〔論　文〕

深層学習による通路領域抽出に基づく進行方向の推定

河野　　央*1

Movement Direction Estimation Based on Extraction of Passage Areas Using Deep Learning

Hiroshi KONO*1

Abstract

In autonomous driving in indoor and outdoor spaces, a moving vehicle coexists with various objects and users. Object recognition in the movement path must detect unknown objects and requires complicated judgments. In object recognition using deep learning, the accuracy of recognizing various moving objects and obstacles depends on the learning data. Alternatively, if we can determine whether a movable passage is accessible, we can recognize the movement path of an unknown object without improving the accuracy of object recognition.

In this study, we investigated a method for estimating the moving route from the passage extracted using the image conversion tool pix2pix, a derivative technology of generative adversarial networks. By obtaining perceptually valid passage-extraction results, the proposed method resolved the problem of image conversion.

Key Words：pix2pix, GAN, movement direction, passage, deep learning

1．研究の背景

モビリティ分野では，各種センサーによる外界の把握，画像処理，AI（Artificial Intelligence）による自動運転が実用化されている．今後，様々な形態のモビリティが増加することが予想されるが，それぞれの移動特性に合った自動運転制御が必要になるであろう．例えば，本学のインテリジェントモビリティ研究所（IML）が研究および社会実装を進めている自動運転車いす[(1)]は，公道だけではなく，室内外の空間においても様々な物体やユーザーと共存しながら移動するため，その移動経路に必要な物体認識は未知の物体も含まれ，より複雑な判断が必要である．

深層学習による物体認識では，様々な，あるいは極端に言えば全ての移動体・障害物を認識する方向性が主であり，学習データにその精度が依存する．そのため，新しく生み出されるような工業製品・未知の物体については認識しないという場面も想定される．一方で，少し見方を変えれば「モビリティが移動可能な通路なのか・そうではないのか」という点を判断できれば，障害物の種類を個別に把握する必要もなく，未知の物体にも対応できる可能性がある．

また，深層学習では，予測・分類に加えて，GANs（Generative Adversarial Networks）[(2)]とよばれる手法が提案されたことにより，画像生成に対応することが可能となった．さらに，その応用としてさまざまな派生技術が生まれ，例えば pix2pix[(3)]が画像変換の手法として提案されている．本研究では，入力画像を目的の出力画像に変換可能な手法として，この pix2pix 着目し，通路抽出に特化した画像処理として利用することで，さまざまなモビリティが室内外の空間において安全に移動する経路を認識する可能性について取り上げる．

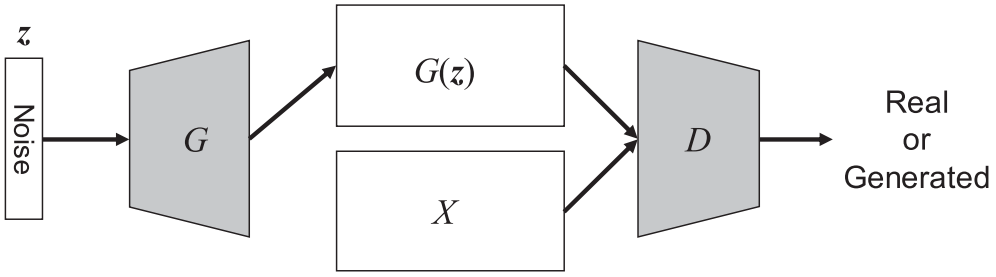
2．関連研究および本研究の着目点

本研究の着想に関連する研究は，深層学習の物体検知と画像生成である．物体検知の分野では，2015年の End-to-end 学習が可能となった Faster-RCNN[(4)]，速度重視の YOLO[(5)]，精度と速度をバランスよく満たす SSD[(6)]，動的にクロスエントロピー誤差を変化させる Focal　Loss が特徴の RetinaNet[(7)]，YOLO v3や RetinaNet よりも高精度な推論が可能な CenterNet[(8)]等，高速かつ高精度な物体検知モデルが発表されている．各種物体検知の手法において比較の対象となる

*1 情報ネットワーク工学科
令和３年10月26日受理

ようなスタンダードな手法はYOLO（You Only Look Once）というモデルであろう．このモデルは，今まで検出と識別を別々の手法で考え2段階構成で学習を行っていた物体検出を1つのディープニューラルネットワークで行う手法で，検出と識別の多段構成を解消し最適化することにより，高速で高精度な検出が可能となった手法である．例として複数のオブジェクトが写っている画像内からオブジェクトの種類，位置を特定することができる．YOLOは改良と共に事前学習済みネットワークも公開されていて，2020年6月にはバージョン5が公開されている．

画像生成の分野では，2014年に発表された敵対的生成ネットワークGANが代表的である．GANは，GeneratorとDiscriminatorと呼ばれる2種のネットワークを用いて学習を行うものである（図1）．この手法は，深層学習によるデータの特徴を捉えた「生成」手法としてインパクトを与えた．GANを基にした派生研究をまとめたGAN ZOO[9]では，2018年10月で更新が止まっているものの，派生研究の量の多さを推察できる．図1のオリジナルGANはVanilla GANとよばれている．



G：Generator，D：Discriminator，z：ノイズベクトル，X：本物の画像，$G(z)$：Generatorが生成した画像

図1 ：GANのモデル構造

Generatorは乱数（ノイズz）の入力から，本物画像の特徴をとらえた画像を生成する役割のネットワークである．Discriminatorは Generatorが生成した偽の画像と本物の画像を識別する役割のネットワークで，Discriminatorへの入力は本物の画像もしくはGeneratorが生成した偽画像である．GANの学習過程はDiscriminatorとGeneratorを交互に更新することの繰り返しで行われる．重要な点は，Generatorは生成した画像をDiscriminatorに本物であると判定されるようにパラメータを更新して学習することに対し，Discriminatorは生成された画像を偽物の画像であると判定するように相手（Generator）のパラメータを更新しない状態で学習する点である．この抑制によってより深い学習をすることが可能になっている．GeneratorはDiscriminatorに正しく識別されないようにパラメータを更新，一方，Discriminatorは正しく識別できるようにパラメータを更新するため互いに対立し競い合うような構造になる．画像データの分布を$P_{\text{data}}(x)$，ノイズベクトルzの事前分布$P_z(z)$とするとき，GeneratorとDiscriminatorの2つの損失関数を合わせたGAN全体の目的関数は，式（1）のようにMinMaxゲームとしてモデル化されている．

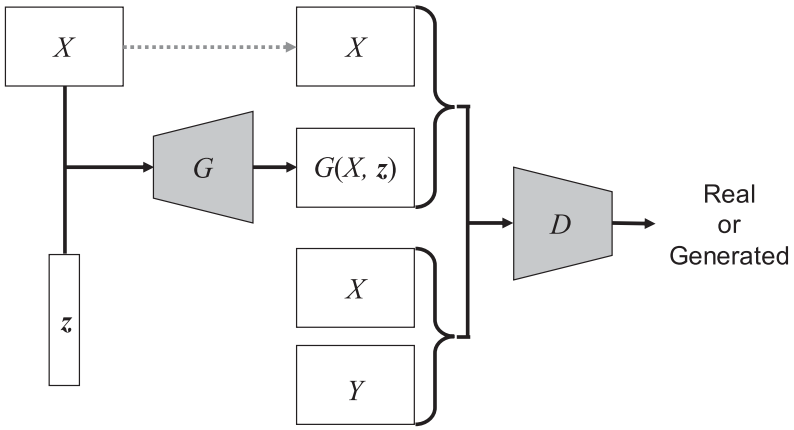
$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim P_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim P_z(z)}[\log(1-D(G(z)))] \qquad (1)$$

$D(x)$は，与えられたデータが学習データXからのものである確率を示す．Dは，本物の画像と生成された画像を識別できるようになることであり，二項分類を実行して検出する．そのため，本物画像を入力とした際は1との差を損失とし，生成された画像を入力とした際は0との差を損失とする．この2つの損失の和をDの損失とし，その損失の和が小さくなるように学習する．つまり，$\log D(x)$と$\log(1-D(G(z)))$を最大化する．Generatorでは，1との差が損失になり，$\log(1-D(G(z)))$を最小化するように学習する．$D(G(z))$の値が大きい場合，Dは$G(z)$がXであると見なしている．

DとGが均衡すると，任意のxに対して$D(x)$は1/2に等しくなる．この状態まで学習が進むと，Dは本物の画像とGの生成画像の見分けがつかないような状態になり，最終的にGは本物とほぼ同じようなデータを生成できているといえる．

pix2pix（図2）は，GANを基としたImage-to-Image Translationの手法である．入力画像・変換画像という2つの画像をペアとした学習を行うことで，入力画像と変換画像のピクセル同士が対応した画像変換を可能としており，ピクセル間で条件付けを行ったGAN，つまりConditionalGAN[10]と捉えることができる．GANではランダムなノイズを入力としているが，pix2pixではGeneratorへの入力は256×256pixelの画像である．そしてGeneratorは画像から画像に変換するモデルで良く利用されるEncoder-Decoder構造のものだけではなく，U-Net[11]というモデルを取り入れることで改良が加えられている．畳み込み処理によるダウンサンプリングを行う際に特徴マップを保持しておき，アップサン

プリング時に特徴マップを使うことで位置情報の復元精度を高めている．Discriminator では，画像全体を本物もしくは生成で識別するのではなく，小さい領域（＝パッチ）に分割してパッチ単位で識別し判定の精度を向上させるといった改良が行われた PatchGAN とよばれる手法を採用している．そして，本物のペア（X と Y）なのか生成されたペア（X と $G(X,z)$）なのかを判断する．



G：Generator，D：Discriminator，z：ノイズベクトル，
X：入力画像，Y：変換画像（アノテーションデータ），$G(z)$：Generator が生成した画像

図 2 ：pix2pix のモデル構造

画像変換にはこの他，CycleGAN[12]が挙げられる．pix2pix と異なり，対でない 2 つの画像群間の写像を変換する手法である．画像全体の画風やテクスチャの変換に特化しているが，形状の異なるドメイン間の変換は対応が難しい．ただし，この問題は通路部分の抽出では特に問題にならないと考えられる．

ここまでいくつかのモデルを選択肢として列挙してきた．物体検知の手法では，精度向上と高速化が実現されているが，学習データとの照合という意味合いが強く明示的なアノテーションを増やす必要があるため，本研究の観点においては未知の物体への対応という点に疑問が生じる．一方，pix2pix や CycleGAN の画像変換では精度や高速化という点よりも「求める出力画像の特徴の学習」がねらいとなっている．未知の物体や場所であっても通路の特徴を捉えて学習されていれば，通路の抽出という点では精度および汎用性が期待できる．

3．研究の目的

移動における通路認識では，セグメンテーションとオブジェクト検出が障害物認識に用いられるが，未知の障害物が多数ある通路認識を考えた場合，同じようなパターンや色調が続く床部分を認識して抽出する場合に Image-to-Image Translation が有効ではないかと考えた．また，通路は入力画像の下方から消失点に向かって存在することが通常であり，ピクセルの位置座標も通路の存在する領域の重要な手がかりとなることを考え，CycleGAN ではなく pix2pix の方がピクセル同士の変換という特徴にピクセルの位置情報が包括されており学習過程として適していると考えた．

そこで本研究では，web カメラを介した入力画像を pix2pix による画像変換処理を行い通路領域の抽出を行う．またその領域から進行方向を推定することで，モビリティ・ナビゲーションへの応用について考察する．

4．研究の方法

本研究では構築する方向導出までデプロイメントパイプラインで実装する．概略を図 3 に示す．はじめに，屋内を中心とした風景画像およびアノテーション画像を学習データとして用意する．次に，pix2pix を用いた学習の実行を行う．次に，リアルタイム処理による通路抽出を行い，最後に，通路抽出した結果から進行方向を導出する．

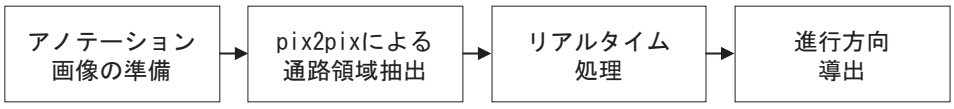


図 3 ：実装の流れ

5．pix2pix による通路抽出

5・1　学習および実行環境

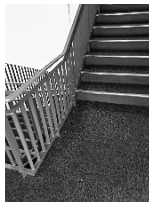
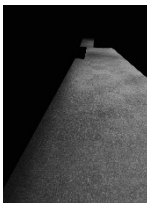
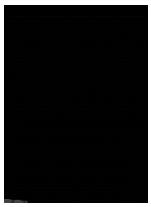
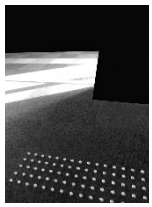
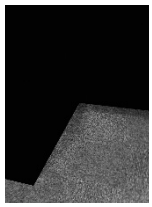
本研究では始めに Google が無償で提供している Web ブラウザ型の開発環境 Google Colaboratory を利用した．ニューラルネットワークモデルの構築には Google が開発しオープンソースで公開されている機械学習ライブラリである Tensor Flow2. 4. 1を用いて実装を行った．学習では最適化アルゴリズムとして Generator と Discriminator のどちらにも Adam を選択した．エポック数については，テスト学習をした際に得られた画像変換結果を踏まえ，100epoch とした．

その後，学習により更新されたパラメータをダウンロードし，ローカル環境でリアルタイム処理について実装を行った．

5・2　学習データセットの作成

学習データとして，久留米工業大学内で撮影した521枚の通路画像を用意した．また，それらに対応する学習時のアノテーション画像として，通路を認識するために通路以外の部分を手動でマスキングした画像を521枚作成した．これらの学習データセットの例を表１に示す．これらの入力画像と対となる変換画像の組を用いて学習を行う．図２に当てはめて説明すると，本物の通路画像（入力画像）は X として，マスク画像（変換画像）は Y として利用する．

表１：学習データセットの一例

	撮影場所A	撮影場所B	撮影場所C	撮影場所D	撮影場所E	撮影場所F	撮影場所G
入力画像							
変換画像							

さらに，入力画像の解像度を256×256pixel に適合する前処理として，過学習などの原因とされる過剰適合を回避させるために学習データセットの画像に揺らぎを与えデータ数を拡張するランダムジッタ処理を取り入れた．例えば，スケールジッタリングによるトレーニングセットの拡張が画像分類の精度を向上させる事例として Simonyan らによる研究[13]で明らかにされている．本研究では学習データの画像サイズを入力サイズより大きな高さと幅（286×286pixel）にリサイズし，そのリサイズした画像を入力サイズ（256×256pixel）にランダムにトリミングし，画像を水平方向にランダムに反転する処理を適用し，見かけ上の学習データセットの拡張を行った（図４）．

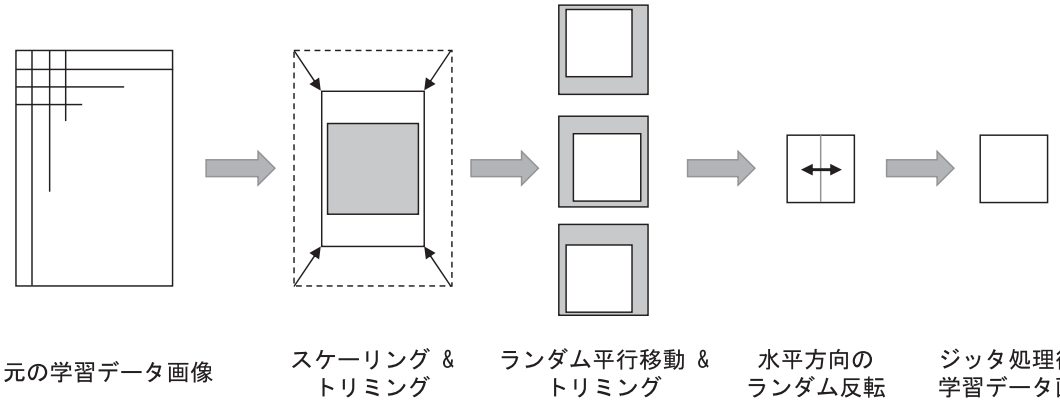


図４：アノテーション画像のランダムジッタ処理

5・3　学習モデル構造

本研究で用いた pix2pix のネットワーク構造を表 2 および表 3 に示す．活性化関数 ReLU の alpha 値は全て0. 2とした．

表 2 ：Generator の構造

Layers	Kernel	Stride	Channels	Width Height	Batch Norm	Activation
Input: Image			3	256		
Convolution: Layer 1	4 x 4	2	64	128		LeakyReLU
Convolution: Layer 2	4 x 4	2	128	64	Yes	LeakyReLU
Convolution: Layer 3	4 x 4	2	256	32	Yes	LeakyReLU
Convolution: Layer 4	4 x 4	2	512	16	Yes	LeakyReLU
Convolution: Layer 5	4 x 4	2	512	8	Yes	LeakyReLU
Convolution: Layer 6	4 x 4	2	512	4	Yes	LeakyReLU
Convolution: Layer 7	4 x 4	2	512	2	Yes	LeakyReLU
Convolution: Layer 8	4 x 4	2	512	1		LeakyReLU
Deconvolution: Layer 9	4 x 4	2	1024	2	Yes	ReLU
Concatenate (Layer 9, Layer 6)						
Deconvolution: Layer 10	4 x 4	2	1024	4	Yes	ReLU
Concatenate (Layer 10, Layer 5)						
Deconvolution: Layer 11	4 x 4	2	1024	8	Yes	ReLU
Concatenate (Layer 11, Layer 4)						
Deconvolution: Layer 12	4 x 4	2	1024	16	Yes	ReLU
Concatenate (Layer 12, Layer 3)						
Deconvolution: Layer 13	4 x 4	2	512	32	Yes	ReLU
Concatenate (Layer 13, Layer 2)						
Deconvolution: Layer 14	4 x 4	2	256	64	Yes	ReLU
Concatenate (Layer 14, Layer 1)						
Deconvolution: Layer 15	4 x 4	2	128	128		Tanh
Output: Generated Image			3	256		

表 3 ：Discriminator の構造

Layers	Kernel	Stride	Channels	Width Height	Batch Norm	Activation
Input: Input and Target Image			3	256	Yes	LeakyReLU
Concatenate (Input, Target)			6	256		
Convolution: Layer 1	4 x 4	2	64	128	Yes	LeakyReLU
Convolution: Layer 2	4 x 4	2	128	64	Yes	LeakyReLU
Convolution: Layer 3	4 x 4	2	256	32	Yes	LeakyReLU
Zero Padding			256	34		
Convolution: Layer 4	4 x 4	1	512	31	Yes	LeakyReLU
Zero Padding			512	33		
Convolution: Layer 6	4 x 4	1	1	30		
Output: Real or Generated pair						

5・4　画像変換の結果

画像変換の学習結果を表 4 に示す．また，モビリティは学習済みの場所だけではなくさまざまな環境における移動を行うため，未学習の場所における画像変換も行った．その結果については表 5 に示す．

表4：画像変換の結果.

各行ごとにそれぞれの入力画像に対する変換結果を表しており，左から入力画像（Input），変換画像（Generated），正解画像（Ground Truth），変換画像と正解画像の差分画像（Difference）を示している.

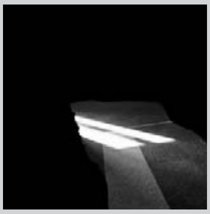
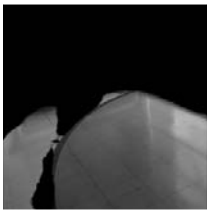
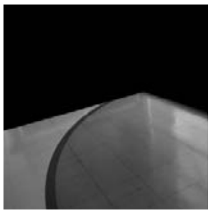
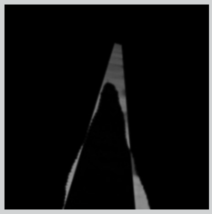
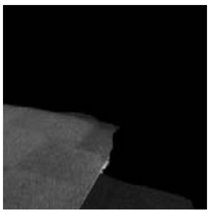
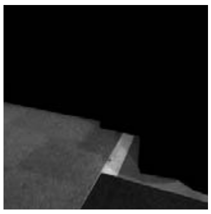
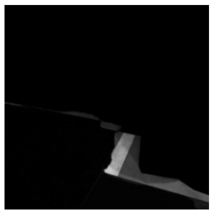
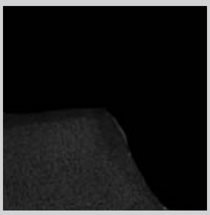
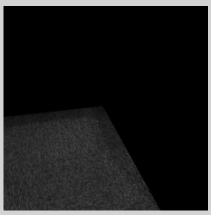
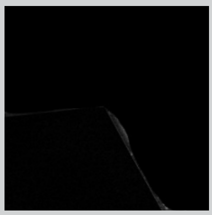
Input	Generated	Ground Truth	Difference
			
			
			
			
			
			
			

表５：未学習の場所における画像変換（画像１～４列目は生涯あんしん住宅）

５・５　考察

表２および表３で示した結果の一部は，通路の境界部分の変換精度が正確ではないが，知覚的に妥当な結果であるように見受けられる．一方で，実際には不適切な画像変換の出力もあった．これらについて例示しながら考察を行う．

・パターンＡ：壁部分の画像変換

図５左では，正面の壁が残っている．ポスター等の掲示物は学習データに含まれていないが，他の結果では窓の風景の変換や遠景の掲示物に対応できていることを踏まえると，通路と連続している部分があると通路の延長として扱っている可能性がある．また，図５右では，廊下の壁を残して画像変換される部分があった．例えば人間であれば，立体知覚の際に表面の陰影等を手掛かりとする．pix2pix でも陰影の変化も学習していると考えられるが，空間構成の認識までは行えない点は当然ながら明確である．しかし，これらの画像を90度回転すると，人間でも窓や照明といった手がかりが無ければ経験的に処理することができなくなり，壁の部分を通路として取り扱う可能性がある．物体の検出と組み合わることで，人間の経験的な認知に近づけて変換できることが考えられる．

図５：壁の誤変換例

・パターンＢ：階段や建造物部分の画像変換

階段などの段差については，学習データに含まれる本学建屋の階段（表４）については対応できていたが，未学習の場所（図６左上）では画像に対して斜め方向の部分が変換されずに残るケースが見受けられた．また，階段と通路が同様の文様（図６右上）の箇所でも変換されない箇所が生じた．階段は，局所的にみると接地面であることは事実であるが，モビリティの移動については障害となる．段差部分については，飛び地が残らないような学習を行う必要がある．また，図６左下のような屋根の構造物については，画角に対して上半分の領域は通路が存在しないとして学習データを工夫することで対応できると推測しているが，図６右下のような手すり部分についての対応は難しいであろう．

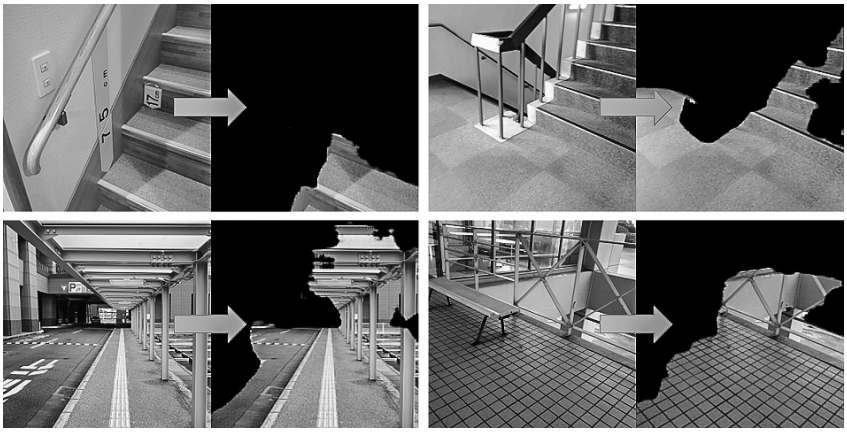


図６：立体情報に非対応の変換例

・パターンＣ：床部分の画像変換

図７は通路部分の画像変換でも特に目立った失敗例である．地面に雨水の浸潤跡や人物の影があると通路部分ではない扱いになり画像変換されている．学習データの通路部分は同じようなパターンが連続する箇所が多いため，このようなテクスチャの変化に弱いと推測している．

図７：通路部分のテクスチャの不規則な変化の変換例

・パターンＤ：類似した入力画像における変換の差異

図８は知覚的に同等の入力画像でも，画像変換の結果に違いがあった例である．左図は画像全体が変換されているが，右図では画面下部の道路部分が表示されている．入力画像における通路部分の面積が影響しているのであれば，Discriminator の特徴であるパッチサイズでの判定が影響している可能性がある．学習データをスケールジッタリングによって拡張すればもう少し精度が上がることが期待される．ただ，この入力画像のパターンは，実際にモビリティにカメラを装着した場合の画角やスケール感と大きく異なることも配慮すれば，このような入力は無いと仮定しても良い．

図８：類似した入力画像間の変換の差異

これらの４つのパターンや5. 3の結果から，pix2pix による通路抽出では，知覚的に妥当な結果が得られているが，通路のエッジ部分の処理の対応が難しいこと，通路部分と連続した壁や天井部分が通路として扱われる場合があること，通路に不規則かつある程度の大きさの模様が存在すると通路ではないとみなされることが分かった．また，入力画像に対する床面積としての大きさも変換の判定に影響している可能性がある．

５・６　リアルタイム処理の適用

実際にモビリティ・ナビゲーションとして利用する場合は，通路認識をリアルタイム処理する必要がある．そこで，車載コンピュータによる処理を想定して，pix2pix の入力画像を web カメラからストリーミング入力しローカルマシンでリアルタイム処理を行った．また，実際のナビゲーションシステムではカメラの設置場所によって必ずしも通路の中

心を的確に捉えた画像を利用できるわけではないことを考慮に入れる必要がある．そのような場合にもリアルタイムで知覚的に妥当な通路認識を行うことができるのかという点について検証した．検証方法は，移動するカメラからの映像データを画像変換モデルへリアルタイム入力し生成画像の確認を行った．

その結果，単一画像からの出力の時と同様，知覚的に高い精度で通路を出力できている場所もあるが，単一画像からの出力では見受けられなかった精度の低い通路認識の画像が存在した．例えば，日差しや照明の影響で白飛びしてしまっている画像や露光不足による黒つぶれ，カメラのパン操作（回転）による画像内の通路方向の大幅な変化時に通路認識の精度が低くなっていた．また，時々白い壁を通路と認識してしまっている場合があり，静止画入力に対する出力結果と同等の問題もあった．この要因として，静止画像の学習データで想定していなかった日差しなどの外的環境が大幅に異なる場合や Web カメラの移動によってブラーのかかった状態の画像が入力されるといった点が挙げられる．この他，特に顕著な問題として，出力画像の更新頻度が10%程度まで低フレーム化する問題（表６）があった．学習用ニューラルネットワークモデルをそのまま利用するのではなく，ネットワークを再構成し処理の最適化が必要であろう．

表６：リアルタイム処理時のフレームレート

CPU	GPU	Webcam Frame Rate	Generated Frame Rate
Intel Core i7-8700K CPU@3.70 GHz	NVIDIA GeForce GTX 1080 Ti	30 fps	3 fps

６．通路認識画像に基づく進行方向の導出

次の段階として，リアルタイム処理の通路認識画像に基づく進行方向の算出・表示について説明する．認識画像から進行方向を導く処理の手順を図９のアクティビティ図で示す．これらの処理は OpenCV（Open Source Computer Vision Library）ライブラリを用いた．はじめに，画像変換処理された画像をグレースケール変換後に２値化し，細かな飛び地やノイズを除去する．次に，輪郭検出を行い，輪郭を構成するピクセルの座標値（x，y）を取得しリスト化する（図10(b)）．次に，リストにあるピクセルから同じ高さの座標値 y を持つピクセルをペアリングし，各ペアの中心点を中心点群とする．最後に，中心点群の中から最大値・最小値 y を持つ点を結んで輪郭に対する中央分割線を描き，方向ベクトルを導出する．最後に，図10(c)のような出力結果を得た．

しかしながら，実際にリアルタイム映像に対して方向導出を実行したところ，処理が停止し対応できない場面があった．この原因として，静止画の入力とリアルタイム映像入力による検証をしたところ，pix2pix の画像変換では図11のような飛び地が発生した場合に進行方向の導出ができないと推測した．そのため，Inpainting 手法を用いて補間生成を試みたが２値画像では画像に含まれる情報が不足し統計的に解析できないこともあり欠損領域としての飛び地を復元できなかった．その他の解決方法として，web カメラのモビリティへの設置箇所が定まれば，入力画像から通路が存在すると想定される領域をパースペクティブに沿ってマスク画像で定義しておき，乗算による画像の演算を適用すること

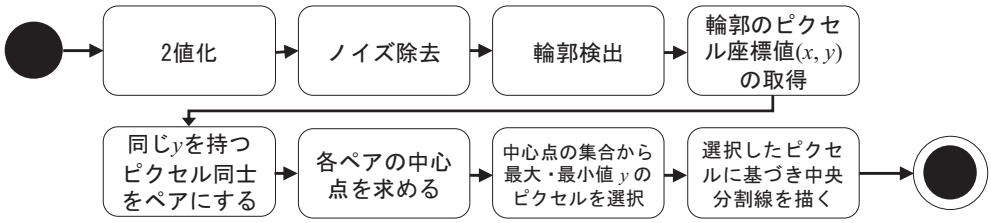


図９：進行方向の導出までの流れ

(a)入力画像　　(b)輪郭検出　　(c)方向の導出

図10：方向導出までの過程

で通路領域の絞り込みを行う方法を試みた（図12）．これにより方向導出が可能な場面も増加したが，エラー解消には至らなかった．完全なエラー解消のためには，pix2pix の出力画像の領域が連続した一つの領域になるようなモデルの構築や学習条件が必要である．

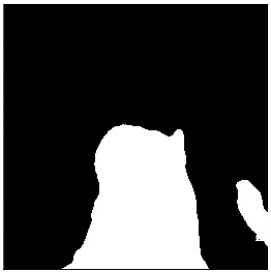


図11：飛び地の発生の例

(a)入力画像

(b)マスク画像

(c)乗算後の画像

図12：通路領域の補正例

7．まとめ

移動における通路認識として，セグメンテーションとオブジェクト検出等の手法がある．本研究では，Image-to-Image Translation の pix2pix を用いてリアルタイム処理による通路領域の抽出を行った．また，その領域から進行方向を導出することをデプロイメントパイプラインにて試みた．その結果，静止画像における画像変換については，知覚的に妥当な結果を得ることができると考えているが，リアルタイム入力では，カメラの移動によって入力画像のダイナミックレンジの変化が生じそれに対応できない場面など本番環境に影響を及ぼす問題や，処理フレームレートの低下の問題点が明らかになった．また，モビリティの進行方向の導出については，飛び地の問題に十分に対応できなかった．

今後の研究では，モビリティに設置されたカメラ入力による実稼働をとおして，正確なアノテーション画像の種類を増やし学習範囲を拡張することや各種センサーとの連携も視野に入れ，さらなる実装を進めて有効性を検証する．

謝　辞

アノテーションデータの作成やプログラム開発に協力していただいた本学大学院修了生の石橋俊二氏，中島滉一氏，および情報ネットワーク工学科卒業生の中垣海氏に感謝いたします．本研究は平成30年度文部科学省私立大学研究ブランディング事業（事業名：先進モビリティ技術で多様な人々が能力を発揮できる，Society5.0に基づく「いきいき地域づくり」）の支援を受けました．私立大学の助成を受けており，謝意を表します．

文　献

(1) 東ら，“人工知能を搭載した対話型自動運転パートナーモビリティの基本システム開発”，久留米工業大学研究報告 No. 43, pp. 2-12, 2021.
(2) Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks, *In Advances in Neural Information Processing Systems (NIPS)*, 2014.
(3) Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp.1125-1134.
(4) Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *arXiv preprint arXiv: 1506.01497*, 2015.
(5) Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, You Only Look Once: Unified, Real-Time Object Detection, *arXiv preprint arXiv: 1506.02640*, 2015.
(6) Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, SSD: Single Shot MultiBox Detector, *CoRR abs/1512.02325*, 2015.
(7) Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár, Focal Loss for Dense Object Detection, *CoRR abs/1708.02002*, 2017.
(8) Xingyi Zhou, Dequan Wang, Philipp Krähenbühl, Objects as Points, *CoRR abs/1904.07850*, 2019.
(9) Avinash Hindupur, *The GAN Zoo*, [URL] https://github.com/hindupuravinash/the-gan-zoo, 最終アクセス2021年10月26日.
(10) M. Mirza and S. Osindero, Conditional Generative Adversarial Nets, *arXiv preprint arXiv: 1411.1784*, 2014.
(11) O. Ronneberger, P. Fischer and T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, *arXiv preprint arXiv: 1505.04597*, 2015.
(12) Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp.2223-2232.
(13) Karen Simonyan and Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, *CoRR abs/1409.1556*, 2014.