

# ディープニューラルネットワークを用いた滞留時間の定量化

Marcelo Marques Ribas\*, Heloisa Benedet Mendes\*, Luiz Eduardo de Oliveira\*,  
Luiz A. Zanolensi†, Paulo Lisboa de Almeida\*

\*Departamento de Informática (DInf), Universidade Federal do Paraná, Curitiba, PR - Brazil  
{marcelomarques, heloisamendes, luiz.oliveira, paulorla}@ufpr.br

†DeepNeuronic, Covilhã - Portugal  
luiz.zanolensi@deepneuronic.com

概要-スマートシティでは、空間の回転性を高め、個々の交通ソリューションの使用を抑制するために、与えられた駐車スペースの最大滞在時間を定義することが一般的である。しかし、画像から個々の車の滞留時間を自動的に決定することは、低解像度カメラから収集された画像、照明のバリエーション、天候の影響などの課題に直面している。本研究では、2つのディープニューラルネットワークを組み合わせ、駐車場における各車の滞留時間を計算する手法を提案する。提案手法は、まず、深い分類ネットワークを用いて、占有と空の間の駐車スペースの状態を定義する。次に、シャムネットワークを用いて、駐車している車が前の画像と同じかどうかをチェックする。データセット横断シナリオに焦点を当てた実験プロトコルを用いて、完全な分類器を使用した場合、提案システムは完全な滞留時間予測の75%を生成し、予測値は車が駐車した時間に正確に一致することを示す。それにもかかわらず、我々の実験は、駐車スペースの状態を予測するために実世界の分類器を使用した場合、予測品質が低下することを示し、完全な予測の49%に達する。これは、提案されたシャムネットワークが有望であるが、パイプラインの最初に使用される分類器の品質に影響されることを示している。索引用語-深層学習、スマートシティ、シャムネットワーク。

## I. INTRODUCTION

画像による駐車スペース監視に関する開発は、ベンチマークとして使用するデータセットの開発[1]~[3]、個々の駐車スペースを空/空に分類するディープラーニングベースのアプローチ[2]~[4]、駐車スペースの位置の自動セグメンテーション[5]、[6]など、近年提案されている。

このようなディープラーニングに基づくイノベーションの系に続き、本稿では、車が駐車スペースに駐車している時間をカウントするアプローチを提案する。そのために、カメラからの画像を使用し、シャムネットワークを使用してこれらの画像を処理する。各車が駐車している時間に関するデータを収集することで、例えば、都市の駐車場の利用状況について有益な情報を得ることができる。

さらに、最大滞在時間(例えば15分)の駐車スペースを設けることは一般的な慣行であり、運転者が駐車している時間をカウントするシステムは、当局が違反者を検知するのに役立つ可能性がある。

本研究は、ブラジル国立科学技術開発評議会(CNPq)-助成金405 511/2022-1、およびブラジル高等教育委員会(CAPES)-財政法001の支援を受けている。

このようなシステムの他の用途としては、放棄された車や壊れた車(例えば、長く駐車している車は放棄されたり壊れたりすることがある)の検出や、違法に駐車している車(例えば、駐車スペースとして定義されていないエリアで長時間静止したままの車)の検出などがある。

駐車場時間の定量化は複雑な問題である。1 - 車両の画像はしばしば遠距離で収集され、低解像度のカメラが設置される(例えば、PKLotデータセットでは、車のバウンディングボックスは平均で56×51ピクセル)、2 - 車両はカメラに対していくつかの位置に駐車することがある、3 - 画像はしばしば低フレームレートで収集され、連続した画像収集の間に何分も経過することがある(例えば、PKLotデータセットでは、5分ごとに1枚の画像が撮影される)、4 - lumの変化の非対称性、オクルージョン、影などにより、異なる画像間で車両を比較することが困難になる可能性がある。図1の同じ車が異なる時間に同じ位置に駐車している例を見てみましょう。図2は、駐車場における様々な車角度と輝度の違いを示している。

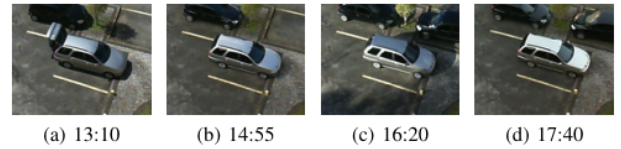


図1. 異なる時間に同じ位置に駐車した車 - PKLotデータセット。

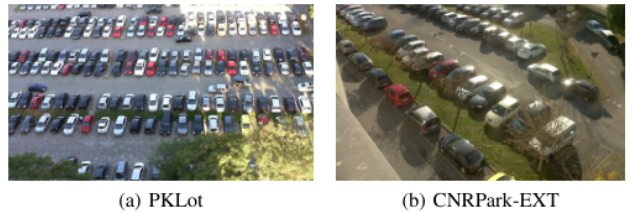


Fig. 2. Image examples from the PKLot and CNRPark-EXT datasets.

本研究では、以下の貢献に焦点を当てる：

- 自動車と比較するためのシャムネットワークを定義する。
- 駐車場の状態を確認し、各車の滞留時間を検証するための完全なパイプラインを定義する。

完全なパイプラインには、駐車スペースを分類する分類ネットワーク、車を比較する畳み込みニューラルネットワーク(CNN)ベースのシャムネットワーク、駐車している各車の滞留時間を更新するために両ネットワークの情報を結合するアルゴリズムが含まれる。

本研究の残りの部分は以下のように構成されている：セクションIIで関連研究を示す。提案するアプローチは、使用される分類とシャムネットワークの説明と、車の滞留時間を計算するために使用されるアルゴリズムを含む、セクションIIIで与えられる。セクションIVでは、実験プロトコルを定義し、ターゲット駐車場からの訓練サンプルがシステムに対して与えられないクロスデータセットシナリオを定義する。セクションVでは、提案手法がPKLotデータセットとCNRPark-EXTデータセットにおいて、それぞれ46分と54分の平均絶対誤差(MAE)に達することを示す。最後に、セクションVIで結論を述べる。

## II. 関連研究

自動車が特定の場所に駐車している時間を計算するには、物体の検出と追跡という2つのコンピュータビジョンタスクを使用する必要がある。ディープラーニングに基づく物体検出は、コンピュータビジョンで最も広く採用されているタスクの1つである。1段階法と2段階法の2種類に分類される。YOLO[7]に代表される1段階法は、関心領域(ROI)の選択に領域提案ネットワーク(RPN)[8]を使用することを控える。その結果、単一のCNNネットワークによって生成されたオブジェクトの位置で、より高速であることが示された。対照的に、Faster R-CNN [8]やMask R-CNN [9]のような2段階の手法は、ROI生成のためにRPNネットワークに依存しており、典型的にはより正確なバウンディングボックスをもたらす。

ディープラーニング技術はトラッキングタスクでもよく使われ、単一オブジェクトトラッキング(SOT)と複数オブジェクトトラッキング(MOT)の2つのカテゴリに大別される。重要な違いは、MOTではターゲットシーン内に複数のオブジェクトが存在することである。その結果、モデルはオブジェクトのオクルージョンや類似した外観を持つオブジェクトの存在など、様々な複雑性に対処しなければならない。

物体追跡の進化は、3つの段階を経て定義することができる。2000年頃に起こった初期段階では、主に古典的なアルゴリズムと機械学習がターゲット追跡に応用されるのを目撃した。これらのアルゴリズムは、低い計算複雑性、迅速な実行、最小限のハードウェア要件といった特徴を誇っていたが、そのロバスト性と精度は比較的低かった。

2010年から2016年にかけて、物体追跡開発の第2段階として、MOOSE[10]やSORT[11]などの手法が台頭し、大きな注目を集めた。このため、研究者は相関フィルタリングに基づくトラッカーを掘り下げ、様々な評価データセットにおいて、速度と精度に顕著な利点があることを明らかにした。SORTはDeep Sort [12]とOCSORT [13]でさらに強化された。最近、Sharmaら[14]は、YoloV8とともにこれら2つの手法を組み合わせ、閉回路カメラを使用して駐車時間の違反を追跡するためのリアルタイムアルゴリズムを作成した。

2016年から現在に至るまで進行中の第3段階は、ディープラーニングアルゴリズム、特にシャムネットワークによって駆動されるターゲットトラッキングの進歩によって特徴付けられる。豊富なデータセットを利用することで、ロバスト性と精度の継続的な向上が達成された。この進歩は、物体追跡の領域における深層学習の強力なエンドツーエンドの学習能力を強調している。視覚追跡のためのディープラーニングの包括的なサーベイは[15]にある。

## III. 提案アプローチ

パイプラインで2つのディープニューラルネットワークを使用することを提案する。まず、分類ネットワークを用いて、駐車スペースを占有と空に分類する。2つ目のネットワークはシャム比較ネットワークで、車を比較するために使われる。分類ネットワークとシャムネットワークの両方について、精度と計算コストのトレードオフのため、バックボーンとしてMobileNetV3 [16]の大規模バージョン(約5Mパラメータ)を採用する[17]。

両ネットワークの画像の入力サイズを $128 \times 128 \times 3$ と定義した。シャムネットワークの入力画像ペア間の類似度を計算するために、Contrastive Loss [18], [19]を用いる。分類ネットワークはCrossentropy損失関数を使用する。駐車場の位置は固定されているので、[1]で説明した戦略を使用し、各画像について、各駐車場の位置を回転した長方形として切り出す。この切り出し画像をネットワークに与え、ステータス(空/占有)を定義し、連続する2つの画像間で駐車車両が同じかどうかをチェックする。

切り出された学習画像とそのステータスは、分類ネットワークの学習に使用される。シャムネットワークを学習するために、学習セット内の個々の車を考慮し、学習ペアを作成する。異なる車の画像であるネガティブペアを生成するために、現在の車の画像と別の車からのランダムな画像を組み合わせる。両方の画像が同じ車に属する正のペアを生成するために、その車の画像と、異なるランダムな時間に撮影された同じ車の画像とを組み合わせる。図3は、ネガティブトレーニングサンプルとポジティブトレーニングサンプルの例である。

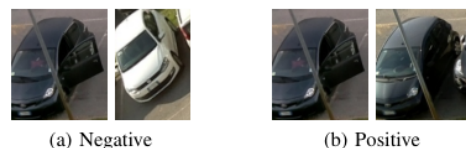


図3. シャムネットワークの学習サンプル例

提案するパイプラインは簡単である。車が駐車している時間をカウントすることに興味があるので、連続した画像を取得する間の時間を考慮する必要がある。k秒ごとに1枚の画像が撮影され、現在時刻tに収集された駐車スペースの画像であることを考慮し、まず分類ネットワークを用いて駐車スペースが占有されているか空であるかを確認する。駐車場が占有されている場合、 $t - k$ で収集された前の画像の駐車場の状態を確認する。

前の画像で駐車スペースが空であった場合、検出された車は新しい車(駐車しているばかりの車)とみなされる。それ以外の場合、時刻 $t$ に収集された駐車スペース(自動車が占有)の画像は、比較(シャム)ネットワークを用いて、 $t - k$ に収集された同じ駐車スペースの画像と比較される。ネットワークが車と同じと判断した場合、車のタイムカウンタは $k$ だけ増加する。そうでなければ、その車は新しい車とみなされる。

図4は、時刻 $t - k$ から $t$ の間の2つの駐車スペースに対するこのパイプラインを示している。完全な手順はアルゴリズム1に描かれている。このアルゴリズムは、分類ネットワークが画像を空か占有かに分類したと仮定している。

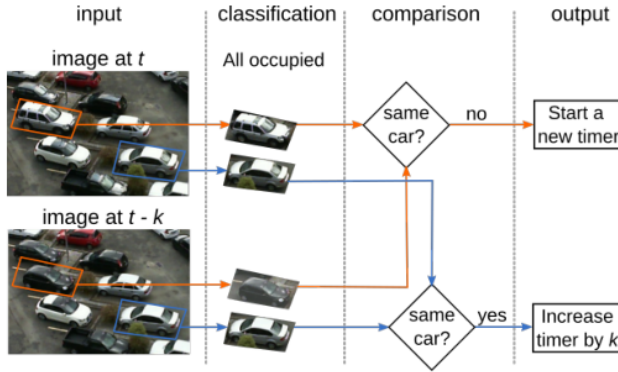


図4. 2つの駐車スペースのフローチャート例。両駐車スペースは $t - k$ から $t$ の間に配置された。オレンジ色の駐車スペースに駐車している車が画像間で同じでなくなり、新しいタイマーが作成される。青い駐車スペースの車はそのまま、タイマーを増やした。

#### Algorithm 1: DWELLTIME( $p, q, k, N$ ).

入力:  $p$ : 現在時刻  $t$  の駐車スペース情報;  $q$ : 前時刻  $t - k$  の駐車スペース情報;  $k$ :  $p$  と  $q$  の間の秒数;  $N$ : 自動車と比較するための学習済みネットワーク。  
結果です: 車の滞在時間が更新される

```

1 if  $p$  is empty then
2   return
3 if  $q$  is empty then
4    $p.time = 0$  // Car just parked
5   return
6 結果 = compare( $N, p.image, q.image$ )
7 if result is same car then
8    $p.time = q.time + k$ 
9 else
10   $p.time = 0$ 

```

## IV. 実験プロトコル

### A. データセットと指標

実験には、よく知られたPKLot [1]とCNRPark-EXT [2]データセットを使用する。これらのデータセットは、カメラが固定位置にあり、駐車スペースの位置が時間と共に変化しないことを前提としている。PKLotデータセットとCNRPark-EXTデータセットの画像例を図2に示す。

連続した画像間の時間 $k$ はPKLotでは5分であるのに対し、CNRPark-EXTでは30分である。画像間の時間が30分というのは、長すぎると考えられる(例えば、車が20分間駐車し、システムがその存在に気づかないまま放置することがある)。ただし、CNRParkEXTの頑健性と、2つの異なるデータセットをテストに含めるために、CNRParkEXTを使用する。

実験用のグラントゥールスを作成するために、各車に個別の識別子を手動でアノテーションし、複数の画像にわたって追跡した。PKLotとCNRPark-EXTデータセットから16,694枚の画像にラベルを付け、これらの新しい識別子を含めることで、21,538台の自動車を追跡することが可能になった。実験中、開発されたアプローチは、駐車スペースに車が駐車している時間をカウントしなければならない。予測値をグラントゥールスと比較するために、平均絶対誤差(MAE)とルート平均二乗誤差(RMSE)値を使用する。

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (2)$$

ここで、 $N$ は予測数、 $y_i$ はグラントゥールスに従って車が駐車した実際の時間、 $\hat{y}_i$ は予測時間である。

1台の車が複数回新しい車と判断された場合、その最初の予測値 $\hat{y}$ を用いてグラントゥールス $y$ と比較し、残りの予測値をゼロと比較する(すなわち、式中の $0 - \hat{y}$ )。同様のアプローチは、車が検出されない場合にも使用され、システムは車について時間を計算しないので、グラントゥールスはゼロと比較される。また、予測された駐車時間がグラントゥールスと完全に一致した車の数を、データセットの駐車車の総数で割ったものとして、達成された精度を示す。

### B. ネットワークと学習手順

分類とシャム(比較)ディープネットワークの両方が、ImageNetデータセットで事前に訓練された。両ネットワークとも、初期学習率0.001、バッチサイズ32のAdamオプティマイザを採用した。[20]で提案されているように、分類ネットワークとシャムネットワークの両方が、クロスデータセット方式で学習される。これは、対象駐車場からの学習サンプルを考慮せずにシステムをテストに投入し、対象シナリオで学習させることなくシステムを展開できる現実的なシナリオを作成しているため、重要である。まず、PKLotデータセットの画像を用いてネットワークを学習し、CNRPark-EXTデータセットでテストした。次に、学習セットとテストセットを逆にし、CNRPark-EXT画像を用いた学習とPKLot画像を用いたテストを行った。学習データセットの各カメラの最初の70%のデイト画像が学習に使用され、残りの30%が検証に使用される。表1に訓練、検証、テストの分割を示す<sup>1</sup>。各エポックにおいて、完全な学習セットは分類ネットワークの学習に使用される。

<sup>1</sup> 新しいサンプルを手動でラベル付けしたため、各データセットで利用可能なサンプル数は元のデータセット[1], [2]とは異なる可能性がある。



シャム比較ネットワークを考慮すると、データセット中の利用可能な車の数に対して、可能な訓練と検証のペアの数は2次元関数的であるため、各エポックについて20,000のランダムな訓練ペアと20,000のランダムな検証ペアを選択する。

TABLE I  
Tトレーニング、検証、テストのスプリット、

CNRPark-EXTを用いたPKLotテストによる学習		
#Training	#Validation	#Test
309,381 occupied	143,529 occupied	83,655 occupied
360,459 empty	113,754 empty	67,972 empty
PKLotを用いたCNRPark-EXTテストによる学習		
#Training	#Validation	#Test
58,625 occupied	25,030 occupied	452,730 occupied
52,958 empty	15,014 empty	474,213 empty

分類ネットワークとシャムネットワークは、それぞれ30エポックと100エポックで微調整される。検証セットで最も損失が少ないモデルを最終モデル<sup>2</sup>として選択する。分類ネットワークでは、各エポックについて、各トレーニング画像に以下のデータ補強技術のいずれかを適用する確率を10%と定義する：水平フリップ、 $5 \times 9$ カーネル、標準偏差 $\in [0.1, 5]$ のガウスぼかし、ランダムクロップ、ランダムオートコントラスト。シャムネットワークでは、各トレーニングペアに対して、輝度変化 $\in [-0.2, 0.2]$ を適用する確率を10%と定義する。分類閾値を定義するために、分類ネットワークと比較ネットワークの両方の検証セットを使用する。最適な閾値は、分類ネットワークのROC(Receiver Operating Characteristic Curve)において、EER(Equal Error Rate)点を生成するものと定義した。シャムネットワークでは、分類しきい値を、異なる車のペアを同じ車のペアと誤認する確率が最大でも5%である最大のしきい値と定義した。

## V. EXPERIMENTS

セクションV-Aでは、複数の画像にわたって自動車を追跡する比較(シャム)ネットワークの能力をチェックするために、完全駐車スペース分類器(駐車スペースが空か占有かを予測する)を考慮した結果を提示し始める。次に、セクションV-Bにおいて、分類ネットワークと比較ネットワークを考慮した完全なパイプラインを考慮した結果を示し、実世界のシナリオにおける提案アプローチの期待される挙動を示す。すべての結果は5回の実行の平均である。

### A. グランドトゥールースを用いた滞留時間の測定

最初の実験では、学習したシャムネットワークが、駐車場の複数の画像に対して同じ車の存在を正しく追跡する能力を分析したい。

また、学習/検証サンプルは各エポックごとにランダムに取られるため、シャムネットワークは少なくとも15エポック学習する必要があり、損失不安定性が生じる可能性があるとして定義した。

空間である。これを達成するために、我々は他の要因の干渉を受けずにシャムネットワークの性能をチェックすることに関心があるので、グランドトゥールースで利用可能な情報を使用して駐車スペースが占有されているか空であるかをチェックする。

まず、シャムネットワークの車比較能力を測定するために、テストセットの各車<sup>3</sup>について、同じ車の画像(異なる瞬間)をランダムに選択して正のペアを作成し、別の車からランダムに画像を選択して負のペアを生成した。このように、 $n$ 台の自動車を含むテストデータセットにおいて、 $2n$ 組のテストペアを生成した。これらのテスト画像における学習済みネットワークの結果を表IIに示す。

TABLE II  
Sシャムネットワークの結果、

Training Set	Testing Set	# Test Pairs	Accuracy (stddev)
PKLot	CNPark-EXT	15,482	93.8% (1.6)
CNPark-EXT	PKLot	27,594	96.2% (1.8)
<b>Total</b>		43,076	95.3%*

\*Weighted average.

表IIの結果から、ネットワークは平均95.3%の精度で良好な汎化を達成したことがわかる。これは興味深い結果である。というのも、学習用にターゲット駐車場からの画像サンプルが与えられていないにもかかわらず、ほとんどの車ペアがネットワークによって正しく認識されたからである。

表IIIでは、セクションIIIで議論したアプローチを用いて、テストセット内の各車の駐車滞留時間を定義するために、学習したシャムネットワークを使用した場合の結果を示す。このように、提案手法は、PKLotテストで69.1%、CNRPark-EXTテストで77.8%と完璧な予測値を示し、有望な結果を得ることができた。MAEとRMSEの結果を分単位で示す。MAE値は、PKLotデータセットとCNRPark-EXTデータセットにおいて、それぞれ平均40.2分と47分の誤差で、提案アプローチが正しい滞留時間を見逃していることを示している。CNRParkEXTデータセットでは、連続する2つの画像間の時間差は30分であることを覚えておく価値がある。したがって、連続する2つの画像から同じ車を検出する際の1回のミスは、少なくとも30分のエラーを発生させる。

完全な予測数が多いことと、MAEが比較的高いことから、システムが滞留時間を正しく予測できない場合、発生する誤差が大きくなる傾向があることがわかる。この結果は、RMSE計算の2次的な性質により、グランドトゥールースと予測値の間の高い差にペナルティを与えるため、高いRMSE値によって裏付けられる。

### B. 分類器を用いた滞留時間の測定

本節では、まず、占有と空の間の駐車スペースを分類するためにネットワークを使用し、車を比較するためにシャムネットワークを使用する、完全な提案アプローチ(セクションIII)の結果を示す。

同じ車が複数の画像に現れることがある。車の画像はランダムに1枚だけ選んだ。

TABLE III  
RESULTS USING THE GROUND-TRUTH TO GET EACH PARKING SPACE  
STATUS (EMPTY OR OCCUPIED).

Test Set	Camera	MAE (stdev)	RMSE (stdev)	% perfect predictions (stdev)
PKLot	UFPR04	52.9 (1.5)	111.6 (2.1)	59.4 (1.0)
	UFPR05	58.1 (1.7)	125.1 (1.3)	58.1 (1.5)
	PUC	29.4 (1.1)	89.4 (1.9)	76.3 (0.6)
	<b>Average</b>	<b>40.2 (1.3)</b>	<b>103.0 (1.7)</b>	<b>69.1 (0.8)</b>
CNRPark-EXT	cam1	47.8 (11.7)	117.1 (13.7)	76.6 (5.9)
	cam2	33.2 (20.4)	93.0 (30.3)	82.7 (9.1)
	cam3	42.2 (12.7)	107.9 (13.8)	78.7 (7.2)
	cam4	44.4 (11.2)	114.0 (12.2)	78.5 (6.1)
	cam5	43.4 (9.9)	114.5 (9.6)	80.1 (5.7)
	cam6	47.5 (9.1)	121.5 (8.1)	77.8 (5.5)
	cam7	53.7 (7.5)	127.5 (6.9)	75.7 (4.5)
	cam8	44.9 (9.2)	113.4 (10.1)	78.6 (4.9)
	cam9	57.5 (12.4)	129.8 (11.2)	73.5 (6.4)
	<b>average</b>	<b>47.0 (10.4)</b>	<b>117.8 (10.8)</b>	<b>77.8 (5.8)</b>

このテストは、駐車スペースの状態を予測する(そしてそれを誤って分類する可能性がある)分類器が、どのように車両滞留時間会計の妨げになるかを答えるために開発された。

まず、表IVに、対象駐車場(テストセット)でテストに投入し、その駐車場空間を占有と空の間に分類したときの分類ネットワークの平均分類結果を示す。これは、データセット横断シナリオ[17]、[20]を考慮した場合に、最先端の手法のほとんどが達成した結果と同様である。

TABLE IV  
分類ネットワークの結果。

Training Set	Testing Set	# Test Samples	Accuracy (stdev)
PKLot	CNPark-EXT	160,298	93.0% (0.24)
CNPark-EXT	PKLot	926,943	92.3% (0.15)
<b>Total</b>		<b>1,087,241</b>	<b>92.4%*</b>

\*Weighted average.

表Vでは、分類ネットワークを使用して駐車スペースが占有されているか空であることを検証し、シャムネットワークを使用して、車が画像間で同じかどうかをチェックし、必要に応じて車両の滞留時間を増加させた、完全なパイプラインの結果を示している。観察できるように、分類ネットワークの精度が高いにもかかわらず、セクションV-Aで利用可能な結果と比較すると、すべてのメトリクスで大幅な低下が見られた。

このようにメトリクスが著しく低下するのは、両データセットにおいて、平均して車の数時間は駐車できるという事実によって説明できる。PKLotデータセットとCNRPark-EXTデータセットでは、それぞれ平均して25枚(2時間5分)と12枚(6時間)の連続画像に同じ車が存在する。図6にその一例を示す。この例を考えると、分類器が車の1枚の画像を空き駐車場として誤って予測した場合、滞留時間の計算に重大な影響を与える可能性がある。

TABLE V  
R完全なパイプラインを使用した結果。

Test Set	カメラ	MAE (標準偏差)	RMSE (標準偏差)	% perfect predictions (stdev)
PKLot	UFPR04	68.7 (5.5)	141.8 (9.2)	21.7 (5.4)
	UFPR05	75.2 (13.9)	151.6 (21.5)	18.3 (5.5)
	PUC	34.5 (2.6)	96.2 (5.9)	56.9 (6.6)
	<b>Average</b>	<b>53.3 (5.4)</b>	<b>124.6 (9.2)</b>	<b>38.4 (4.8)</b>
CNRPark-EXT	cam1	107.9 (6.4)	175.2 (6.5)	43.8 (2.2)
	cam2	46.2 (3.9)	114.2 (5.6)	75.6 (2.3)
	cam3	82.8 (3.9)	158.2 (3.5)	57.3 (2.5)
	cam4	66.5 (3.5)	136.5 (4.4)	61.5 (1.1)
	cam5	86.2 (4.2)	156.2 (4.9)	52.0 (1.5)
	cam6	80.2 (3.6)	151.6 (4.1)	56.8 (0.9)
	cam7	101.6 (5.0)	170.0 (4.5)	46.5 (2.6)
	cam8	74.4 (1.7)	144.0 (2.4)	58.8 (0.7)
	cam9	88.4 (5.7)	156.9 (7.4)	51.9 (1.7)
	<b>average</b>	<b>84.5 (1.4)</b>	<b>155.0 (1.8)</b>	<b>54.3 (0.7)</b>

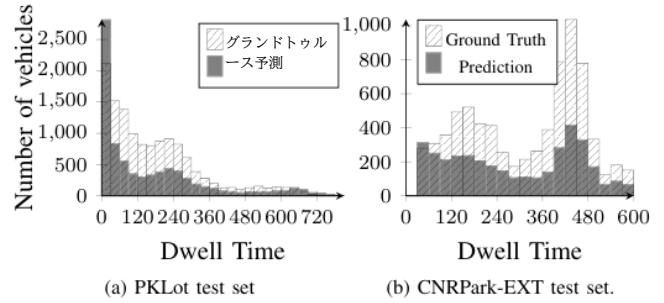


図5. a) PKLotデータセットとb) CNRParkEXTデータセットをテストセットとして用いた車の滞留時間ヒストグラム。

特に、誤分類された画像がシーケンスの先頭付近にある場合、現在の車のタイマーがあるべきよりずっと早く停止され、タイマーが(誤って)新しい車のために開始される可能性があるためである。



図6. 数時間駐車した車 - CNRPark-EXTデータセット。

この挙動は図5で見ることができ、PKLotとCNRPark-EXTデータセットの両方を考慮して、異なる時間スパンで駐車した車の台数のヒストグラムを示す。提案手法は、PKLotデータセットとCNRPark-EXTデータセットの両方において、0分から30分の間の低い滞在時間を持つ車数を過大評価した。これは、分類ネットワークの誤分類やシャムネットワークによって発生したエラーによって、車のクロノメーターが早期に停止されたためと説明できる。

### C. 教訓

連続画像間の時間スパンは結果に影響を与える：本研究で議論したように、連続する2つの画像間の時間スパンをkとする。kの値が大きいと、撮影・処理する画像数が分単位で減るため、より軽量のシステムになる可能性がある。

とはいえ、1つの誤分類画像は少なくともk秒の誤差につながる可能性があるため、kの値が大きいと誤差が大きくなる可能性がある。さらに、kの値が大きいくほど、連続する画像間でより多くの変化が起こり、シャムネットワークが自動車を比較する能力を妨げる可能性がある。

**The classification network is a key point in generating** より良い結果を得ることができる：比較的誤差の小さい分類ネットワーク(我々の実験では7.6%)でも、車が数時間駐車しているため、滞留時間の推定を著しく減少させる可能性があり、1回の誤分類で、ある車のクロノメーターを早期に停止し、現実には存在しない新車のクロノメーターを開始する可能性がある。20]で議論されているように、取り組むべき重要な点は、100%に近い精度を持つ汎用分類器の開発であり、これは我々のシステムの性能を大幅に向上させる可能性がある。ほぼ100%の精度を達成する分類器が存在する [1], [2], [20]. とはいえ、このような分類器には、対象駐車場からの学習サンプルが必要である。本研究では、対象駐車場からの学習サンプルを必要としない一般的なアプローチに焦点を当てた。ターゲット駐車場からの学習サンプルが利用可能な場合、仮想的な完全分類器が使用されたセクションV-Aで示された結果に近い結果を期待できるはずである。システムが物事を間違えると、本当に間違えてしまう：RMSE値はすべての実験においてMAE値よりはるかに大きい。これは、いくつかの難しいサンプルでは、システムは大きなエラーを生成できるが、ほとんどのサンプルでは、エラーは比較的小さいことを示している(RMSEはその2次的な性質のために大きなエラーにペナルティを与える)。これらの困難なサンプルは、今後の研究において考慮されるべきものである。

## VI. CONCLUSION

本研究では、駐車空間における車両の滞留時間を計算するために、深い分類ネットワークとシャムネットワークを含む完全なパイプラインを示した。提案手法を、ビジョンベースの駐車スペース監視システム[20]にとって最も困難なシナリオである、ターゲット駐車場からの学習サンプルが与えられないクロスデータセットシナリオで評価した。我々の知る限り、これは、データセット横断シナリオや、大量のデータとMAEやRMSEのようなく知られたメトリクスを使用を用いた実験を含む、ロバストなプロトコルを定義する画像を使用して、車の滞留時間を計算するパイプラインを提案した最初の研究である。

我々の実験は、画像間の自動車比較を担当するシャムネットワークに関して、提案アプローチに有望な結果を示した。我々の結果は、仮想的な完全分類ネットワークが各駐車スペースが占有されているか空であるかを見分けることができることを考えると、このようなシステムがPKLotとCNRParkEXTデータセットの69%と78%の車の滞留時間を完全に計算できることを示している。多くの分類ネットワークは、ターゲット駐車場からの学習サンプルが与えられていることを考えると、100%に近い精度に達することができることは言及に値する[1], [2], [20]。データセット横断シナリオに焦点を当てるため、完全なパイプラインに関する実験も行った。ここでは、ターゲット駐車場から学習サンプルを受け取っていない分類ネットワークを採用し、滞留時間を計算する前に各駐車場を分類する。

我々の結果は、比較的誤差の小さいネットワーク(我々のテストでは7.6%)であっても、車のクロノメーターをあるべきよりもはるかに早く停止させることができるため、結果を著しく悪化させる可能性があることを示している。このことは、[1]で示唆されているように、対象駐車場の訓練インスタンスに依存しない駐車場分類ネットワークは、少なくとも対象駐車場で訓練されたものに近い精度に到達するように改善されるべきであることを示している。

今後の課題として、アルゴリズムIIIを考慮し、データセット横断シナリオでの誤差とパイプライン全体の誤差を低減するために、シャムネットワークと分類ネットワークの両方の学習を改善する予定である。

## REFERENCES

- [1] P. R. Almeida, L. S. Oliveira, A. S. Britto Jr, E. J. Silva Jr, and A. L. Koerich, "Pklot—a robust dataset for parking lot classification," *Expert Systems with Applications*, vol. 42, no. 11, pp. 4937–4949, 2015.
- [2] G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Meghini, and C. Vairo, "Deep learning for decentralized parking lot occupancy detection," *Expert Systems with Applications*, vol. 72, pp. 327–334, 2017.
- [3] R. M. Nieto, Á. García-Martín, A. G. Hauptmann, and J. M. Martínez, "Automatic vacant parking places management system using multicamera vehicle detection," *IEEE ITS*, vol. 20, no. 3, pp. 1069–1080, 2018.
- [4] V. Dhuri, A. Khan, Y. Kamtekar, D. Patel, and I. Jaiswal, "Real-time parking lot occupancy detection system with vgg16 deep neural network using decentralized processing for public, private parking facilities," in *ICAECT*, 2021, pp. 1–8.
- [5] P. Almeida, L. S. Oliveira, and J. Honorio, "Vehicle occurrence-based parking space detection," in *2023 IEEE International Conference on Systems, Man, and Cybernetics*, 2023, pp. 3603–3608.
- [6] R. Grbić and B. Koch, "Automatic vision-based parking slot detection and occupancy classification," *Expert Systems with Applications*, vol. 225, p. 120147, 2023.
- [7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE CVPR*, 2016, pp. 779–788.
- [8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *NIPS*, 2015, pp. 91–99.
- [9] K. He, P. D. G. Gkioxari, and R. Girshick, "Masked R-CNN," in *Int. Conf. on Computer Vision*, 2017, p. 2961–2969.
- [10] D. Bolme, J. Beveridge, B. Draper, and Y. Lui, "Simple online and realtime tracking," in *CVPR*, 2010, pp. 2544–2550.
- [11] A. Bewley, F. R. Z. Ge, L. Ott, and B. Upcroft, "Simple online and realtime tracking," in *ICIP*, 2016.
- [12] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *ICIP*, 2017, pp. 3645–3649.
- [13] J. Cao, J. Pang, X. Weng, R. Khiroudkar, and K. Kitani, "Observation-centric sort: Rethinking sort for robust multi-object tracking," in *CVPR*, 2023, pp. 9686–9696.
- [14] N. Sharma, S. Baral, M. P. Paing, and R. Chawuthai, "Parking time violation tracking using yolov8 and tracking algorithms," *Sensors*, vol. 23, 2023.
- [15] S. Marvasti-Zadeh, L. Cheng, H. Ghanei-Yakhdan, and S. Kasaei, "Deep learning for visual tracking: A comprehensive survey," *IEEE ITS*, vol. 23, pp. 3943–3968, 2021.
- [16] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan et al., "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314–1324.
- [17] A. G. Hochuli, J. P. Barddal, G. C. Palhano, L. M. Mendes, and P. R. Lisboa de Almeida, "Deep single models vs. ensembles: Insights for a fast deployment of parking monitoring systems," in *ICMLA*, 2023, pp. 1379–1384.
- [18] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 539–546.

- [19] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2006, pp. 1735–1742.
- [20] P. Almeida, J. Alves, R. Parpinelli, and J. Barddal, "A systematic review on computer vision-based parking lot management applied on public datasets," *Expert Systems With Applications*, vol. 198, p. 116731, 2022.