移動対象の属性と環境情報を導入したLSTMによる経路予測 *

箕浦大晃** 平川 翼** 山下隆義** 藤吉弘亘**

Path Predictions by LSTM Using Object Attributes and Semantic Environment

Hiroaki MINOURA, Tsubasa HIRAKAWA, Takayoshi YAMASHITA and Hironobu FUJIYOSHI

Path prediction methods with deep learning architectures consider the interactions of pedestrians with the feature of the surrounding physical environment. However, these methods process all pedestrian targets as a unified category, making it difficult to predict a suitable path for each category. In real scenes, both pedestrians and vehicles must be considered. Predicting the path that corresponds to a target type is possible by considering the types of multiple targets. Therefore, to achieve path prediction compatible with individual categories, we propose a path prediction method that simultaneously represents the target type as an attribute and considers physical environment information. Our method inputs feature vectors that represent i) past object trajectory, ii) the attribute, and iii) the semantics of the surrounding area into a long short-term memory, making it possible to predict a proper path for each target. Experiments prove that our approach can predict a path with higher precision. Also, we analyze its effectiveness by introducing the attribute of the prediction target and the physical environment's information.

Key words: computer vision, path prediction, deep learning, long short-term memory, convolutional neural network

1. 序　　論

コンピュータビジョンで古くから人を対象とした研究が様々行われている．人を対象とした問題の一つに経路予測がある．経路予測とは，歩行者や自動車などの移動対象が未来にどのような道筋を移動するかを予測する技術である．経路予測は自動運転における交通事故の防止[1)~3)]やロボットの自律制御[4)~7)]など，将来様々な分野への応用が期待されている[8)~13)]．そのため，近年盛んに取り組まれている問題の一つである．

すでに様々な予測手法が提案されているが，特に近年では深層学習の発達により，Convolutional Neural Network (CNN)[14)]および，Long Short-Term Memory (LSTM)[15)]を用いた予測手法が数多く提案されている[7)16)~21)]．正確な予測経路の獲得には様々な情報が必要になる．例えば，歩行者同士の衝突を避けるモデルの導入[17)18)22)~25)]やシーン中のセマンティックな情報[18)26)27)]を用いることで予測精度の向上を図っている．

しかしながら，これらの予測手法では全ての予測対象を同一種類の予測対象として扱っているという問題が存在する．現実のシーンでは歩行者だけでなく，自動車や自転車などの異なる移動物体が存在する環境下で予測を行う必要がある．その際，対象の種類により移動速度や移動する領域が異なることが考えられる．

そのようなシーンにおいて，歩行者や自動車といった複数種類の対象の経路を同時に予測する場合，上記の予測手法では同一種類の対象として予測しているため，予測対象の種類に応じた経路を予測することが困難であると考えられる．

上記の問題を解決するためには，対象の種類毎にモデルを作成し，予測を行うことが考えられる．しかしながら，対象の種類が増加するにつれ，扱うモデルの数が増加するため，計算コストの面から現実的とは言えない．

そこで本研究では，予測対象の種類および，予測対象の周囲の環境情報を導入した経路予測手法を提案する．具体的には，歩行者や自動車といった予測対象の種類を対象が保有する属性とみなし，対象の属性情報を one-hot vector として表現する．そして，予測を行うシーンに付与されたシーンラベルを CNN へ入力することで，予測対象の周囲の環境に関する特徴ベクトルを抽出する．対象の移動情報に加え，属性および環境に関する特徴ベクトルを LSTM へ入力することで，出力として次の時刻の対象が存在する移動情報を得る．予測時には，ネットワークの出力を次の時刻の入力として逐次的に入力することで，異なる対象の種類に対する速度の違いや移動する領域の違いを考慮した経路予測を実現する．また，移動情報に 2 つの連続する座標の差から得られる相対座標を使用する．相対座標をネットワークへ入力することで，予測結果が学習したシーンに対する依存を防ぐことができるため，複数の異なるシーンの経路を予測することができる．

評価実験において，予測対象の種類と周囲の環境情報を導入することによる精度の変化を検証する．また，属性毎の予測精度の違いや入力するシーンラベルを変更した比較実験を行う．

本研究の新規性は，属性情報のより詳細な拡張が可能な点である．歩行者や自動車といった属性情報を one-hot vector としてコンパクトに表現することで，従来手法のように対象の種類毎にモデルを作成することはない．そのため，前述の属性だけでなく，大人や子供，男性と女性など，より詳細な属性情報を one-hot vector の次元数のみ上げることで拡張できると考えられる．

本論文は次のように構成される．まず，2.にて経路予測の関連研究について説明する．3.で提案手法の具体的な処理について説

* 原稿受付 令和 2 年 3 月 27 日
掲載決定 令和 2 年 6 月 15 日
** 中部大学大学院（愛知県春日井市松本町 1200）

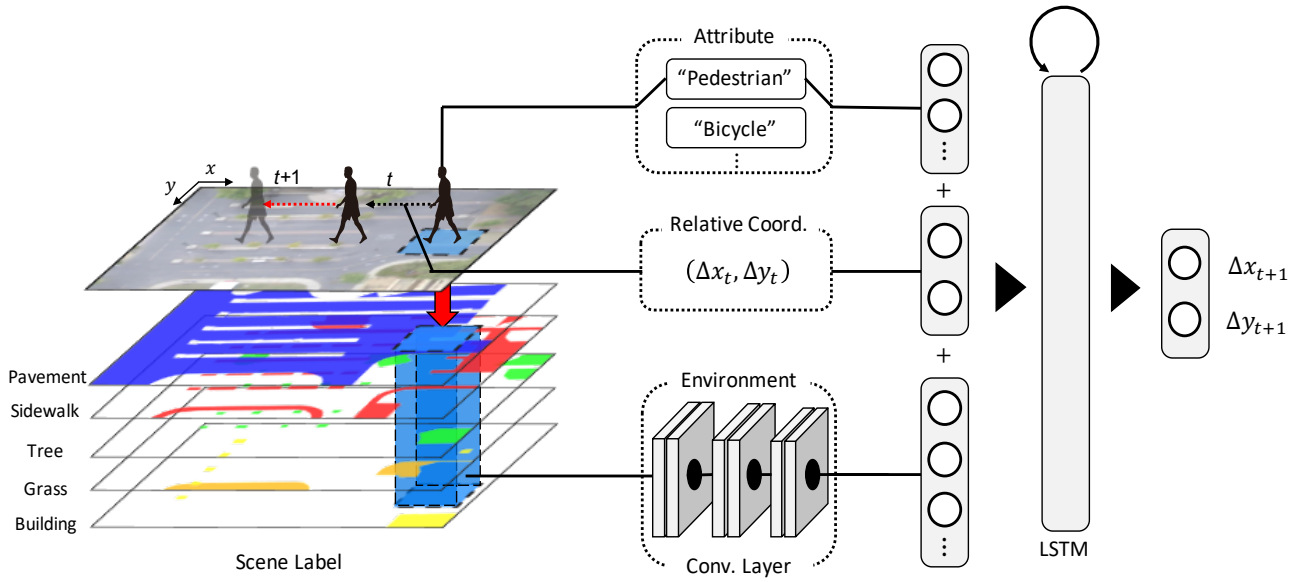

Fig. 1 The overview of the proposed method

明する．4.で評価実験について説明および考察を述べる．最後に5.でまとめる．

2. 関 連 研 究

経路予測は古くから扱われている問題であり，現在までに数多くの研究がされている．本章では，代表的な予測モデルについて説明する．

2.1 ベイズモデルに基づく予測手法

代表的なアプローチとしてベイズモデルに基づく手法が存在する [1)3)27)]．Scheneider ら [1)]は車載カメラに映った車両前方の歩行者の移動経路を予測することを目的として，拡張カルマンフィルタに基づいた経路予測手法を提案している．この手法では，歩行者の速度や加速度などの対象の移動に関する情報をいくつか変更することで，予測に必要な情報を検討している．また，Kooij ら [3)]は Dynamic Bayesian Network (DBN) [28)]を用いて，歩行者が車道を横切る際の移動を予測している．この手法では，歩行者の頭部の向きや車両との距離，縁石までの距離を対象および，周囲の環境の情報として用いている．それらを DBN の観測に使用し，歩行者が止まるか横断するかというモードを推定することで，予測精度の向上を図っている．一方，提案手法では，対象の種類を対象が保有する属性として導入し，セマンティックな領域ラベルを周囲の情報として導入する．

2.2 深層学習を用いた予測手法

近年では深層学習，特に LSTM を用いた経路予測手法が提案されている [17)~20)]．深層学習を用いた経路予測は歩行者同士の衝突を避けるインタラクション情報を考慮することで，正確な予測経路を獲得している．Alahi ら [17)]は，シーン中の複数の歩行者の移動経路を予測する手法を提案している．その際，インタラクションを考慮するために，Social-Pooling (S-Pooling)と呼ばれるプーリング層を提案している．自身の周辺に存在する歩行者の LSTM の中間層の出力を S-Pooling に入力し，その出力を次の時刻の LSTM の入力として用いることで，歩行者同士の衝突を避けるようなインタラクションを考慮した経路予測を実現している．Gupta ら [21)]は，Generative Adversarial Networks (GAN) [29)]を用いた予測手法を提案している．この手法は，未来に辿りうる経路は 1 つではない可能性から，GAN を用いることで複数の予測経路をサンプリングしている．同時に，対象間のインタラクション情報をコンパクトに表現したモジュールを提案している．Lee ら [18)]は RNN Encoder-Decoder [30)]および，Conditional Variational Auto-encoder [31)]を用いた経路予測手法を提案している．この手法では，対象間のインタラクションに加えて，周囲の環境情報を考慮することで高い予測精度を実現している．このように LSTM を用いた予測手法では，対象の周囲の環境に関する特徴ベクトルを記述し，活用することで精度の向上を図っている．一方で，提案手法では，周囲の環境に加えて対象自身の情報を導入することで予測精度の向上を図る．

2.3 属性を考慮した予測手法

対象の属性を考慮した手法として，Ma ら [25)]は歩行者の性別や年齢を考慮した手法を提案している．この手法では，歩行者の年齢および，性別によって歩行速度が異なることを仮定し，シーン中に存在する複数の歩行者の移動経路を同時に予測している．この手法では，歩行者のみが存在するシーンで予測を行なっており，環境に関する情報は用いられていない．一方，本研究では，対象の属性および，周囲の環境情報を考慮した経路予測を行う．

3. 提 案 手 法

本章では，予測対象の種類および，対象周囲の環境情報を導入した経路予測手法について述べる．提案手法のネットワーク構成を**図 1** に示す．まず，歩行者や自動車といった予測対象の情報から，その属性を表現するベクトルを生成する．次に，対象の移動情報を抽出する．本研究では，前の時刻の位置から現在までの移動量，すなわち相対座標をネットワークへの入力として用いる．そして，シーン中の予測対象を中心とする静的なシーンラベルを CNN に入力し，予測対象周囲の環境に関する特徴ベクトルを抽出する．予測対象の属性情報，移動情報および，環境に関する特徴ベクトルを連結し，LSTM へ入力する．LSTM の出力として，次の時刻の移動情報を出力する．この時，LSTM は内部状態を記憶するメモリセルの働きにより，過去の情報を保持することがで

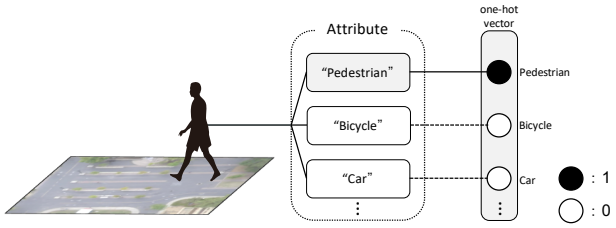

Fig. 2 The representation of the attribute of a target object

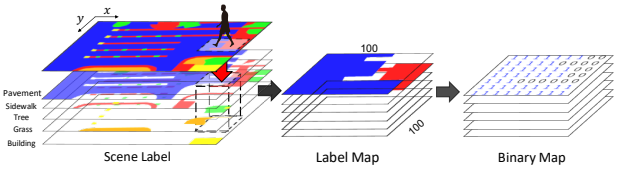

Fig. 3 The representation of the physical environment surrounding a prediction target

きる. そのため, LSTM は上記の情報を逐次的に入力することで, 予測対象の将来の位置を予測することが可能となる.

提案手法では，対象の属性情報，過去の移動情報および，環境情報を抽出するための表現方法が重要となる．以下に，各入力情報の表現方法および，ネットワークへの入力について説明する．

3.1 問題設定

本研究では，シーン中の複数の対象種類を予測することを目的とする．ネットワークには予測対象$i = 1, 2, \ldots, N$の観測時刻$t = 1, 2, \ldots, T_{obs}$の移動経路を入力し，予測時刻$t = T_{obs} + 1, T_{obs} + 2, \ldots, T_{pred}$における予測経路を出力する．

3.2 属性情報

複数の対象種類の経路を予測するために，入力に対象の種類に関する追加情報を導入する．対象の種類をネットワークへ入力するために，本研究では対象の属性情報を one-hot vector で表現する(図 2)．具体的には，与えられた対象の属性情報をベクトル$\boldsymbol{e}$に埋め込まれる．具体的な処理は，属性数をnと表した時，第n成分のベクトルの要素e_nを 1，それ以外を 0 とすることで予測対象の属性を表現する. 例えば図 2 のように属性情報が pedestrian の場合，第 1 成分のベクトルの要素e_1に 1，それ以外の要素を 0 とする．この処理により，対象の属性毎に LSTM を構築する必要がなく，コンパクトな表現にできる．また，このベクトルを入力すると速度と方向に関して一意の経路を予測することが可能になる．さらに，この one-hot vector と静的環境を表す特徴ベクトルと組み合わせることにより，対象が移動する傾向がある領域も考慮できる．そのため，対象の種類に応じた経路を予測することが期待できる．

3.3 移動情報

複数のシーンの予測を行う場合，予測シーン毎に建物や木などの障害物，歩道や車道などの移動領域が異なる．そのため，移動情報に過去から現在までの位置情報，すなわち相対座標を使用する．相対座標は式(1)で定まる．

$$\Delta x_t = x_t - x_{t-1}$$
$$\Delta y_t = y_t - y_{t-1} \tag{1}$$

ここで，tは現時刻，$t-1$は前時刻を示す．式(1)で求めた相対座標を LSTM に入力することで, 次の時刻の相対座標を取得することができる．相対座標を用いることで，対象の現在地を常に基点とすることができるため，建物などの障害物や歩道などの移動領域の位置情報に依存することなく，複数のシーンでの経路の予測を行うことが可能となる．

Table 1 The detailed network architecture of the proposed method

layer	kernel size	output size	remarks
input (attribute)		6	
input (coordinate)		2	
input (environment)		(100, 100, 7)	
conv1	(5, 5)	(48, 48, 16)	ReLU, stride=2
norm1		(48, 48, 16)	batch norm.
pool1	(2, 2)	(24, 24, 16)	max pool.
conv2	(5, 5)	(20, 20, 32)	ReLU, stride=1
norm2		(20, 20, 32)	batch norm.
pool2	(2, 2)	(10, 10, 32)	max pool.
conv3	(5, 5)	(6, 6, 32)	ReLU, stride=1
pool3	(2, 2)	(3, 3, 32)	max pool.
concat		296	
LSTM		128	
ouput		2	

3.4 環境情報

環境情報は予測精度を向上させるためにも不可欠な要素である. 従って，歩道や建物などのシーンに付与されたセマンティックなシーンラベルを用いて環境情報に関する特徴マップの抽出を行う. 環境情報の導入手順を図 3 に示す．はじめに，シーンラベルから予測対象を中心とする 100 x 100 pixel の領域ラベルを抽出する. 次に，抽出したシーンラベル毎に分割したラベルマップを作成する．そして，ラベルマップを 0 または 1 で表現されたバイナリマップに変換する．作成したバイナリマップを CNN へ入力することで，環境に関する特徴ベクトルV_tを抽出する．環境情報に関する特徴は式(2)で定まる.

$$V_t = CNN(I_t ; W_{cnn}) \tag{2}$$

ここで，I_tはクロップしたバイナリマップ，W_{cnn}は重みパラメータを示す．求めた環境に関する特徴ベクトルをネットワークへ入力することで，障害物の有無や対象の属性に応じた領域を考慮した経路予測が可能になる.

3.5 ネットワークへの入力方法

上記の属性情報，移動情報および，環境情報を LSTM へ入力することで, 次の時刻における対象の位置を予測する. 具体的には, 予測対象の各情報を観測データとして用いることで予測を行う. 観測データは予測対象が実際に移動した真値を用いる．LSTM に

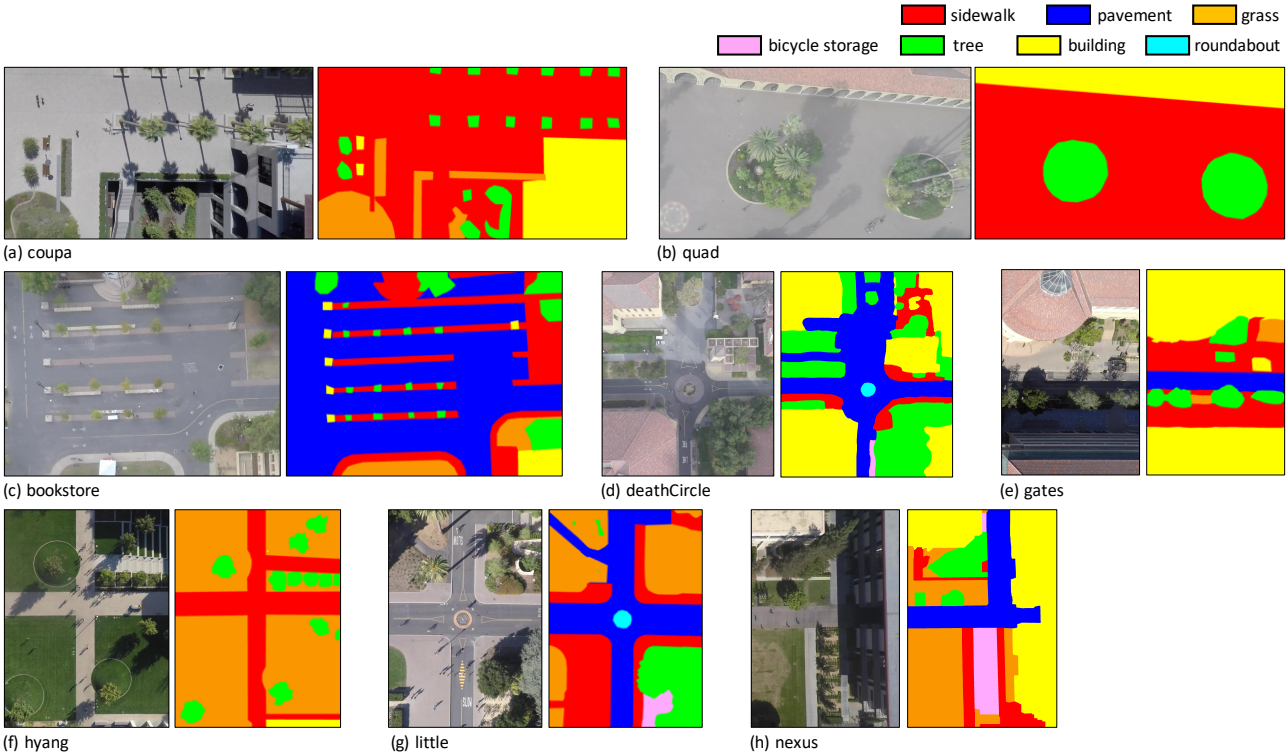

Fig. 4 Examples of original scene (left) and annotated scene labels (right) in the SDD

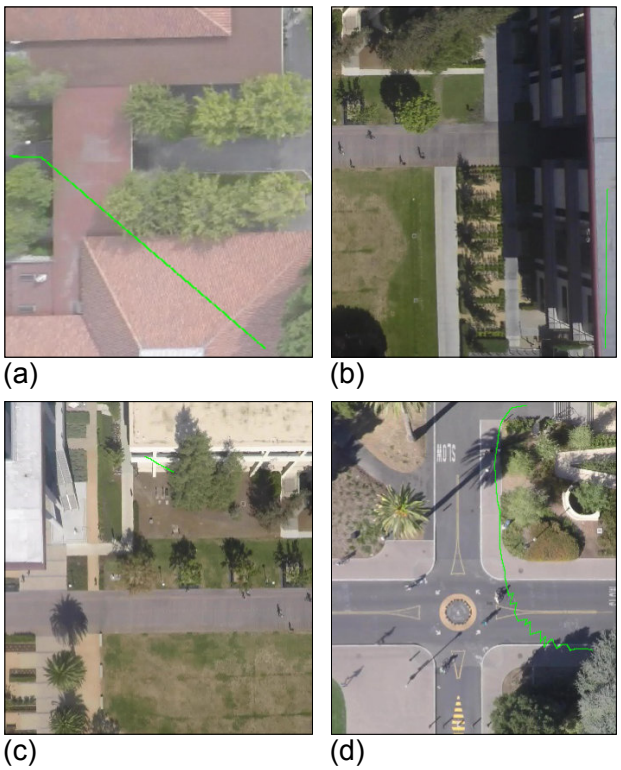

Fig. 5 Examples of incorrect annotations in the SDD

は観測データを予測開始直前のフレームまで逐次的に入力する．予測時には LSTM の出力である予測値を次の時刻の入力として逐次的に入力する．その処理を予測終了時刻まで行い，予測の実現をする．式(3)に予測経路の算出式を示す．

$$h_t = LSTM\left((\Delta x_t, \Delta y_t), \boldsymbol{e}, V_t, h_{t-1}\,; W_{LSTM}\right) \tag{3}$$

$$[\Delta x_{t+1}, \Delta y_{t+1}] = \phi(h_t\,; W_c)$$

Table 2 Training and test data details

		train	test
Number of scenes		52	8
attribute	bicycle	2,369	545
	pedestrian	2,696	500
	cart	71	15
	car	75	5
	bus	17	2
	skateboarder	137	15

ここで，hは LSTM の出力ベクトル，ϕは単一の全結合層を示す．**表 1** にネットワーク構成を示す．表 1 より提案手法のネットワーク構成は 3 層の CNN と 1 つの LSTM および，単一の全結合層から成るシンプルな構造である．convolution layer で環境に関する入力を受け取る．そして，convolution layer から抽出された特徴ベクトルと属性に関するベクトルおよび，経路情報を LSTM へ入力し，次時刻の予測経路を出力する．

4. 評 価 実 験

提案手法の有効性を評価実験により検証する．具体的には，属性および，環境情報の有無による予測精度の変化について検証する．

4.1 データセット

データセットには，Stanford Drone Dataset (SDD) [24]を用いる，SDD は bookstore, coupa などの 8 つの異なる予測シーンから構成される．各シーンには異なる日時で撮影された動画が複数含まれており，合計で 60 本の異なる動画から構成される．**図 4** に予測シーンの例を示す．

Table 3 Quantitative results for prediction methods (unit : pixel)

Metric	KF	S-LSTM	trajectory	trajectory + attribute	trajectory + environment	trajectory + attr. + env.
Final disp. error	174.42	206.22	196.13	173.04	172.12	**109.44**
Avg. disp. error	116.02	125.41	86.42	76.32	76.01	**53.20**

提案手法では，環境情報を表現するために，シーンラベルを入力として用いる．しかしながら，SDD にはシーンラベルの情報は含まれていないため，全てのシーンに対してシーンラベルを付与した. 付与したシーンラベルの種類は sidewalk, pavement, grass, bicycle storage，tree，building，roundabout の合計 7 種類である．図 4 に付与したシーンラベル例を示す．各図の左はオリジナルのシーン例，右図はオリジナルのシーン例をアノテーションしたシーンラベル例を示す．シーンラベルはシーンの視覚的な外観だけでなく，歩行者の移動経路などに基づいて，右上の注釈に従ったクラスをシーン毎に慎重にアノテーションを行った．

SDD にはいくつかの不正確な経路が含まれている. **図 5** に不正確な経路例を示す．緑線はアノテーションされた経路を示す．これらの例では，建物上または建物を横切る経路(図 5(a, b, c))や不正確にアノテーションされた経路(図 5(d))を示す．このような経路を学習および評価に使用すると，予測精度が大きく低下し，公正な比較が困難になる．そのため，本研究では正確で正しい経路のみを選択して使用する．

SDD は bicycle，pedestrian，cart，car，bus，skateboarder の異なる 6 種類の移動対象と座標データから構成される．予測対象の数は学習用に 5,365，評価用に 1,082 を使用する．使用するデータの内訳を**表 2** に示す. SDD はフレームレート 30fps で撮影されており，本実験では 20 フレーム毎の座標データを用いて実験を行う．すなわち，予測の 1 ステップは約 0.66 s に対応する．本実験では，このうちの 5 ステップ，すなわち約 3.3 秒間を観測として用いる．また，8 ステップ，すなわち約 5.3 秒間を予測として用いる．

4.2 評価指標と比較手法

定量的評価のために，本実験では 2 種類の評価指標を用いる．1 つ目は予測の最終フレームにおける真値と予測値のユークリッド距離である Final Displacement Error (FDE)，2 つ目は各予測フレームにおける真値と予測値のユークリッド距離の平均である Average Displacement Error (ADE)を用いる. FDE および, ADE は式(4)，式(5)で表される．

$$FDE(i) = \sqrt{(\hat{x}_i^T - x_i^T)^2 + (\hat{y}_i^T - y_i^T)^2} \tag{4}$$

$$FDE = \frac{1}{n}\sum_{i=1,2,\cdots,n} FDE(i)$$

$$ADE(i) = \frac{1}{T}\sum_{t=1,2,\cdots,T} \sqrt{(\hat{x}_i^t - x_i^t)^2 + (\hat{y}_i^t - y_i^t)^2} \tag{5}$$

$$ADE = \frac{1}{n}\sum_{i=1,2,\cdots,n} ADE(i)$$

ここで，真値は(x, y)，予測座標は$(\hat{x}, \hat{y})$，時刻はt，予測対象はiである．各指標はシーン内の予測対象i毎に求められる．

また，比較手法として，状態空間モデルを用いて内部の状態を効率的に推定するカルマンフィルター (KF) [32]および，深層学習モデルの代表的な手法である Social LSTM (S-LSTM) [17]をベースラインとして用いる．

4.3 実験条件

学習条件として，最適化手法に RMSprop [33]を用いる．RMSprop の初期学習率を 0.01，α=0.99，ε=10e-8 として学習する．また，全ての予測モデルはバッチサイズを 10 に設定し，100 エポックで学習する．学習時には各時刻，すなわち観測の開始時刻から予測最終時刻までを通して過去の対象の経路として真値をネットワークに入力する．予測時には，観測最終時刻で得られた最初の予測時刻を逐次ネットワークへ入力し予測値を得る．損失関数を真値と予測値との平均二乗誤差とする．フレームワークは Chainer，GPU に Nvidia Titan Xp を利用し End-to-end で学習および，評価する．

4.4 評価結果

本節では，予測対象の属性および，環境情報を導入した場合の予測精度の変化について確認する．予測精度の比較結果を**表 3** に示す．表 3 より，属性と環境情報のどちらか一方をネットワークへ入れることで予測精度が向上している．また，属性と環境情報の両方をネットワークへ導入した提案手法が最も予測精度が向上していることが確認できる． **図 6** に予測結果例を示す．KF の予測経路では，過去の経路が観測とみなされるため，予測結果は障害物領域の有無に限らず線形の予測となる．LSTM ベースの予測手法では，pedestrian の予測経路はどの手法も真値と似た経路が得られた(図 6(d, e, f))．しかしながら，経路のみを利用した予測結果では，KF の予測結果より予測経路が正確に獲得できていないことが確認できる(図 6(a, c, g))． また，表 3 より S-LSTM は KF より予測精度が低下している．結果を再現するために，パラメータを慎重に選択したが，妥当な結果を得ることができなかったため, 図 6 には S-LSTM の予測結果を記載していない. S-LSTM の結果について，同じ報告が [21]で報告されている．

表 3 に示すように，LSTM に他の補助情報を導入することで，予測精度の向上が確認できる．特に，静的な環境情報を導入することで，障害物を避けた経路を正確に予測することができる(図 6(h))．属性と環境情報を追加した手法では，定量的評価の観点からの改善は KF と比較しても小さいが，提案手法である trajectory+attr.+env.は，他の手法と比較し精度向上が確認できる．また，提案手法では，図 6(a, b, c, g) の真値に近い経路を予測していることが確認できる．図 6(d, e, f)は pedestrian の経路を示している．この場合で得られた結果は，全ての経路予測手法が bicycle よりも歩行間隔が狭く pedestrian の経路を容易に予測できるため, 真値に似た経路を追跡できることを示している. 図 6(g)は，予測対象が車道に沿って進行する car の経路を示している．ただし，KF，trajectory，trajectory+attribute，trajectory+environment がネットワークの入力として使用され

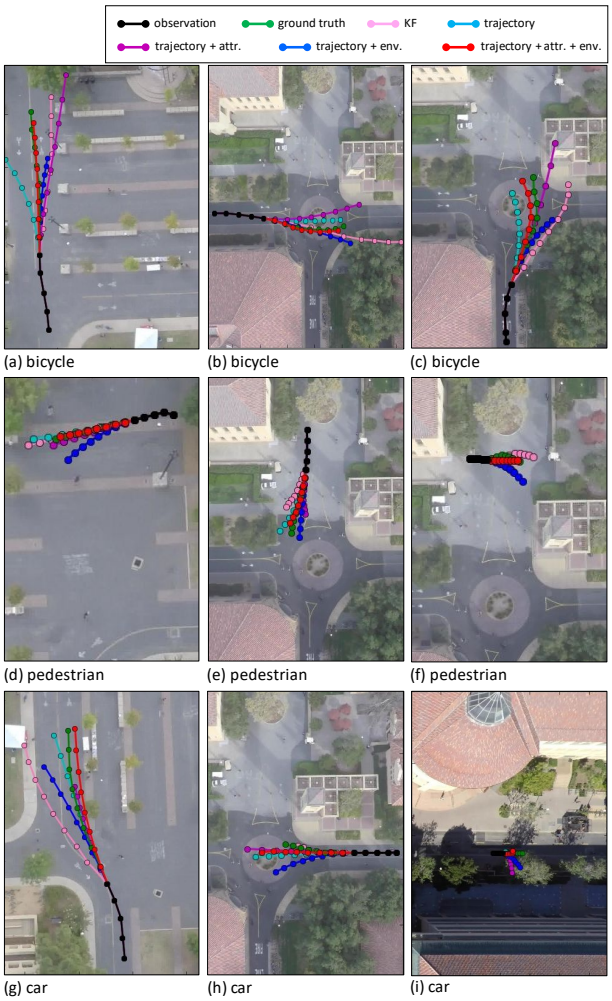

Fig. 6 Examples of prediction results of each method on SDD

Table 4 Quantitative results of each attribute

input	Final disp. error	Avg. disp. error
bicycle	113.82	51.25
pedestrian	43.22	23.14
cart	85.52	53.68
car	129.53	58.68
bus	151.74	76.34
skateboarder	132.79	61.67

る場合，予測結果は直進している．一方で，環境と属性を同時に導入することで，真値に似た経路を予測することができる．ただし，図 6(h, i)に示すように，環境情報を導入すると，予測結果は真値とは異なる予測経路を獲得する場合もある．

以上の結果から，本研究の手法は比較した経路予測手法の中で最も精度が高いことが確認できる．KF は線形に辿る経路の予測の精度は良いが，障害物回避の場合のような非線形経路を予測することは困難である．経路を正確に予測するには，属性と環境情報を対象の経路に導入する必要がある．

4.5 異なる対象毎の評価結果

表 4 に，属性と環境情報を導入するときの各対象に関する予測誤差を示す．表 4 は属性と環境情報を両方考慮した結果を示す．表 4 より，bicycle，car および，skateboarder は他の属性より速く移動するため，予測誤差は大きい．また，表 1 に示すように，cart，car，bus，skateboarder のサンプルは他のサンプルよりも少ないため，予測誤差が大きくなる傾向が見られる．

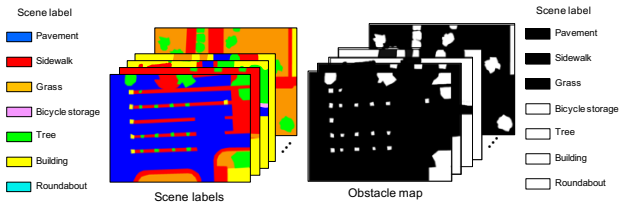

Fig. 7 Examples of obstacle map

Table 5 Quantitative results introducing obstacle map or scene

input	Final disp. error	Avg. disp. error
Obstacle map	130.12	59.42
Scene label	**109.44**	**53.20**

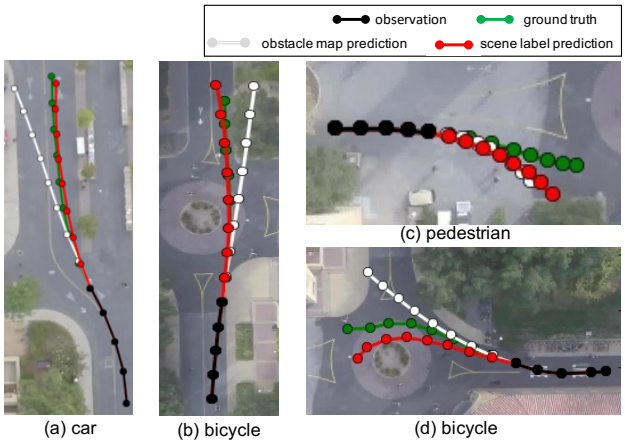

Fig. 8 Examples of prediction results with different environment information

4.6 入力が異なるシーンラベルを用いた検証実験

環境情報としてアノテーションされたシーンラベルを使用して実験を行なった．セマンティックなラベルの有効性を評価するために，アノテーションされたシーンラベルを障害物エリアとして設定した bicycle storage，tree，building および，roundabout と他の領域に分割して作成した obstacle map を使用する実験を行う．**図 7** に obstacle map の例を示す．obstacle map は 1 に設定された障害物領域と 0 に設定された移動可能領域のバイナリマップとして表現される．異なる環境情報をネットワークへの入力として用いることで，将来の経路を予測するのに適した環境情報を分析する．

入力が異なるシーンラベルを用いた定量的評価結果を表 5 に示し，予測結果例を**図 8** に示す．**表 5** は属性と環境情報を両方考慮した結果を示す．表 5 より，シーンラベルが obstacle map よりも優れていることを示す．図 8(a)の car の予測結果では，obstacle map を使用すると，sidewalk に向かって移動する経路を予測していることを示す．また，図 8(b)の bicycle の予測結果では，grass に向かった経路を予測していることが確認できる．こ

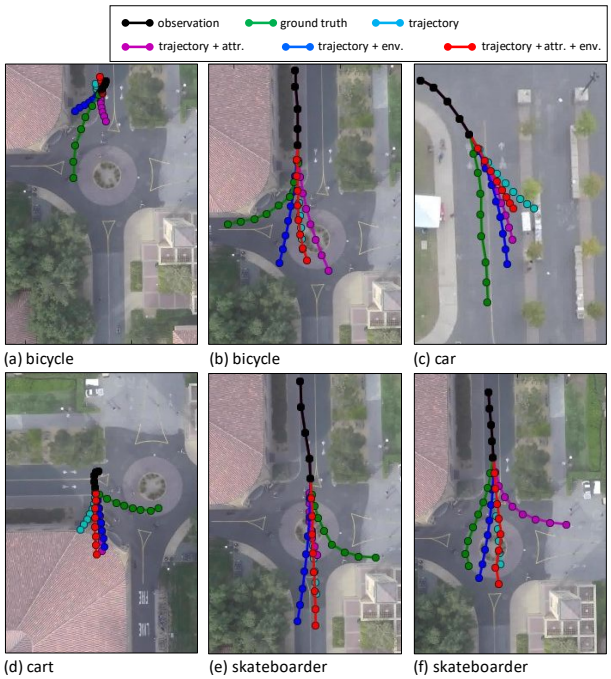

Fig. 9 Selected failed prediction results

れらのような予測結果となったのは，obstacle map が障害物領域のみを区別するため，対象によっては sidewalk や grass など通常移動することがない領域を移動可能領域として予測してしまう．その結果，正確な経路の予測が困難になったと考えられる．

4.7 誤った予測結果例

図 9 に，適切な予測が行えなかった例を示す．図 9(a)より，観測データが移動間隔の狭い移動経路から急激に速度が変化する場合では，急速な動きに対応することができず真値と異なる動きを予測している．また，図 9(b)では観測データが直進しているため，予測結果も直進した動きを予測している．このことから，分岐した経路の予測は困難であると考えられる．予測対象が cart の場合を図 9(c, d)に示す．図 9(d)では，真値は左折しているが，予測結果は建物方面に直進していることが確認できる．これは，cart のデータサンプルが少なく，学習が不充分だったため建物に衝突する動きを予測したと考えられる．予測対象が skateboarder の場合を図 9(e, f)に示す．どちらの予測結果も障害物上を移動していることが確認できる．これは，cart の場合と同様でデータのサンプル数が少なかったため，学習が不充分で障害物に衝突する動きとなったと考えられる．

以上より，移動対象の突発的な動きや分岐した経路の予測には対応できず，真値とは異なる予測結果となる問題が確認された．突発的な動きについて，本研究の手法が予測対象毎に特徴的な動きを考慮した経路予測を実現できるが，予測対象同士の衝突に関するインタラクション情報は考慮していないため，突発的な行動の予測に対応できないためだと考えられる．そのため，予測対象同士の衝突を防ぐインタラクション情報をネットワークへ導入することで，これに対処する．また，データ数が少ない属性の場合における経路予測は学習が不充分であったため，予測するのは困難となる問題が確認された.このようなデータに対応するためには，幾何変換などでデータを増やす Data Augmentation を活用することで，上記の問題を解決できると考えられる．

5. 結　　言

本研究では，予測対象の属性および，周囲の環境情報を導入した経路予測手法を提案した．提案手法では，予測対象の属性を one-hot vector で表現し，周囲の環境情報のシーンラベルを CNN へ入力することで表現した．これらの情報を LSTM へ逐次的に入力することで，予測を実現した．SDD を用いた評価実験により，位置情報のみを入力する場合と比較して属性および，環境情報を導入した場合における予測精度が高い結果となった．しかしながら，突発的な行動やデータサンプルが少ない属性では，充分な学習を行うことができず，適切な予測を行うことが困難であった．

今後の課題として，属性情報を拡張することによる応用，それに伴いデータセットを新たに作成することで，より現実的な問題に対処する．また，予測対象同士の衝突を防ぐインタラクション情報，すなわち動的な環境を考慮することが挙げられる．

謝　　辞

この研究は JSPS 科研費助成金番号 JP16H06540 によって部分的に支援された．そして，この研究に使用された Titan Xp GPU の寄付による NVIDIA Corporation のサポートに感謝します．

参 考 文 献

1) N. Schneider and D.M. Gavrila. Pedestrian path prediction with recursive bayesian filters: A comparative study. In Gearman Conference on Pattern Recognition, (2013) 174.
2) C. G. Keller and D. M. Gavrilla. Will the pedestrian cross? A study on pedestrian path prediction. IEEE Transactions on Intelligent Transportation Systems, **15**, 2, (2014) 494.
3) J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila. Context-based pedestrian path prediction. In European Conference on Computer Vision, (2014) 618.
4) Br. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa. Planning-based prediction for pedestrians. In Internarional Conference on Intelligent Robots and Systems, (2009) 3931.
5) V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto. Intent-aware long-term prediction of pedestrian motion. In International Conference on Robotics and Automation, (2016) 2543.
6) A. Vemula, K. Muelling, and J. Oh. Modeling cooperative navigation in dense human crowds. In International Conference on Robotics and Automation, (2017) 1685.
7) A. Vemula, K. Muelling, and J. Oh. Social attention: Modeling attention in human crowds. In International Conference on Robotics and Automation, (2018).
8) E. Rehder and H. Kloeden. Goal-directed pedestrian prediction. In Workshop on International Conference on Computer Vision, (2015) 139.
9) S. Huang, X. Li, Z. Zhang, Z. He, F. Wu, W. Liu, J. Tang, and Y. Zhuang. Deep learning driven visual path prediction from a single image. IEEE Transactions on Image Processing, **25**, 12, (2016) 5892.
10) D. Xie, S. Todorovic, and S. C. Zhu. Inferring 'Dark Matter' and 'Dark Energy' from videos. In International Conference on Computer Vision, (2013) 2224.
11) J. Walker, A. Gupta, and M. Hebert. Path to the future: Unsupervised visual prediction. In Computer Vision and Pattern Recognition, (2014) 3302.
12) H. S. Park, J. J. Hwang, Y. Niu, and J. Shi. Egocentric future localization. In Computer Vision and Pattern Recognition, (2016) 4697.
13) S. Su, J. P. Hong, J. Shi, and H. S. Park. Predicting behaviors of basketball players from first person visions. In Computer Vision and Pattern Recognition, (2017) 1502.
14) Y. Lecun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel. Backpropagation Applied to handwritten Zip Code Recognition. Neural Computation, **1**, 4, (1989) 541.
15) S. Hochreiter. LONG SHORT-TERM MEMORY. Neural Computation, **9**, 8, (1997) 1735.
16) S. Yi, H. Li, and X. Wang. Pedestrian behavior understanding and prediction with deep neural networks. In European Conference on Computer Vision, (2016) 263.
17) A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social LSTM: Human Trajectory Prediction in Crowded Spaces. In Computer Vision and Pattern Recognition, (2016) 961.
18) N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. S. Torr, and M. Chandraker. Desire: Distant future prediction in dynamic scenes with interacting agents.

In Computer Vision and Pattern Recognition, (2017) 336.
19) T. Fernando, S. Denman, S. Sridharan, and C. Fookes. Soft+hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection. arXiv preprint arXiv:1702.05552, (2017).
20) T. Fernando, S. Denman, A. McFadyen, S. Sridharan, and C. Fookes. Tree memory networks for modeling long-term temporal dependencies. arXiv preprint arXive:1703.04706, (2017).
21) A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In conference on Computer Vision and Pattern Recognition, (2018) 2255.
22) D. Helbing and P. Molnar. Social force model for pedestrian dynamics. Physical review E, **51**, 5, (1995) 4282.
23) K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg. Who are you with and where are you going? In Computer Vision and Pattern Recognition, (2011) 1345.
24) A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes. In European Conference on Computer Vision, (2016) 549.
25) W. C. Ma, D. A. Huang, N. Lee, and k. M. Kitani. Forecasting inteactive dynamics of pedestrians with fictitious play. In Computer Vision and Pattern Recognition, (2017) 774.
26) K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert. Activity forecasting. In European Conference on Computer Vision, (2012) 201.
27) L. Ballan, F. Castaldo, A. Alahi, F. Palmieri, and S. Savarese. Knowledge transfer for scene-specific motion prediction. In European Conference on Computer Vision, (2016) 697.
28) J. W. Robinson and A. J. Hertemink. Non-stationary dynamic Bayesian networks. In Advances in Neural Information Processing Systems, (2009) 1369.
29) L. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. In Advances in Neural Information Processing Systems 27, (2014) 2672.
30) K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv: 1406.1078, (2014).
31) D. P. Kingma, D. J. Rezende, S. Mohamed, and M. Welling. Semisupervised leraning with deep generative models. In Advances in Neural Information Processing Systems, (2014) 3581.
32) R. E. Kalman. A new approach to linear filtering and prediction problems. Journal of basic Engineering, **82**, 1, (1960) 35.
33) T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the gradint by a running average of its recent magnitude. COURSERA: Neural networks for machine learning, (2012).