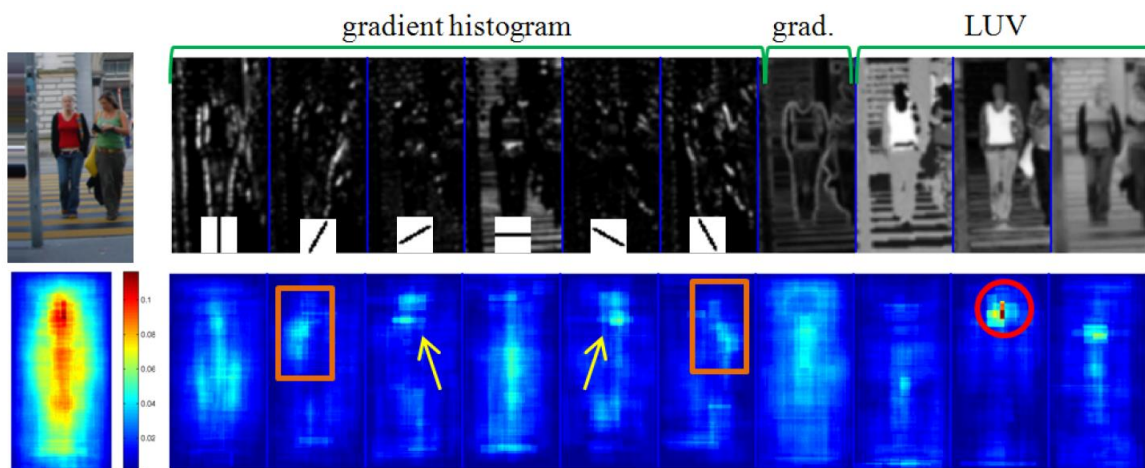


図 2.2 : 各特徴抽出における特徴強度[9]

## 2.2 昼間環境での物体認識(DeepLearning)

RGB カメラを利用した物体認識手法として 2012 年より高精度な成果が報告されている DeepLearning を利用した物体認識手法がある[10-19]. Krizhevsky らの研究[10]は ILSVRC(Image Net Large Scale Visual Recognition Challenge)で群を抜いた性能を示し, DeepLearning を利用した物体認識の火付け役となった. この DeepLearning のネットワークは Alexnet と呼ばれ, 5 層の畳み込み層と 3 つの全結合層を持つ畳み込みニューラルネットワーク(Convolution Neural Network, CNN)となる. Alexnet は DeepLearning の基本となる以下の 4 つの技術が使われている.

- ReLU 活性化関数
- マルチ GPU での学習
- Data augmentation(データ拡張)
- Dropout

表 2.1 : Alexnet の概要[10]

レイヤーの順番	種類	カーネルサイズ
1	conv	11x11x3x96
2	max-pooling	-
3	conv	5x5x48x256
4	max-pooling	-
5	conv	3x3x256x384
6	conv	3x3x192x384
7	conv	3x3x192x256
8	max-pooling	-
9	FC-4096	1x1x4096
10	FC-4096	1x1x4096
11	FC-4096	1x1x1000
12	softmax	-

## R-CNN: *Regions with CNN features*

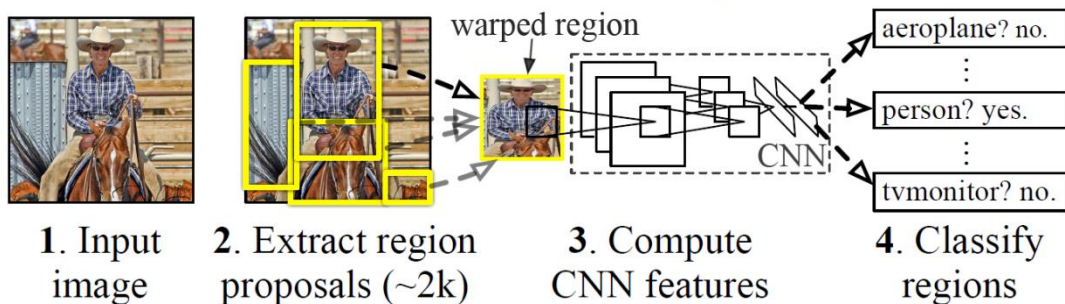


図 2.3 : R-CNN 概要図[11]

Girshick ら[11]の R-CNN は Edge Boxes や Selective Search を利用して候補領域を検出する。この領域検出としておよそ以下の手順となる。

1. 画像の特徴量(色合いや濃淡勾配など)を元に画像をいくつかの領域に分けていく。
2. 類似度を元に領域を結合していく。
3. 適度な大きさになるまで結合を続ける。
4. 分けた領域を元に候補となる領域(四角形)を生成していく

このとき検出される候補領域は約 2000 個の領域となる。

候補領域検出後、CNN に入れ、特徴を抽出する。CNN 特徴量を得た後、従来の分類器であるサポートベクターマシン(Support Vector Machine, SVM)で領域内の分類を行う。

Ren ら[12]は処理速度向上の為、物体の候補領域の検出と認識で同じ特徴マップを利用して物体の特徴抽出と位置を推定する Faster R-CNN を提案した。大まかな流れを以下に述べる。

1. 画像内のある領域が物体なのか背景なのか学習する。
2. 1 で検出した場所に存在する物体が具体的に何なのか学習する。

Faster R-CNN の特徴として、1 のある領域検出する処理を外部の処理(Selective Search 等)を使用せず、CNN の特徴マップから行ったことである。この処理を Region Proposal Network(RPN)という。Faster R-CNN 以前の DeepLearning による物体認識は候補検出処理→分類の 2 段型なのに対し、Faster R-CNN は候補領域検出と分類を同じネットワーク内で行う 1 段型、End to End ができる最初のモデルである。

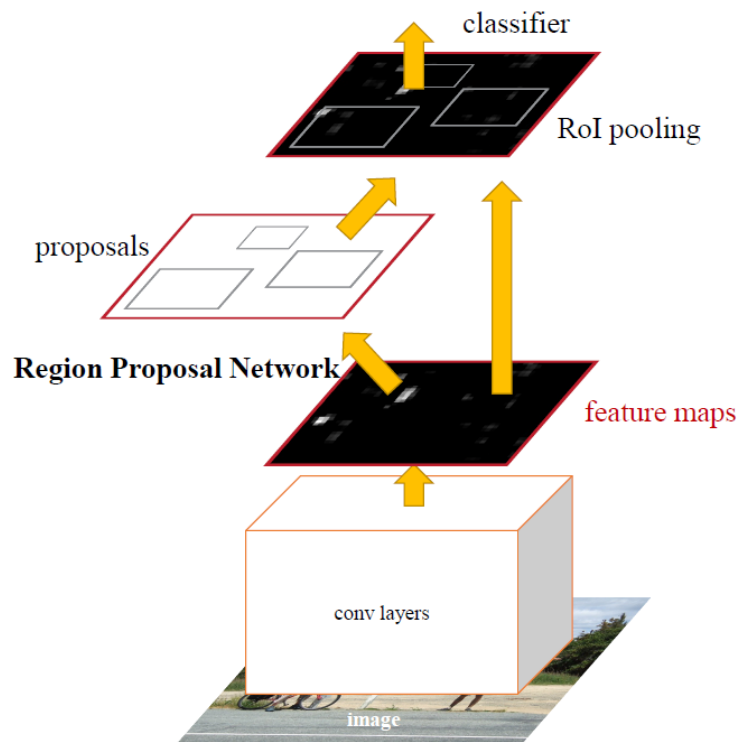


図 2.4 : Faster R-CNN 概要[12]

近年, 実時間での処理や画像奥に存在する小さい物体を検出する手法が開発されており, End to End の手法として Joseph らが考案した YOLOv3[15]がある. 小さい物体の認識精度向上の為, YOLOv3 では ResNet と Feature Pyramid Net(FPN)を利用している.

#### ResNet(Residual Net)

ResNet[16]では Residual Block と呼ばれる従来のネットワークに shortcut path を加えた構造を言う. 本来, ネットワークの層が深くなると, 学習が難しくなるが, shortcut path を加えることによって, ある層で求める最適な出力を学習するのではなく, 前層の入力を参照した残差関数を学習することで, 特徴量の学習をしやすくする. この結果, もともと複雑な特徴量は, 古い特徴量に新しく学習した残差を足し合わせて学習することが出来る. これにより学習の難易度が下がり, 深い層まで細かく学習し, 精度向上することができる.

### Feature Pyramid Net(FPN)

画像中の物体の大きさ変化に対して, YOLOv3 は FPN 構造[17]を利用する. FPN 構造はトップダウンの方向と潜在的な結びつき(lateral connection)により, 低解像だが意味的に強い(semantically strong)特徴と高解像だが意味的に弱い(semantically weak)特徴の両方を利用している. この2種類の特徴マップを足し合わせることで小さい物体の分類に強い feature pyramid を作成することが可能になる.

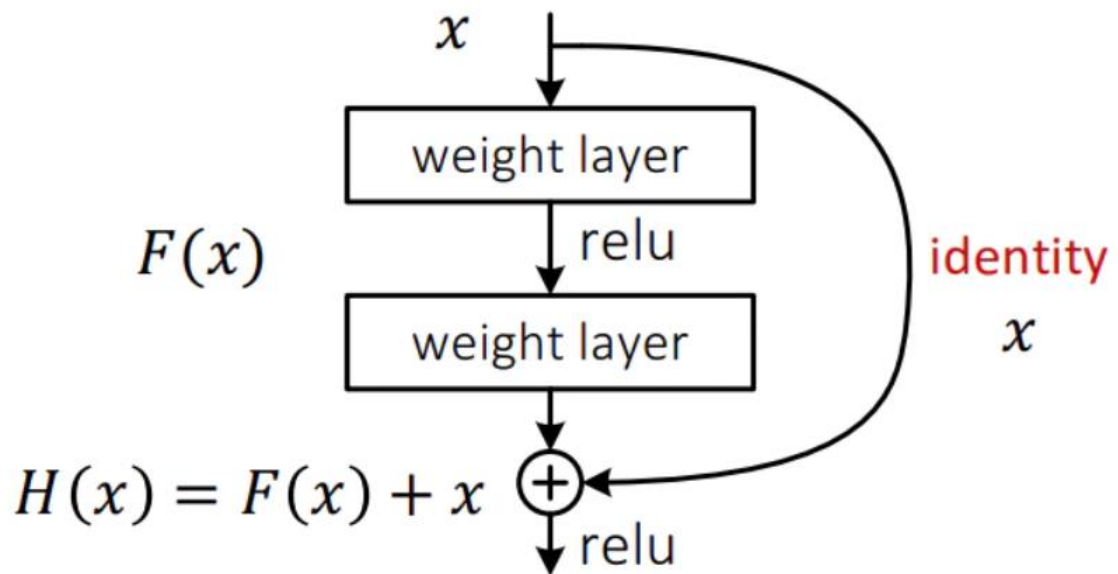


図 2.5 : ResNet 概念図[16]

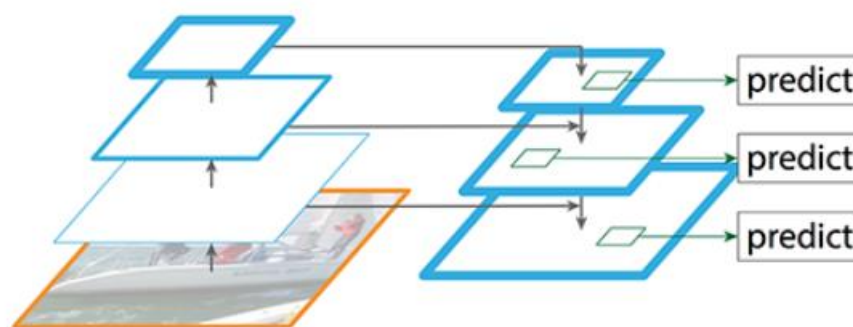


図 2.6 : Feature Pyramid Networks の概念図[17]



また、2019年に研究された物体認識手法として M2Det[18]がある。M2Det の概要図を図 2.7 に記す。M2Det の特徴として以下がある。

- Backbone network
- Multi-Level Feature Pyramid Network(MLFPN)
- Prediction layers

#### Backbone network

Backbone network は特徴抽出器であり、DeepLearning で使用される従来のネットワークを使用する。使用されるネットワークは VGG-16 や ResNet-101 が使用される。

#### Multi-Level Feature Pyramid Network(MLFPN)

MLFPN は以下の 3 つのモジュールで構成される。

- Feature Fusion Module(FFM)

Backbone network で得られた特徴マップのうち異なる解像度の特徴マップを合わせることで以降の処理のベースとなる Base feature を生成する。

- Thinned U-shape Module(TUM)

TUM はエンコーダとデコーダの構成になっており、FFM からの出力を受け取り、再度マルチスケールに対応する為の特徴ピラミッドを生成する。

- Scale-wise Feature Aggregation Module(SFAM)

SFAM は複数の TUM から得られた特徴ピラミッドを統合する処理である。具体的には、各特徴ピラミッドをチャンネル方向に結合して、チャンネル毎に Global Average Pooling を適用する。この処理により各チャンネルの情報を圧縮し、それらを全結合層に適用することで各チャンネルに応じた重みに変換する。この重みを各チャンネルにすることで特徴ピラミッドを生成し、Prediction layers に引き渡す。

#### Prediction layers

Prediction layers は、MLFPN から得られる特徴ピラミッドに Convolution 層を繋げて物体の認識と位置の推定を行う。得られた特徴マップに対し、6 種類のアンカーボックスと 3 種類のアスペクト比を適用し、推定スコアが 0.05 以上となるものを最終的な出力とする。

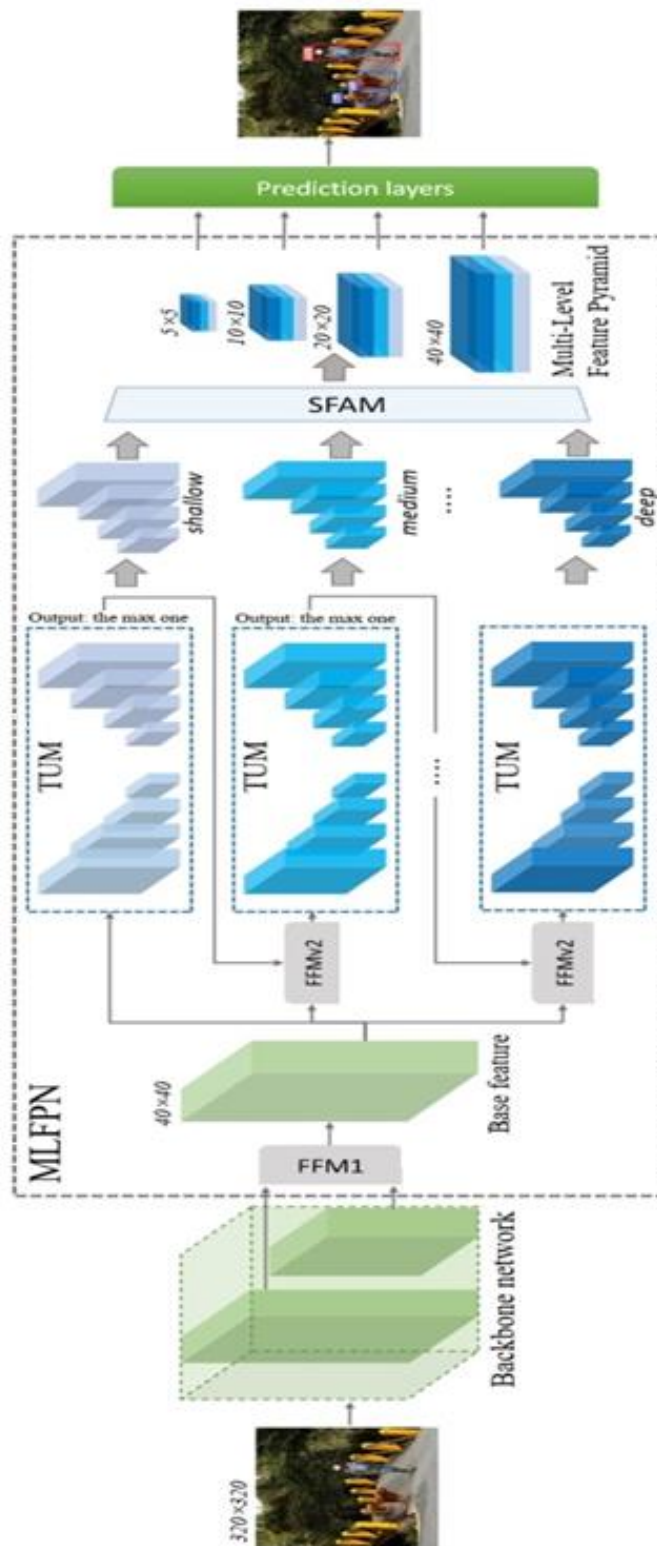


図 2.7: M2Det のネットワーク概要図[18]

これら DeepLearning による手法は学習データとして大量の歩行者画像を準備して学習を行い、認識の高精度化を実現している。昼間の歩行者認識では、Web 上に各研究機関が公開している画像群がある為、高い認識率となる。昼間と同様に大量の夜間の歩行者画像を準備することで夜間歩行者認識へ対応することができるが、夜間歩行者の頻度は昼間の歩行者と比較して圧倒的に少ない為、様々なパターンの歩行者画像の収集に膨大なコストが必要であり、現実的でない。

### 2.3 赤外線を利用した夜間物体認識

夜間の歩行者認識として近赤外線や遠赤外線カメラを利用した研究が行われている。赤外線とは可視光の波長  $0.36\mu\text{m}\sim 0.83\mu\text{m}$  より長い  $0.83\mu\text{m}\sim 1\text{mm}$  の波長を言う。そして波長  $1\text{mm}$  以上のものを電波と言う。更に赤外線の波長の中には、近赤外、中赤外、遠赤外があり以下の表 2.2 のように区分される。赤外線カメラの撮像にはアクティブ型とパッシブ型があり、アクティブ型は撮影機器側から撮影対象(歩行者等)に向けて赤外線カメラの波長に応じた赤外光を照射する。近赤外光は歩行者から放射されることが少ない為、近赤外光を照射する装置を必要とする。遠赤外線カメラは歩行者から放射される遠赤外光を検出して撮影を行う。

赤外線カメラを使用した物体認識手法として手法[20-21]がある。前掲ら[20]は、近赤外線カメラ画像用の HOG 特徴量と Nearest First Traversing Graph(NFTG)を利用し夜間歩行者の認識を行う。青木ら[21]では遠赤外線カメラを利用した研究であり、赤外線カメラを利用することで比較的容易に夜間歩行者を認識することが可能である。

表 2.2：赤外線の波長による区分け

赤外線の種類	波長
近赤外(near infrared)	$0.83\mu\text{m}\sim 3\mu\text{m}$
中赤外(intermediate infrared)	$3\mu\text{m}\sim 6\mu\text{m}$
遠赤外(far infrared)	$6\mu\text{m}\sim 1\text{mm}$

しかし、赤外線カメラの活用では以下の課題が残る。近赤外線カメラでは、夜間道路走行時、近赤外光源と近赤外線カメラを搭載した2台の自動車がすれ違うシーンで双方の自動車がそれぞれ自車前方に向けて近赤外線光源を点灯すると、近赤外線カメラに近赤外線を向け合う形になり、カメラ画像ではホワイトアウトが発生、物体認識が困難になる課題がある。又、近赤外光を照射するにあたり、人間の目は赤外線に感度が無い為、赤外線の存在を感知することが出来ない。その為、近赤外線の照射を受けることで網膜を損傷する可能性がある。遠赤外線カメラでは、カメラ本体価格がまだ40万円と高価であり、普及の妨げとなっている[22]。

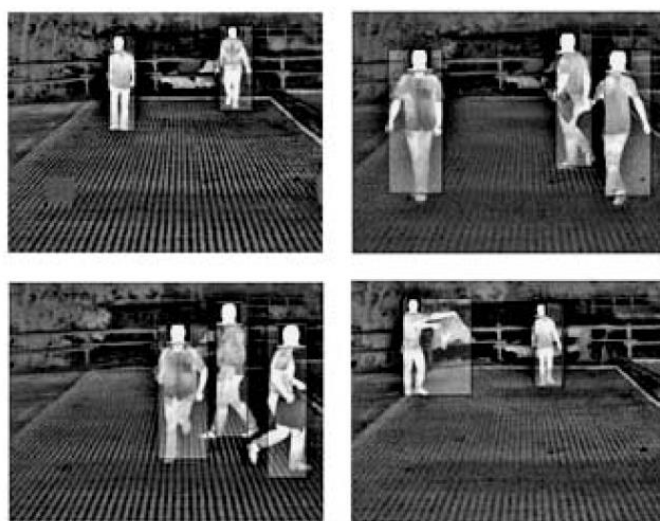


図 2.8：赤外線画像による歩行者認識[21]

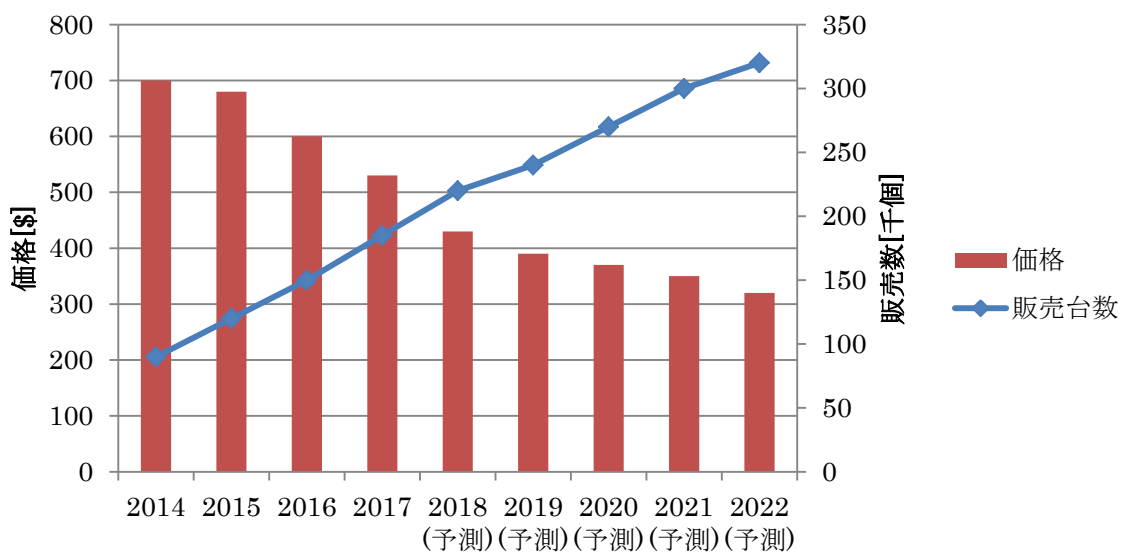


図 2.9：遠赤外線カメラの販売台数と単価[22]

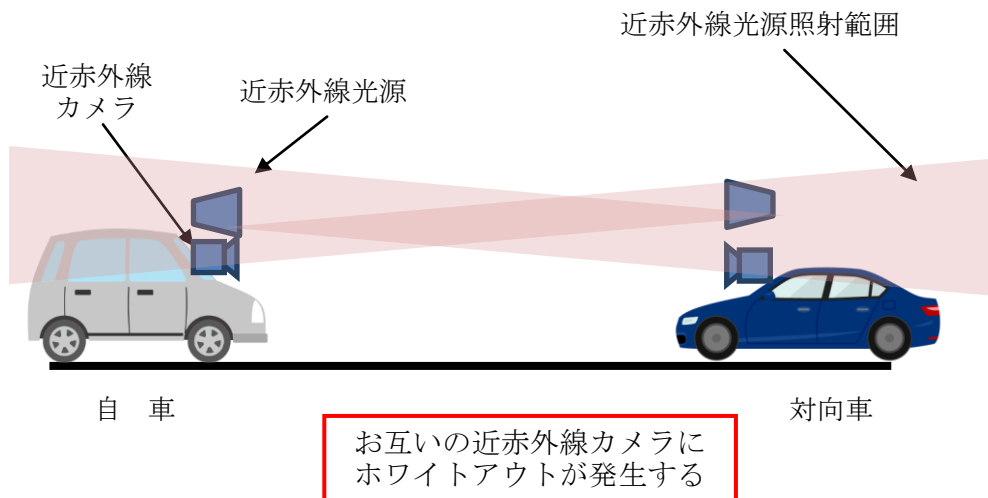


図 2.10: 近赤外線カメラによるホワイトアウトの発生

## 2.4 DeepLearning を利用した画像変換手法

2.2 節に記したように、夜間の歩行者を認識する為に夜間歩行者用の学習データを収集することは収集コストが膨大である為、夜間画像を昼間画像に近づくように画像変換し、その後、既存の昼間の歩行者データを使用した物体認識モデルの作成、物体認識を行うアプローチが考えられる。

画像を変換する研究として、2014年に Ian. Goodfellow ら考案した敵対的生成ネットワーク Generative Adversarial Network(GAN)の研究がある[23]。GANは1つのネットワークではなく、「Generator」、「Discriminator」と呼ばれる2つのネットワークを使うという特徴がある。Generatorは画像生成を行う為のネットワークであり、Discriminatorは、入力された画像が本物なのか、またはGeneratorが生成した画像なのか判定を行うネットワークである。これらのネットワークを学習する際は、GeneratorはDiscriminatorを騙せる画像を生成しようとし、Discriminatorはより本物と偽物を正確に判別できるように同時に学習させていく。以降、GANは様々な改良が行われ、成果が報告されている。

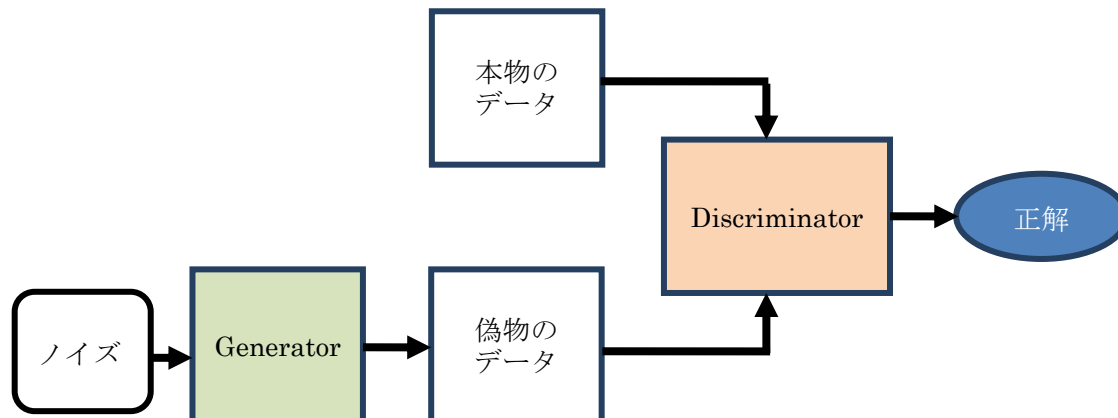


図 2.11: GAN の概念図

Gatys ら[24]は GAN を使用して画風を変換する手法を提案した。GAN は、画像生成モデルの一種であり、データから特徴を学習することで実在しないデータを作成したり、存在するデータの特徴に沿って変換したりすることができる処理である。Gatys らは GAN により画風の特徴を残しておき、物体等の情報を他の画像に置き換えることで任意の画像の画風を変換している。その為 GAN では、歩行者認識に必要な輪郭の再現はできず、夜間画像の視認性向上には向かない。又、Iizuka ら[25]はモノクロ画像から各画素の彩度を推論する Neural Network を構築し、入力モノクロ画像と統合することでモノクロ画像をカラー画像にする手法を提案した。しかし、昼間のモノクロ画像のカラー化はできるが、夜間画像は彩度の情報が少ない為、適切に変換することができない。Liu ら[26]は学習データに変換前後のペア画像を用意しなくとも変換に必要な特徴量を学習する手法を提案した。Liu らの手法は昼間の画像を夜間の画像に変換することは可能だが、夜間画像から昼間画像の変換のように画像情報を付加する変換は不得意である。また、Anoosheh ら[27]は夜間画像から昼間画像に変換するネットワークを検討しており、自動走行の位置情報取得に活用している。Huang ら[28]も夜間での物体認識の研究を行っており、昼間画像を夜間画像に変換し夜間の物体の学習データとして使用している。



図 2.12 : Gatys らによる画像変換結果[24]



図 2.13 : Iizukaらによる色付け結果例[25]

入力画像

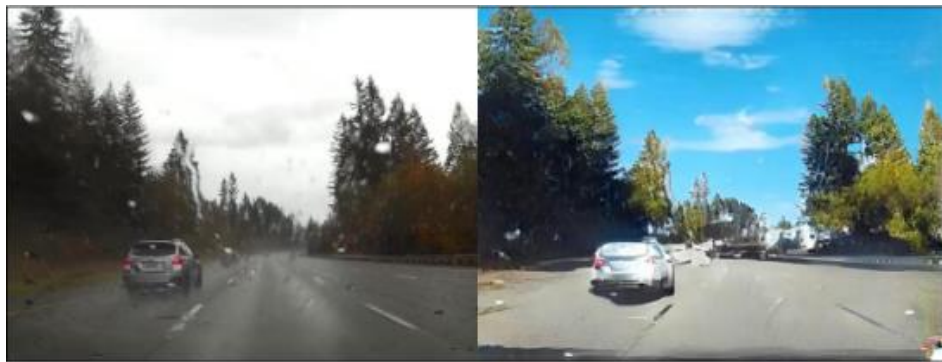
変換画像



シーン 1



シーン 2



シーン 3

図 2.14 : Liu らによる画像変換結果例[26]





図 2.15 : Anoosheh らによる夜間画像の変換結果例[27]

## 第3章 画像変換を使用した夜間歩行者認識の概要

本章では、4章、5章で提案手法を述べるにあたり、先に一般的な画像中の歩行者認識手法の処理の流れを述べ、提案手法との処理の違いを比較し提案手法の特徴を明らかにする。

### 3.1 一般的な歩行者認識手法の処理の流れ

一般的な歩行者認識では、構築したニューラルネットワークに歩行者の特徴を重みとして抽出する為に、以下の手順で認識モデルを作成する。

- ① 実際に車両にカメラを取り付け、数千から数万枚の歩行者データを収集する。
- ② 画像中から歩行者の存在する領域を特定する為に、手作業でアノテーション作業（正解データ作成）を行う。
- ③ 収集した画像と作成したアノテーションを学習プログラムに入力して学習を行う。
- ④ 認識モデルにテスト画像を入力して評価する。

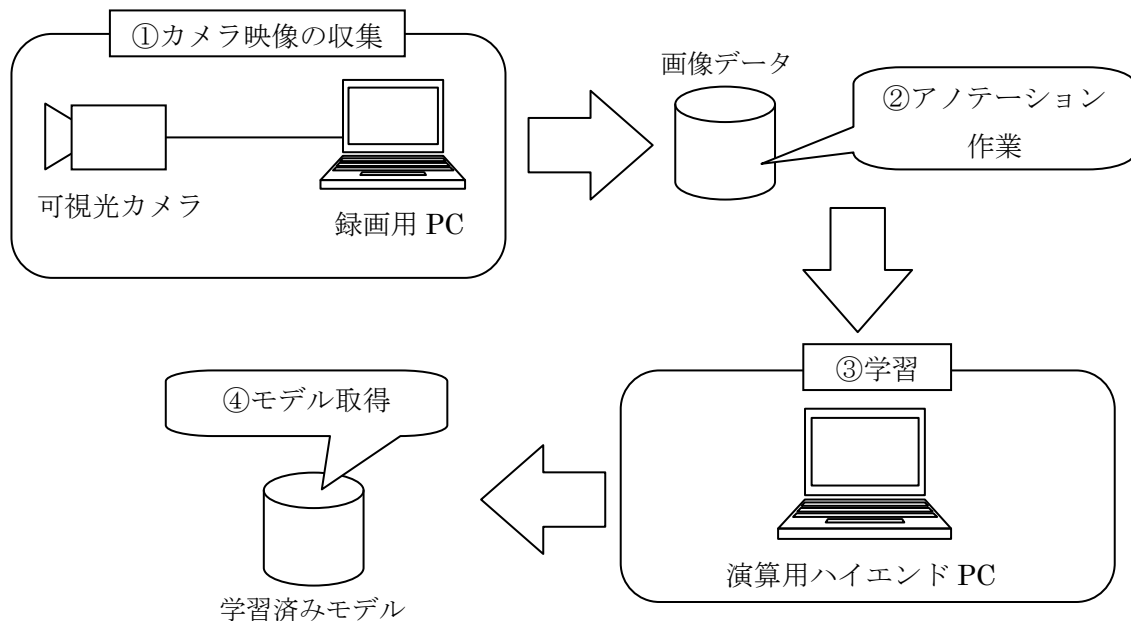


図 3.1：従来の物体認識手法の処理の流れ

### 3.2 提案手法による歩行者認識手法の流れ

本節では、提案手法による夜間歩行者認識の概要を述べる。

- ① 画像変換をする為に昼夜のシミュレーション画像を作成する。  
(夜間画像を変換する為の正解画像はシミュレーションを使用することで容易に作成することができる。)
- ② シミュレーション画像を画像変換ネットワークに入力し、画像変換モデルを学習する
- ③ 画像変換モデルと既存の物体認識手法を繋げて歩行者認識を行う。

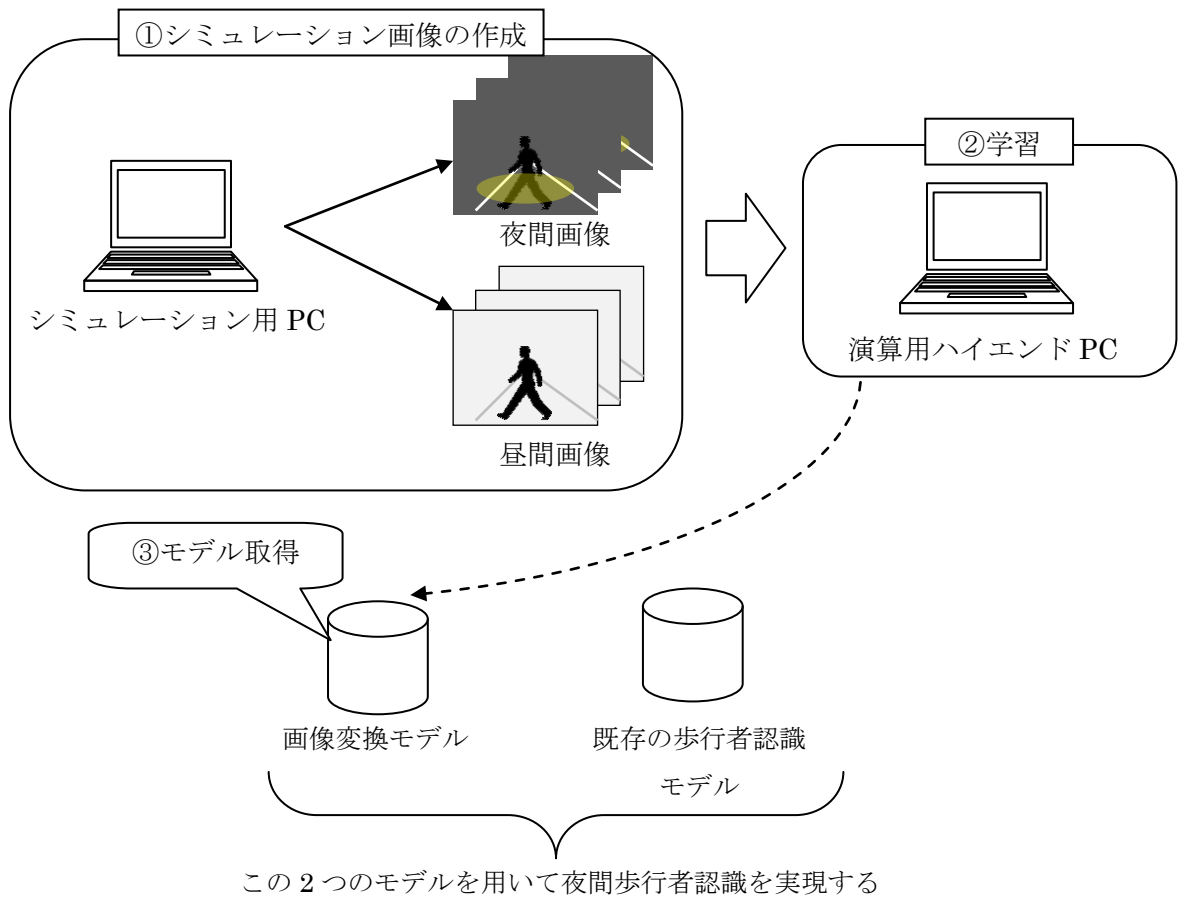


図 3.2: 提案手法の全体処理の概要

### 3.3 本研究の目的

3.1 節, 3.2 節で一般的な歩行者認識手法と提案手法の全体の処理の流れを述べた。これに基づいて一般的な歩行者認識手法と提案手法を比較した結果を表に記す。提案手法と一般的な歩行者認識手法を比較すると、一般的な歩行者認識手法では、夜間の歩行者データの取得及びアノテーション作業が必要になる為、膨大なコストが必要になる。アノテーションコストは数万枚の画像だと、数十時間かかり、作成するのは現実的ではない。

一方、提案手法のアプローチでは、歩行者認識には既存の学習データとアノテーションデータを使用する為、そのコストはほぼ無しとなる。しかし、夜間の画像を昼間の画像に近づける為のニューラルネットワーク用に学習データが必要になる。本研究では、このコストを抑える為に画像変換に使用する画像をシミュレーション画像で準備し、対応する昼間画像もシミュレーションで作成する為、作業コストを小さくするようにした。図に本研究で検討するニューラルネットワークの位置付けを記す。図上段に学習時の処理フローを記し、図下段に学習したモデルを使用して推論する処理フローを記す。図 3.3 の通り、本研究で収集する学習データは画像変換に用いる昼夜の画像であり、それらにより学習した学習モデルを保存、推論時に読み込んでテスト画像の変換を行う。変換された画像に対し、広く公開されている歩行者画像(アノテーションデータ含む)を使用して学習した物体認識アルゴリズムを適用して夜間歩行者認識を行う。

5 章, 6 章では、夜間画像から昼間画像に変換するニューラルネットワークについて述べる。

表 3.1 : 一般的な歩行者認識手法と提案手法の作業コストの比較

	一般的な歩行者認識手法	提案手法
画像変換用学習データの種類	-	シミュレーション画像
画像変換用学習データ 収集コスト	-	小 (シミュレーションで作成)
画像変換用アノテーション 作成コスト	-	小 (シミュレーション画像 使用の為)
物体認識用学習データ 収集コスト	数十時間 (カメラ設置, 車両準備等)	- (既存の学習データ使用)
物体認識用アノテーション 作成コスト	数十時間 (数万枚の画像に対し行う為)	- (既存の学習データ使用)

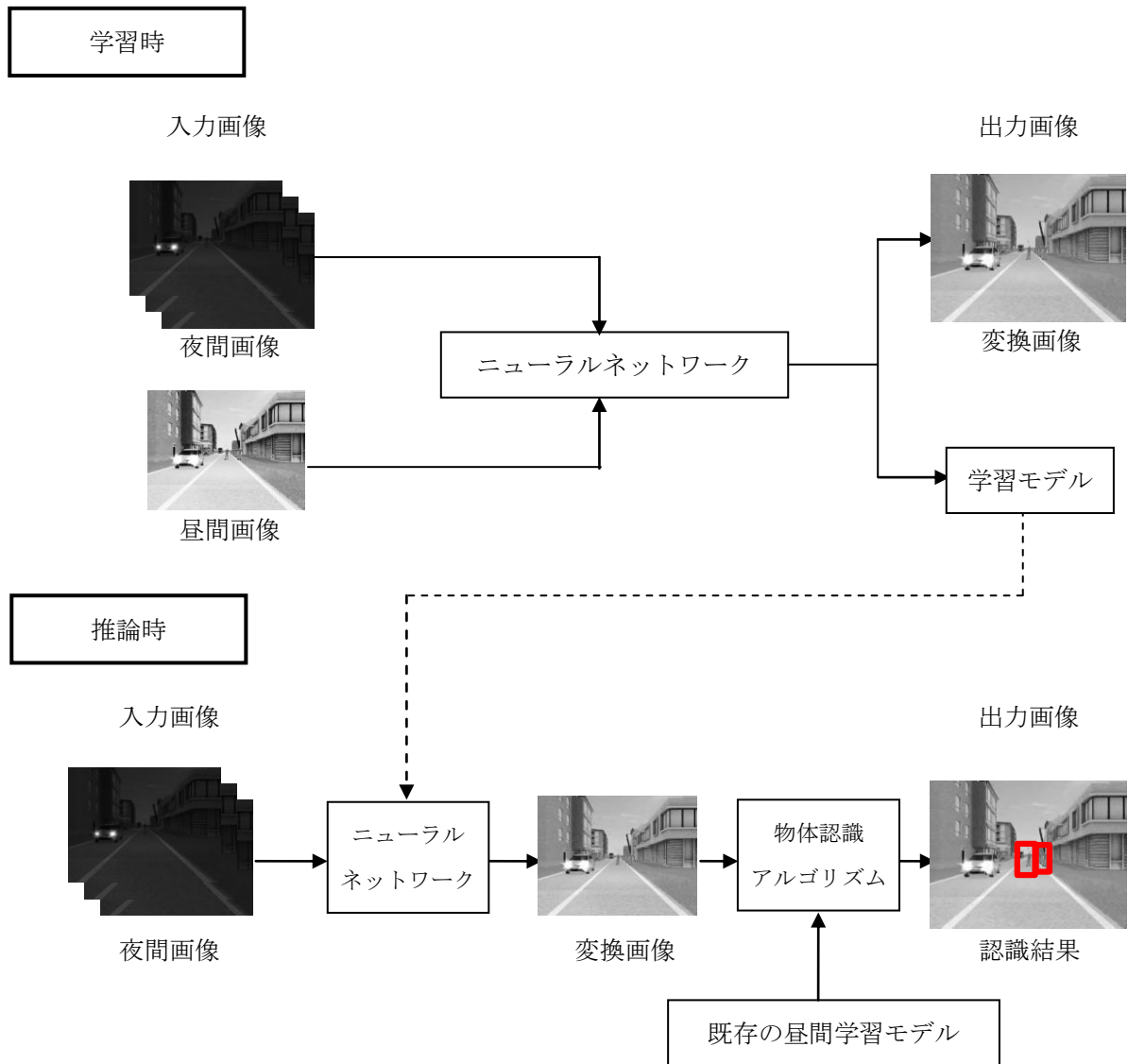


図 3.3：本研究における提案手法の大枠

## 第4章 画像データ，物体認識手法の準備

本研究の実験はシミュレーション画像とカメラ画像に対して行う。本章では，実験に使用するシミュレーション画像とカメラ画像について記載する。

### 4.1 シミュレーション画像の収集

本節では，実験にて使用したシミュレーション画像について述べる。シミュレーションで使用したソフトは PreScan(v8.5.0) というドライビングシミュレータである。PreScan を作成したのは TASS international で，ドイツに本社を置く Siemens のグループ会社である。学習用のシミュレーション画像は著者が作成した。学習画像に建物や木等のオブジェクトを配置して不自然が無いようにした。シミュレーション用のカメラの設置高さは現在市販されている ADAS カメラと同様にフロントガラスに設置した。詳細は表 4.1 に記す。

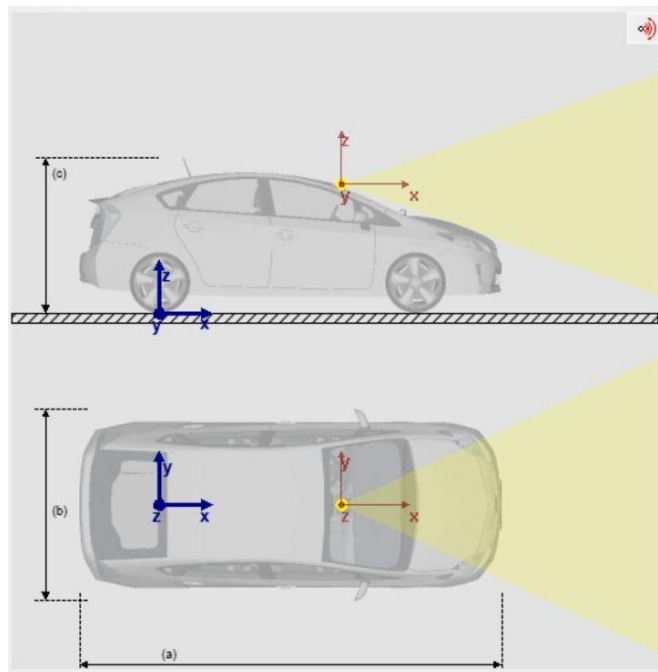
尚，PreScan は海外製ソフトの為，自車両は右側走行とする。またシミュレーション及び実車カメラ画像の撮影において，カメラの水平方向の設置位置について車両の中央にした為，ドライバーの座席の位置の影響はないものとする。



図 4.1: シミュレーションで作成した道路の鳥瞰図

表 4.1: シミュレーションでのカメラ設定

使用車両	トヨタ プリウス
カメラ設置高さ	1.3[m]
カメラ設置水平位置	車両中央 (左端から 872 mm, 右端からも同様)
撮影解像度	960×720[pix]
撮影フレームレート	20[fps]



2020 copyright@siemens

図 4.2: シミュレーションでのカメラ設置位置



歩行者 1



歩行者 2



歩行者 3

図 4.3: シミュレーションで撮影した歩行者例



## 4.2 カメラ画像の取集

本節では、実験に使用したカメラ画像について述べる。カメラ画像の撮影は実際の車両にカメラを取り付けて行った。カメラの取付高さはシミュレーション画像取得時と同様にフロントガラス、高さ 1.3m に設定した。撮影後、動画データを画像ファイルに分割し演算用 PC に移動、学習を行った。

本研究で使用するカメラについて下表に記す。

表 42: カメラ仕様

イメージャ	SENTEC 製 STC-MCS231U3V
レンズ	タムロン製 M118FM08
固定用治具	TP-JVA
接続ケーブル	NU3MBASU3S

走行道路の状況を下表にまとめる。

表 43: カメラ画像撮影場所

走行日時	2017 年 12 月 20 日, 21 日 17:00~21:00
走行場所	静岡県清水区狐ヶ崎駅周辺及び清水駅周辺
撮影時間	両日とも 4h ずつ
天候	両日とも晴れ
撮影車両	5 人乗り 普通乗用車 トヨタポルテ(コンパクトカー)
カメラ設置高さ	1.3[m]
カメラ設置水平位置	車両中央 (左端から 847 mm, 右端からも同様)
撮影解像度	960×720[pix]
撮影フレームレート	30[fps]

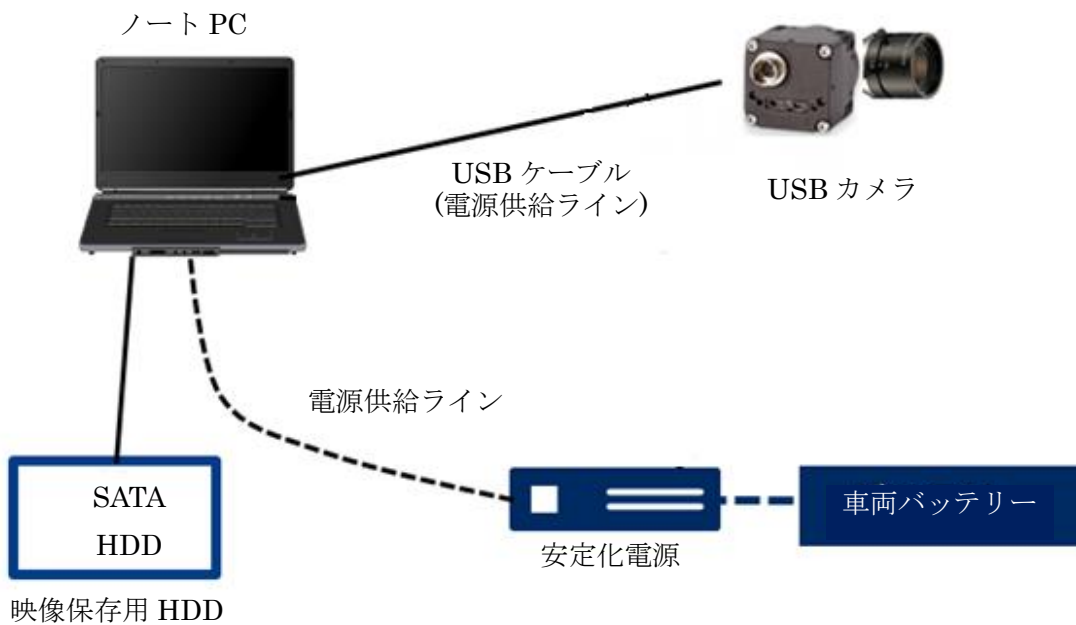


図 4.4: カメラ録画システム構成図



※個人情報保護の為、該当箇所は黒塗り実施

図 4.5: カメラ撮影実験におけるカメラ設置位置



図 4.6: カメラ画像撮影機器

### 4.3 歩行者認識手法の準備

本節では実験にて画像変換後に適用する歩行者認識手法について述べる。本研究で使用する歩行者認識手法は実験室で検討したアルゴリズムや学習データではなく、過去、一般に公開されている物体認識手法を使用する。一般に公開されているアルゴリズムを使用して認識率を評価することで、特殊な歩行者認識手法を使用しなくても夜間歩行者認識ができることを示す。使用する物体認識アルゴリズムは 2 章でも説明した Faster R-CNN と YOLOv3 を使用する。学習データと評価指標については次節で述べる。

#### 4.3.1 学習データ(PascalVOC)

PascalVOC は正式名称 Pattern Analysis Statistical Modelling and Computational Learning で欧州の研究コミュニティが主催するコンペティションで 2005 年から 2012 年まで開催されていた。PascalVOC のデータセットが生まれた経緯として、それまでの Caltech-101(2004 年にカリフォルニア工科大学が公開)や Caltech-256(カリフォルニア工科大学が 2007 年に公開)等のデータセットは色やテクスチャーのバリエーションには富んでいるものの、物体の大きさや向きを揃えている為、理想的な配置となり、現実の画像認識とは難易度が異なるという課題があった。この課題を解決する為に、Pascal が複数の物体が様々な条件の下で撮影された現実的なデータセットとして公開した。PascalVOC で公開された各年のデータセットは画像共有サイトである Flickr の画像を利用している。コンペティションのタスクとして、物体認識の他に動作識別、セマンティックセグメンテーション、人体部位のレイアウト推定等が年ごと実施された。2012 年以降、DeepLearning 向けのデータセットの公開が増え、PascalVOC ではクラス数、データ数の不足からコンペティションが終了した。本研究では、歩行者認識に注目している為、データ数、クラス数が適量で準備しやすい PascalVOC を使用することにした。

PascalVOC では使用するクラスは 20 個であり、以下の通りである。画像中に存在する物体毎に物体を矩形で囲うようにバウンディングボックスの位置座標とクラスがアノテーションとして配布される。バウンディングボックスの位置座標は左上の x 座標と y 座標、バウンディングボックスの横幅と縦幅が格納される。

表 4.4: 既存の歩行者認識モデルの作成で使用したデータ(PascalVOC2012)

インデックス	クラス名	train		validation	
		画像数	物体数	画像数	物体数
1	飛行機	327	432	343	433
2	自転車	268	353	284	358
3	鳥	395	560	370	559
4	ボート	260	426	248	424
5	ビン	365	629	341	630

6	バス	213	292	208	301
7	自動車	590	1013	571	1004
8	猫	539	605	541	612
9	椅子	566	1178	553	1176
10	牛	151	290	152	298
11	テーブル	269	304	269	305
12	犬	632	756	654	759
13	馬	237	350	245	360
14	バイク	265	357	261	356
15	人	1994	4194	2093	4372
16	鉢植え	269	484	258	489
17	羊	171	400	154	413
18	ソファー	257	281	250	285
19	電車	273	313	271	315
20	テレビ	290	392	285	392
	合計	5717	13609	5823	13841

#### 4.3.2 評価指標

歩行者認識にて認識した結果を評価する指標として Precision と Recall を使用する.

$$Precision [\%] = \frac{TP}{TP + FP}$$

$$Recall [\%] = \frac{TP}{TP + FN}$$

TP は各手法が歩行者と推測したデータのうち、実際に歩行者である数、FP は歩行者と推測して実際は歩行者以外である数、FN は歩行者以外と推測して実際は歩行者である数を表す.

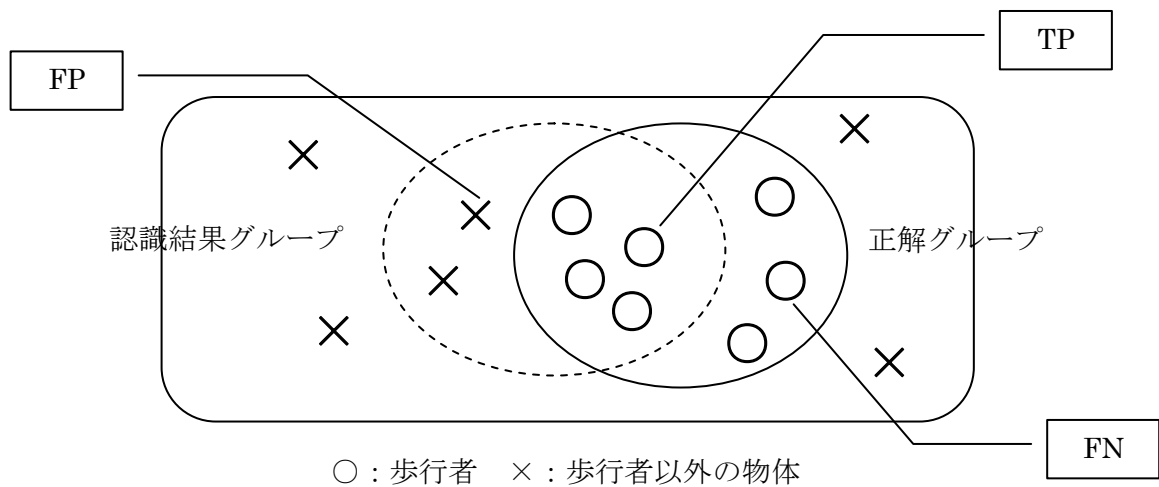


図 4.7 : Precision, Recall 概念図

#### 4.4 認識限界距離

本実験で認識する歩行者の限界距離を決定する為に、最も外乱が少なく、認識しやすいシミュレーション画像に対し、物体認識処理を昼間の歩行者画像に適用した。シミュレーションソフト PreScan で直線道路に歩行者を配置する。自車両を 10km/h で進行させて歩行者を認識した地点を計測した。シミュレーション上の歩行者のテクスチャーによる認識の違いを確認する為、5人の歩行者を配置して計測した。以下にシミュレーションでの配置の概要図及び、実際に PreScan で作成したマップでの鳥瞰図を記す。

上記、5人の歩行者に認識速度は速くないが認識精度が高い Faster RCNN を適用して認識した。

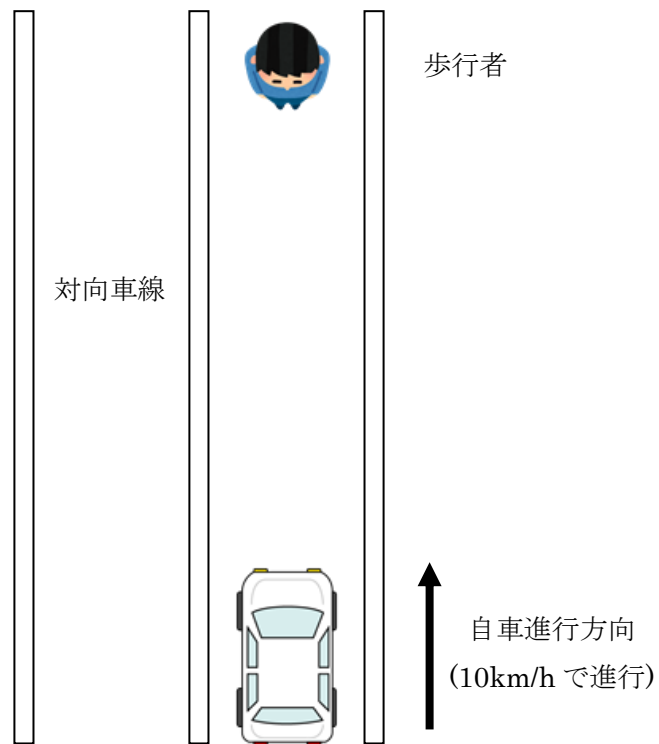


図 4.8：認識限界距離の計測の為の昼間における実験(模式図)

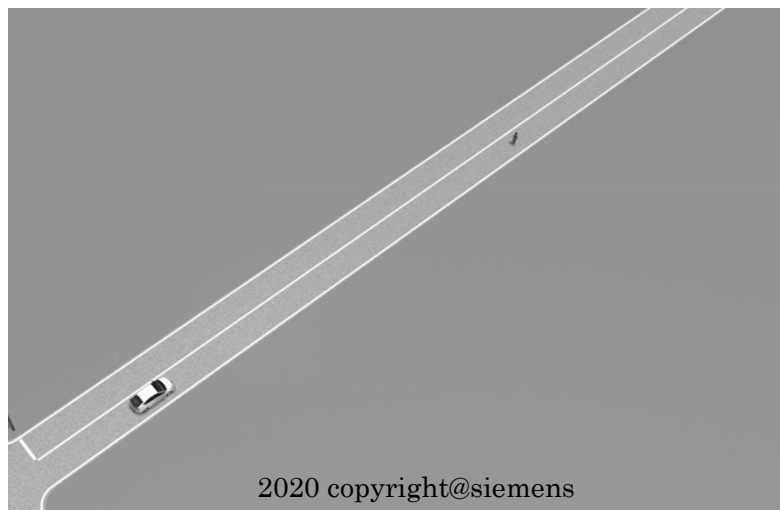


図 4.9：認識限界距離の計測の為の昼間における実験(鳥瞰図)



図 4.10：認識限界距離の計測の為の昼間における実験(側面図)



図 4.11：認識限界距離の計測の為の昼間における実験(正面図)



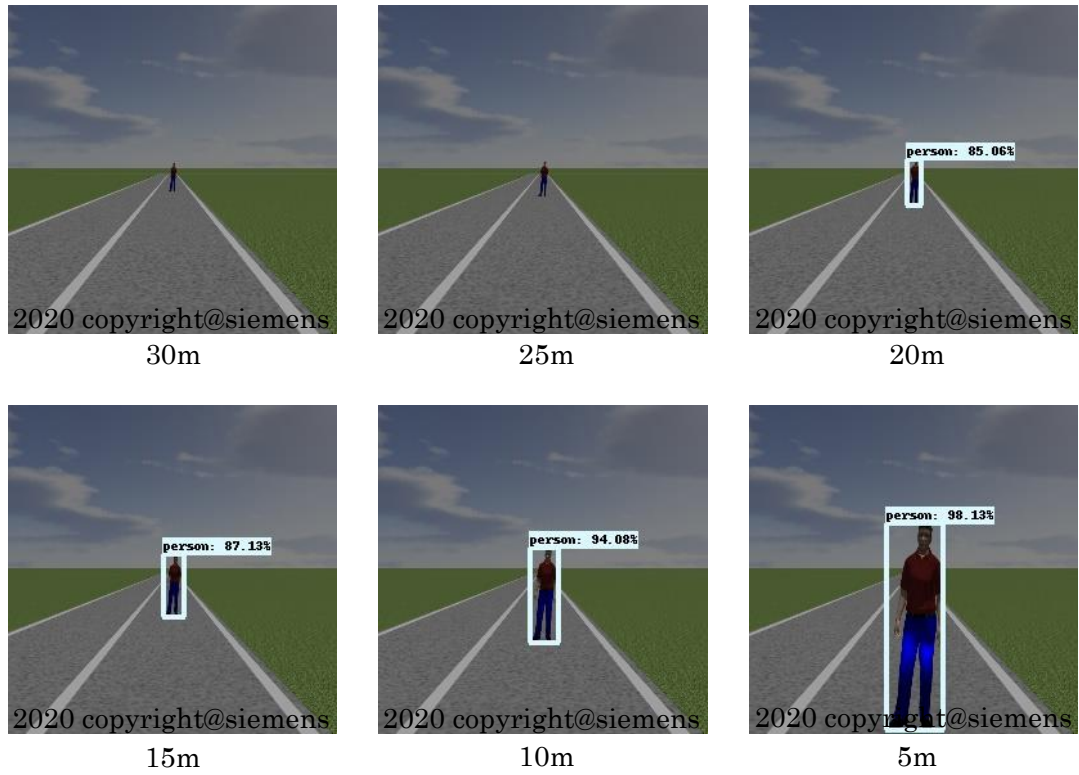


図 4.12：昼間環境での歩行者認識(歩行者 1)

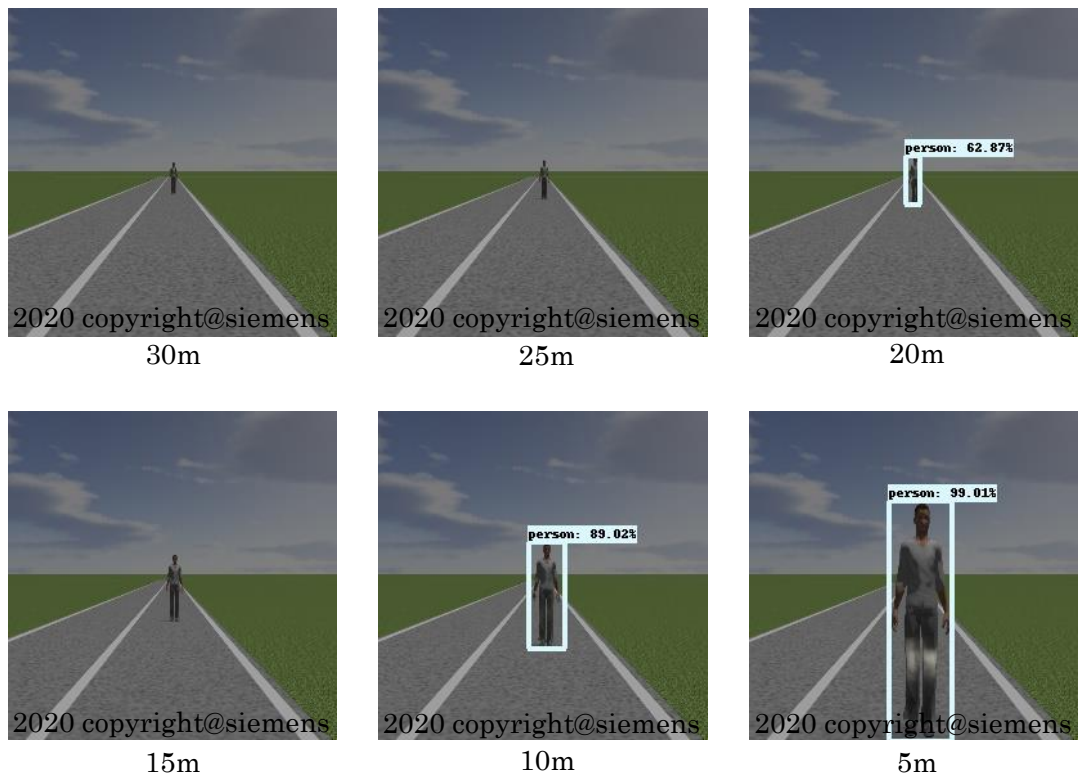


図 4.13：昼間環境での歩行者認識(歩行者 2)

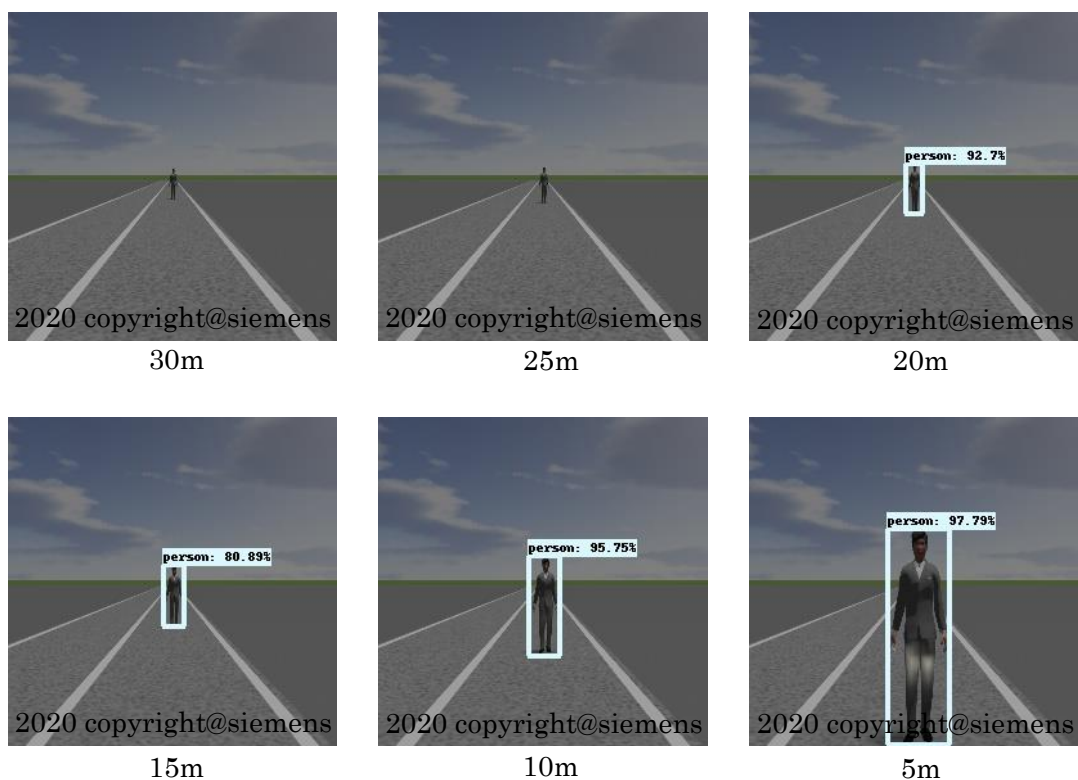


図 4.14: 昼間環境での歩行者認識(歩行者 3)

歩行者の認識率は各距離において以下の式を用いて算出した。

$$\text{Accuracy} = \frac{a}{N}$$

$a$ は各距離で歩行者と認識できたフレーム数,  $N$ は各距離の歩行者の数を表す。

その結果, 自転車から 20m(歩行者の大きさが横: 8pix, 縦: 36pix)先までの歩行者を安定して認識することを確認した. このことから今回の認識対象距離を 20m とし, 20m 以遠の歩行者は認識対象外とする.

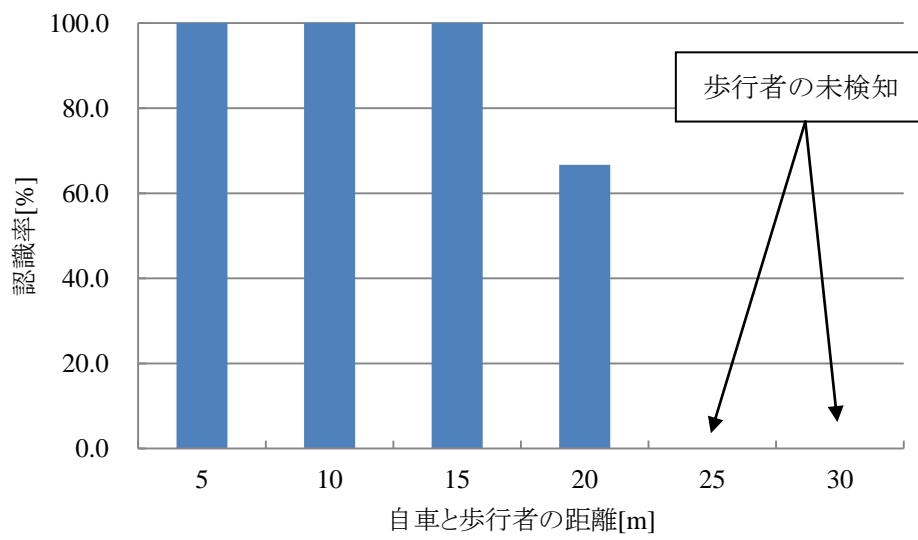


図 4.15：昼間環境での歩行者認識率

## 第5章 夜間直線道路における画像変換手法の検討

本章では、直線道路において夜間画像を昼間画像に変換する手法について述べる。5.1 節では、提案手法の概要を述べ、5.2 節でニューラルネットワークの中間層について詳細を述べる。その後、5.3 節にて有効性を確認する為の実験について述べる。

### 5.1 提案手法の概要

本節では、夜間直線道路における画質改善手法について述べる。図に提案手法の概要フローチャートを示す。

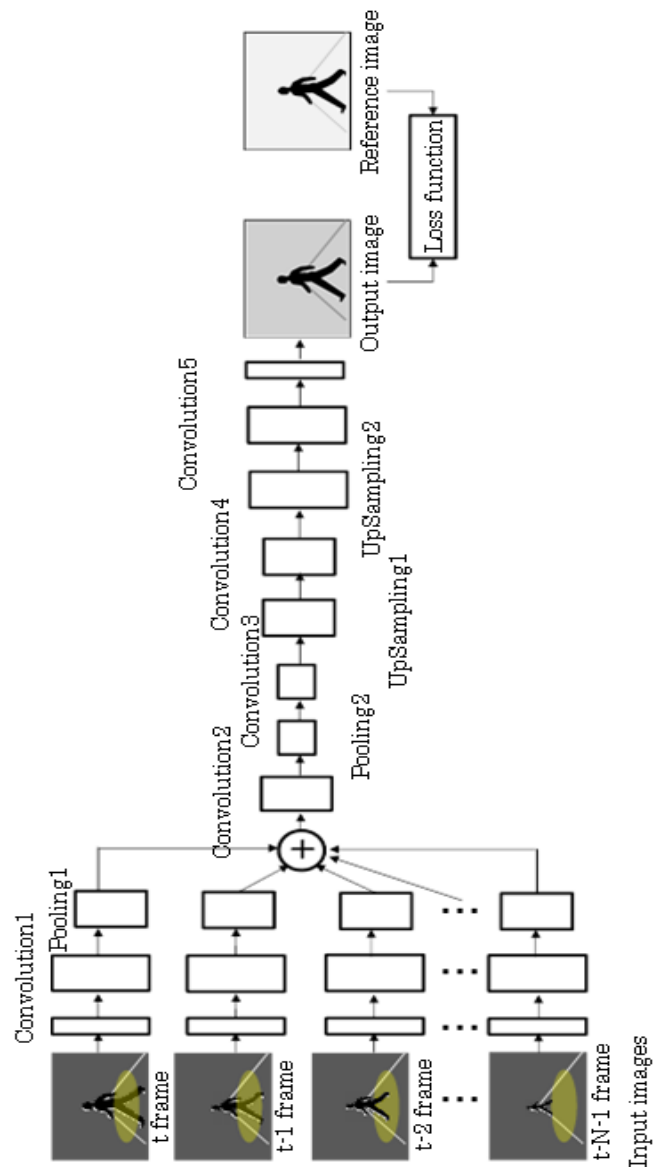


図 5.1: 直線道路における画像変換ニューラルネットワークの概要図

## 5.2 提案手法の詳細

本節では、提案手法のアプローチについて記載する。一般的に暗い画像のコントラストを改善する手法として、カメラのレンズの設定値を調整できる場合、露光時間を長くして明るい画像の撮影が可能である。同様の効果を画像撮影後のソフトウェア上でも可能であり、撮影した連続フレームを足し合わせることで明るくすることができる。しかし、本研究の対象では、静止状態ではなく、カメラ及び対象物が動く車載環境である為、撮影した連続フレームを足し合わせるとフレーム間の物体の移動に伴うブレが発生する。そこで本節では、DeepLearningのConvolutionやPoolingによる位置変化や大きさ変化への頑健性に着目してネットワーク上で連続フレームの画素値の足し合わせを行うこととした。

提案手法は入力画像として連続するNフレームの夜間画像を入れ、出力画像を1フレーム作成するConvolutionが5層、Poolingが2層、UpSamplingが2層の計9層のNeural Networkとなる。Pooling1まで入力画像のNフレームを分けて処理し、Pooling1以降、Nフレームの値を統合し、ConvolutionとPoolingを繰り返し、特徴量を抽出する。そして変換画像を作成し、昼間の正解画像と比較して損失関数にて誤差計算を行う。正解画像は連続する夜間画像の内、最新フレームと対応した昼間画像を準備する。これを実現する為に本研究では、学習画像にシミュレーション画像を使用する。シミュレーションで同期の取れた昼夜の画像を作成し提案手法に投入し、学習を実現する。

又、一般的に暗い画像のコントラストを改善する手法として1フレームだけでなく複数のフレームを畳み込む手法がある。その為、提案手法でも入力画像に連続する複数フレームを活用する。

### 〈2・2〉 複数フレームの統合処理

提案手法では、自車の車速が学習外の画像になっても画像変換できるように連続画像の計算値を処理の途中で統合する。連続する夜間画像を1層目でConvolutionを行い、2層目でPoolingを行う。その後、Nフレーム分の畳み込み結果を統合する。ここで自車の速度の違い(フレーム間の位置の違い)に対応する為に、本手法ではConvolution1の結果を加算する場所について検討しPooling1に着目した。一般的にPoolingは微小な位置ずれや大きさの違いに対応することが出来る。その性質を利用して、Convolution1で得られたNフレームの出力値を2層目のPooling1に投入し、Pooling1で得られたNフレーム分の計算値を加算する。その結果、学習で使用したデータの自車速度に依存せず、速度変化に対してロバスト性を向上させる。式(1)、(2)にConvolution1及びPooling1を表す式を記し、式(3)にPooling1後のNフレームを統合する計算式を示す。

$$o_{ij}^{(k)} = \sum_{s=0}^{m-1} \sum_{t=0}^{n-1} \omega_{st}^{(k)} x_{(i+s)(j+t)}$$

Oは畳み込みフィルタを通した後の出力データを表し、i,jは入力画像の注目画素を

表す.  $m, n$  は畳み込みフィルタのサイズ,  $k$  はフィルタの番号,  $\omega$  は重み,  $x$  は  $N$  フレームの入力画像の内のある 1 フレームの夜間画像を表す.

$$u_{ij}^{(k)} = \max_{p, q \in P_{i,j}} (o_{pq}^{(k)})$$

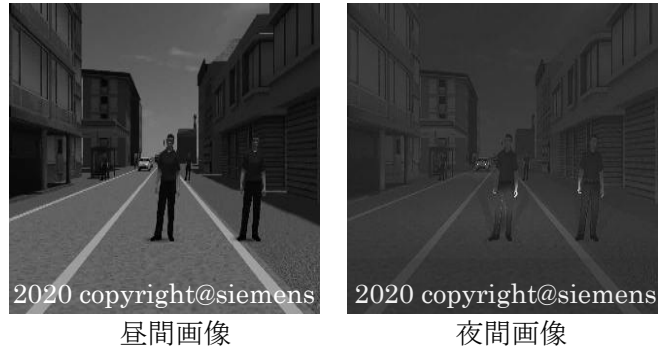
$u$  は Pooling1 後の出力データを表す.  $P_{ij}$  は注目画素を中心に  $H \times H$  に取った領域である.  $p, q$  は  $P_{ij}$  内の各画素を表す.  $O$  は式(1)の畳み込み結果を表し,  $k$  はフィルタ番号を示す.

$$z_{ij}^{(k)} = \sum_{f=0}^N u_{ijf}^{(k)}$$

$z$  は統合後の値を示し,  $f$  は入力フレーム数を表しており, 本手法では  $N$  フレームの入力があることを示す.  $i, j$  は Pooling1 後の注目画素を表す.

表 5.1: 直線道路における画像変換ネットワークの構造

layer	type	Output size
1	Convolution1	$256 \times 256 \times 256 \times 3$
2	Pooling1	$128 \times 128 \times 256 \times 3$
3	Convolution2	$128 \times 128 \times 128$
4	Pooling2	$64 \times 64 \times 128$
5	Convolution3	$64 \times 64 \times 128$
6	UpSampling1	$128 \times 128 \times 128$
7	Convolution4	$128 \times 128 \times 256$
8	UpSampling2	$256 \times 256 \times 256$
9	Convolution5	$256 \times 256 \times 1$



(a) シーン 1



(b) シーン 2



(c) シーン 3



(d) シーン 4

図 5.2: シミュレーションによる昼夜の画像例

## 5.3 実験

本手法の有効性を確認する為に、5.3.1 節と 5.3.2 節では以下の 2 つの実験を行った。

### 5.3.1 学習データテストデータで同等の車速の場合

#### (1) 画像変換モデルの学習

5.3.1 節では、学習画像に 60km/h で走行するシミュレーション画像、テスト画像には学習外の 60km/h のシミュレーション画像と実際のカメラ画像の 2 パターンを使用して歩行者認識率について検証した。5.3.2 節では、自車の車速(自車速)を変更した時の画像変換について記述する。5.3.1 節、5.3.2 節共に入力フレーム数はコントラストの改善性能と遅延時間を考慮して、3 フレームとした。

シミュレーション画像のカメラの設定値を表 5.2 に記す。今回画像のカラー化に伴い 1 枚当たりのデータ量が 3 倍に増えることから学習時のネットワークのサイズが大きくなりメモリ不足が発生した。その為、今回は画像をグレースケール画像にし、データ量を減らすことで対応した。また同様の理由で画像サイズを 256×256pix とした。学習に使用した歩行者シーンを表 5.3 に記す。学習データは 4.1 節でも記述したドライビングシミュレータである PreScan を使用して収集した。PreScan で作成した道路マップ上に歩行者を配置した。この時、配置した歩行者は表 5.3 の通り、最低身長が約 160 cm～180 cmと実在する歩行者と同等になるようにした。表 5.4 に学習時のハイパーパラメータを設定する。Epoch 数は学習が進みある程度収束する 300 とした。Batch サイズ、Optimizer(最適化手法)、Activation function(活性化関数)は高い精度が報告されているものを使用する。

表 5.2: カメラ設定値

Camera location	Top of windshield
Camera height	150[cm]
Channel	1 [channel]
image size	256×256[pix]

表 5.3: 学習データ(シミュレーション画像)

Number of pedestrian scene	45[people]
Number of images	7048[frame]
Maximum height	180[cm]
Minimum height	160[cm]



表 5.4: 学習条件

Epoch	300
Batch	32
Optimizer	Adam <sup>[29]</sup>
Activation function	Relu <sup>[30]</sup>
Loss function	MSE

Adam は Optimizer(最適化手法)の 1 つで, 他の Optimizer として SGD, AdaGrad, Adadelta, RMSprop がある. Adam は AdaGrad と RMSprop を合わせたものであり, ニューラルネットワークの学習において良く使われる手法である. AdaGrad は学習率を全てのパラメータで一律に使用するという処理を改善して, 勾配の二乗で減らし, パラメータ毎に学習率を更新する手法を提案している. また RMSprop は直近の勾配の値を使用して過去の平均的な勾配と比較して学習率を調整する.

Relu()は Activation function(活性化関数)の 1 つであり, DeepLearning で発生する勾配消失問題を解決する為に使用される. 活性化関数は他にシグモイド関数, Tanh 関数, ステップ関数がある. 勾配消失問題とは, ニューラルネットワークの層が深くなるにつれて学習の為に勾配が消えてしまうことである. 初めに考案されたシグモイド関数は微分係数が 0.0~0.25 であり, シグモイド関数を使用すればするほど勾配値が小さくなってしまふ. Relu の微分係数は 0.0 か 1.0 である為, 勾配が小さくなりにくい.

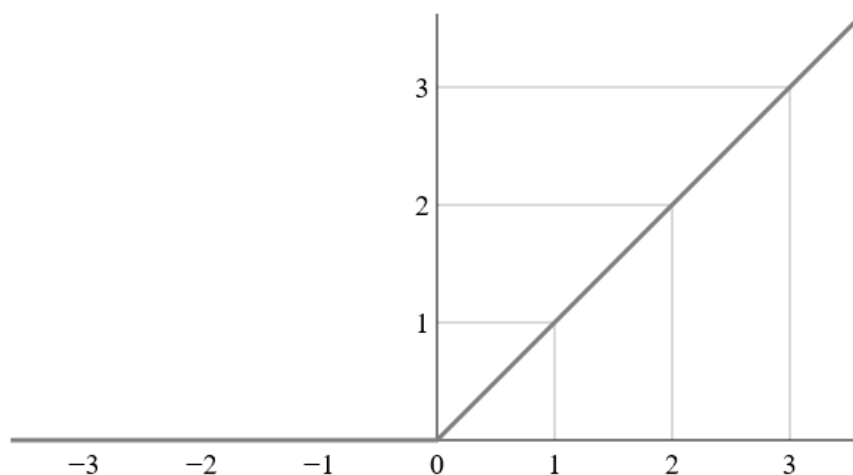


図 5.3: Relu 関数

〈2〉 シミュレーション画像でのテスト

本節では、評価画像としてシミュレーション画像を使用したときの認識結果を記す。そして、提案手法による画像変換の有効性を確認する為に、以下の変換手法と比較する。下記変換結果に対し、3種類の物体認識手法を適用する。

表 5.5: 比較手法のインデックスと概要

No	比較画像の概要
-	変換前の入力画像(Original image)
①	提案手法に入力する入力画像を1フレームにした時
②	画像のコントラスト改善手法として知られている Zuiderveld[31]の CLAHE(Contrast Limited Adaptive Histogram Equalization)による変換
③	Iizuka ら[24]の画像変換ネットワーク
④	Anoosheh ら[26]の画像変換ネットワークから得られた出力画像
⑤	Huang ら[27]の手法を用いて、昼間画像から夜間画像へ変換

表 5.6: テストデータ(シミュレーション画像)

Number of pedestrian scene	15[people]
Number of images	1050[frame]
Maximum height	180[cm]
Minimum height	160[cm]

### 〈3〉 実験結果

夜間画像における歩行者認識の流れを図 5.4 に示す。提案手法における歩行者認識では、変換前の夜間画像ではコントラストが低く歩行者の認識が困難であるが(図 5.4(a)), 画像変換により背景、歩行者の輝度の増加及び、白飛びが少なく背景と歩行者のエッジが見やすくなることで歩行者認識が容易となる(図 5.4(b)).

夜間歩行者認識した結果を図 5.5 に記す。歩行者認識手法は 2.2 節で述べた Faster RCNN, YOLOv3, M2det を用いた。その結果, 提案手法による画像変換結果が最も歩行者を認識率が高いことを確認した。表 5.7 より提案手法による歩行者認識率の平均は, Precision=92.4[%], Recall=84.3[%]であった。

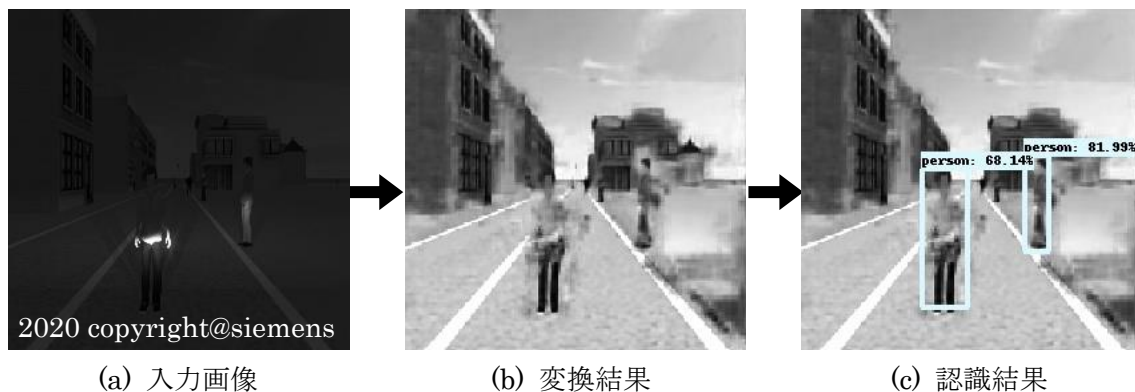
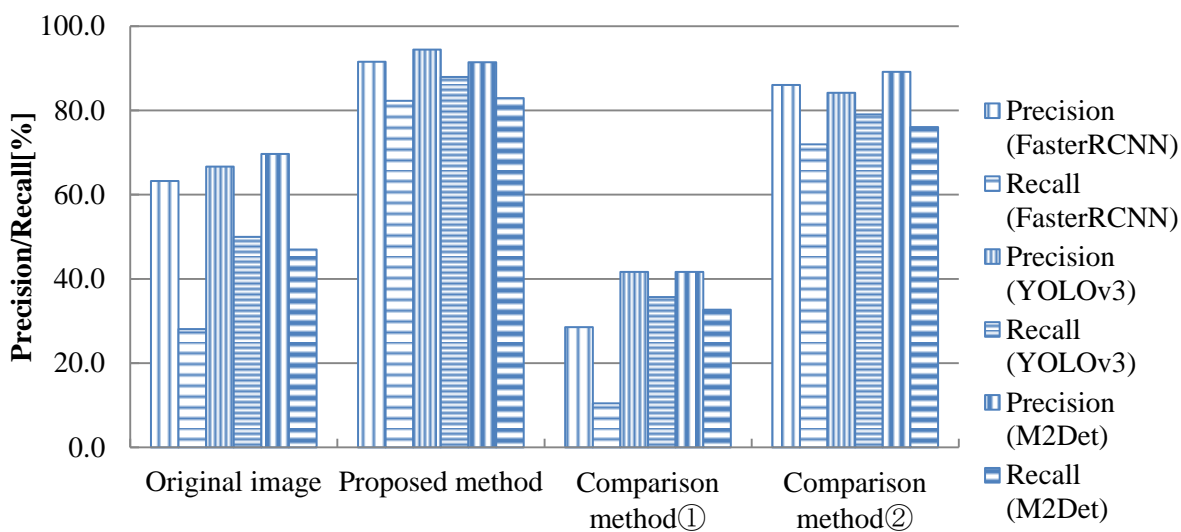
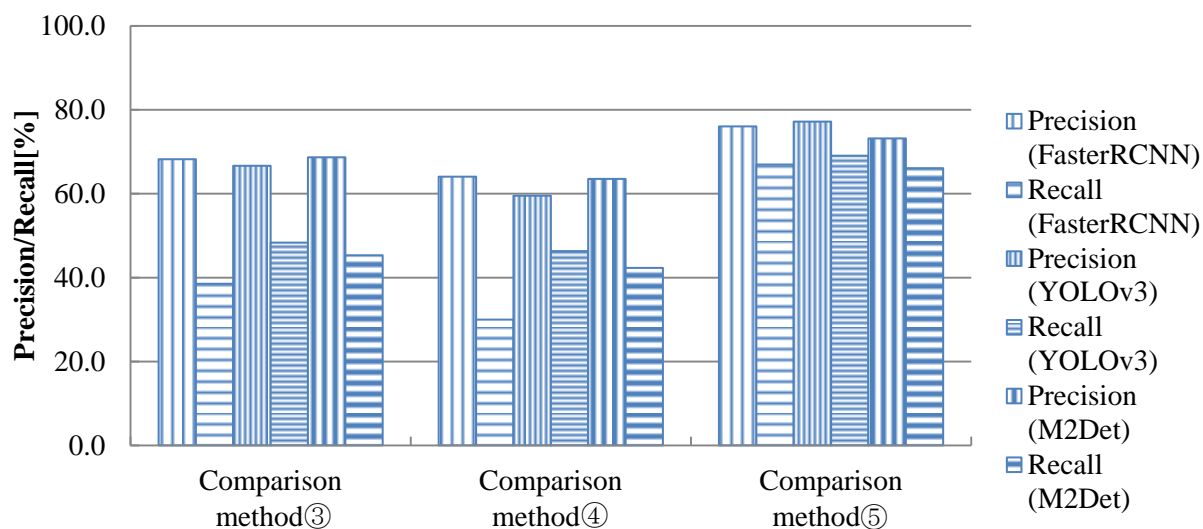


図 5.4：夜間歩行者認識の流れ



(a) The test image is simulation image.



(b) The test image is simulation image.

図 5.5：直線道路でのシミュレーション画像を使用した時の歩行者認識結果

表 5.7：直線道路でのシミュレーション画像を使用した時の歩行者認識結果

画像変換手法	Precision			Recall		
	Faster RCNN	YoloV3	M2Det	Faster RCNN	YoloV3	M2Det
入力画像	63.2	66.7	69.7	28.1	50.0	47.0
提案手法	91.5	94.4	91.4	82.3	87.9	82.9
提案手法に入力する画像を1フレームにした時	28.6	41.7	41.7	10.4	35.7	32.7
CLAHE[31]	86.0	84.2	89.2	72.0	79.1	76.1
Iizuka[24]らの画像変換ネットワーク	68.2	66.7	68.7	38.5	48.3	45.3
Anooshehら[26]の夜間画像から昼間画像への変換	64.1	59.5	63.5	30.0	46.3	42.3
Huangら[27]の手法を用いて、昼間画像から夜間画像	76.0	77.2	73.2	67.0	69.1	66.1

原画像の夜間画像に対し、画像変換及び物体認識(Faster RCNN)した結果例を図 5.6 に記す。(a)に原画像に対して Faster RCNN した結果を記す。(b)~(f)には夜間画像に対し、それぞれ提案手法及び比較手法を適用し、Faster RCNN を適用した認識結果例を記す。

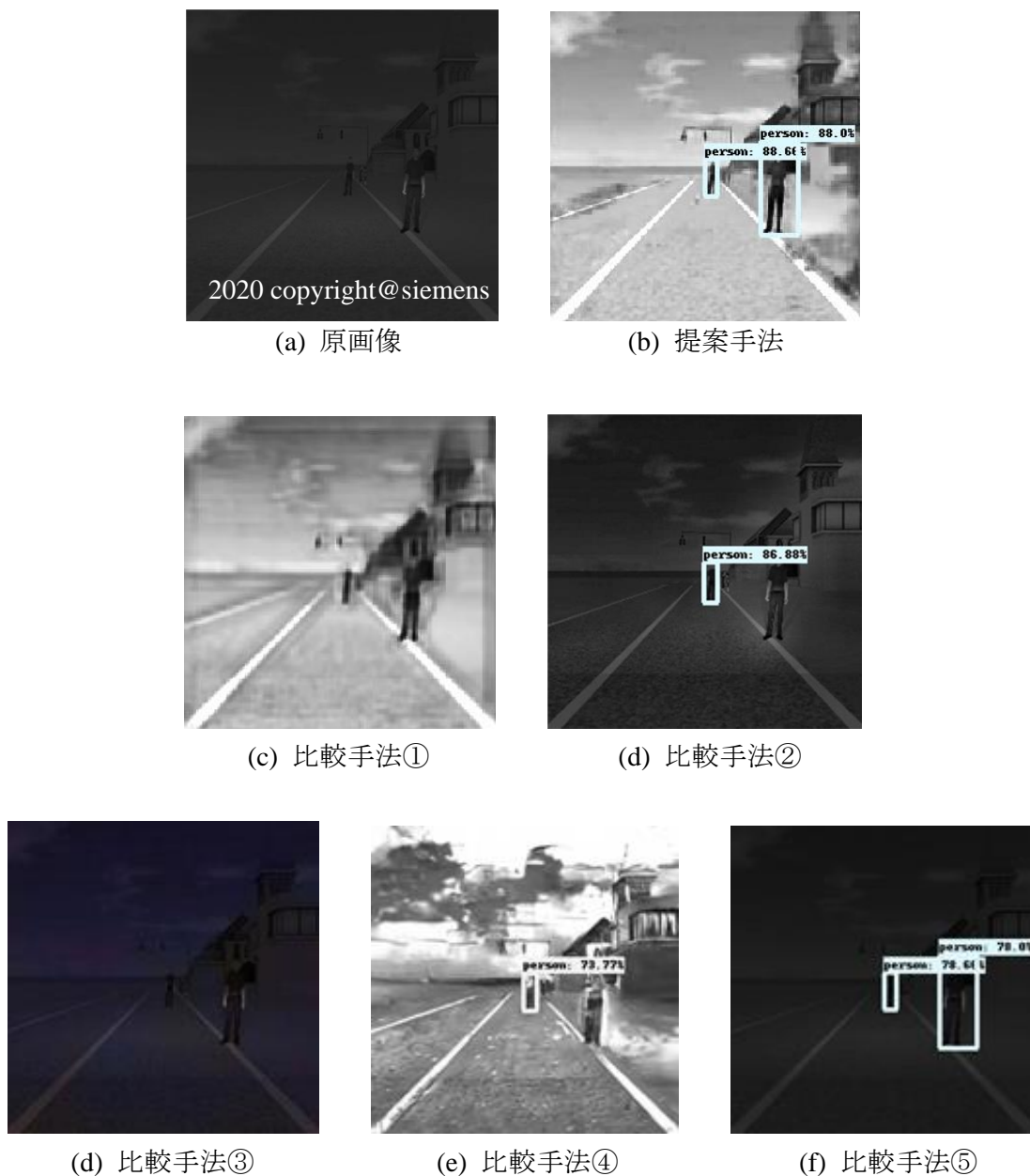


図 5.6: 直線道路における画像変換結果例(シミュレーション画像)

#### 〈4〉 カメラ画像でのテスト

5.3.1 節 〈2〉 ではテスト画像にシミュレーション画像を使用して評価した。本節では 4.2 節で述べた USB カメラを使用して撮影した画像を入力し画像変換及び歩行者認識を行う。テスト画像のカメラ画像には実際の車両のフロントガラスにカメラを設置して撮影を行った。カメラの撮影条件は表 5.8 に記し、撮影した道路環境は前節表 4.2, 表 4.3 に記す。テスト画像の歩行者はロバスト性を検証する為、一般道路上を通行する歩行者を撮影した。

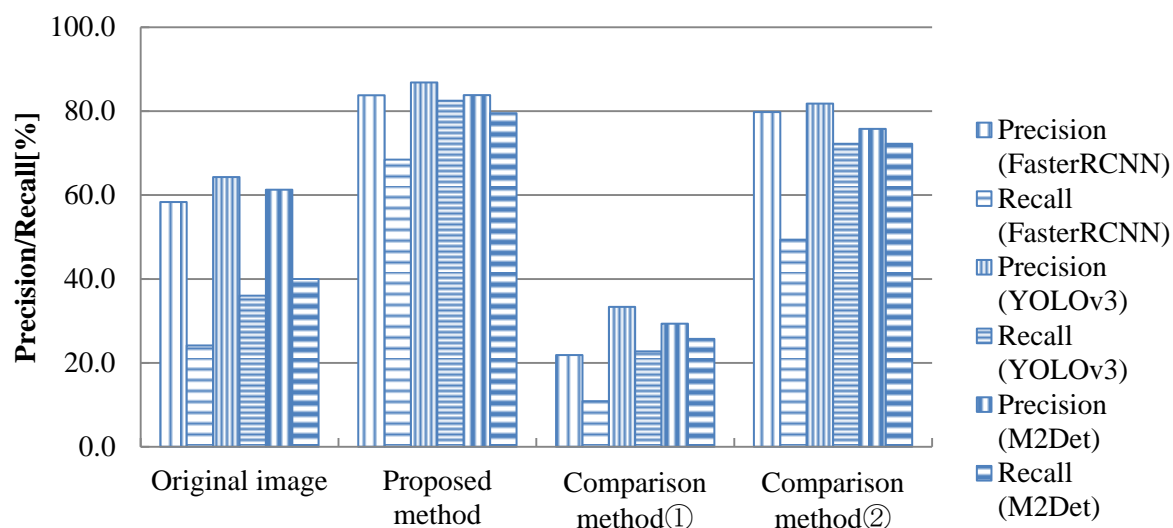
撮影した夜間画像を 5.3.1 節 〈2〉 と同様に提案手法と物体認識手法に入力し、歩行者の認識率を算出する。

表 5.8: テストデータ(カメラ画像)

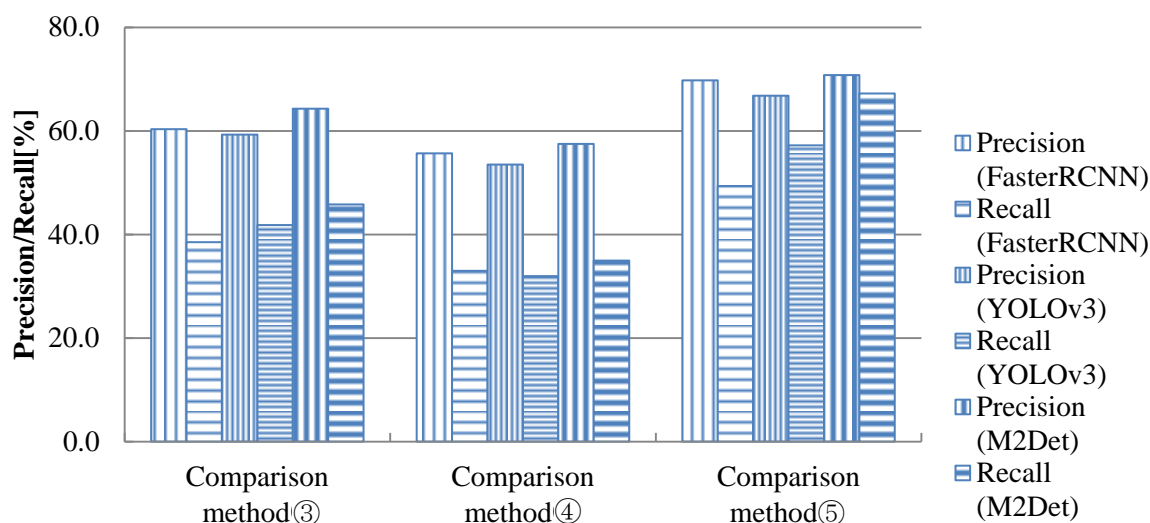
テストフレーム数	1500[フレーム]
歩行者人数	15[人]

#### 〈5〉 実験結果

夜間歩行者認識した結果を図 5.7 に記す。歩行者認識手法は 2.2 節で述べた Faster RCNN, YOLOv3, M2det を用いた。その結果、提案手法による画像変換結果が最も歩行者を認識率が高いことを確認した。表 5.9 より提案手法による歩行者認識率の平均は、Precision=84.4[%], Recall=76.8[%]であった。



(a) The test image is camera image.



(b) The test image is camera image.

図 5.7: 直線道路においてカメラ画像を使用した時の歩行者認識結果

提案手法による歩行者認識率が向上した理由について考察する. 5.3.1 節<1>のテスト画像にシミュレーション画像を使用した場合には, 学習・テスト画像でシミュレーション画像を使用している. 提案手法の **Pooling** 後の値を統合することで複数フレームの値を足して画質の向上ができたと考える. **Pooling** は画像の解像度を小さくして, 物体の位置ずれや大きさの違いに頑健にするメリットがある. 今回, このメリットを利用して過去フレームと現在フレームの歩行者の位置ずれや大きさのずれの影響を小さくして各フレームの画素値を加算が上手くいったと考える. 同様の理由で 5.3.1<4>テスト画像にカメラ画像を使用したときにも言え, 直線道路において **Pooling** 後の値を利用することで, 微小なずれを吸収して複数フレーム内の画素値の統合ができ, 画質改善できたと考える.

表 5.9: テスト画像にカメラ画像を使用した時の歩行者認識率

画像変換手法	Precision			Recall		
	Faster RCNN	YoloV3	M2Det	Faster RCNN	YoloV3	M2Det
入力画像	58.3	64.3	61.3	24.1	36.0	40.0
提案手法	82.7	86.8	83.8	68.4	82.5	79.5
提案手法に入力する 画像を1フレームに した時	21.9	33.3	29.3	10.9	22.7	25.7
CLAHE[31]	79.8	81.8	75.8	49.4	72.2	72.2
Iizuka ら[24]の画像 変換ネットワーク	60.3	59.3	64.3	38.5	41.8	45.8
Anoosheh ら[26]の夜 間画像から昼間画像 への変換	55.7	53.5	57.5	33.0	31.9	34.9
Huang ら[27]の手法 を用いて, 昼間画像か ら夜間画像	69.8	66.8	70.8	49.4	57.2	67.2



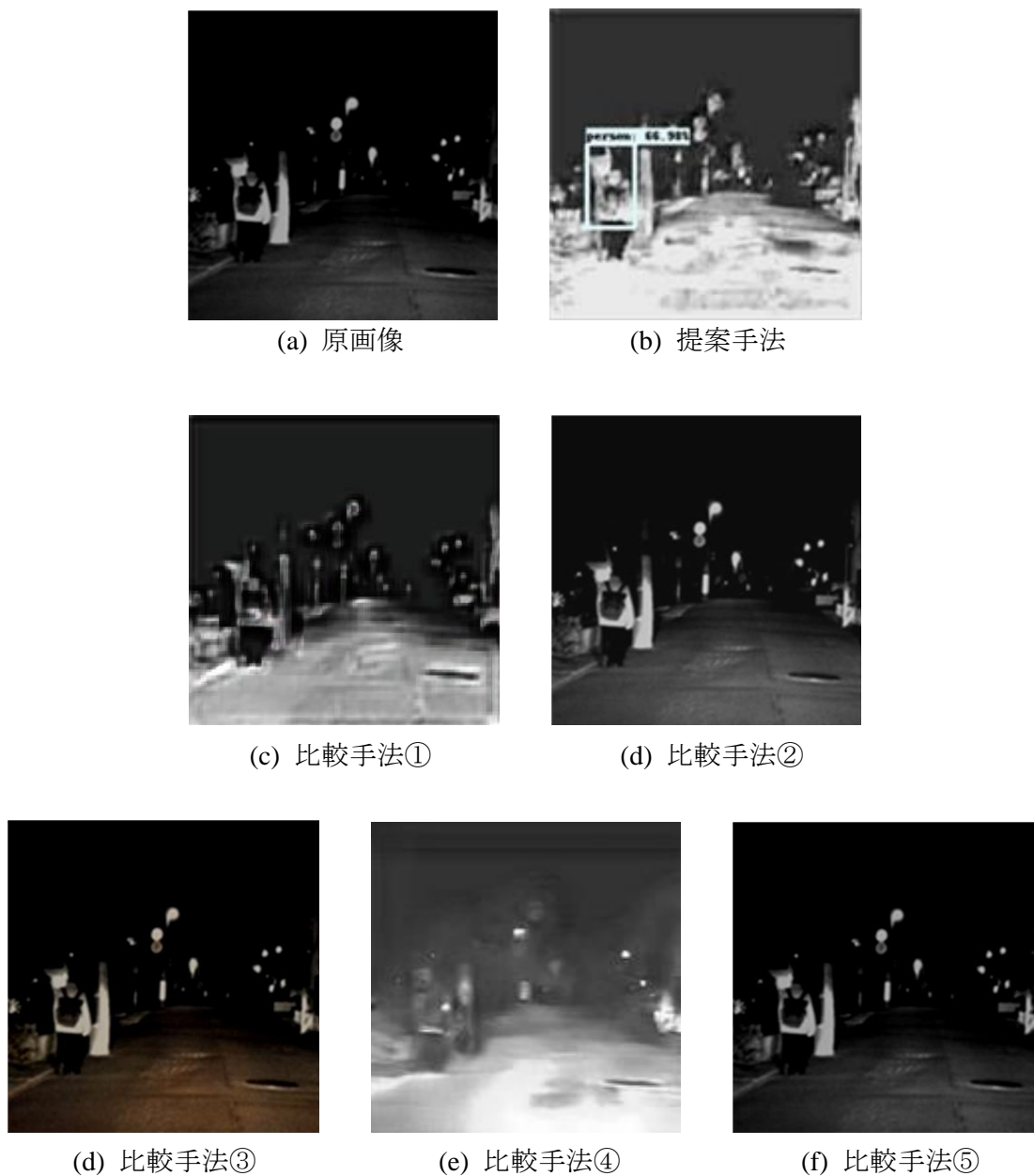


図 5.8: 直線道路における画像変換結果例(カメラ画像)

画像変換及び物体認識(Faster RCNN)した結果例を図 5.8 に記す. (a)に原画像に対して Faster RCNN した結果を記す. (b)~(f)には夜間画像に対し, それぞれ提案手法及び比較手法を適用し, Faster RCNN を適用した認識結果例を記す.

### 5.3.2 自車速の変化に対する検討

#### 〈1〉 5.3.1 節との実験条件違い

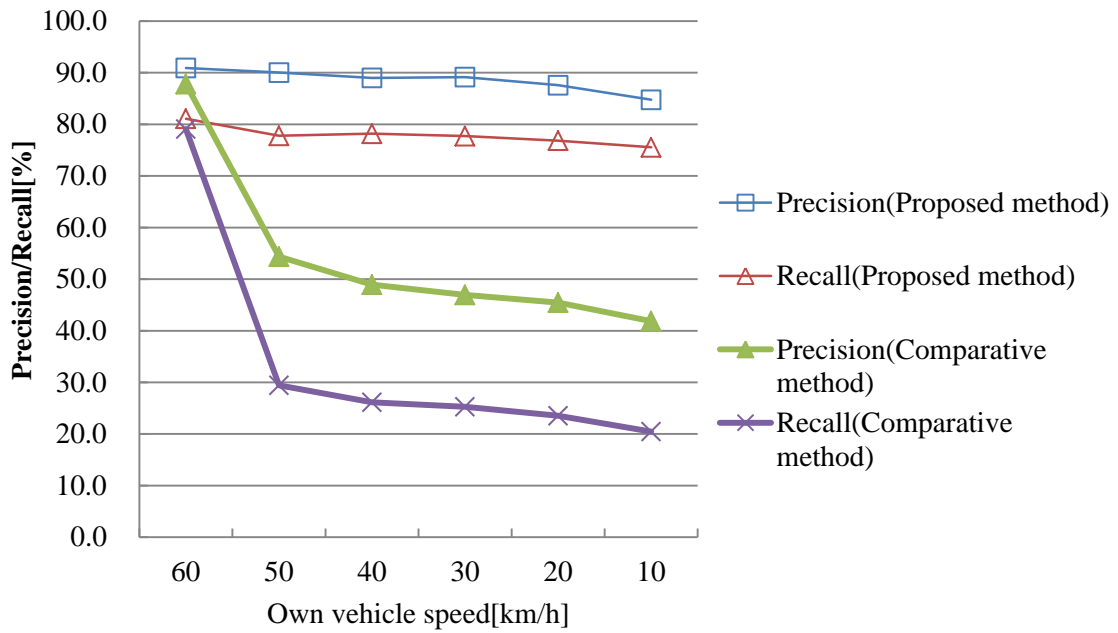
本節では、自車速を変更した時の画像変換について実験を行い、物体認識率がどのように変化するかを確認する。5.3.1 節の基礎実験にて、学習データとテストデータの自車速を同じ車速にした時では、自車速が同じである為、フレーム間の物体の移動量がほぼ同等になり、画像変換した結果、夜間画像を昼間画像に変換することができた。その結果、歩行者の認識率の向上を確認することができた。本節では、テスト画像で使用する自車速を学習で使用した 60km/h 以外の自車速に変更した時に画像変換を行い、その後歩行者認識率を算出する。5.3.1 節と本節の共通部分、変更部分を下表に記す。

表 5.10: 実験条件の共通部分と変更部分

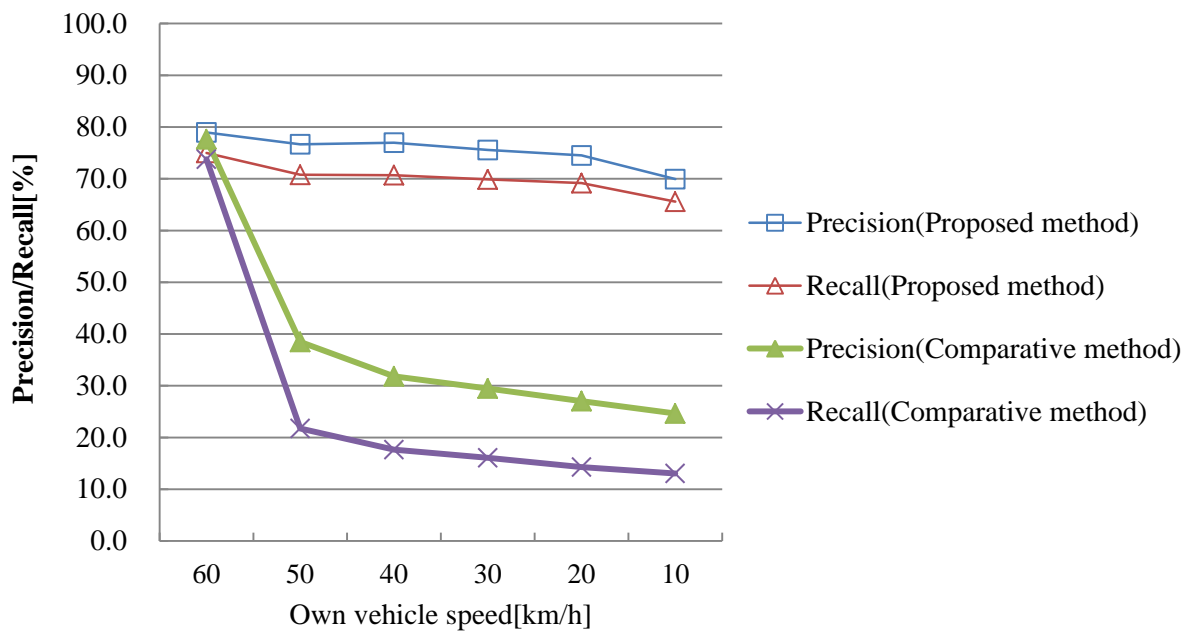
共通部分	<ul style="list-style-type: none"><li>• 画像変換ニューラルネットワーク</li><li>• 学習に使用した画像データ</li><li>• 画像変換後の歩行者認識モデル</li><li>• テスト画像に使用する画像はシミュレーション画像とカメラ画像</li></ul>
変更部分	<ul style="list-style-type: none"><li>• テスト画像の自車速(自車速を 10km/h ~ 60km/h の間で 10km/h 毎変更する)</li><li>• 比較手法として複数フレームの計算値を単純に足し合わせた手法</li></ul>

#### 〈2〉 実験結果

自車速を学習外の車速に変更したときの歩行者認識率を図 5.7 に示す。図 5.7 に示す通り、比較手法では学習データと同じ自車速である 60km/h では画像変換が上手くいき、歩行者認識率が高いのに対し、学習外データ(テスト画像が自車速 10km/h から 50km/h)では画像変換時に徐々に画像にぼやけが発生し、その結果、既存の物体認識手法では物体認識率が低下する結果となった。一方、提案手法では、学習データと同じ自車速である 60km/h では最も認識率が高く、自車速 10km/h に変化していくに連れて、微減していく結果となった。しかし、認識率は約 70% 後半となり十分な認識率であることを確認した。これらの結果から提案手法が学習外データの自車速にも頑健なニューラルネットワークであることがわかり、実際の車両に搭載し、様々な車速で走行したときにも夜間画像を昼間画像に変換できることを示唆した。



(a) The test image is simulation image.



(b) The test image is camera image.

図 5.9：直線道路における自車速を変更したときの歩行者認識結果

自車速を変更したときの歩行者認識結果例を図 5.10 に記す. 図 5.10 では自車速を 30km/h に変更したときの認識結果例を記す. 図内(a)比較手法では複数枚の画像を足し合わせた為, 輝度は改善されているが画像中の歩行者のエッジがぼやけており, 歩行者の未検知が発生している. しかし, (b)提案手法では, エッジが(a)比較手法と比較して改善されており, 歩行者検知できていることがわかる.

提案手法が自車速の変化に応じて改善された理由として, 複数フレームの Pooling 後の中間層を統合することで, 微小な位置ずれや大きさの変化の影響が小さくなった中間層を統合する為, 提案手法では, 夜間画像を明るく, 且つ物体のエッジのブレを抑制したまま変換することができた. 一方で, 比較手法は Pooling 層の前の中間層を統合している為, 複数フレームに存在する歩行者の位置や大きさにずれが発生して, 夜間画像が明るくならないもの, 歩行者のエッジにブレが発生していることが確認できる.

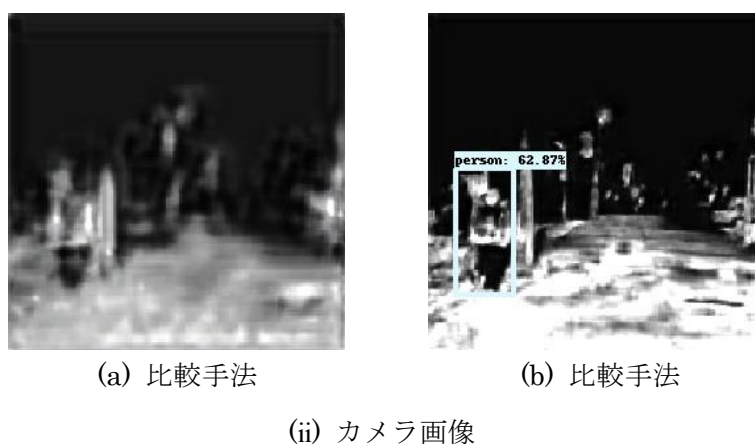
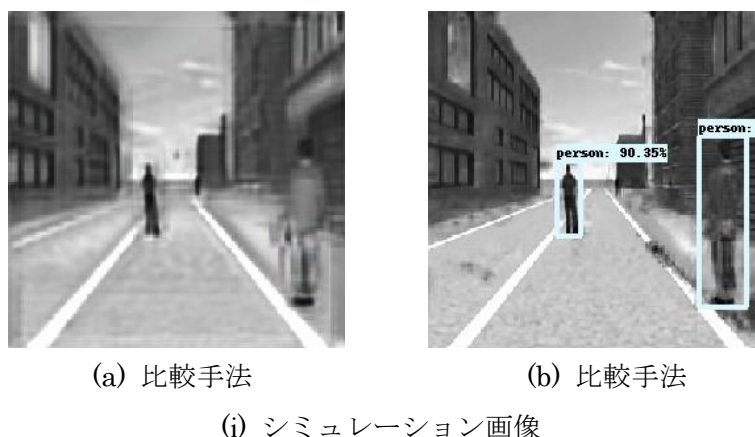


図 5.10: 直線道路における自車速を変更したときの歩行者認識結果例

### 5.3.3 過去フレームの枚数の違いによる検討

本節では提案手法で使用するフレーム枚数について検証した。使用する過去のフレーム枚数を変更して最も歩行者認識可能な最適なフレーム枚数を決定する。テストデータとして実験条件の変更が容易且つノイズの少ないシミュレーション画像を使用する。

#### 〈1〉 シミュレーション画像でのテスト

検証に使用する実験条件を以下に記す。基本的には 5.3.1 節記載の実験と同じ実験条件で行い、ニューラルネットワークに一度に入れるフレーム枚数を変更する。テスト画像も 5.3.1 節と同様のテスト画像を使用する。

表 5.11: フレーム枚数検証実験条件

画像変換	モデル	図 5.1 記載のニューラルネットワーク
	学習データ	PreScan で作成した歩行者シーン 表 5.3 に詳細記載
	入力枚数	1 枚, 2 枚, 3 枚, 4 枚, 5 枚
物体認識	モデル	Faster RCNN
	学習データ	PascalVOC
テストデータ	5.3.1 節と同じ画像	

#### 〈2〉 実験結果

図 5.11 及び表 5.12 に画像変換を行い、歩行者認識した認識率を記す。入力枚数を横軸に、各フレームでの歩行者認識率を縦軸に記す。認識率の推移を見ると、入力するフレーム枚数が 1 枚や 2 枚では Precision 及び Recall が他の枚数の時より低いことを確認した。これは、入力するフレーム枚数が少ないことで夜間画像の明るさが改善されずに誤検知や未検知が発生したことが考えられる。入力画像を 3 枚から 5 枚にしたときは、画像の明るさ及び変換に伴う歩行者の輪郭のブレが改善されることで Precision, Recall 共に改善したことが確認できた。しかし、入力するフレーム枚数が多くなるにつれて、画像変換時に使用される過去の情報が多くなる為、輪郭のブレが再度発生し認識率が低下する結果となった。

上記の結果より、本研究ではニューラルネットワークに入力する画像枚数は 3 枚以上が望ましいと考える。且つ、より早い段階で歩行者認識を実現する為にはより少ない枚数で実現した方が良い為、以降の検証では入力枚数を 3 枚にして検討を行う。

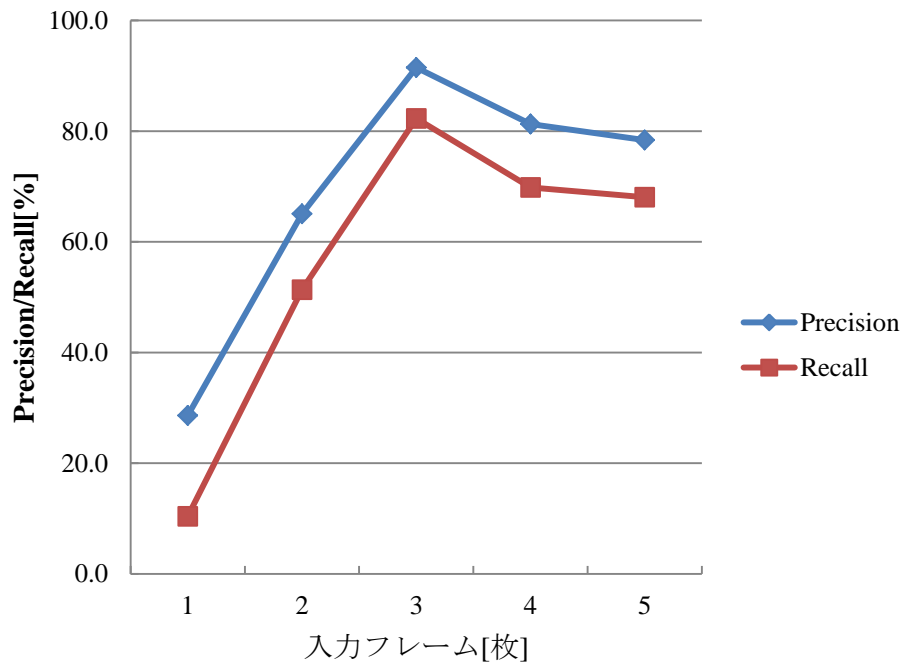


図 5.11：入力フレームの枚数違いによる認識率の推移

表 5.12：入力フレームの枚数違いによる認識率

フレーム枚数	1	2	3	4	5
Precision[%]	28.6	65.1	91.5	81.3	78.4
Recall[%]	10.4	51.4	82.3	69.8	68.1

## 第6章 夜間カーブ道路における画像変換手法の検討

本章では、直線道路において夜間画像を昼間画像に変換する手法について述べる。6.1 節では、提案手法の概要を述べ、6.2 節でニューラルネットワークの中間層について詳細を述べる。その後、6.3 節にて有効性を確認する為の実験について述べる。

### 6.1 提案手法の概要

本節では、夜間カーブ道路における画質改善手法について述べる。図 6.1 に提案手法の概要フローチャートを示す。

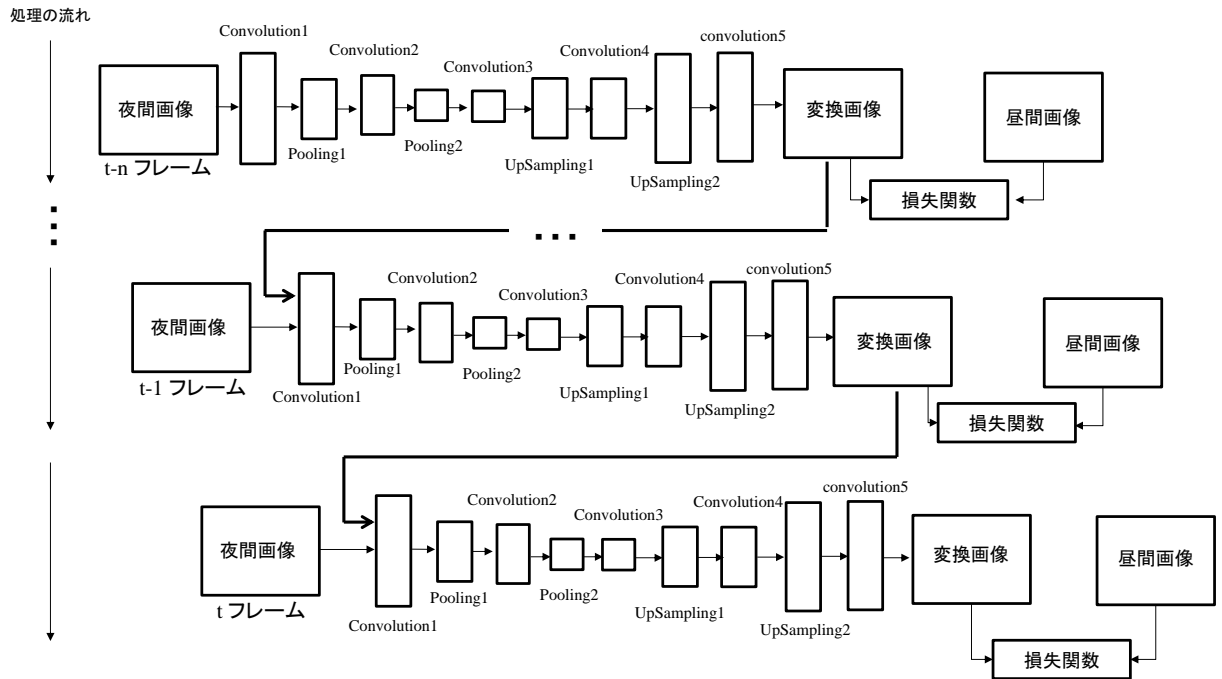


図 6.1: カーブ路における画像変換ニューラルネットワークの概要図

## 6.2 提案手法の詳細

本節カーブ路における提案手法では、5章の直線道路での画質改善と同様に夜間走行画像を使用して画像の明るさ改善を行う。5章における提案手法では、ニューラルネットワークの Pooling1 において複数フレームで処理した計算値を足し、画像の明るさの改善を行う。この時、同じタイミングで足し合わせを行う為、物体の移動量が大きいカーブ路において複数フレームを入力したとき、物体の輪郭のボケが発生する。

図 6.2 に 5 章の直線での画像変換手法をカーブ道路で適用したときの画像変換結果を記す。変換後の結果を見ると、道路上の歩行者や背景の輪郭でブレが発生していることを確認した。その為、ニューラルネットワークの改善を行う必要がある。

そこで本節では以下のニューラルネットワークを提案する。提案するカーブ路に対応するニューラルネットワークでは、現在のフレームより過去フレームの変換結果を活用する。カーブでは、道路上の物体のフレーム間の移動量が大きいことから、DeepLearning の Convolution や Pooling による位置変化、大きさ変化を吸収し、且つ過去フレームの変換結果との入力によってカーブ道路に対して逐次変換を行う。

まず、現在の時刻  $t$  フレームより過去の  $n$  フレームまで遡って変換画像を作成し、逐次、夜間画像と一緒にニューラルネットワークに入れてその時点での変換画像を作成する。例えば  $t-n$  フレームでは、 $t-n$  フレームの夜間画像のみをニューラルネットワークに入力して変換画像を作成する。変換画像は  $t-n$  フレーム目の昼間画像をリファレンスとして使用し、差分を計算し重みを更新する。次の時刻  $t-n+1$  フレームでは  $t-n+1$  フレームの夜間画像と  $t-n$  フレームの変換画像を入力して変換画像を作成する。これを繰り返す。 $t$  フレーム目では、 $t-1$  フレーム目の変換画像と  $t$  フレーム目の夜間画像をニューラルネットワークに入力して最終的な変換画像を得る。

過去フレームの変換された画像を  $t$  フレーム目と一緒に畳み込みとプーリングを適応することで、カーブにおける物体の移動に対応できると考える。

表 6.1 にカーブ路における画像変換ネットワークの構造を記す。提案手法では現在のフレームと過去 1 フレーム前の変換画像を入力する為 Convolution1 に 2 枚分の画像を表す。





(a)入力画像



(b)変換結果

図 6.2 : 5 章における提案手法でのカーブ道路変換結果

表 6.1 : カーブ路における画像変換ネットワークの構造

layer	type	Output size
1	Convolution1	$256 \times 256 \times 256 \times 2$
2	Pooling1	$128 \times 128 \times 256$
3	Convolution2	$128 \times 128 \times 128$
4	Pooling2	$64 \times 64 \times 128$
5	Convolution3	$64 \times 64 \times 128$
6	UpSampling1	$128 \times 128 \times 128$
7	Convolution4	$128 \times 128 \times 256$
8	UpSampling2	$256 \times 256 \times 256$
9	Convolution5	$256 \times 256 \times 1$

### 6.3 実験

本手法の有効性を確認する為に、6.3.1 節から 6.3.3 節にて実験を行った。6.3.1 節ではカーブの大きさ(曲率)の違いによる画像変換結果を記す。曲率を任意の値にする為にシミュレーションを使用してテスト画像を作成する。6.3.2 節では実際の車載カメラ画像で得られたカーブ画像に対して画像変換を行い、歩行者認識を実施する。6.3.3 節では、本節の提案手法による画像変換がどれほどのカーブまで対応できるかのシミュレーション画像を用いて検証を行う。比較手法として 5 章で提案した手法(比較手法①)と CLAHE(比較手法②)を使用した。

#### 6.3.1 曲率の違いによる画像変換結果(シミュレーション画像)

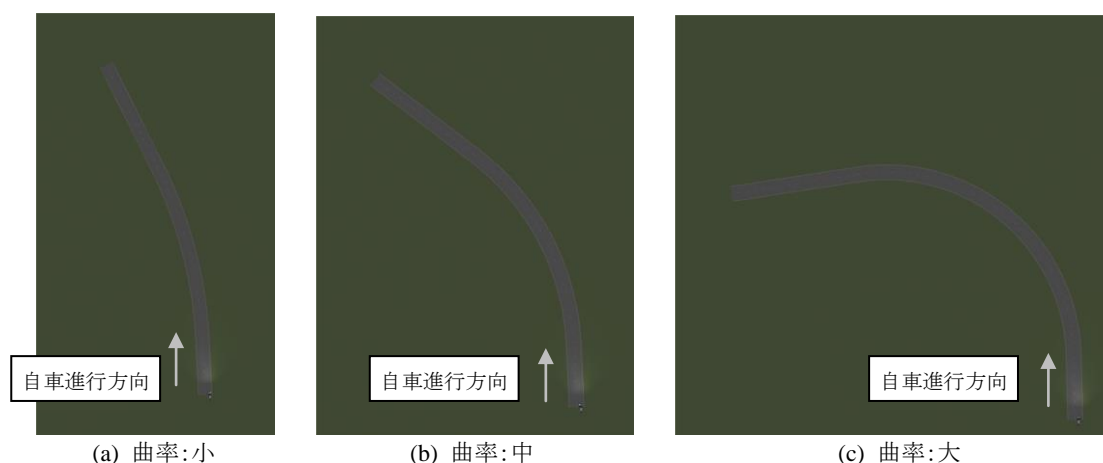
##### (1) 実験条件

本節では、曲率の異なるカーブ路をテスト画像として入力し、画像変換・歩行者認識を行い精度評価する。学習に使用するデータ(カメラ設定値、学習データ)は 4.3 節で作成したシミュレーションデータ(表 4.2, 表 4.3)を使用する。表 6.2 に本節で使用するテストデータを記す。

表 6.2: テストデータ(カーブ路)の概要

Number of pedestrian scene	10[people]
Number of images	800[frame]
Maximum height	180[cm]
Minimum height	160[cm]

図 6.3 にシミュレーションで作成したカーブ路の外観図(印刷用に明るさ+10%)を示す。



2020 copyright@siemens

図 6.3: シミュレーションによるカーブ画像取得時の鳥瞰図

## 〈2〉 実験結果

カーブ路における歩行者認識率を図 6.4 に示し、認識結果例を図 6.5, 図 6.6 に示す。

図 6.4 より、原画像、提案手法、比較手法①、比較手法②で得られた変換画像を各物体認識手法に入力したとき、Precision(適合率)が高いことは確認した。これは画像から認識した物体は歩行者である確率が高いことを表し、他のタグと間違えることなく認識できていることを示す。一方、Recall(再現率)においては、原画像、比較手法①、比較手法②を入力したときは、概ね 50%以下であり、提案手法を入力したときは 70%を超えていることを確認した。

この結果から原画像、比較手法①、比較手法②を入力したときは歩行者の未検知が発生しており、提案手法を入力したときは未検知が低減していることを表している。原画像において未検知が発生する理由は直線道路の場合と同様に画像全体が暗く歩行者と背景のコントラストが小さく認識しにくいことが考えられる。比較手法①では 4 章における手法であり、複数フレームを活用する部分は同じだが、シンプルに計算値を足し合わせているだけなのでカーブ路での物体の移動に対応できず、歩行者の輪郭にブレが発生して認識できず Recall が下がったと考えられる。比較手法②(CLAHE)では、明るさが改善されたが昼間画像のように明るくはならず再現率を向上することができなかった。

提案手法では、ニューラルネットワーク中の Pooling により複数フレームの解像度を落とし、フレーム間の物体の移動量の差を小さくすることができた為、複数フレームの画素値の統合が上手くいったと考える。更に比較手法①では一度に複数フレームの画素値の統合を行ったが、提案手法では過去フレームから逐次画像変換した結果を足し合わせることで複数フレームの移動量の差を小さくし、変換後の歩行者のエッジのブレを抑制できたと考える。

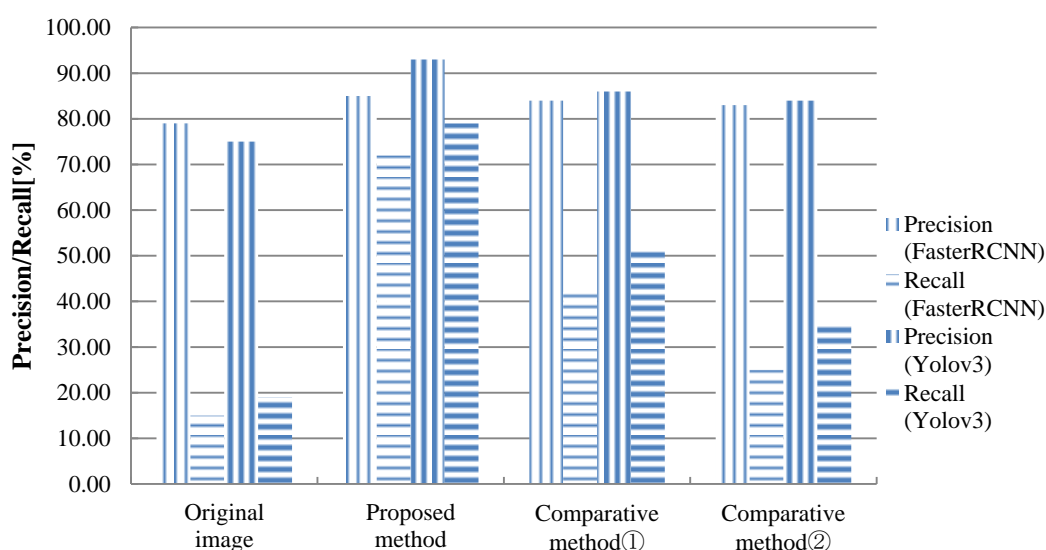


図 6.4: カーブ路における歩行者認識率の比較



(a) 小カーブ

(b) 中カーブ

(c) 大カーブ

(i) original image

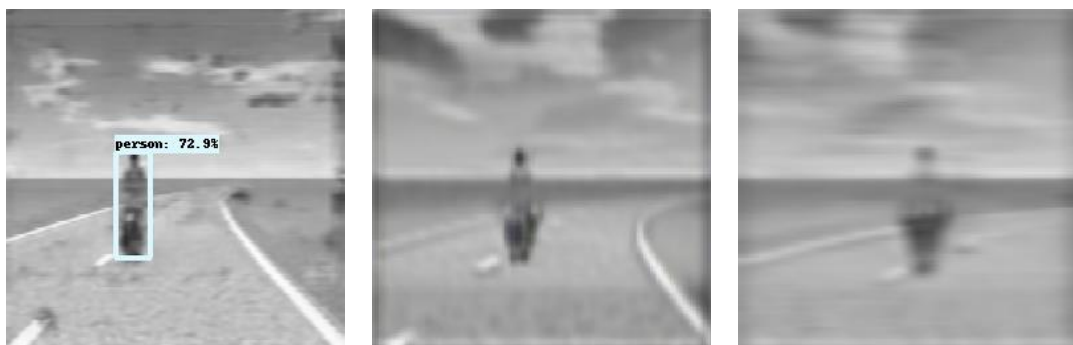


(a) 小カーブ

(b) 中カーブ

(c) 大カーブ

(ii) proposed method



(a) 小カーブ

(b) 中カーブ

(c) 大カーブ

(iii) comparison method 1

図 6.5: カーブ路における歩行者認識結果例



(a) 小カーブ

(b) 中カーブ

(c) 大カーブ

(i) original image

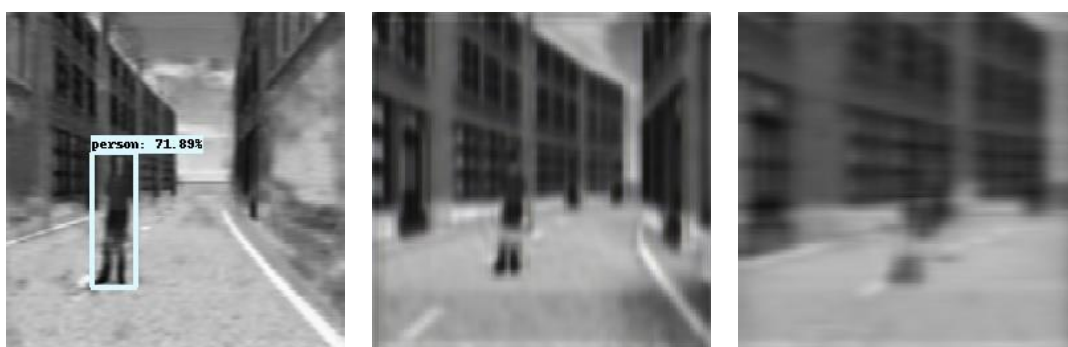


(a) 小カーブ

(b) 中カーブ

(c) 大カーブ

(ii) proposed method



(a) 小カーブ

(b) 中カーブ

(c) 大カーブ

(iii) comparison method 1

図 6.6: カーブ路における歩行者認識結果例

### 6.3.2 実車走行画像(カメラ画像)

#### 〈1〉 実験条件

本節では、実車走行実験で撮影したカーブ路をテスト画像として入力し、画像変換・歩行者認識を行い精度評価する。本節の実験は実車のカメラ撮影で得られたシーンの中からカーブしているシーンを抽出して適用した。画像変換の学習に使用するデータ(カメラ設定値, 学習データ)は4.3節で作成したシミュレーションデータ(表4.2, 表4.3)を使用する。表6.3に本節で使用するテストデータを記す。

表 6.3: テストデータ(カメラ画像)

シーン概要	片側1車線の住宅地におけるカーブシーン
テストフレーム数	1000[フレーム]
歩行者人数	10[人]

#### 〈2〉 実験結果

以下に夜間カメラ画像を変換した結果及び物体認識した結果を記す。図6.7に認識率を示し、図6.8, 6.9に結果例を示す。結果が示すようにテスト画像をシミュレーション画像からカメラ画像に変更すると、提案手法ではカーブによる悪影響は少ないものの、実フィールドで発生する外乱光により変換後の画像で濃淡のバランスがくずれ、見やすさがシミュレーション画像のときより向上しなかった。しかし、入力の変換前画像よりは画像中の歩行者が見やすくなっていることから提案手法による効果があると考え。画質の改善が出来た理由として、Pooling層ではMaxpoolingを使用しており、解像度が小さくなったときに最も特徴となる画素が残る。この特徴マップを複数重ね合わせている為、エッジのブレを抑制して明るさの改善をすることができたと考え。

又、比較手法①では過去フレームの夜間画像をまとめて入力しているが、過去フレームの変換は逐次行い入力している為、現フレームを入力した時の中間層で特徴量を抽出しやすく、画質向上に貢献したと考える。

本結果を受けて、5節で提案した手法より6節で提案した手法の方が直線道路だけでなくカーブへの対応が可能である為、実運用での効果が期待できる。

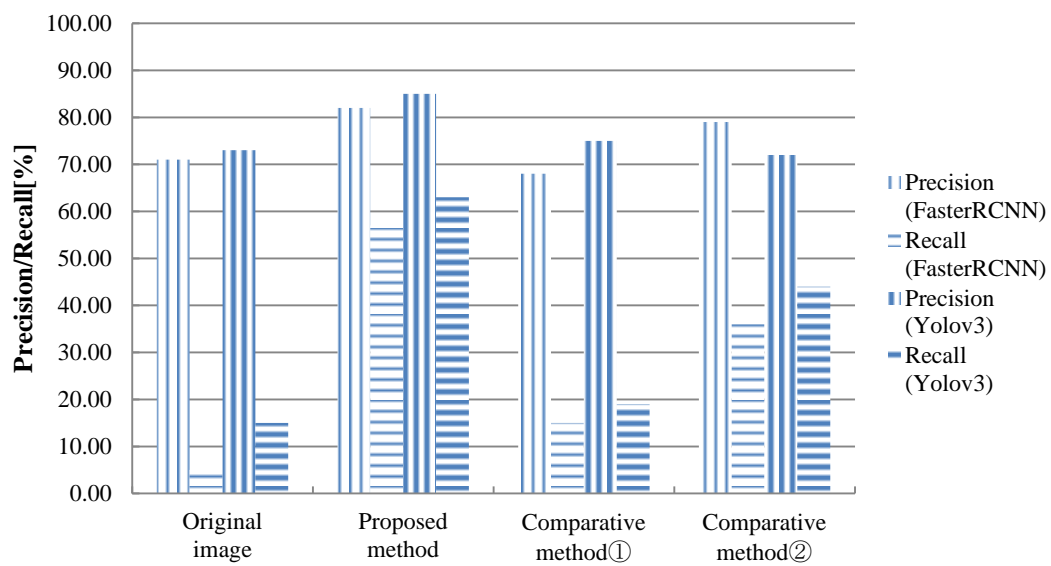


図 6.7: カメラ認識率の比較

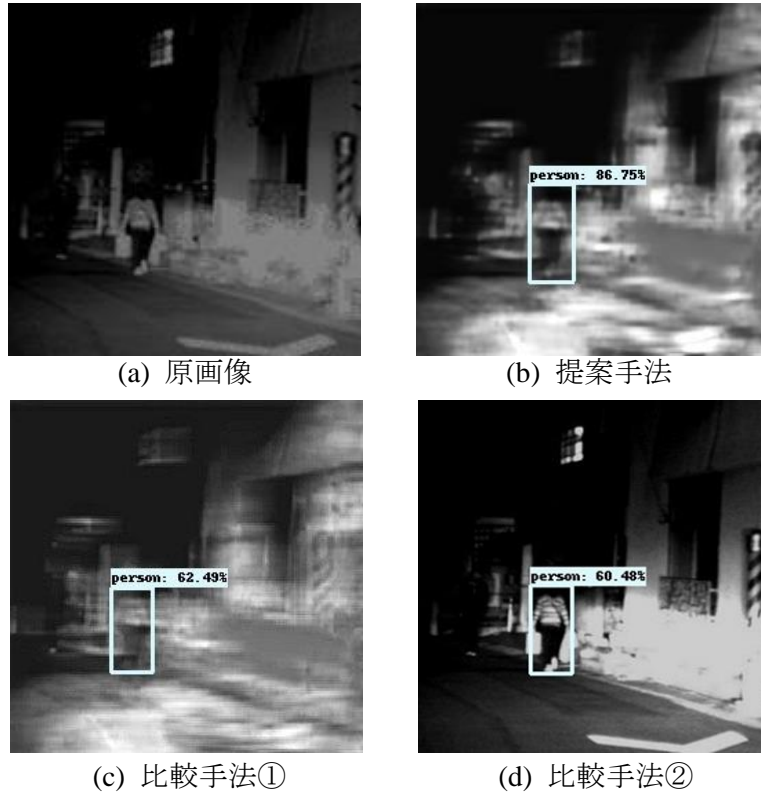


図 6.8: カメラ認識結果例(シーン 1)

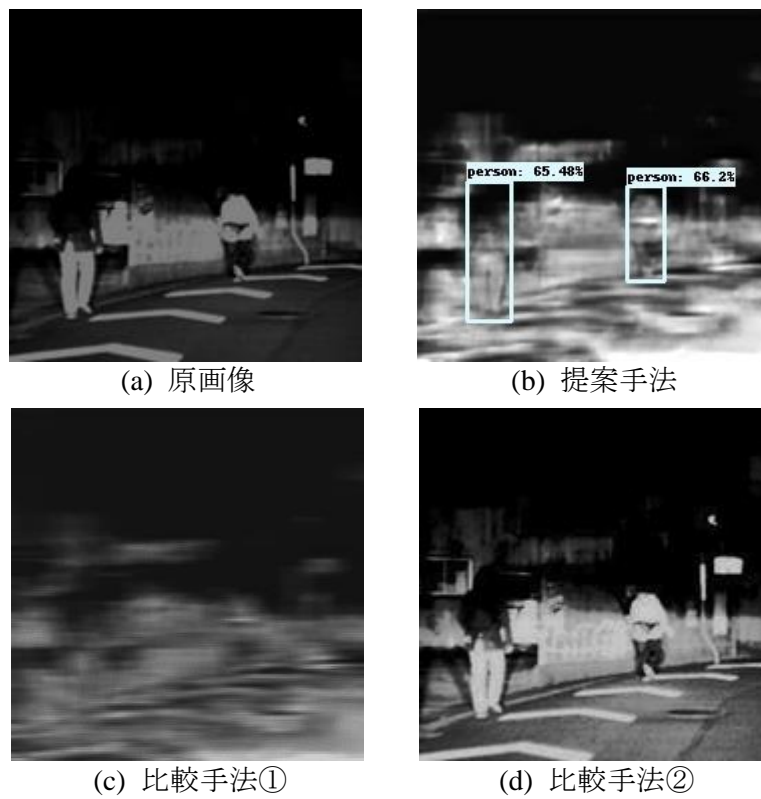


図 6.9: カメラ認識結果例(シーン 2)



### 6.3.3 十字路走行時の対応(限界性能)

本節提案手法の限界性能を把握するにあたり、十字路(右左折角度  $90^\circ$ )カーブにおいて画像変換がどのようになるか検証した。本節におけるテスト画像はシミュレーションで作成したものを使用する。

#### (1) 実験条件

表 6.4 に撮影条件を記し、図 6.10 に走行する十字路を示す。自車両がスタート地点から直進し、右折するパターンと左折するパターン、十字路上に配置した歩行者の種類を変えて実験を行った。図 6.11 に十字路に進入したときの自車両と歩行者の位置関係を鳥瞰図で示す。又、画像変換後の歩行者認識にはこれまでの実験で使用した既存の昼間の歩行者データで学習した Faster RCNN 及び YOLOv3 を使用する。

表 6.4: 十字路走行実験のパラメータ

走行経路	2[経路](右折・左折)
歩行者	5[人]
自車両	トヨタ プリウス
自車速	10[km/h](徐行速度)

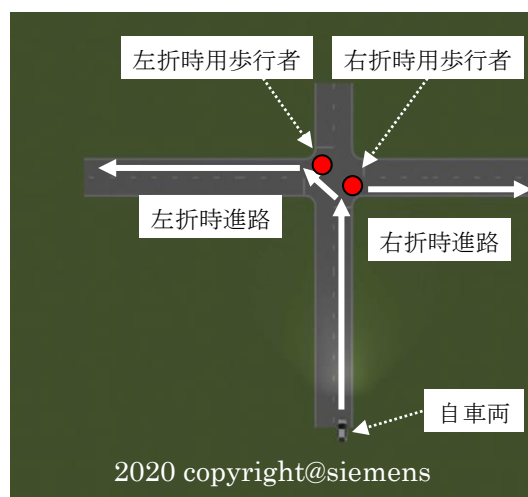


図 6.10: 十字路走行実験のオブジェクト配置図



図 6.11：十字路進入時の位置関係

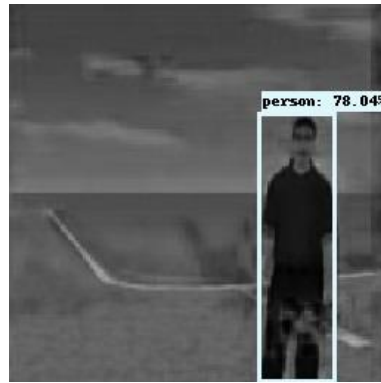
## 〈2〉 実験結果

図 6.12 に十字路における夜間画像の変換結果を記す。十字路におけるカーブでも画像変換を行うことができ、歩行者認識できていることが確認できた。また図 6.13 に認識率を表す。前節と同様に夜間の入力画像では画像のコントラストが低く、未検知となっていたシーンにおいて、昼間のように明るく変換することができたので、歩行者認識率の向上を確認することができた。これは、十字路において自車速は徐行速度まで減速する為、直角カーブによるフレーム間の物体の移動量が大きくなることを低減でき、6.1 節、6.2 節での緩やかなカーブと同様のシチュエーションを再現できたと考えられる。その為、提案手法のニューラルネットワークで歩行者の移動量を吸収でき画質改善を行うことができたと考える。

この結果より、提案手法による画質改善が直線道路及び緩やかなカーブ、十字路でのカーブという基本的な道路形状で実現することができ、提案手法の有効性を確認することができた。



(a) 入力画像の t フレーム

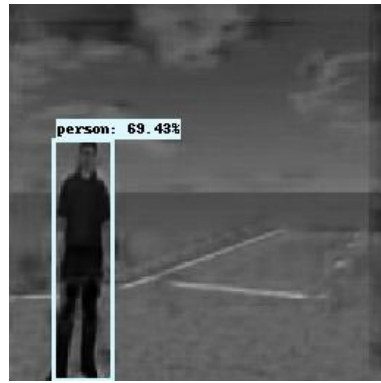


(b) 提案手法による変換結果画像

(i) シーン1



(a) 入力画像の t フレーム



(b) 提案手法による変換結果画像

(ii) シーン2

図 6.12: 十字路走行時の画像変換結果例

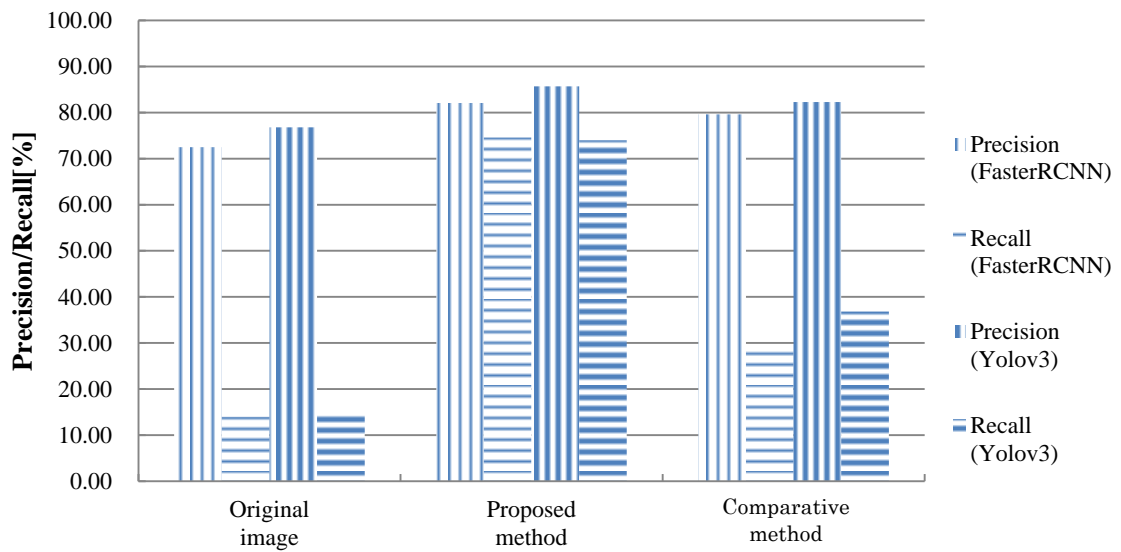


図 6.13: 十字路における認識率の比較

## 第7章 結論

本論文では、夜間における自動車運転時の歩行者事故を低減する為に、コストが安く大衆車まで普及している単眼カメラを用いて、歩行者認識手法の確立を目指した。従来の物体認識手法では、認識対象の画像収集を行い、手動にてアノテーション作業が必要であった。その為、夜間歩行者を認識する為には実車実験を行い夜間歩行者のデータを大量に収集、その後手動でアノテーション作業をする必要があり、膨大なコストと時間が必要であった。そこで本研究では、物体認識モデル作成のコスト削減かつ認識性能が低下しないモデル構築の為、既存の昼間の歩行者認識モデルを活用及び、ドライビングシミュレータを活用したアプローチを試みた。その結果、以下の知見を得た。

第1章では、昼夜の交通事故統計やシーン別交通事故統計を述べ、夜間の歩行者保護の必要性を明らかにした。次に自動運転及び先進運転支援システム(ADAS)の定義及び市場動向、自動運転システムの歴史を述べ、最後に自動運転に必要な各センサの特徴、メリット・デメリットを説明することで、本研究で用いるカメラセンサの特徴を明らかにすることができた。

第2章では、関連研究として物体認識及び画像変換の概要を説明した。2.1節 DeepLearning を使用していない従来の画像処理による物体認識手法について説明し、2.2節 2014年以降に発表された DeepLearning を使用した物体認識手法について述べた。2.3節では夜間物体認識に強みを持つ赤外線を利用した物体認識システムについて述べた。そして2.4節では DeepLearning を利用した画像変換手法について説明した。

第3章では、本研究で検討した夜間歩行者認識手法の概要を述べた。データ収集・アノテーション作業が必要である一般的な物体認識手法と夜間歩行者の認識を可能にし、且つモデル作成コストの削減を目的とした提案手法の処理の流れを示し、比較を行い提案手法のメリットを示した。そして、従来手法の問題点及び本研究の目的を明らかにした。提案手法では夜間画像をニューラルネットワークを使用して昼間画像に近づくよう変換し、得られた変換画像に対し既存の歩行者認識モデルを適用することで夜間歩行者認識を実現する。

第4章では、第3章で述べた提案手法について詳細を述べた。本研究では段階的にアルゴリズムの改善を行い、第4章では直線道路に着目して夜間画像の変換手法を検討した。暗い画像の画質改善を行う手法として複数の画像を利用する手法がある。今回車載動画画像がターゲットである為、連続する連番画像を利用した。また、自車の移動によりフレーム間の対象物体の移動に対応する為に、Pooling 処理の後に複数フレームの画素値の統合を行った。これにより、複数フレーム統合時に発生する歩行者のぶれを低減す

る。提案手法の有効性を確認する為に、シミュレーションと実車カメラから取得したテスト画像を使用して歩行者認識率の向上を確認した。

第5章では、カーブ路に着目して夜間画像の変換手法を検討した。現在のフレームから過去のフレームを画像変換し、その変換結果と現在の夜間画像を入力するニューラルネットワークを検討した。これにより、カーブ路走行時、フレーム間の歩行者の移動量が大きいシーンにおいて、逐次画像変換を行い歩行者のブレを抑えて画像変換することが可能となった。

第6章では、本研究の実証実験で使用するデータ、認識アルゴリズムについて述べた。まず、シミュレーション画像及び、実車カメラ画像の取得方法について述べた。続いて画像変換後に使用する既存の歩行者認識アルゴリズムについて述べ、評価指標や認識限界距離について述べた。

以上の検討・実証実験を通じて **DeepLearning** を利用した画像変換手法及び物体認識手法についての知見を得ることができ、先進運転支援システムにおける夜間歩行者保護に貢献することができたと言える。

## 謝辞

本論文は、筆者が公立ほこだて未来大学大学院システム情報科学研究科博士(後期)課程、システム情報科学専攻在学中の研究成果をまとめたものである。

本論文の執筆および研究の遂行にあたり、丁寧なご指導を賜りました公立ほこだて未来大学長崎健教授に格別なる感謝の意を表します。修士までは従来の画像処理に関する研究をしていた私にとっては機械学習の基礎から学ぶことからの出発でしたが、長崎健教授から賜りました数々のご助言やご指導のおかげで、只今こうして博士論文を執筆するに至っております。長崎健教授の下で博士課程の研究生活を終えるにあたり、心より感謝の意を表すとともに、これまでの貴重な体験を今後の研究生活に活かしていきたいと考えています。

同様に、本研究の初期段階から始終丁寧なご指導を賜りました公立ほこだて未来大学松原仁特任教授に別格なる感謝をいたします。私は長崎健教授の研究室に所属しているにも関わらず、在学中、研究打合せの時間を確保して下さい、多くのご指導とご助言を下さりました。松原仁特任教授の下で培った研究力を今後の技術開発に活かしてまいります。

又、公立ほこだて未来大学三上貞芳教授、同大学鈴木恵二教授には、本博士論文の審査委員をご快諾戴き、又、数々のご助言とご指導を賜りました。深く感謝致します。

本研究は、株式会社小糸製作所の上司、先輩、同僚の協力によって支えられてきました。カメラセンサ等機材の準備や実車走行実験、ドライビングシミュレータソフト、高性能GPU搭載PCなど、研究を円滑に遂行できる恵まれた環境を提供して頂きました。深く感謝致します。IEEE GCCEやIW-FCV、電気学会の関係者および聴衆者の皆様から数々の有益なご意見を戴きました。発表および質疑応答を通して、研究の質を高めることができました。

公立ほこだて未来大学学長を始め、諸先生方、事務局、情報ライブラリなど、大学関係者の皆様のご尽力によって、日々の研究活動を快適に過ごすことができました。ここに謹んで感謝の意を表明します。最後に、大学院での研究活動を応援してくれた家族に心より感謝申し上げます。

## 参考文献

- [1] 内閣府,“平成 30 年版交通安全白書”, p.14, 2019.
- [2] 警視庁,“令和元年度 歩行者事故発生状況”, p.4, 2020.
- [3] 矢野経済研究所,“自動運転システムの世界市場に関する調査”, 2019.
- [4] 津川定之,“自動運転技術の発展”, 国際交通安全学会誌, Vol.40, No.2, pp.82-90, 2015.
- [5] 城殿清澄, 渡邊章弘, 内藤貴志, 三浦純 : 「高解像度レーザレーダによる歩行者識別」, 日本ロボット学会誌, Vol.29, No.10, pp.963-970, 2011.
- [6] N. Dalal, B. Triggs : “Histograms of Oriented Gradients for Human Detection” , Computer Vision and Pattern Recognition, pp.886-893, 2005.
- [7] 渡辺友樹, 伊藤聡, 横井謙太郎, : 「画像認識による歩行者検出のための特徴量抽出技術の開発」, 映像情報メディア学会誌, Vol.71, No.1, pp.J28-J34, 2017.
- [8] P. Dollar, Z. Tu, P. Perona, S. Belongie : “Integral Channel Features” , British Machine Vision Conference, pp.91.1-91.11, 2009.
- [9] P. Dollar, R. Appel, S. Belongie, P. Perona : “Fast feature pyramids for object detection” , IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.36, No.8, pp.1532-1545, 2014.
- [10] A. Krizhevsky, I. Sutskever, G. E. Hinton : “ ImageNet Classification with Deep Convolutional Neural Networks” , NIPS, Vol.1, pp.1097-1105, 2012.
- [11] R. Girshick, J. Donahue, T. Darrel, J. Malik : “ RichFeature Hierarchies for Accurate Object Detection and Semantic Segmentation” , Computer Vision and Pattern Recognition, pp.580-587, 2014.
- [12] S. Ren, K. He, R. Girshick, J. Sun : “Faster R-CNN: towards real-time object detection with region proposal networks” , NIPS, pp.91-99, 2015.
- [13] W. Ouyang, X. Wang : “Joint deep learning for pedestrian detection” , International Conference on Computer Vision, pp.2056-2063, 2013.
- [14] P. Sermanet, K. Kavukcuoglu, S. Chinatala, Y. LeCun : “Pedestrian detection with unsupervised multi-stage feature learning” , Computer Vision and Pattern Recognition, pp.3626-3633, 2013.

- [15] J. Redmon, A. Farhadi : “Yolov3: An incremental improvement”, arXiv preprint arXiv:1804.02767, 2018.
- [16] Kaiming He. Xiangyu Zhang. Shaoqing Ren. Jian Sun,” Deep Residual Learning for Image Recognition”, arXiv:abs/1512.03385, 2015.
- [17] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie,” Feature Pyramid Networks for Object Detection”, arXiv:abs/1612.03144, 2016.
- [18] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling : “M2Det: A single-shot object detector based on multi-level feature pyramid network”, arXiv preprint arXiv: 1811.04533, 2018.
- [19] B. Fardi, U. Schenert, G. Qanielik , “Shape and Motion-based Pedestrian Detection in Infrared Images”, A Multi Sensor Approach, Proc. IV2005 IEEE Intelligent Vehicles Symposium, pp.18-23, 2005.
- [20] 前渕啓材, 吳海元, 和田俊和 : 「幾何制約付 HOG と NFTG による夜間歩行者検出」, 電子情報通信学会技術研究報告(信学技報) , Vol.111, No.380, pp.291-296, 2012.
- [21] 青木正喜, 安田升 : 「車載赤外線カメラを用いた歩行者検出」, 情報処理学会論文誌, Vol.47, No.5, pp.1-9, 2006.
- [22] 中條博則, “2018 車載カメラ徹底解説”, 共創企画, p.81, 2018.
- [23] Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, “Generative Adversarial Networks” Proc. Adv. Neural Inf. Process. Syst. 27, pp. 2672–2680, 2014.
- [24] L. A. Gatys, A. S. Ecker, M. Bethge : “A Neural Algorithm of Artistic Style”, Computer Vision and Pattern Recognition, pp.2414-2423, 2015.
- [25] S. Iizuka, E. Simo-Serra, H. Ishikawa : “Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification”, C ACM Transactions on Graphics, Vol.35, No.4, pp.1-11, 2016.
- [26] M. Liu, T. Breuel, J. Kautz : “Unsupervised Image-to-Image Translation Networks”, NIPS, pp.1-11, 2017.
- [27] Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, L. V. Gool : “Night-to-Day Image Translation for Retrieval-based Localization”, ICRA, 2019.
- [28] S. W. Huang, C. T. Lin, S. H. Chen, Y. Y. Wu, P. H. Hsu, S. H. Lai : “AugGAN: Cross



Domain Adaptation with GAN-Based Data Augmentation”, ECCV, 2018.

- [29] D.Kingma, J. Ba: “Adam: a method for stochastic optimization”, In the 3rd International Conference for Learning Representations ICLR 2015, 2015.
- [30] X.Glorot, A.Bordes, Y.Bengio : “Deep Sparse Rectifier Neural Networks”, Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistic, Vol.15, pp.315-323, 2011.
- [31] K. Zuiderveld : “Contrast limited adaptive histogram equalization” , Graphics gems IV, pp.474-485, 1994.

## 図目次

図 1.1 : 状態別交通死亡事故発生状況.....	9
図 1.2 : 2018 年 昼夜別交通事故負傷者数 .....	10
図 1.3 : 2018 年昼夜別交通事故死者数.....	11
図 1.4 : 歩行者の状態別事故件数.....	11
図 1.5 : ADAS/自動運転システムの世界市場規模予測(概要).....	15
図 1.6 : 路面下に設置された誘導ケーブルとパンパーに装着されたコイル[4] .....	17
図 1.7 : 単眼カメラ例 : STC-MCS241U3V(オムロンセンテック製).....	21
図 1.8 : 単眼カメラ例 : デンソー製 Toyota Safety Sense .....	22
図 1.9 : ステレオカメラ例 : デンソー製ステレオカメラ .....	22
図 1.10 : IR カメラの例 : JVC ケンウッド製遠赤外線カメラ .....	23
図 1.11 : LiDAR 例 : Velodyne 製 VLS-128-AP.....	24
図 1.12 : LiDAR による 3 次元点群データの取得結果(引用元 Velodyne HP より) .....	24
図 1.13 : ミリ波レーダー例 : 富士通テン性ミリ波レーダー.....	25
図 1.14 : 超音波センサ例 : タムラ製作所製超音波センサ .....	26
図 2.1 : ICF による複数の特徴抽出器による特徴画像[8].....	28
図 2.2 : 各特徴抽出における特徴強度[9] .....	28
図 2.3 : R-CNN 概要図[11] .....	30
図 2.4 : Faster R-CNN 概要[12] .....	31
図 2.5 : ResNet 概念図[16].....	32
図 2.6 : Feature Pyramid Networks の概念図[17] .....	32
図 2.7 : M2Det のネットワーク概要図[18] .....	34
図 2.8 : 赤外線画像による歩行者認識[21].....	36
図 2.9 : 遠赤外線カメラの販売台数と単価[22] .....	36
図 2.10 : 近赤外線カメラによるホワイトアウトの発生.....	37
図 2.11 : GAN の概念図 .....	37
図 2.12 : Gatys らによる画像変換結果[24] .....	38
図 2.13 : Iizuka らによる色付け結果例[25].....	39
図 2.14 : Liu らによる画像変換結果例[26].....	40
図 2.15 : Anoosheh らによる夜間画像の変換結果例[27].....	41
図 3.1 : 従来の物体認識手法の処理の流れ .....	42
図 3.2 : 提案手法の全体処理の概要 .....	43
図 3.3 : 本研究における提案手法の大枠 .....	45
図 4.1 : シミュレーションで作成した道路の鳥瞰図 .....	46

図 4.2: シミュレーションでのカメラ設置位置 .....	47
図 4.3: シミュレーションで撮影した歩行者例 .....	48
図 4.4: カメラ録画システム構成図 .....	50
図 4.5: カメラ撮影実験におけるカメラ設置位置 .....	50
図 4.6: カメラ画像撮影機器 .....	51
図 4.7: Precision, Recall 概念図 .....	54
図 4.8: 認識限界距離の計測の為の昼間における実験(模式図) .....	55
図 4.9: 認識限界距離の計測の為の昼間における実験(鳥瞰図) .....	55
図 4.10: 認識限界距離の計測の為の昼間における実験(側面図) .....	56
図 4.11: 認識限界距離の計測の為の昼間における実験(正面図) .....	56
図 4.12: 昼間環境での歩行者認識(歩行者 1) .....	57
図 4.13: 昼間環境での歩行者認識(歩行者 2) .....	57
図 4.14: 昼間環境での歩行者認識(歩行者 3) .....	58
図 4.15: 昼間環境での歩行者認識率 .....	59
図 5.1: 直線道路における画像変換ニューラルネットワークの概要図 .....	60
図 5.2: シミュレーションによる昼夜の画像例 .....	63
図 5.3: Relu 関数 .....	65
図 5.4: 夜間歩行者認識の流れ .....	67
図 5.5: 直線道路でのシミュレーション画像を使用した時の歩行者認識結果 .....	68
図 5.6: 直線道路における画像変換結果例(シミュレーション画像) .....	69
図 5.7: 直線道路においてカメラ画像を使用した時の歩行者認識結果 .....	71
図 5.8: 直線道路における画像変換結果例(カメラ画像) .....	73
図 5.9: 直線道路における自車速を変更したときの歩行者認識結果 .....	75
図 5.10: 直線道路における自車速を変更したときの歩行者認識結果例 .....	76
図 5.11: 入力フレームの枚数違いによる認識率の推移 .....	78
図 6.1: カーブ路における画像変換ニューラルネットワークの概要図 .....	79
図 6.2: 5章における提案手法でのカーブ道路変換結果 .....	81
図 6.3: シミュレーションによるカーブ画像取得時の鳥瞰図 .....	82
図 6.4: カーブ路における歩行者認識率の比較 .....	83
図 6.5: カーブ路における歩行者認識結果例 .....	84
図 6.6: カーブ路における歩行者認識結果例 .....	85
図 6.7: カメラ認識率の比較 .....	87
図 6.8: カメラ認識結果例(シーン 1) .....	88
図 6.9: カメラ認識結果例(シーン 2) .....	88
図 6.10: 十字路走行実験のオブジェクト配置図 .....	89
図 6.11: 十字路進入時の位置関係 .....	90

図 6.12: 十字路走行時の画像変換結果例 .....	91
図 6.13: 十字路における認識率の比較.....	91

## 表目次

表 1.1: 自動運転レベルの住み分け定義 .....	12
表 1.2: 自動運転レベル別定義 .....	12
表 1.3: ADAS/自動運転システムの世界市場規模予測(レベル別詳細).....	15
表 1.4: 自動運転システムの歴史概要.....	16
表 1.5: ITS プロジェクトにおける各機関の取り組み .....	18
表 1.6: Grand Challenge と Urban Challenge の概要.....	20
表 1.7: 単眼カメラのメリット及びデメリット .....	21
表 1.8: ステレオカメラのメリットとデメリット.....	22
表 1.9: IR カメラのメリットとデメリット .....	23
表 1.10: LiDAR のメリットとデメリット .....	24
表 1.11: ミリ波レーダーのメリットとデメリット.....	25
表 1.12: 超音波センサのメリットとデメリット .....	25
表 1.13: 自動運転に必要なセンサ機能.....	26
表 2.1: Alexnet の概要[10] .....	29
表 2.2: 赤外線波長による区分け .....	35
表 3.1: 一般的な歩行者認識手法と提案手法の作業コストの比較.....	44
表 4.1: シミュレーションでのカメラ設定 .....	47
表 4.2: カメラ仕様.....	49
表 4.3: カメラ画像撮影場所.....	49
表 4.4: 既存の歩行者認識モデルの作成で使用したデータ(PascalVOC2012) .....	52
表 5.1: 直線道路における画像変換ネットワークの構造 .....	62
表 5.2: カメラ設定値.....	64
表 5.3: 学習データ(シミュレーション画像).....	64
表 5.4: 学習条件.....	65
表 5.5: 比較手法のインデックスと概要 .....	66
表 5.6: テストデータ(シミュレーション画像) .....	66
表 5.7: 直線道路でのシミュレーション画像を使用した時の歩行者認識結果.....	68
表 5.8: テストデータ(カメラ画像) .....	70
表 5.9: テスト画像にカメラ画像を使用した時の歩行者認識率 .....	72
表 5.10: 実験条件の共通部分と変更部分 .....	74
表 5.11: フレーム枚数検証実験条件 .....	77
表 5.12: 入力フレームの枚数違いによる認識率 .....	78
表 6.1: カーブ路における画像変換ネットワークの構造 .....	81

表 6.2: テストデータ(カーブ路)の概要 .....	82
表 6.3: テストデータ(カメラ画像) .....	86
表 6.4: 十字路走行実験のパラメータ .....	89

## 研究業績

### 学術雑誌（査読有）

小倉亮太, 長崎健, 松原仁, 「夜間における車載カメラを使った歩行者認識の為の夜間画像の視認性向上」, 電気学会論文誌 C 部門, vol.140, no. 5, pp.531-538, 2020.

### 国際学会（査読有）

R. Ogura, T. Nagasaki : “ Improve Visibility of Nighttime Images for Pedestrian Recognition by In-Vehicle Camera”, IEEE Global Conference on Consumer Electronics, pp28-26, 2018.

R. Ogura, T. Nagasaki, H. Matsubara : “ Improve Visibility of Nighttime Images for Pedestrian Recognition on Curve Road”, The International Workshop on Frontiers of Computer Vision, 2020.

R. Ogura, T. Nagasaki, H. Matsubara : “ Improve Visibility of Nighttime Camera Images for Pedestrian Recognition by In-Vehicle Camera”, IEEE Global Conference on Consumer Electronics , 2020.