

街路の全方位画像ビッグデータに基づく深層学習を用いた地域分析手法

沖 拓弥*・小川 芳樹**

A Method for Regional Analysis Using Deep Learning Based on Big Data of Omnidirectional Images of Streets

Takuya Oki and Yoshiki Ogawa

In this paper, we propose a method for regional analysis using image recognition technology based on deep learning and big data of street images captured by omnidirectional cameras on vehicles. Specifically, we first construct a classification method of regions based on street images using a pre-trained deep learning model (VGG16) for image recognition as a feature extractor. Next, we develop a method to evaluate the landscape and safety of streets based on the ratio of street components (such as buildings, roads, fences, vegetations, sky, street lights) at each shooting point, which is calculated by semantic segmentation.

Keywords: 全方位画像 (omnidirectional image), 深層学習 (deep learning), ビッグデータ, (big data), セマンティックセグメンテーション (semantic segmentation), クラスタリング (clustering)

1. はじめに

地域の特徴づけや地域分類の方法として、これまでに様々なものが提案されてきた。しかし、「高い空間解像度」かつ「広域」を対象に分析を行うことは困難であった。昨今、計測機器の進歩やIoT技術の発達により、多種多様な建築・都市ビッグデータが容易に入手可能となりつつある。また、得られたデータを活用するための人工知能(AI)技術の発展も目覚ましい。したがって、こうしたデータや技術を活用した新たな地域分析手法を検討することの意義は大きい。

そこで本稿では、車両に設置した全方位カメラで撮影した街路画像ビッグデータをもとに、深層学習による画像認識技術を用いた地域分析の手法を提案する。具体的には、まず、画像認識用の学習済み深層学習モデルを特徴量抽出器として使い、街路画像に基づく地域分類手法を構築する。次に、画素単位の自動意味づけ (Semantic segmentation) により、各撮影地点における街路

構成要素 (建物、道路、塀、植栽、空、街灯など) の構成比を算出し、これをもとに街路の景観や安全性等を評価する手法を構築する。

2. 分析の方法

2.1. 街路の全方位画像の仕様

本稿では、株式会社ゼンリンから提供された全方位画像ビッグデータを分析に用いる。屋根部分に360度カメラを搭載した専用車両が地域を巡回し、2.5m間隔で街路の全方位画像を撮影したものである。各画像には、GPSにより取得された撮影地点の位置情報 (緯度・経度) のほか、車両の方向 (北が0度、東が90度、南が180度)、撮影日時、撮影車両IDなどの基本的な情報が付属している。オリジナル画像は、図1に示すように横5,400 pixel × 縦2,700 pixelで、正距円筒図法 (equirectangular) に基づいていることから、画像の上部あるいは下部に向かうほど、歪みが大きくなるという特徴を有する。このとき、画像の下

* 正会員 東京工業大学 環境・社会理工学院 (Tokyo Institute of Technology)

〒152-8550 東京都目黒区大岡山 2-12-1 E-mail: oki.t.ab@m.titech.ac.jp

** 正会員 東京大学 空間情報科学研究センター (The University of Tokyo)



図1 街路の全方位画像（オリジナル）および変換後の画像の例

部約 35% あるいは約 45%（車両に依存）の部分には、撮影車両の屋根部分が映り込んでいる。また、全方位画像からは、キューブマップ、円周魚眼画像、パース画像などへと自由に変換することができる。本稿では、Google Street View と同様のパラメータ（視野 120 度、仰角 0 度）に基づき、全方位画像をパース画像に変換し、全方位画像とパース画像の分析結果の違いを考察する（図 1）。

2.2. 分析対象データ

世田谷区全域（一部、区外も含む）で撮影された 178 万 9,821 枚の全方位画像をデータセット①とする。また、①から一部の画像をサンプリングし、詳細な分析を行う。具体的には、三軒茶屋駅付近で 1 車両によって同一日に撮影された連続する 4,689 枚（データセット②）、同じく連続する 1 千枚（データセット③）、および、世田谷区全域から無作為に抽出した 1 千枚（データセット④）を分析対象とする。

2.3. 画像の特徴量抽出方法

本稿では、ImageNet データセットで事前にパラメータを学習した VGG16 モデル（Simonyan & Zisserman, 2014）を特徴量抽出器として用いる。VGG16 モデルは畳み込みニューラルネットワークの一種で、本来、1 千クラスの画像分類を目的としたモデルであるため、画像の特徴抽出性能に優れる。そこで、VGG16 モデルの全結合層（FC1 層）で得られるベクトル（4,096 次元）を画像の特徴ベクトルとして用いる。

2.4. 特徴ベクトルを用いた地域分類方法

画像の特徴ベクトルを用いれば、コサイン類似度に基づき街路画像間の類似度を算出したり、主

成分分析や多次元尺度構成法、t-SNE などを用いて各画像の位置関係を 2 次元平面上へプロットしたり、非階層クラスター分析により地域分類を行うことができる。

ただし、画像枚数が多くなったり、画像サイズが大きくなると、計算負荷の観点から、一般的な k-means 法などのクラスタリング手法の適用は困難となる。本稿では、Matsui et al. (2017) による PQk-means 法を適用することで、画像枚数が約 180 万枚に上るデータセット①についても、適切な地域分類が行えることを示す。

2.5. 街路評価の方法

特徴ベクトルは機械学習モデルが画像の特徴を捉えたものであるが、この多次元ベクトルを人間が解釈することは困難である。そこで、画素単位での自動意味づけ手法である Semantic segmentation を用いることで、各画像に何がどの程度映っているかを機械的に算出し、これをもとに街路評価を行う。具体的には、CityScapes データセット（Cordts et. al., 2016）を学習させた DeepLab V3+ モデル（Chen et. al., 2018）を用いて、各街路画像中の要素の構成比を算出する。

3. 類似街路検索と地域分類の結果

3.1. 各画像の位置関係の可視化

ここでは、連続する 1 千枚（データセット③）

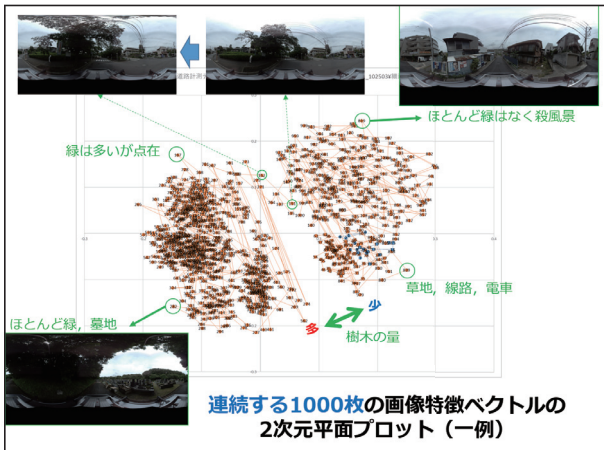


図2 連続する1千枚の画像特徴ベクトルの2次元平面プロットの例

の全方位画像について特徴ベクトルを求めた上で、多次元尺度構成法により2次元平面上に画像の位置関係を付置することを試みた(図2)。この例では、画像が大きな2つのクラスタに分かれ、樹木や植栽の量や分布の違いがプロットに表れている。連続する2枚の画像であっても、樹木に接近することで画像間の距離が大きく異なるケースも見受けられる。

3.2. 街路類似度検索の結果

画像特徴ベクトル間のコサイン類似度に基づき、データセット②・④のパス画像で類似度検索を試みた結果を図3に示す。

人間は画像全体の総合的な印象から判断するため、連続する2地点(間隔2.5m)の類似度は高いように思えるが、コサイン類似度を算出すると類似度は約80%にとどまる(図3(a))。一方、不連続な地点の画像であっても、類似度が80%を超えるケースも見受けられる(図3(b)・(c))。抽出された画像を仔細に見ると、樹木の存在や空・道路の構成、電線や道路線など、全体の印象には違いがあっても、個々の構成要素には共通点が見出せる。これは、畳み込みニューラルネットワークの仕組みによるものと考えられる。

この後の分析結果を解釈する上で、人間とAIとの間に、こうした感覚の違いが存在することには留意する必要がある。

3.3. 地域分類の結果(データセット②)

3.1節および3.2節の分析を通じて、全方位



図3 コサイン類似度に基づく類似画像検索の例

画像とパス画像それぞれについて、VGG16モデルで抽出した画像特徴ベクトルの有用性を確認することができた。これをふまえ本節では、連続する画像(データセット②)を例に、一般的なk-means法を用いて、画像特徴ベクトルに基づく地域分類を試みる。ここでは、全方位画像を用いた場合とパス画像を用いた場合で、分類結果がどう変わるかを検証する。

クラスタ数を10に指定したときの、地域分類の結果を図4に示す。全方位画像を用いた場合(図4(a))とパス画像を用いた場合(図4(b))で大きな違いは見られず、同一の通り(街路)は概ね同一のクラスタに分類されている様子が見てとれる。ただし、同一街路内で所属ク

クラスタが頻繁に変動するケースも一部見受けられる。また、三軒茶屋駅の東側を南北に走る茶沢通りに着目すると、全方位画像の場合は概ね1つのクラスタに分類されているが、パース画像の場合には途中でクラスタが変わっている。各クラスタに含まれる画像を具体的にみると、北側のクラスタでは建物の間口が比較的広いのに対し、南側のクラスタでは建物の間口が狭く、建物ファサードも煩雑な傾向が見られる。

全方位画像の場合、画像中で建物部分の占める面積が比較的小さくなるため、こうした建物の特徴の違いは見過ごされやすい。一方、パース画像の場合、1枚の画像からは1方向の特徴しか考慮することができない。分析の目的にあわせた使い分けが必要だと考えられる。

3.4. 地域分類の結果 (データセット①)

分類対象の画像枚数が多くなると、前稿で用いた k-means 法の適用は困難になる。しかし、今後の街路画像ビッグデータ時代を見据えると、大規模データセットに対する地域分類手法を検討しておくことは重要である。ここでは、VGG16 モデルで抽出した 4,096 次元の特徴ベクトルを、4 次元などの低次元に圧縮可能な PQk-means^{注1)}

(Matsui et al., 2017) を用いて、データセット①の画像約 180 万枚を 10 個のクラスタに分類することを試みた。

データセット②と同じ画像に着目すると (図 5), PQk-means では概ね 3 つのクラスタに分類された。1 つのクラスタ (赤色) に属する画像の割合が高いが、圧縮後の次元数を増やす (特徴量の損失を減らす) ことで、この傾向は変わる可能性がある。

次に、世田谷区全域におけるクラスタ 0 ~ 9 の空間分布を図 6 に示す。クラスタ 3, 8, 9 が世田谷区東側に偏在しているのは、全方位画像中の車両の映り込みの程度の違いによる (クラスタ 3, 8, 9 の画像は下部 45%, その他の画像は下部 35% に車両が映り込んでいることを確認した)。

地域分類の妥当性を検証するために、各クラスタから無作為に 100 枚ずつ画像を抽出した上で、

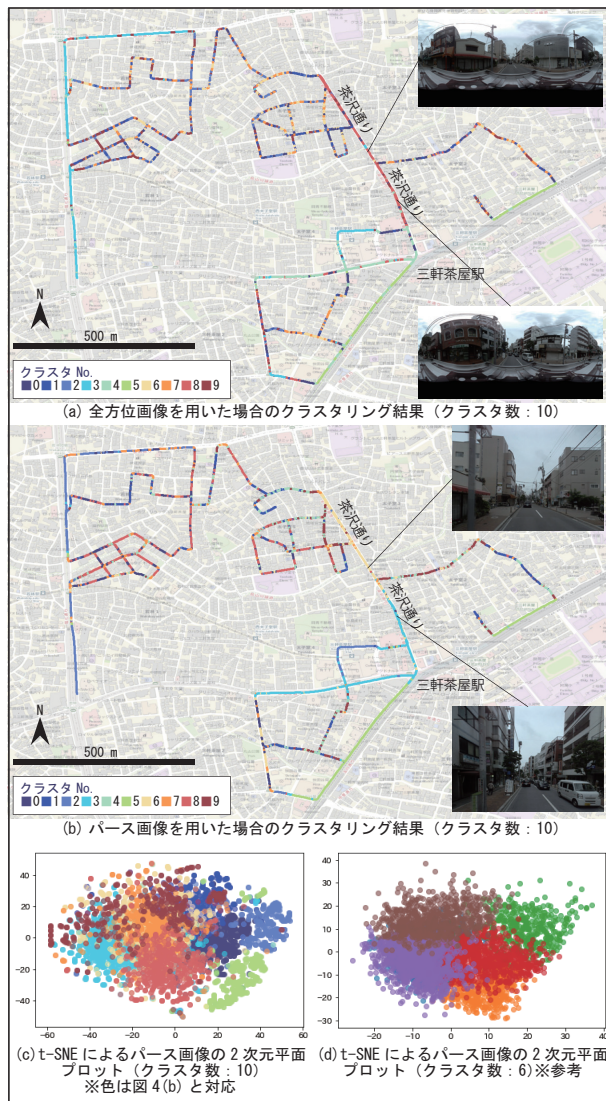
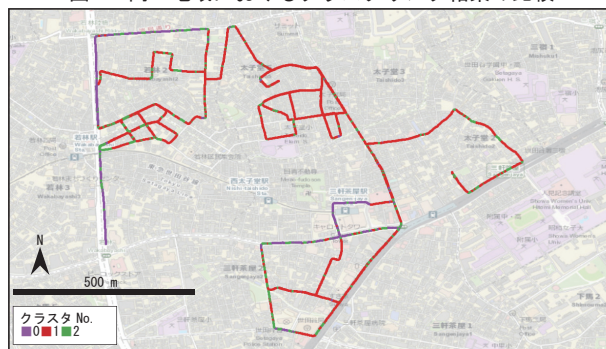


図 4 同一地域におけるクラスタリング結果の比較



2.5 節で述べた Semantic segmentation を各画像に適用し、クラスタ間で街路構成要素の分布にどのような違いが見られるかを考察する (図 7)。

まず、車両の映り込みが少ないクラスタ 7 つについての解釈を以下に記す。

【クラスタ 0】 Road・Sidewalk・Sky の割合が大

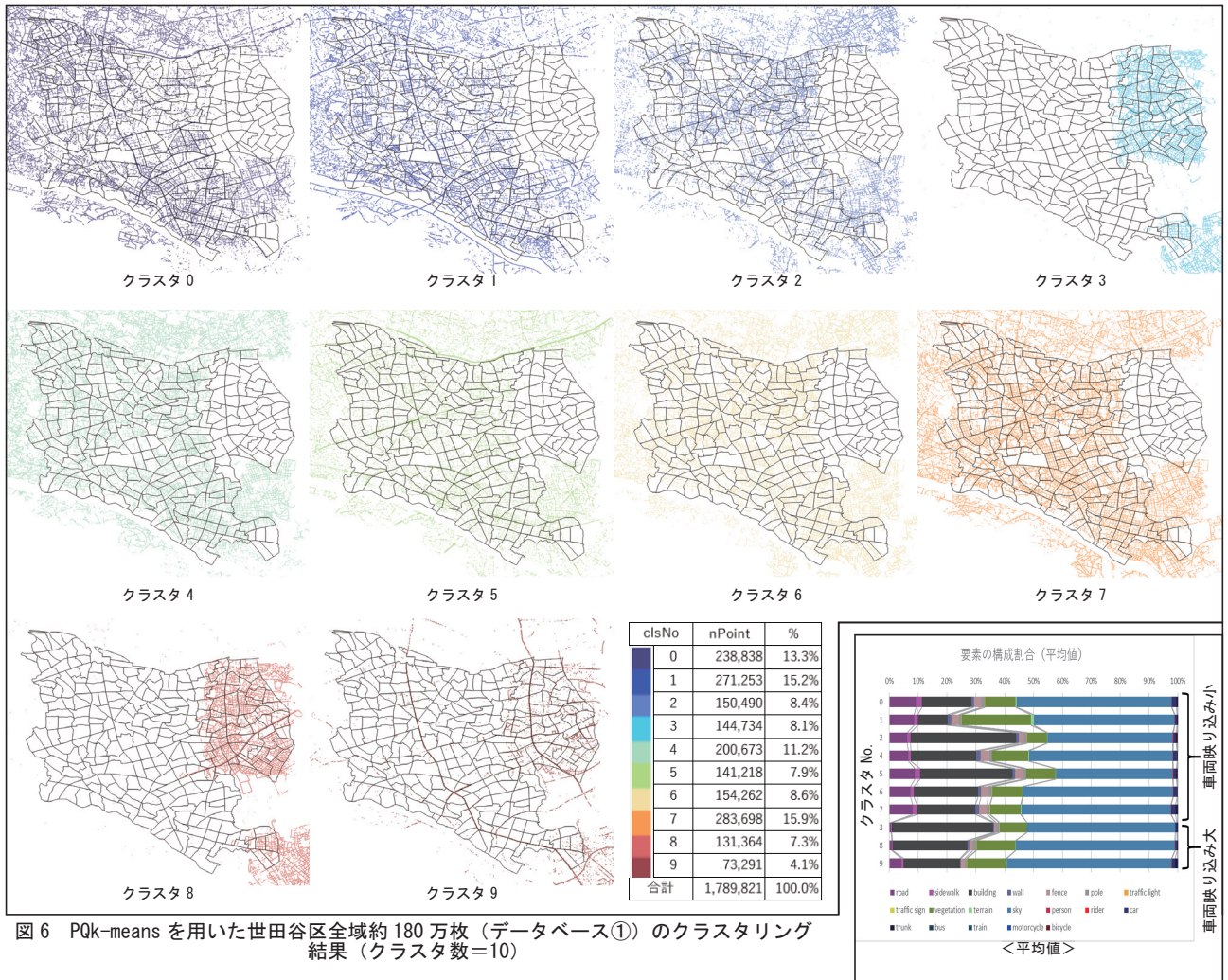


図6 Pqk-means を用いた世田谷区全域約 180 万枚 (データベース①) のクラスタリング結果 (クラスタ数=10)

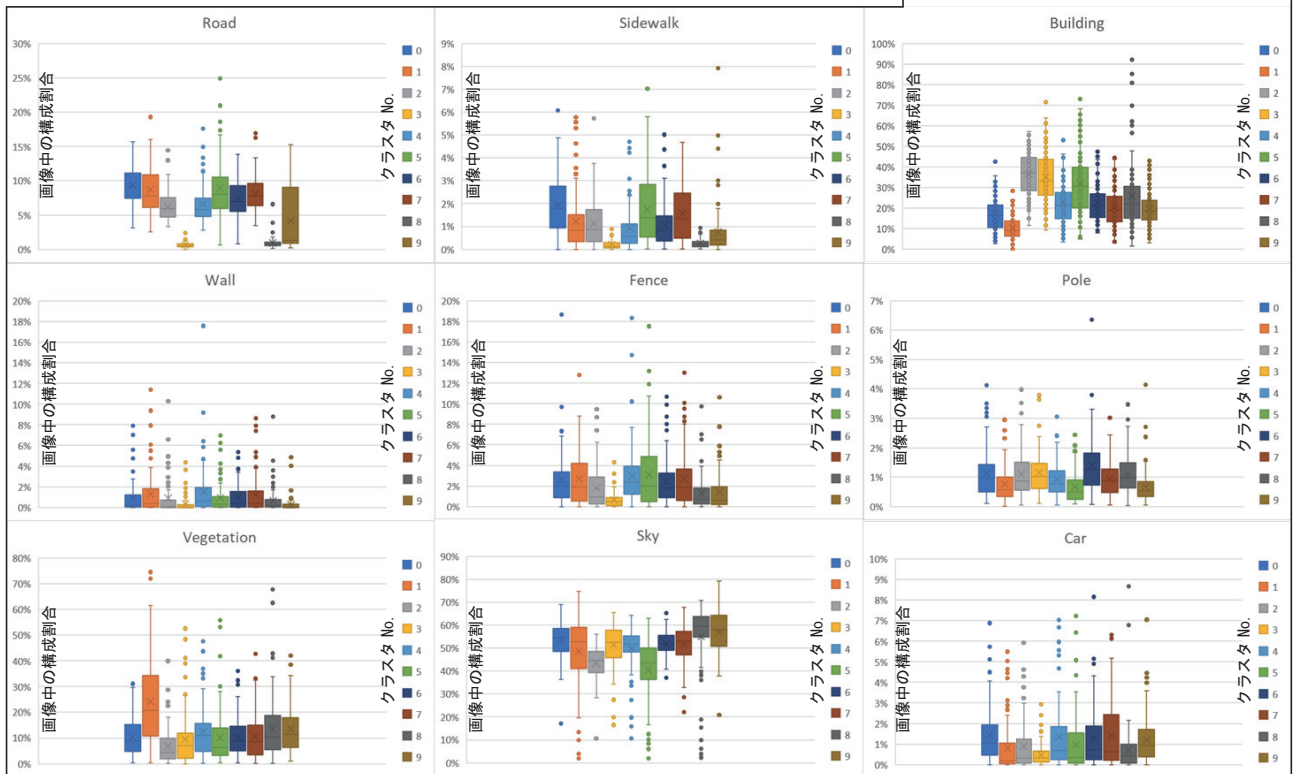


図7 Semantic segmentation に基づく街路構成要素分布のクラスタ間比較

きく、Building の割合がやや小さいことから、幹線道路や広幅員道路沿いである可能性が高い。

【クラスタ 1】Vegetation や Terrain の割合が大きく、比較的緑豊かな地域と解釈できる。

【クラスタ 2】Road・Fence・Wall・Vegetation・Sky の割合が小さく、Building の割合が大きいことから、住宅密集地域の可能性が高い。

【クラスタ 4】Road・Sidewalk の割合は小さいが、Fence・Wall の割合は大きい。道は狭いが塀はしっかりと存在する、特徴の少ない地域である。

【クラスタ 5】Road・Sidewalk・Building・Fence の割合が大きく、Sky の割合が小さい。整備された住宅街、あるいは、中高層住居専用地域の可能性も考えられる。高架下においても道路上空の高架が建物と誤判別され、このクラスタに分類される傾向がある。

【クラスタ 6】Pole の割合が大きく、Sky の割合もやや大きい。特徴の少ない地域である。

【クラスタ 7】Sidewalk・Sky の割合がやや大きい。標準的な地域である。

次に、車両の映り込みが大きいクラスタ 3 つについての解釈は以下の通りである。

【クラスタ 3】Road・Vegetation・Car の割合が小さく、Building の割合が大きい。車所有率低い住宅密集地域の可能性がある。クラスタ 2 に類似している。

【クラスタ 8】比較的標準的な地域である。

【クラスタ 9】Road・Sidewalk・Vegetation・Sky・Car の割合が大きく、Building の割合が小さい。整備された住宅街と考えられ、クラスタ 1 に類似している。環七通りや環八通りで撮影された画像も多くがこのクラスタに属していることがわかる。

4. 街路景観や安全性評価への応用可能性の検討

4.1. 分析の趣旨

前章では、あくまで一種の AI モデルが機械的に画像の特徴を抽出し、それを類似度検索や地域分類に応用した結果を示した。しかし、分類結果の妥当性を判断する上では、Semantic segmentation

を用いて各画像の街路構成要素割合を算出し、それをもとに議論する必要があった。このように、画像に意味づけを行えば、人間による画像理解や人間がもつ感覚との比較もより容易になると考えられる。

そこで本章では、街路画像ビッグデータと Semantic segmentation を用いた、簡便な景観評価や安全性評価の可能性について検討する。ここでは、連続する画像であるものの様々な特徴の街路を含む、データセット②のパース画像を例に分析を行う。

4.2. 街路の特徴づけの結果と考察

図 8 に、Semantic segmentation に基づく街路構成要素割合の可視化結果を示す。パース画像は人間の視野に近いことから、例えば緑 (Vegetation) の割合で見ると、必ずしも樹木や植栽の絶対量が多い街路で値が高くなっているとは限らず、街路を通行中にアイレベルで感じる緑の多少が抽出できている。他にも例えば、

- ・歩道の整備度合い ((b) Sidewalk) ,
- ・建物から受ける圧迫感の度合い (既存指標では D/H に相当, (c) Building)
- ・地域の耐犯犯性あるいは倒壊危険性のあるブロック塀の多さ ((d) Wall および (e) Fence)
- ・電柱本数や街路灯の整備度合い ((f) Pole)
- ・天空率 (西尾・伊藤 (2020) , (h) Sky)
- ・街路のにぎわいや群集密度 ((i) Person)
- ・地域の自動車保有率あるいは路上駐車の数 ((j) Car)
- ・自転車利用者の多さや放置自転車の多さ ((k) Bicycle)

といったように、様々な地域分析 (ウォーカビリティ評価, 防犯・防災性能評価など) に応用できる可能性がある。実際に、犯罪情報マップ (警視庁, 2016) で公表されている身近な犯罪の空間分布 (図 9) と比較すると、子どもや女性に関する犯罪の発生分布と、Vegetation や Sky の量の空間分布には関連があるように見える。また、街路画像ビッグデータから人間による印象評価を推定する深層学習モデル (木澤・沖, 2021) と組み合わ

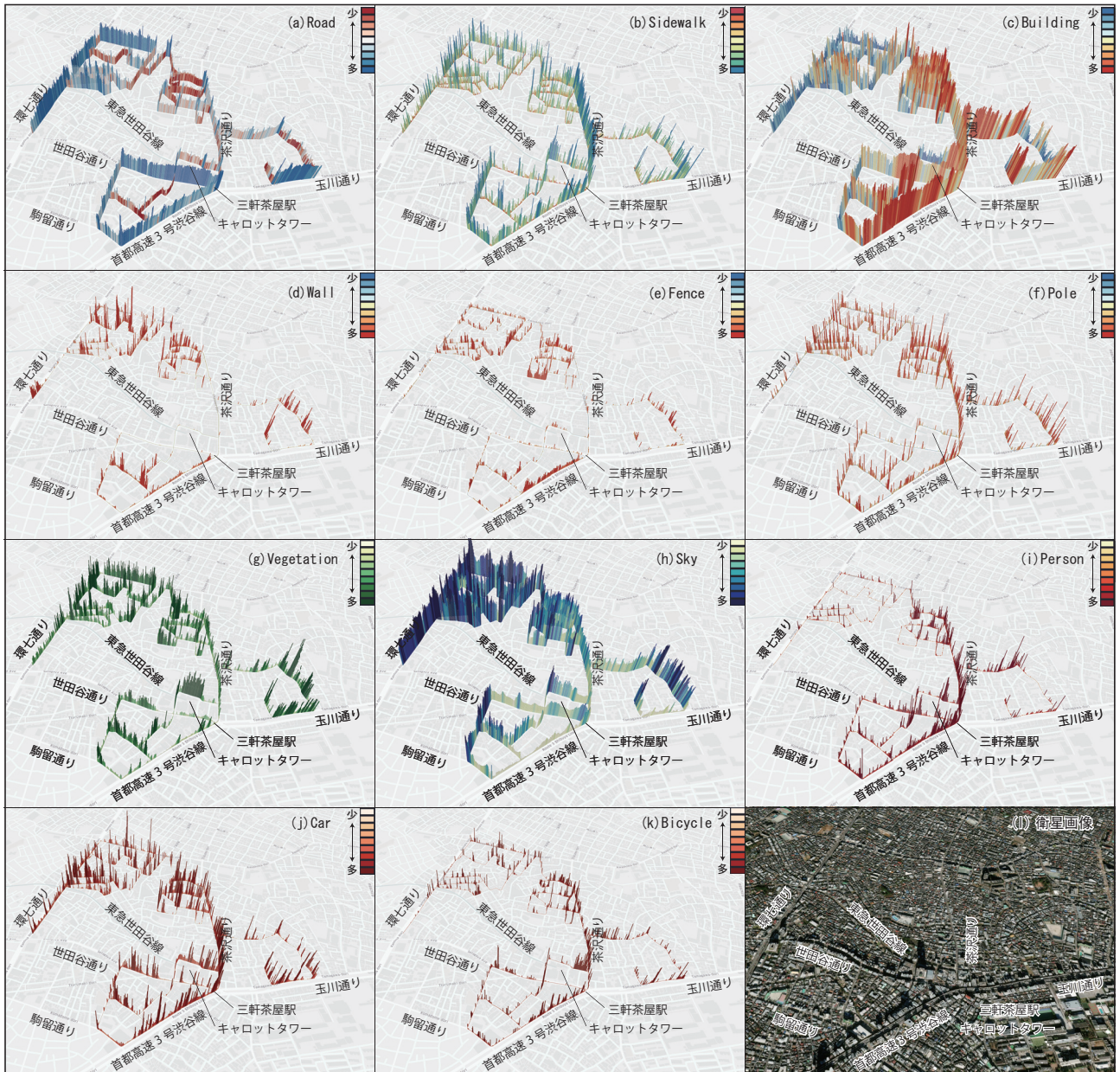


図8 Semantic segmentationによる街路の特徴の可視化（棒の長さとは、画像中の各構成要素の多さに対応）

せることで、応用の幅はさらに広がることが期待される。

本節で行ったように、走行車両で撮影した全方位画像をパース画像に変換し、Semantic segmentationで街路構成要素割合を算出する方法を用いれば、広範囲の街路景観評価や安全性評価を迅速かつ効率的に行うことができる。その一方で実用上の課題も存在する。例えば、Semantic segmentationの精度を高めるためには出来るだけ高解像度の画像を用いることが望ましいが、逆に処理に要する時間が長くなる。学習に用いて



図9 犯罪情報マップ（警視庁）でみる身近な犯罪の空間分布

いる CityScapes データセットが、特に住宅街などでは、日本特有の景観構成要素に対応せず、segmentation の精度が低下するケースも見られる。日本の街路空間画像を用いた学習用データセットの整備が待たれる。また、木造住宅密集地域のように、自動車が通行できないような細街路での街路画像を、いかに効率的に収集するかについても、検討する必要がある。

5. まとめ

本稿では、車両から撮影された街路全方位画像ビッグデータを活用した地域分析の手法を示し、その有用性と課題を明らかにした。

謝辞

本研究は、2020年度東工大挑戦的研究賞に関する研究課題の一部である。また、株式会社ゼンリンには本研究で必須となる全方位画像データを提供して頂いた。東京大学の山崎俊彦准教授と松井勇佑講師からは、PQk-means を実行する上で技術的な助言を頂いた。さらに、画像処理の一部の作業では、山中理沙氏（東京工業大学 環境・社会理工学院 沖研究室）の協力を得た。この場を借りて関係各位に謝意を表す。

注

- 1) ここでは、VGG16 モデルの FC1 層から出力された 4,096 次元の画像特徴ベクトルを、4 次元に超圧縮している。圧縮後の次元数を増やすことで、一般にクラスタリング精度は向上するが、処理負荷は高まる。
- 2) DeepLabV3+ モデル (Chen et. al., 2018) の学習に用いた CityScapes データセット (Cordts et. al., 2016) で定義されている以下の 19 分類に基づく : road, sidewalk, building, wall, fence, pole, traffic light, traffic sign, vegetation, terrain, sky, person, rider, car, trunk, bus, train, motorcycle, bicycle.

参考文献

K. Simonyan and A. Zisserman (2014) Very deep

convolutional networks for large-scale image recognition, arXiv:1409.1556.

Y. Matsui, K. Ogaki, T. Yamasaki and K. Aizawa (2017) PQk-means: Billion-scale clustering for product-quantized codes, ACM International Conference on Multimedia (ACMMM).

L. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation, arXiv:1802.02611.

M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth and B. Schiele (2016) The cityscapes dataset for semantic urban scene understanding, arXiv:1604.01685 (2020.8.31 参照) <<https://www.cityscapes-dataset.com>>

木澤佐椰茄・沖拓弥 (2021) 街路画像ビッグデータとクラウドソーシングによるアンケートを用いた街路の印象評価構造分析, 「地理情報システム学会講演論文集」 Vol. 30.

西尾尚子・伊藤史子 (2020) 天空率と空間構成要素の関係および天空図形状による空の見え方の規定要因, 地理情報システム学会大会ポスター発表 (P-25)

警視庁 (2016) 犯罪情報マップ. (2021年8月31日参照) <<http://www2.wagmap.jp/jouhomap/Map?mid=2&mpx=139.648190513433&mpy=35.634099649473576&bsw=1519&bsh=664>>