畳み込みニューラルネットワークと全方位空間画像による都市空間の特徴抽出と評価手法

○衣川 雛[*1]，瀧澤 重志[*2]

キーワード：都市景観，全方位画像，深層学習，CNN オートエンコーダー，CAM，次元圧縮

1. はじめに

空間の印象を評価するために，建築や都市計画の分野では，画像を用いた分析が広く行われている．例えば，空の見える量を示す天空率 [1]や緑の量を示す緑視率 [2]などは比較的よく用いられている．一般に画像を使って空間の評価を行う場合は，画像から何かしら意味のある計算可能な指標を抽出する必要があると考えられている．しかし，空間の印象を左右する特徴は無数に考えられ，明示的に定義できるとは限らない．

近年，画像から特徴量を自動的に抽出する深層学習を使った，景観等の印象評価の研究が行われている．それらには例えば，矢吹ら [3]，阿部ら [4]，山田ら [5]，Stephen[6]ら，Liu ら [7]のものなどがある．筆者らも既往研究 [8]で，深度情報を含む 3 次元の全方位画像を，ゲームエンジンを用いて仮想空間からリアルタイムに取得する空間評価システムを開発した．そして，VR を用いた印象評価実験を行い，その評価結果を CNN ベースの画像判別モデルで学習し，検証データの予測精度を把握し，提案手法の有効性を検証した．この研究では，画像から特徴量を事前定義することなく学習で抽出することになるが，目的変数を設定した上でモデル全体を学習させる，すなわち教師あり学習を行うので，目的変数が変化したら，ファインチューニングなどの方法はあるにせよ，原則，都度モデル全体を学習させる必要がある．一般に深層学習の学習には時間がかかるし，場合によってはニューラルネットワーク以外の学習モデルを使って，判別規則を明示化するといった用途も考えられるので，現状の教師有り学習の枠組みでは，画像データを都市の多様な分析に生かすことには限界があると考えられる．

そこで本研究では，教師無し学習の方法である CNN オートエンコーダーを使って，事前に全方位画像の高次元情報を目的変数に依らず次元圧縮し，その後，以前行った印象評価実験の結果を学習させ，既報のモデルと精度や注目領域部分の差異を比較する．

2. 既報 [8]の内容のまとめ

本研究では 3 次元空間評価システムの基盤として，ゲームエンジンとして普及が進んでいる Unity を用いる．以下に本研究に関連する既報の内容をまとめる．

2.1 対象空間の設定

ゼンリンから公開されている Unity 用の 3 次元空間データ ZENRIN City Asset Series[9]の難波エリア（Japanese Naniwa City）を評価対象の空間とした．このエリアは難波エリアを中心とした，東西約 700m 南北約 600m の範囲をモデル化したものである（図 1）．

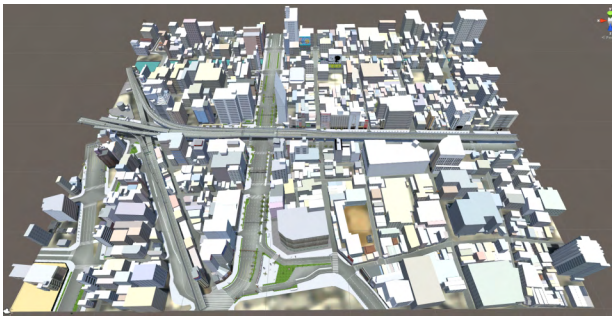


図 1 Unity 用の 3 次元マップ：Japanese Naniwa City[9]

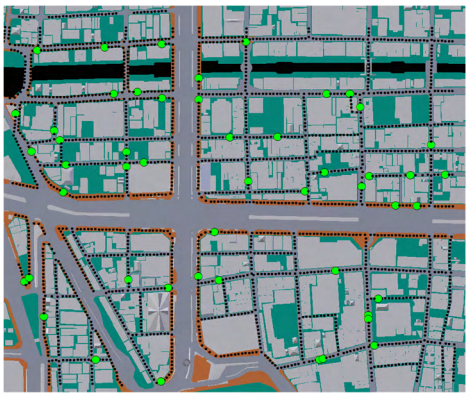


図 2 観測点（黒○）と評価点（緑○）

印象評価実験では，歩行者が普段歩く場所から街を眺めた際の印象を質問するので，道路上の歩行者が歩く場所で空間特徴量を計測する必要がある．場所の選定の基準として，歩道がある区域では歩道幅の中心，歩道がなく車道で交通量が多い地域では，道路の両端，歩道がなく交通量が少ない地域では道路の中央を歩行者の通行場所と仮定し，それらの場所を 5m 間隔で座標を収集した．その結果，図 2 に示す 2,029 点の観測点が得られたが，それらの点をすべて使って印象評価実験を行うのは困難なので，それらの中から図 2 に示す 50 地点を無作為に選択し，印象評価実験における評価点とした．

2.2 空間特徴量の抽出

各観測点の特徴を抽出するために，Unity のアセットである Spherical Image Cam（現在は販売終了）を利用し，

カメラが存在する場所の全方位画像を，リアルタイムで撮影・保存できるように，Unity 上でスクリプトを作成した．全方位の RGB 画像の例を図 3 に示す．なお，既報では RGB 画像の他に深度画像も出力していたが，本研究では（時間の制限により）RGB 画像のみで学習を行った．

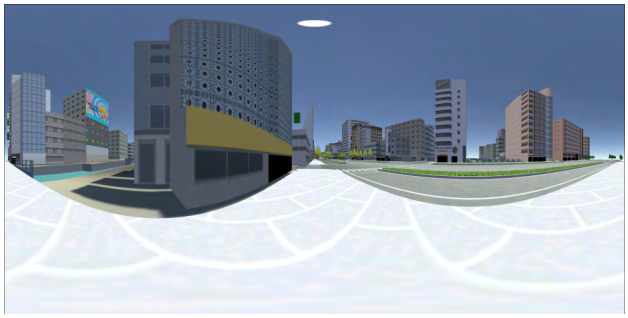


図 3 全方位画像の例

2.3 VR による印象評価実験システムの構築と実験

Unity に Oculus Rift を接続し，印象評価実験を行うためのインターフェースを作成した（図 4）．

図 4 VR による印象評価実験の様子

2016 年の冬に建築系の大学生・大学院生 8 名を対象として印象評価実験を行った．評価は 1（悪い），2（少し悪い），3（少し良い），4（良い）の 4 段階を，直感で判断するよう教示を行った．いずれの被験者も，評価値が 4 の場所は少なく，2 前後が多かった．そのため 4 クラス分類は難しいので，評価値が 1, 2 の場所を Class 0，評価値が 3,4 の場所を Class 1 とした 2 分類問題として学習を行うことにした．表 1 に 2 クラス分類の場合の各クラスの度数分布を示す．全体的に低評価の地点の方が多い．

表 1 被験者ごとの評価クラスの分布

評価値	被験者ごとのクラスの集計							
	A	B	C	Đ	E	F	G	H
1 or 2: Class 0	42	34	23	29	30	40	32	26
3 or 4: Class 1	8	16	27	21	20	10	18	24

2.4 CNN による学習

深層学習のライブラリとして Chainer[10]を用い，そのサンプルファイルの一つである train_imagenet.py をカスタマイズし，GoogLeNet[11]（Batch normalization 版）を用いて学習を行った．

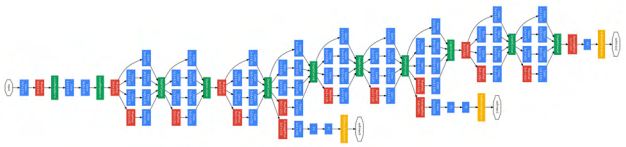


図 5 GoogLeNet の構造 [11]

Unity を用いて，学習に使う画像を各地点で撮影する．画像を扱う深層学習では，人工的にできるだけ多くのデータを用意する必要がある．本研究では各観測点について，図 6 に示すように，カメラを鉛直軸に対して 0 度から 20 度刻みで 340 度まで回転させて，同一地点で 18 枚ずつ撮影した．通常の全方位画像は縦横比が縦：横=1：2 の正距円筒図法だが，CNN に入力するデータの制限から，画像サイズを 256×256 ピクセルの正方形として JPG 形式で出力した．

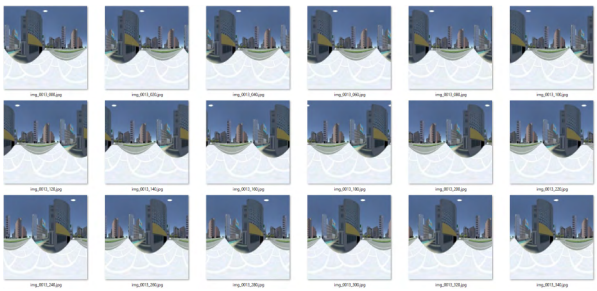


図 6 カメラの回転による同一地点の全方位画像の増量

さらに，学習用と検証用データがおおよそ3:1になるように，各クラスの比率を考慮して，被験者ごとにランダムに地点を分割した．以上の設定に基づき，各被験者で学習用の画像データセットでDCNNの学習を行い，得られたモデルに対して検証用の画像データセットを適用して，それらの誤答率を調べた．

3. CNN オートエンコーダーによる画像情報の次元圧縮

ここからは，本研究で行った CNN オートエンコーダー（以下 CNNAE と訳す）について説明する．基本となるオートエンコーダーは，高次元の入力データを何層かの隠れ層で次元圧縮した後（Encode），そこから反対に高次元化（Decode）して，元のデータとできるだけ同じような出力を得ようとする手法である．これを画像処理データでできるように，畳み込み層で行うようにしたのが CNNAE である．図 7 に CNNAE の概念図を示す．入力画像に関して畳み込み演算（CONV）が行われ，中間の最も画像サイズが小さくなる層が特徴表現層である．そこから逆畳み込み（DCONV）を行って，元の画像にできるだけ類似した画像を出力できるように，各層のパラメータを学習させる．特徴表現層では，大きなサイズの入力画像が，多数の小さな画像に変換されている．この層の情報を使って，画像の特徴ベクトルを抽出できる．具体

的には，特徴表現層の各チャンネルの画像全体を平均プーリング（GAP）することで，フィルタ数の次元の特徴ベクトルを得ることができる．このようにして得た特徴ベクトルを，本研究では単純パーセプトロンによって，次に述べる評価実験の学習に用いる．

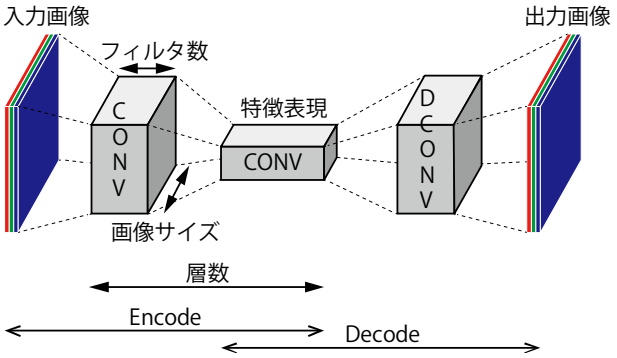


図 7 CNN オートエンコーダーの概念図

4. 評価実験結果の学習

表 2 に学習で用いる CNNAE の緒言を示す．表 2 に示すように 4 つのモデルを作り，2.1 で説明した 2,039 個の観測点で撮影された全方位画像 16,232 枚について，それぞれの CNNAE を学習させた．

表 2 実験で用いる CNNAN の緒言

	CNNAE1	CNNAE2	CNNAE3	CNNAE4
層数	3	3	2	2
フィルタ数	16, 32, 40	8, 16, 32	16, 32	16, 32
フィルタ幅	5, 5, 3	5, 5, 3	5, 5	8, 8
ストライド	3, 3, 2	3, 3, 2	3, 3	4, 4
特徴表現層の画像サイズ	13	12	27	14

表 3 テストデータの正解率

被験者	GoogLeNet	CNNAE1	CNNAE2	CNNAE3	CNNAE4
A	0.71	0.85	0.85	0.85	0.85
B	0.80	0.69	0.71	0.77	0.72
C	0.55	0.69	0.48	0.62	0.58
D	0.81	0.75	0.65	0.72	0.72
E	0.64	0.68	0.72	0.59	0.70
F	0.96	0.85	0.87	0.90	0.85
G	0.86	0.75	0.78	0.73	0.68
H	0.77	0.75	0.70	0.61	0.61
Ave.	0.76	0.75	0.72	0.72	0.71
Std.	0.12	0.06	0.11	0.11	0.09

次に，50 の評価点の画像 900 枚について，学習したそれぞれの CNNAE を使って特徴量を抽出し，2 章の分類問題のデータを，単純パーセプトロンを使って学習させ，10-fold 交差検証により，平均正答率を求めた．それらの結果と，2.で行った過去の GoogLeNet の正解率を比較したものを表 3 に示す．最も構造が複雑な CNNAE1 の平均精度は GoogLeNet に近く，その標準偏差は GoogLeNet を凌駕している．精度については，CNNAE による教師無し学習でも，教師あり学習のモデルを使った場合と比較して安定して良好な結果を得ることができた．

5. Class Activation Mapping(CAM)[12]による注目領域の比較

CAM は，CNN ベースのニューラルネットで分類問題を解いた時に，分類モデルが画像のどこに注目したのかを可視化する手法である．この概念図を図 8 に示す．CAM では CNN の最終層の画像の画素値を平均プーリングして，得られた特徴ベクトルの重み付き和として分類モデルを構築する．最終層の特徴マップの画素をその重みで積を取った上で，和を取って拡大することで，CNN の注目地点を可視化できる．

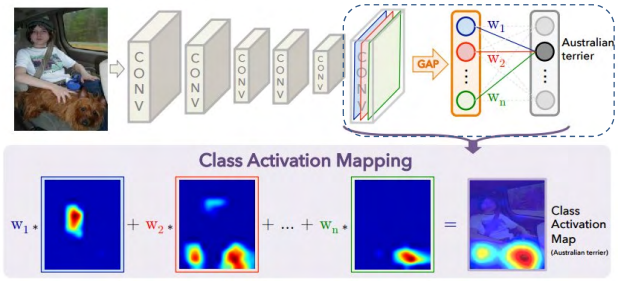
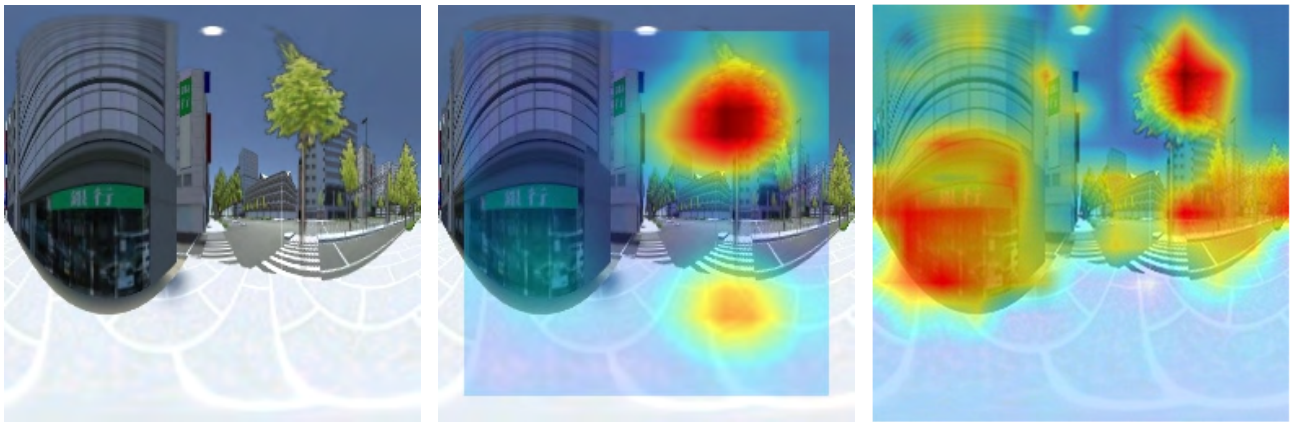


図 8 CAM の概念図 [12]

この方法を使って，被験者 F の GoogLeNet と CNNAE1 モデルの注目領域を可視化した例を図 9，10 に示す．GoogLeNet の場合，高評価地点では，街路樹のあたりにはっきりとした注目領域が形成されている．一方，低評価地点の注目領域は奥行の方向に延びている．CNNAE1 の場合，高評価地点では木や手前の建物の 1 階のファサードなどに，低評価地点では全体的に注目領域が広がっている．CNNAE の方は注目領域が分散する傾向があり，GoogLeNet と比較して解釈が難しいと結論付けられる．

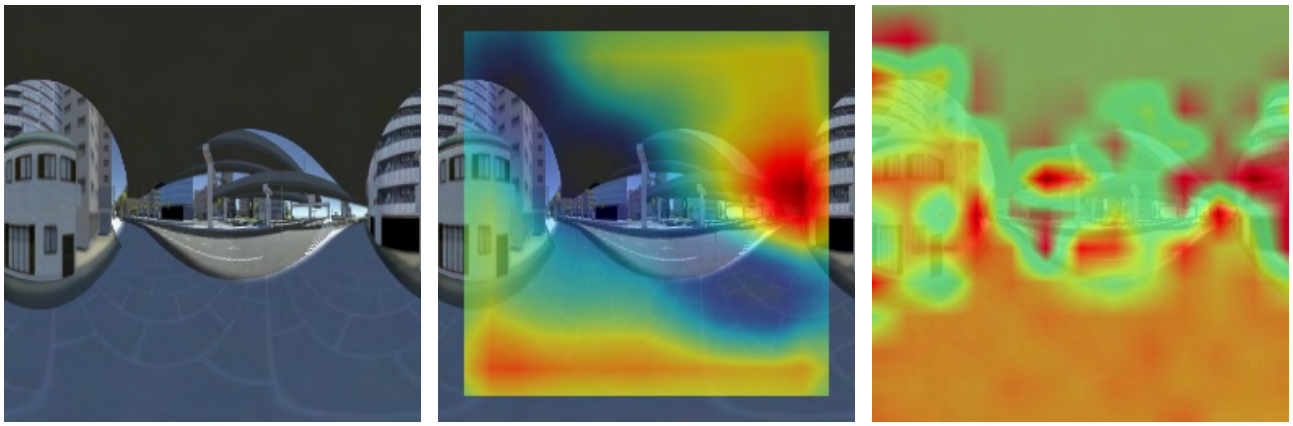
6. おわりに

本研究では，教師無し学習の方法である CNNAE を使って，事前に全方位画像の高次元情報を目的変数に依らず次元圧縮し，その後，以前行った印象評価実験の結果を学習させ，既報のモデルと精度や注目領域部分の差異を比較した．その結果，精度や安定性は教師有学習と比較して遜色ないことを確認した．一方，CNNAE の注目領域は解釈が難しいため，その理由はモデルの構造などさらなる分析が必要と考えられる．



(a) 元画像　　(b) GoogLeNet　　(c) CAE1

図 9 CAM による CNN の注目点（高評価地点）



(a) 元画像　　(b) GoogLeNet　　(c) CAE1

図 10 CAM による CNN の注目点（低評価地点）

謝辞

本研究の一部は科研費基盤 C(16K06652), 基盤 A(V16H01707) の補助の下で行われました．研究を主に担当した，当時学部 4 年生の古田愛理さんと上田健之祐君に感謝します．

［参考文献］

1) 国土交通省住宅局: 建築物に対する景観規制の効果の分析手法について, 2007, http://www.mlit.go.jp/jutakukentiku/jutaku-kentiku.files/keikan/keikankisei.pdf
2) 藤井健史，山田悟史，廣瀬徳郎，及川清昭：CG モデルによる全方位緑視率の計量手法とその応用可能性，日本建築学会技術報告集, 19(43), pp.1067-1072, 2013.
3) 矢吹和也，安福健祐，阿部浩和: Deep Learning を用いた景観評価の手法に関する基礎的研究，日本図学会大会学術講演論文集, pp.23-28, 2015.
4) 阿部浩和，李ロウン，安福健佑: 街路空間評価におけるディープラーニングの適用可能性，日本図学会大会学術講演論文集, pp.85-90, 2016.
5) 高橋秀彬，山田悟史：Deep Learning を用いた街並み画像の分類と感性評価の推定，日本建築学会第 40 回情報・システム・利用・技術シンポジウム論文集：報告，pp.329-329，2017.
6) S. Law et al.: An application of convolutional neural network in street image classification: the case study of London, ACM GeoAI'17 Proc. 1st Work. on Artif. Intell. Deep. Learn. for Geogr. Knowl. Discov, 2017.
7) L. Liu, et al.: A machine learning-based method for the large-scale evaluation of the qualities of the urban environment, Computers, Environment and Urban Systems, 65, pp.113-125, 2017.
8) A. Takizawa and A. Furuta, 3D Spatial Analysis Method with First-Person Viewpoint by Deep Convolutional Neural Network with Omnidirectional RGB and Depth Images, eCAADe 2017, Sapienza University of Rome, Rome, Italy, pp.693-702, 22 Oct. 2017.
9) Zenrin 株式会社：ZENRIN City Asset Series, http://www.zenrin.co.jp/product/service/3d/asset/
10) S. Tokui et al.: Chainer: a Next-Generation Open Source Framework for Deep Learning, Proceedings of Workshop on Machine Learning Systems in The Twenty-ninth Annual Conference on Neural Information Processing Systems, 2015.
11) A. Krizhevsky et al.: ImageNet Classification with Deep Convolutional Neural Networks, Pereira, F. et al. (eds), Advances in Neural Information Processing Systems 25, Curran Associates, Inc.,pp.1097-1105, 2012.
12) B. Zhou, et al.: Learning Deep Features for Discriminative Localization, Computer Vision and Pattern Recognition, 2016.

*1 大阪市立大学生活科学部　学部生
*2 大阪市立大学生活科学研究科　教授　博士（工学）