

Deep Learning を用いた印象評価推定 AI の作成と検証

- 街並み画像の街路名と訪問意欲を対象に -

DEVELOPMENT AND VERIFICATION OF THE IMPRESSION DEDUCTION MODEL FOR CITY LANDSCAPE WITH DEEP LEARNING

- Street names city landscapes and desire/no desire or degree of desire to visit -

山田 悟史*, 大野 耕太郎**

Satoshi YAMADA, Kotaro ONO***

This research utilizes artificial intelligence (AI) by applying deep learning in architecture and urban design. The study's novelty is clarifying the applicability of deep learning to impressions of city landscapes as visual elements of design to differentiate "desire/no-desire to visit" and "degree of desire to visit". Street names were therefore classified as constituents of image consciousness. Thus, deduction AI that could achieve high precision from the viewpoint of psychological and medical treatment statistics is developed, and the confidence interval of 100 kinds of AI developed with this method is confirmed to be small.

Keywords : *Artificial intelligence, Deep learning, Impression, Sensibility, City landscape*

人工知能 (AI), Deep Learning, 印象, 感性, 街並み

1. はじめに

近年, Deep Learning を基盤とする人工知能 (AI) が急速に社会的な関心を集め, 様々な分野において応用が期待されている。それを反映するように書籍も多数発表され, 具体的な技術を解説する技術書だけでなく, 事例紹介や実務上での活用の可能性を示唆するような一般向けの本も多数出版されている。契機の一つには, 世界的な統計競技サイト「ILSVRC (ImageNet large scale visual recognition challenge)」での成果が挙げられる。「ILSVRC」には画像の「クラス分類」の精度を競う部門がある。この部門は, 公開されているクラス分類付き画像のクラスを画像情報から推定する精度を競う部門である。2012 年にこの部門で Deep Learning を使った AI (AlexNet¹⁾) が以前の成果を大きく上回ったことが, Deep Learning を用いた AI に対する注目度が近年に高まった契機の代表例と言える。Deep Learning の利点の一つを画像の分類を例に挙げると, 画像を表現する特徴量を人が指定せずに, 予め用意した分類に近づくように特徴量自体を学習させられる点が挙げられる。このような特徴量自体を学習する特徴を活かし, 対象の適切な特徴量記述及び結果との因果関係の解明が複雑であった「画像認識」「音声認識」「自然言語処理」などの分野において, 既存の結果を上回る成果が報告され大きな展開を生み出した。

上述の例に限らず Deep Learning という萌芽的な技術の適用可能性は建築・都市分野にも様々あり得ると考えられ, 本研究が着目し

ているのは, デザインに対する感性・印象評価の学習と推定である。視覚を通して抱く感性・印象はデザインの重要な目的関数であり, デザインは当然ながら感性・印象評価が高くなるよう思考される。課題の重要性を反映するように, 学術的にも列挙し切れない程多くの既往研究がある。筆者らも, 画像中の特定要素の割合や全方位視認率 (緑視率) を空間の定量的な表現 (特徴量) に採用し, 自己組織化マップやページアンネットワークを用いて印象評価の推定に取り組んできた^{2) 3)}。筆者らの研究に限らず既往研究では, 空間の特徴量は因果関係の仮説にも基づき分析者が用意してきた。しかし, 視覚を通して認識する空間の特徴量選択と感性・印象の因果関係という複雑な関係は, 今日でも学術的にも実務的にも未解明な部分が残る。そのため, 入力と出力の構造的な把握に基づくデザインに対する感性・印象評価の推定が困難な側面は残っている。このような課題をふまえ, Deep Learning を用いて成果を挙げている分野を俯瞰すると, 前掲の分野を例に, 対象の特徴量選択と結果の因果関係の構造的な把握が複雑であるという類似の性質を持つ分野がある。そのため, 建築・都市のデザインにかかわる感性・印象に対しても, Deep Learning を用いた AI という萌芽的な技術の適用可能性があり得ると考えられる。

上述をふまえ, 本研究では萌芽性の検討対象を, 「街並み画像」から「街路名と訪問意欲」を「推定する AI」の作成, とする。街並みに対する訪問意欲は当然ながら重要である。加えて街並みは多

* 立命館大学 理工学部 建築都市デザイン学科 任期制講師・博士 (工学)

** 立命館大学 理工学部 建築都市デザイン学科

* Lecturer, Dept. of Architecture and Urban Design, Ritsumeikan Univ., Dr. Eng.

** Undergraduate, Dept. of Architecture and Urban Design, Ritsumeikan Univ.

数の要素から構成されており, 視覚的な特徴量と訪問意欲との因果関係は複雑である。このような複雑な因果関係によって出力される結果の推定には, 上述したような Deep Learning の特徴を活かした AI が適している可能性があり, 他分野で成果が報告されている。

上述の Deep Learning の街路景観に対する適用可能性に関する研究には, Lun Liu⁴⁾ らや Stephen Law⁵⁾ らの研究があり, 本研究と内容は異なるが街路景観への Deep Learning の適用可能性が示唆されている。また VR 技術を併用した研究も行われており, 瀧澤ら⁶⁾ は VR 空間で実施した印象評価実験を VR 空間の 360 天球画像から推定することに Deep Learning を適用している。いずれも新規的で萌芽性の高い研究である。しかし, 異なる都市の街路の街並み画像と印象(訪問意欲の有無と度合)の分類・推定に Deep Learning の適用を試みた研究は筆者が知る限りでは見当たらず, この点が本研究の新規性に該当する。

やや発展的な内容になるが推定する AI とは別の視点からも本研究の位置付けを示す。公開されている Deep Learning を用いた AI 技術には, 分類以外にも建築・都市に関するデザイン分野への応用が考えられる AI がある。「コンテンツ生成 AI」である。具体的には画像を対象とした「GAN(DCGAN)」⁷⁾ である。これは対立的な関係にある生成器と判定器を組み合わせた AI であり, 学習データの持つクラス(例えば「顔」)に対応した新たな画像を写真のような精度で生成できることが報告されている。既往研究のクラスは「顔」「車」「橋」といった物体認識を対象としたクラスであるが, 特定の建築家をクラスにした画像生成という萌芽性が考えられる。しかしこれは, 人々や特定の建築家に対する判定器(作品・デザイン感の判定)を生成できることが前提となる。本研究の「推定する AI」の検証は, このような萌芽性に対する基礎研究にも位置付けられる。

上述の内容に限らず AI 活用に社会的な関心が高まっているが, 活用には新規的な技術の萌芽性を体系化する基礎研究が必要である。本研究では, 上述の「推定する AI」と「コンテンツ生成 AI」という二つの萌芽性を視野に, 建築・都市に関するデザイン分野での Deep Learning を用いた AI 活用の基礎研究として, 「街並み画像」への適用可能性を検証する。具体的には, 「街並み画像」の「街路名(21クラス分類)」、「訪問意欲の有無(2クラス分類)」、「訪問意欲の度合い」の推定に対する Deep Learning を用いた推定 AI の適用可能性を明らかにすることを目的とする。なお本稿は参考文献⁸⁾を再分析・加筆したものである。

2. 研究概要

2.1 研究の流れ

本研究は「街並み画像」を対象に「街路名」と「訪問意欲」を Deep Learning が学習し推定できるかを検証する。研究の流れを図 1 に示す。「街並み画像」は街路に立った時に見える街並みを想定した画像からなる学習データセットである。まずは都市を選定し学習データセットを作成した。次に「訪問意欲」の推定に先立ち, 顔認識に代表される物体認識が街並み画像では可能であるかを検証するために「街路名」の推定を行なった。物体認識としての精度を確認した後に「訪問意欲」の推定を行なった。なお「訪問意欲」は二種類である。一種類目は「訪問したい」「訪問したいとは思わない」の二者択一の「訪問意欲の有無(2クラス分類)」推定である。二

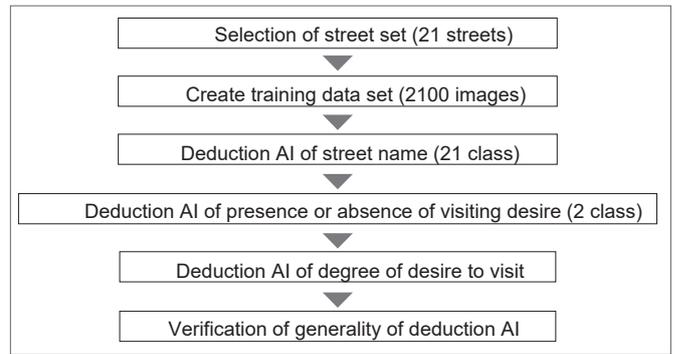
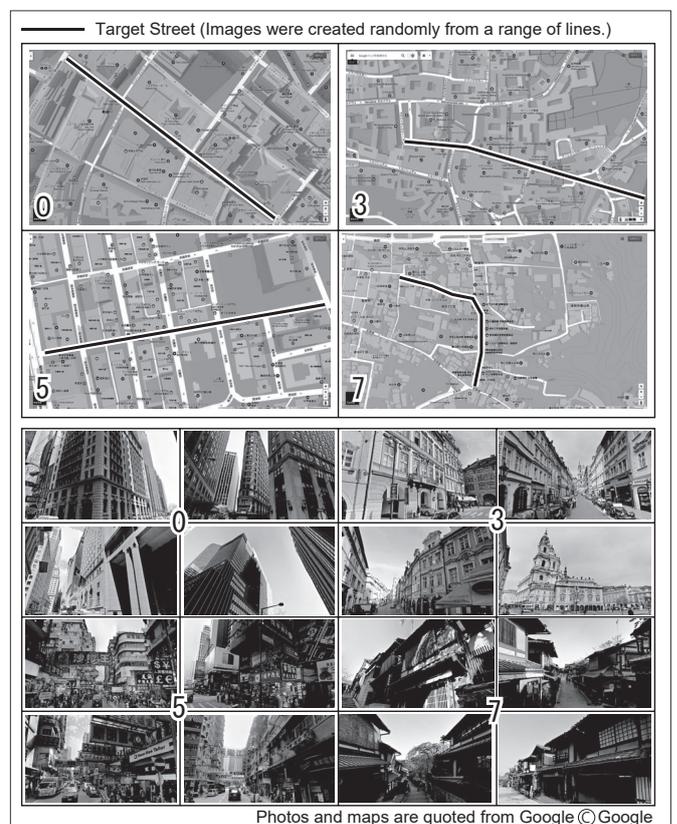


Fig.1 Flow of research

Table 1 List of target streets

Class	Country	City	Name
0	America	New York	Wall Street
1	Australia	Melbourne	Hosier Ln
2	Canada	St John's	Cabot Street
3	Czech	Prague	Mostecka
4	England	London	Oxford Street
5	Hong Kong	Mong Kok	Shandong Street
6	Italy	Firenze	Via dei Calzaiuoli
7	Japan	Kyoto	Sanneizaka Street
8	Japan	Osaka	Dotonbori Street
9	Japan	Tokyo	Takeshita Street
10	Mexico	Guanajuato	Pocitos
11	Peru	Cusco	Questa de Santa Anna
12	Portugal	Porto	R.de Sao Joao
13	Russia	Moscow	Tuverskaya Street
14	Scotland	Edinburgh	Cockburn Street
15	South Africa	Cape Town	Dorp Street
16	Korea	Seoul	Bukchon Hanok Village
17	Spain	Barcelona	Plaça de Catalunya - Gran Teatre del Liceu
18	Taiwan	Chiufen	Chiufen area
19	Thailand	Bangkok	phraeng Nara Rd
20	Arab	Dubai	Al Worood 1 Street



Photos and maps are quoted from Google © Google

Fig.2 Image example

Table 2 Mixture matrix of verification data (street name)

[y_] is the correct class, [y'_] is the Deduction class																						
	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_10	y_11	y_12	y_13	y_14	y_15	y_16	y_17	y_18	y_19	y_20	Recall(再現率)
y_0	57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0.9661
y_1	1	52	0	0	0	0	0	0	0	3	1	0	2	0	0	0	0	1	1	0	0	0.8524
y_2	0	0	51	0	0	0	0	0	0	0	0	1	2	0	0	1	0	0	0	0	0	0.9272
y_3	4	0	0	53	6	0	4	0	0	0	1	3	1	2	0	0	0	0	0	0	0	0.7162
y_4	0	0	0	4	39	0	1	0	0	0	0	0	2	2	1	0	0	0	0	0	0	0.7959
y_5	0	0	0	1	0	51	0	0	1	2	1	0	1	0	0	0	0	0	1	1	0	0.8644
y_6	4	0	0	6	0	0	48	0	0	0	1	0	5	0	2	0	0	4	0	0	0	0.6857
y_7	0	0	0	0	0	0	0	55	0	0	0	0	0	0	0	0	0	3	1	0	0	0.9322
y_8	1	0	1	5	0	2	0	0	47	6	1	0	0	0	0	0	0	0	0	0	0	0.746
y_9	0	0	0	1	0	2	0	0	0	55	1	0	1	0	0	0	0	0	1	0	0	0.9016
y_10	2	0	1	0	0	0	0	0	0	0	51	4	0	0	0	3	0	0	0	0	0	0.836
y_11	0	0	0	1	0	0	0	0	0	0	0	60	2	0	0	0	0	0	0	0	0	0.9523
y_12	1	0	0	2	0	1	3	0	0	0	0	0	53	0	0	0	0	3	0	0	0	0.8412
y_13	1	0	0	0	0	0	1	0	0	0	0	0	0	44	2	0	0	1	0	0	0	0.8979
y_14	0	0	0	0	0	2	3	0	0	0	0	0	1	1	52	0	0	2	0	0	0	0.8524
y_15	0	0	3	0	0	0	0	0	0	0	1	3	0	0	0	52	0	0	0	0	0	0.8813
y_16	0	0	0	0	0	0	0	1	0	0	0	1	0	2	0	0	55	0	0	0	0	0.9322
y_17	5	0	0	1	1	0	1	0	0	0	0	0	3	3	0	0	0	46	0	0	0	0.7666
y_18	0	2	2	0	0	3	0	0	1	0	0	0	1	0	0	0	2	2	42	0	0	0.7636
y_19	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	59	0	0.9672
y_20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	59	1
Precision(予測適合率)	0.75	0.9629	0.8644	0.7162	0.8478	0.8225	0.7868	0.9821	0.9591	0.8333	0.8793	0.8333	0.7162	0.8148	0.9122	0.9285	0.9649	0.7187	0.913	0.9833	1	
F-Measures	0.8444	0.9042	0.8946	0.7162	0.821	0.8429	0.7327	0.9564	0.8392	0.8661	0.8571	0.8888	0.7736	0.8543	0.8812	0.9042	0.9482	0.7418	0.8316	0.9751	1	
Accuracy(正確性)	0.8579		Avg.Precision		0.8661		Avg.Recall		0.8609		Avg.F-Measures		0.8607		K-Factor(Non-Weigut)		0.8579		P-Value		2.2e-16	

種類目は「訪問したい度合い」の推定である。そして最後に AI という方法の一般性について検証した。

2.2 研究対象・学習データセット・学習モデル

研究対象と街並み画像は大都市や観光情報^{注1)}から選定した表1に示す21街路から作成した。画像は街並みの画像であるため、画像が単一の建築物・地面・空に占められることのないように配慮しながら Google Earth のストリートビューから作成した。枚数は既往事例と予備実験を参考に1街路100枚ずつの2100枚とした。図2に21都市中4都市の対象街路と画像例を示す。なお印刷の都合でグレースケール画像を掲載しているが実際にはRGBカラー画像(8bit)である。学習モデルの具体的な構成は次章にて示すが、共通する方針を本章にて示す。学習モデルには小さいサイズのフィルターによる畳み込みを繰り返すことで精度を向上させる Network In Network 型の「VGG」^{注2)}を参考に学習を実行した^{注3)}。

3. 街路名の分類

印象評価の推定に先立ち、建築・都市に関する画像への適用可能性の基礎分析として、他分野で成果を挙げている物体認識が本分野でも可能であるかを検証した。具体的には、街並み画像の街路名をクラスとする21クラス分類についての適用可能性を検証した。

学習モデルを図3に示す。「VGG」を参考に研究対象に併せて少し変更している。学習データセットは左右反転により水増し処理を行った4200枚の3チャンネル・200ピクセル*338ピクセルの街並み画像である(ホールドアウト法におけるランダムな画像の分割確率は学習用70%・検証用30%)。各層における畳み込みの設定は図中のCNNと示した文字の下部にある[]内が示しており、左から順にフィルター数、フィルター高さ、フィルター幅である。活性化関数には「CRReLU^{注4)}」^{注5)}、Poolingには「MaxPooling^{注6)}」を用い、図中の位置にて「DropOut^{注7)}」を適用している。全結合の後には活性化関数「SoftMax^{注8)}」を適用した後にクラスを決定し誤差を「Categorical Cross Entropy^{注9)}」にて算出し学習する構造である。学習の更新方法は「Adam(Alpha = 0.001, Beta2 = 0.9, Beta1 = 0.999)^{注10)}」, 「Batch size = 64^{注11)}」で学習を行った。

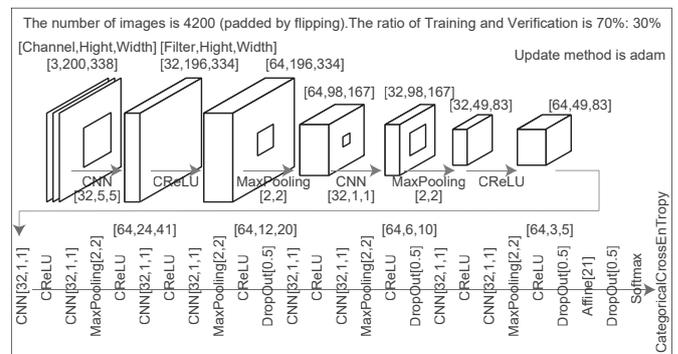


Fig.3 Learning model (street name)

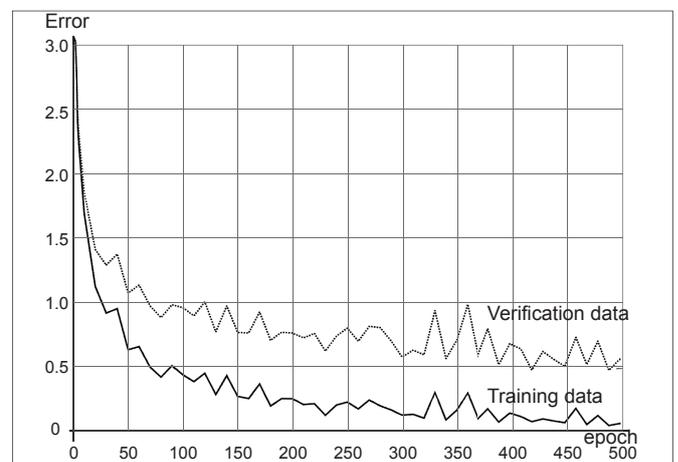


Fig.4 Learning transition (street name)

学習の推移を図4に示す。グラフの縦軸はAIの誤差(Categorical Cross Entropy)の大きさを示している。一方は作成する際に使用した学習用のデータセットに対する誤差の大きさであり、他方はAI作成の学習には用いていない検証用データセットの誤差の大きさを示している。横軸は学習の回数(epoch)である。誤差の推移をみると、既往研究を参照すれば大きいと言及できる約3の誤差がepoch500までに約0.1に収束していることが読み取れる。そこでepoch500までの最適モデルを用いて推定を行った具体的な結果を表2に示す^{注12)}。Recallは再現率であり、あるクラスにおい

て、実際の枚数に対して何枚を正しいクラスに推定したかを示す。Precision は予測適合率であり、あるクラスにおいて推定した枚数に対して何枚を正しいクラスに推定したかを示す。Accuracy は正確性であり全画像枚数のうち何枚を正しいクラスに推定したかを示す。なお Recall と Precision はトレードオフ関係にあることから、このような推定精度の検証には両者の調和平均である F-Measure も用いられる^{注13)}。各クラスの Recall と Precision と F-Measure の値を見る約 70% から約 90% となっておりクラス毎の精度に約 20% の違いがみられる。全体の精度を見ると、Recall の平均値、Precision の平均値、Accuracy、F-Measure の平均値は全て約 86% である。各指標の値の散らばりが小さいことから作成した推定 AI の精度は約 86% であると考えられる。人でも学習を行えば同程度の精度で街並み推定は可能である。特に専門家であれば同程度以上の推定も可能であると考えられる。しかし約 30 分で学習し、人に依存せずに推定 AI を誰でも再現・利用できるという観点をふまえると、比較的高い精度と考えられる。次に統計的な精度評価を示す。用いた統計指標は K 係数^{注14)} 及び統計的仮説検定である。K 係数は 0.8579 となり、P-Value は $2.2e-16$ となった。K 係数 0.8579 は、医療統計の分野の知見を引用すれば高い一致率と言及することができる。また P-Value からも有意な結果となった。

以上のように本章では、AI の建築・都市に関する画像への適用可能性の萌芽性について、街並み画像の街路名の推定 AI については、F-Measure の観点からは約 86%、K 係数の観点から 0.8579 (P-Value= $2.2e-16$) という精度であることを明示した。学習モデルは日々進化しており、さらに高い精度をもつ推定 AI を作成する余地はあり得ると考えられるが、統計的にも高いと肯定できる精度を確認したことから、街並み画像の街路名推定という物体認識に対しても推定 AI は明示した程度には適用可能であると考えられる。そこで次章以降は訪問意欲の推定 AI を作成する。

4. 訪問意欲の有無の推定

本章では印象評価の推定への適用可能性について検証を行う。画像は前章と同様であるが、本章のクラスは「訪問したい」「訪問したいとは思わない」という印象評価の 2 クラスである。教師となる分類クラスには 1 名の被験者（建築系学科所属の 4 回生）による街路を対象とした被験者実験の結果を用いた。被験者実験では画像を一覧した後に直感でクラスに分類して貰った。通常このような被験者実験には複数人の代表値を用いることが多いが、個人（例えば建築家）のデザイン感に基づく感性・印象評価を学習・推定できるかという観点から被験者は 1 人とした。

学習モデルを図 5 に示す。前章からの変更点は損失関数を「Squared Error^{注15)}」とした点である。クラス分類に採用可能な損失関数は他にもあり、2 分類に特化した損失関数もある。しかし次章において訪問意欲の度合いの推定を予定していることから、分類クラスに該当する得点が 0 から 1 の連続変数で得られる「Squared Error」を採用した。学習の推移を図 6 に示す。誤差の推移をみると epoch 500 までに収束していることが読み取れる。そこで epoch 500 までの最適モデルを用いて推定を行った結果を表 3 に示す。実際の訪問意欲の有無の画像の割合は「有」が 56.66% であり、推定した割合 56.03% との差は小さい（比率の差の効果量（対応あり）

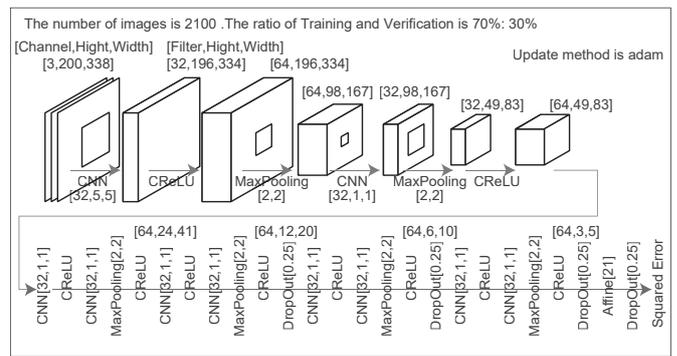


Fig.5 Learning model (desire to visit)

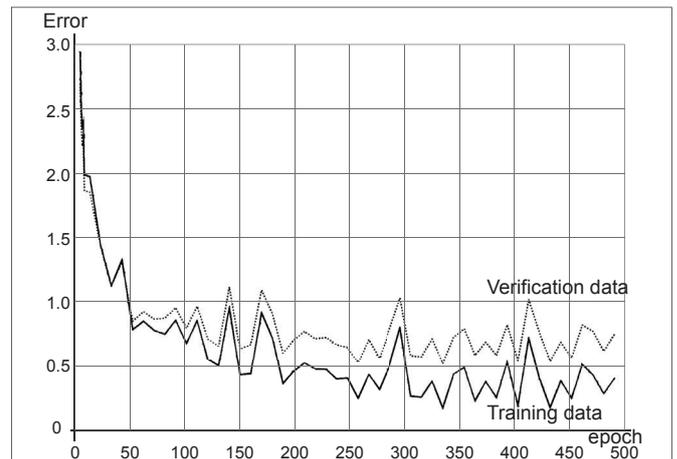


Fig.6 Learning transition (desire to visit)

Table 3 Mixture matrix of Verification data(desire to visit)

[y₋] is the correct class, [y'₋] is the Deduction class

	y' = 0 (Want)	y' = 1 (Don't want)	Recall(再現率)
y = 0 (Want)	328	9	0.9732
y = 1 (Don't want)	25	268	0.9146
Precision(予測適合率)	0.9291	0.9675	
F-Measures	0.9506	0.9403	

Accuracy (正確性)	0.9460	Effect Size	0.0539
Avg.Precision	0.9483	K-factor	0.8920
Avg.Recall	0.9439	P-Value	2.2E-16
Avg.F-Measures	0.9454		

で表現すれば約 0.05)。画像ごとの対応を含めて確認するために、Recall と Precision と F-Measure を算出すると表のような値になった。全体の精度として、評価が一致した割合である Accuracy をみると約 94% (630 枚中の 596 枚が一致) であり、Recall・Precision・F-Measure の平均値も約 94% である。K 係数は 0.8920、P-Value は $2.2e-16$ となり、高い一致率かつ有意な結果となった。

以上より、前章と同じくより高精度な推定 AI 作成の余地はあり得るが、本研究は、被験者の訪問意欲の有無と高い一致度をもつ訪問意欲の有無を推定する AI を作成することができたと言える。

5. 訪問意欲の度合いの推定

本章では前章にて作成した推定 AI を援用することで、訪問意欲の度合いの推定が可能かを検証する。前章では誤差に「Squared Error」を採用した。これにより、訪問意欲の有無のクラス分類の推定に際して、「訪問したい」を 0、「訪問したいとは思わない」を 1 とする連続変数が得られている。本章ではこれを訪問意欲の度合いとして解釈可能であるかを検証した。検証方法は推定値と新たな

被験者実験の比較である。被験者実験では, 前章と同じ被験者に「訪問したい」度合いを 10 段階で評価して貰った。画像枚数は, 2100 枚に 10 段階の度合いを回答することは被験者の負担が大きいと考え, ランダムに各都市から抽出した 50 枚とした。本章はこの回答値 (Human と表記) と推定値 (AI と表記) の比較である。値の並行ジッタープロット, ヒストグラム, プロット図を図 7 に示す^{注16)}。

まず両者の代表値の差についてである。分布の形状を考慮して中央値を代表値として比較すると, 並行ジッタープロット・ヒストグラムから中央値は Human と AI どちらも約 75 で近似であるが, AI の方が僅かに高い (差は 1.56)。この差の統計的な検証として効果量を算出した。算出した効果量は, ヒストグラムから正規性は仮定できないと判断できることから Wilcoxon の符号付順位検定 (対応のあるノンパラメトリックデータを対象とする検定) の効果量^{注17)}である。効果量は約 0.18 (観測値の単位だと約 1.56) となった。心理学分野の研究では 0.2 以下を small と言及する。本研究も本章は心理に関するものであるためこれを引用すれば, 両者の差は小さいと言及することを肯定できる。次に統計的仮説検定であるが, 標本結果の母集団・一般性に適用の解釈において有意差が認められたとしても, 差の効果量としても意義としても小さいと考えられることから Wilcoxon の符号付順位検定は実施しなかった。

次に度合いの高低の傾向の推定についてである。ヒストグラムから値の分布の範囲は近似していることが読み取れる。サンプル毎の Human と AI の値の対応を示すプロット図からは, 全体的に正の相関関係であるが, Human の回答値が低いサンプルを AI が実際より高く推定していることが読み取れる。この度合いの高低の傾向の推定の精度について相関係数を算出した。算出した相関係数は, 分布の形状から両変数はノンパラメトリックと考え, Spearman の順位相関係数^{注18)}を選択した。Spearman の順位相関係数は 0.7564 となった。心理学分野の研究においては 0.5 以上を Large と言及する。これを引用すれば, 相関関係を高く推定できていると言及することを肯定できる。しかし相関係数が高くても, 標本結果の母集団・一般性に適用の解釈において無相関である可能性があるため, 無相関仮説の検定^{注19)}を行なった。p 値は 0.0005742 となり 1% 水準において有意な結果となった。このことから相関関係の母集団への適用可能性と一般性についても肯定できると考えられる。

以上より, 前章と同じく改善の余地はあり, 記した実験条件下ではあるが, 被験者の訪問意欲の度合いと差が小さく, 高い相関関係を有する度合いの推定が可能で AI を作成することができたと言える。

6. Deep Learning の一般性の検証 (度合いの推定)

最後に, Deep Learning を用いて作成した AI という方法論の一般性について検証した。これは前章にて示した AI が偶然優れた精度を持って生成された AI である可能性があるためである。これは学習過程に含まれるランダム性に起因する。この点について検証するために同一の方法で 100 種の AI を作成し, その結果をサンプルとらえた 95% 信頼区間を図 8 のように確認した。効果量において 0.2 を跨がず, 相関係数においても 0.6 を跨がない結果となった。このことから, 本研究の実験条件下において, Deep Learning を用いた作成した AI という手法は, 原稿にて明示した精度程度に一般性のある方法であると言える。

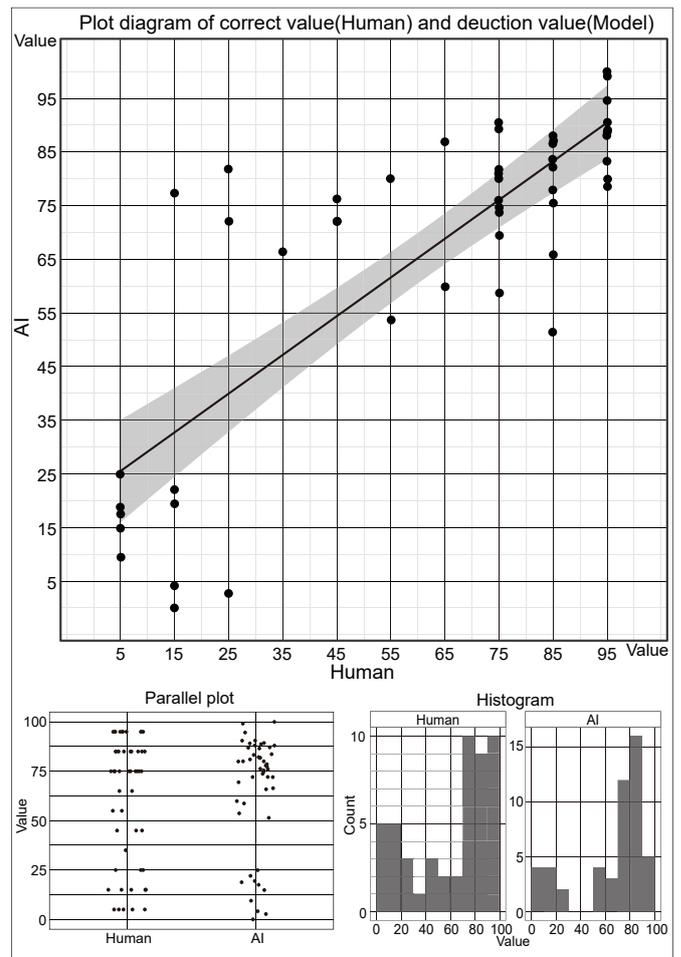


Fig.7 Result of deduction (degree of desire to visit)

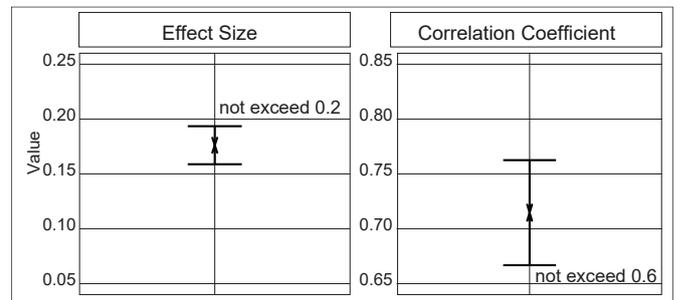


Fig.8 95% Confidence interval (Sample size (Number of AI) = 100)

7. 考察とまとめ

本研究は建築・都市に関するデザイン分野での Deep Learning を用いた AI 活用という萌芽性を体系化するための基礎研究として, 3 種の適用可能性について検証した。下記にまとめる。

まず「街路名のクラス分類 (21 クラス)」の推定において比較的高い精度の推定 AI を作成した。その精度を統計的に表現すれば, F 値が約 86%, K 係数が 0.8579 (p 値=1.6e-16) である。

次に「訪問意欲の有無 (クラス分類)」において, 被験者の訪問意欲の有無と高い一致度をもつ推定 AI を作成することができた。その精度を統計的に表現すれば, K 係数が 0.8920 (p 値=2.2e-16) である。そして「訪問意欲の度合いの推定」においては, 被験者の訪問意欲の度合いと差が小さく, 高い相関関係を有する推定が可能で AI の作成に成功した。同じく統計的に精度を表現すれば, Wilcoxon の符号付順位検定の効果量が 0.18, Spearman の順位相関

係数は 0.7564, (p 値 =0.0005742) である。

最後に, Deep Learning を用いて作成した AI という方法論の一般性として, 同一の方法で作成した 100 種の AI をサンプルととらえた 95% 信頼区間が狭いことを確認した。具体的には, 効果量が 0.2 を超えず (効果量が small と言及される閾値は 0.2), 0.6 を下回らない (相関係数が Large と言及される閾値は 0.5) 結果となった。なお本研究の主眼は印象評価の推定により景観デザインの支援であるため, 街路名の推定 AI の検証の掲載は割愛したが頑健な精度であったことを最後に補足する。

以上のように本稿では, Deep Learning を用いて作成した AI による推定という手法について, 実験条件下かつ頻度主義の観点からではあるが, 前掲のような精度と一般性を明示した。これにより建築・都市デザイン分野における Deep Learning・AI 適用という萌芽性を体系的に整理するための一端を示すことができた。

学習モデルは日々進化しており, 物体認識においてはさらに高い精度の結果も報告されている。印象の推定に関する事例は多くないが適用が進むと考えられる。本研究においても, さらに高い精度をもつ推定 AI を作成することを今後の課題とする。他にも, CAM などによる推定根拠の解明, 都市を変えた検証, 被験者・質問項目を変えた検証, 複数被験者の共通傾向の推定, 街並み以外や推定対象を写真以外に変えた検証に取り組む予定である。加えて, 本研究の AI は推定 AI であるが, 一定の精度を確保できるクラス分類が可能であることを把握したことは, 1 章にて示したようにコンテンツ生成 AI の可能性を示したとも言える。本研究の結果を援用し, GAN (DCGAN) によるコンテンツ生成 AI も試行する予定である。

謝辞

データセットの作成は 2017 年度 建築情報研究室の卒業生の高橋秀彬君の協力によるところが大きい。記して謝意を表する。

注

- 注 1) 街路を選定するために参考にした一般的な情報は, 美しい街並みの都市ランキング (wondertrip), 何か懐かしく温かく感じるアジアの混沌とした町並み (JapanTimesAggregator), などである。但し Google ストリートビューで閲覧不可能な場所は候補から除外した。
- 注 2) 参考文献 10) にて提案されて深層学習の構造である。「VGG」は 2014 年のコンペティションにおいて, 「GoogLeNet」に次ぐ 2 位の成績を収めた学習モデルである。「VGG」は精度では「GoogLeNet」に劣る結果となったが, 畳み込みを繰り返すという単純な構成により, 対象とクラスの複雑な関係を表現する変数を増やすことで精度を確保し得るという利点がある。この深さを重視するシンプルな構造ながら 2 位ではあるが高い精度が報告された。このことから, 学習層の深さは異なるが多くの研究者・技術者によって応用が試行されている。そこで本研究でも「VGG」を参考にして学習を実行した。
- 注 3) 学習には GTX1080 (ビデオメモリ 8GB, CUDA コア数 2560) 一台を用いた。PC 側のメモリは 64GB, CPU は Core i7-8700 CPU 3.2Ghz である。用いた学習フレームワークは Neural Network Libraries である。
- 注 4) CReLU: 活性化関数の一種であり, 参考文献 11) にて提案された関数である。自然画像 (2 値画像ではない) を学習する際の精度向上を意図し, 下記の注にある ReLU を基本として拡張された関数である。ReLU は負の値を 0 として恒等写像する関数であるが, CReLU は負の値に対しても特徴を消さないことを意図して拡張された関数である。CReLU の具体的な中身はコードで表現すれば下記によって表される。
CReLU : $f(x) = (\max(0, x), \max(0, -x))$
また下記のようにも表現できる
CReLU: $f(x) = (\text{ReLU}(x), \text{ReLU}(-x))$
つまり負の値も絶対値として大きな値は特徴として重視する関数であ

る。比較的浅い層で ReLU に代わり使用することが提案されているが, 本研究では複数パターンを試行し, 全ての活性化関数を CReLU とした。
注 5) ReLU: 活性化関数に用いられる関数の一つの種類である (参考文献 12))。この関数は恒等写像に似ているが, 入力値が 0 を超えていれば, その入力をそのまま出力し, 0 以下であれば 0 を出力する関数である。単純な関数であるが汎用性の観点から画像認識分野において多く用いられている。但し本研究は 2 値画像ではなくカラー画像であることから CReLU を採用した。なお ReLU 関数は次式によって表される。

$$y(x) = \begin{cases} x & (x > 0) \\ 0 & (x \leq 0) \end{cases} \quad \begin{array}{l} y(x): \text{活性化された値} \\ x: \text{入力値} \end{array}$$

- 注 6) MaxPooling: 畳み込み層の後に, 特徴量マップから指定範囲毎に範囲内の最大値を抽出することで情報を圧縮する処理である。
- 注 7) Dropout: 過学習の抑制する手法として参考文献 13) にて提案された手法である。学習する過程においてある層の中のノードのうちいくつかを指定する割合で無効にすることを, 学習の更新毎に繰り返す手法である。これにより学習時のネットワークの自由度が強制的に小さくなり, 汎用度が高くなる (検証用データに対しても同様の精度が期待できる) ことが報告されている。位置については全結合層だけでなく畳み込み層においても有効であることが示されており, 割合については 50% 程度が用いられることが多い。本研究でもこれを参考に学習モデルの構成を複数検討した。
- 注 8) SoftMax: 各画像に対して出力された得点を 0 以上 1 以下で成分の和が 1 であるように変換する処理である。0 から 1 の確率に変換することで得点が扱いやすくなるため Deep Learning の全結合層において頻繁に用いられる。具体的な算出式は下記である。

$$y_n = \frac{\exp(a_n)}{\sum_{i=1}^k \exp(a_i)} \quad \begin{array}{l} k: \text{出力層のユニット数} \\ \exp(a_n): \text{入力信号 } a_n \text{ の指数関数} \end{array}$$

出力層が全部で k 個あるとして n 番目の出力 y_n を求める式
分母は全ての入力信号の指数関数の和

- 注 9) Categorical Cross Entropy: クラス分類用の損失関数として用いられる関数である。

$$E_n = - \sum_{k=1}^K t_{nk} \log y_{nk} \quad E = \sum_{n=1}^N E_n$$

n: サンプル番号, K: クラス数,
 y_{nk} : n サンプル目のクラス k の出力データ (確率),
 t_{nk} : n サンプル目のクラス k の教師データ
one-hot 表現 (0 (正解) または 1 (正解以外))

- 注 10) Adam: 損失関数により算出された誤差からネットワークを更新する方法として参考文献 14) にて提案された方法である。主たる設定は既往研究を参考に複数パターンを試して決定した。
- 注 11) Batch Size: 深層学習の分野においては, 学習を効率的に実行する手法として Mini Batch 学習が頻繁に採用されている。Mini Batch 学習は, 学習データをいくつかの Mini Batch に分け Mini Batch ごとに学習を行い, 段階的に更新を進める方法である。本研究でも Mini Batch 学習を採用しており, Batch Size (Mini Batch Size) は Mini Batch のデータサンプル数のことである。Batch Size には「16, 32, 64, 128, 256, 512」などが用いられており, 本研究では実行に必要なメモリも勘案して複数の設定を試し「64」とした。なお Mini batch を作る際には Mini Batch に対して画像をクラス毎にランダムに振り分けている。
- 注 12) 学習用と検証用の画像は AI 作成のために確率的に 7:3 になるようにランダムに分割している。割合ではないため, 近似的には 7:3 となるが, 必ずしも一街路の画像数が 140 枚と 60 枚となる訳ではない。このため Table. 2 のある行の全ての列の値を足し合わせた数値は必ずしも 60 枚とはならない。なお AI 作成の学習回数の総数は 500 回であるが, 掲載している結果は, 最も誤差が小さい回の AI を用いた結果である。また本稿では街路名の分類は AI 作成を繰り返しても精度が頑健であったため一般性検証の掲載を割愛した。表 2 は掲載を割愛した複数の AI からランダムに取り出したある AI の結果である。
- 注 13) Recall, Precision, Accuracy, F-Measure: 正しいクラスと推定したクラスの混合行列から推定の精度を定量的に評価する際に用いられる指標である。本研究の意図である推定や予測の分野においては一般的に用いられる指標であるが, 建築・都市分野において頻繁には用いられないためクロス集計表を例に算出式を示す。

		Deduction						Sum
		Class_1	Class_2	..	Class_j	..	Class_K	
Correct	Class_1	n11	n12	..	n1j	..	n1k	n1.
	Class_2	n21	n22	..	n2j	..	n2k	n2.

	Class_i	ni1	ni2	..	nij	..	nik	ni.

	Class_k	nk1	nk2	..	nkj	..	nk.	nk.
Sum	n.1	n.2	..	n.j	..	n.k	N	

$$\text{Recall}(\text{Class}_i) = n_{ij} / n_{i.} \quad (i=j)$$

$$\text{Precision}(\text{Class}_j) = n_{ij} / n_{.j} \quad (i=j)$$

$$\text{Accuracy} = \sum_i^k n_{ii} / N$$

$$F\text{-Measure} = 2 * \text{Recall} * \text{Precision} / (\text{Recall} + \text{Precision})$$

注 14) K 係数：医療統計や心理統計に用いられている統計量である。医療統計を例にすれば 2 名の診断者の一致率を示す指標として統計的仮説検定と併せて用いられる。心理統計では 2 名の被験者の回答の一致率の解釈として用いられる。本研究はサンプルサイズが大きいが p 値も参考までに併記する。
算出式は前掲のクロス集計表を例に示す（参考文献 15）16）。

$$K = \frac{\sum_i^k p_{ii} - \sum_i^k p_i. p_{.i}}{1 - \sum_i^k p_i. p_{.i}} \quad p: \text{表の値の発現割合}$$

分子はいわば「見かけ上の一致率のうちで偶然によらない一致率」であり、分母は「全体の一致率のうちで偶然によらない一致率」である。この比である係数の値の解釈については参考文献 17) と参考文献 18) が用いられることが多い。本研究でもこれを引用した。
参考文献 17) における解釈は下記である。

- 0.0 ~ 0.2: (slight agreement)
 - 0.21 ~ 0.40 (fair agreement)
 - 0.41 ~ 0.60 (moderate agreement)
 - 0.61 ~ 0.80 (substantial agreement)
 - 0.81 ~ 1.0 (almost perfect or perfect agreement)
- 参考文献 18) における解釈は下記である。
- 0.67 未満 (discounted)
 - 0.67 ~ 0.80 (conclusions tentatively)
 - 0.80 以上 (definite conclusions)

注 15) Squared Error：損失関数の一つの種類である。数値に置き換えた正しいクラス (0 と 1) と画像に対して推定した得点の誤差を二乗する損失関数である。Deep Learning の損失関数として当初から用いられているベーシックな損失関数である。

$$E_n = \frac{1}{2} (y_n - t_n)^2 \quad E = \sum_{n=1}^k E_n$$

n：サンプル番号, y_n ：出力データ, t_n ：教師データ, k: サンプル数

注 16) 被験者負担を勘案して枚数を減らしたことにより、5 章の回答値は街路を満遍なく見た上での訪問意欲ではなく、ある地点の視覚的な景観評価を通じた訪問意欲の回答となっている。このような実験計画の背景には、街路景観を計画する際に、訪問したいと思つて貰えるような街路の特性を表現する象徴的な景観を計画する、という実際に景観デザインを思考する際のデザイン行為がある。また一般の人が訪問を検討する際に目にする機会が多いメディア等においても、このような街路の特性を表現する象徴的な景観画像がキービジュアルに多く用いられている。被験者の負担と上述の考えから原稿の被験者実験とした。なお 5 章は新たな AI を作成している訳ではないため学習・検証という区別はない。

注 17) Wilcoxon の符号付順位検定の効果量：ノンパラメトリック・対応ありの場合に用いられる 2 群の代表値の差に着目した統計的仮説検定の効果量である。なお値の解釈 (言及) には心理統計において頻繁に引用される Cohen (1988) を本研究でも引用した (参考文献 19)。

注 18) 順位相関係数：2 変数間の相関関係を表現するノンパラメトリックな指標である。Spearman と Kendall の 2 種が提唱されているが両者は近似の値を示すことが知られている。本研究では Spearman の順位相関係数を採用した。これは建築・都市分野において頻繁に用いられるパラメトリックな相関関係を示す Pearson の積率相関係数を援用する

Spearman の方が分かりやすいと考えたためである。なお値の解釈 (言及) には上記と同じく Cohen (1988) を引用した。

注 19) 無相関仮説の検定：帰無仮説を「2 変数間に相関関係は無い」とする統計的仮説検定である。

参考文献

- 1) Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks, NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, pp. 1097-1105, 2012. 12
- 2) Satoshi YAMADA, Takeshi FUJI, Shinsaku Munemoto: Development and Verification of The Impression Deduction Model for Green Space with Ratio of Omnidirectional Visibility Green Space-Using Self-Organizing Map with ratio of omnidirectional visibility green space and impression evaluation experiment-, Journal of Architecture and Planning (Transactions of AIJ), Vol. 81, No. 727, pp. 2083-2093, 2016. 9 (in Japanese)
山田悟史, 藤井健史, 宗本晋作: 方位緑視率を用いた緑地環境に対する印象評価推定モデル作成と検証 - 全方位緑視率と印象評価実験をもとにした自己組織化マップを用いて -, 日本建築学会計画系論文集, 第 81 巻, 第 727 号, pp. 2083-2093, 2016. 9
- 3) Shinsaku Munemoto, Satoshi YAMADA: A Study on Probabilistic Model of Evaluation of Green Environment Based on Ratio Of Visible Green Using Bayesian Network, Journal of Architecture and Planning (Transactions of AIJ), Vol. 80, No. 708, pp. 2411-2419, 2016. 11 (in Japanese)
宗本晋作, 山田悟史: ベイジアンネットワークを用いた緑視率に基づく緑環境評価の確率モデルに関する研究, 日本建築学会計画系論文集, 第 81 巻, 第 729 号, pp. 2411-2419, 2016. 11
- 4) Lun Liu, Elisabete A. Silva, Chunyang Wu, Hui Wang, : A machine learning-based method for the large-scale evaluation of the qualities of the urban environment-, Computers, Environment and Urban Systems, Volume 65, pp. 113-125, 2017. 06
- 5) Stephen Law, Yao Shen, Chanuki Seresinhe, : An application of convolutional neural network in street image classification -the case study of london-. ACM GeoAI'17 Proceedings of the 1st Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery, pp5-9, 2017. 11
- 6) Atsushi Takizawa and Airi 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, 2017. 10
- 7) Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen: Progressive Growing of GANs for Improved Quality, Stability, and Variation, Under review as a conference paper at ICLR, 2017. 12
- 8) Satoshi Yamada: Estimation of City Name and Sensibility Evaluation of Cityscape Images Using Deep Learning, AIJ Kinki Chapter Research Meeting, Vol. 58, pp. 257-260, 2018. 7 (in Japanese)
山田悟史: Deep Learning を用いた街並み画像の都市名と訪問意欲の推定, 日本建築学会 近畿支部研究報告集, 第 58 号, 計画系, pp. 257-260, 2018. 7
- 9) Hideaki Takahashi, Satoshi Yamada: Classification of Cityscape Images Using Deep Learning and Estimation of Sensibility Evaluation, Proceedings of the 40th Symposium on Computer Technology of Information, Systems and Applications (Report) (Transactions of AIJ), pp. 329-332, 2017. 12 (in Japanese)
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, 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. 12, 日本建築学会
- 10) Karen Simonyan, Andrew Zisserman: VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION, arXiv:1409.1556, 2014. 9
- 11) Wenling Shang, Kihyuk Sohn, Diogo Almeida, Honglak Lee: Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units, arXiv:1603.05201, 2016. 3
- 12) Xavier Glorot, Antoine Bordes and Yoshua Bengio: Deep sparse rectifier neural networks, Proceedings of the 14th International

- Conference on Artificial Intelligence and Statistics, PMLR, pp. 315-323, 2011. 4
- 13) Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, The Journal of Machine Learning Research, Volume 15 Issue 1, pp. 1929-1958, 2014. 6
 - 14) Diederik P. Kingma, Jinyang Ba.: Adam: A Method for Stochastic Optimization, arXiv:1412.6980, 2014. 12
 - 15) Jacob Cohen: A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, pp. 37-46, 1960
 - 16) Jacob Cohen: Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit, Psychological Bulletin, 70, 213-220, 1968
 - 17) Klaus Krippendorff: Content Analysis: An Introduction to Its Methodology (2nd Edition), Sage Publications, 2003
 - 18) Landis JR, Koch GG.: The measurement of observer agreement for categorical data. Biometrics, 33(1):159-174, 1977
 - 19) Jacob Cohen.: Statistical power analysis for the behavioral sciences (2nd Edition), Routledge, 1988
 - 20) Yasuki Saito: zerokaratukuruDeepLearning (Create Deep Learning from 0), oreilly, 2016. 9 (in japanese)
斎藤康毅:ゼロから作る Deep Learning —Python で学ぶディープラーニングの理論と実装, オライリージャパン, 2016. 9
 - 21) Kazuya Fujita, et al.: zissōdēpurāningu (Implementation Deep learning), OHM, 2016. 11 (in japanese)
藤田一弥, 高原歩, 株式会社フォワードネットワーク:実装ディープラーニング, オーム社, 2016. 11
 - 22) Hiroyuki Shinnou: Chainev2niyoruzissensinsōgakusyū (Practice of deep learning with Chainev 2), 2017. 9 (in japanese)
新納浩幸:Chainev2による実践深層学習, オーム社, 2017. 9
 - 23) Matia Okubo, Kensuke Okada: tutaerutamenosinritōkei, kōkar yō · sinraikukan · kenteiryoku (Psychological statistics for telling), Keisou Shobo, 2012. 1 (in Japanese)
大久保衛亜, 岡田謙介:伝えるための心理統計, 効果量・信頼区間・検定力, 勁草書房, 2012. 1

DEVELOPMENT AND VERIFICATION OF THE IMPRESSION DEDUCTION MODEL FOR CITY LANDSCAPE WITH DEEP LEARNING

- Street names city landscapes and desire/no desire or degree of desire to visit -

*Satoshi YAMADA**, *Kotaro ONO***

*Lecturer, Dept. of Architecture and Urban Design, Ritsumeikan Univ., Dr. Eng.

**Undergraduate, Dept. of Architecture and Urban Design, Ritsumeikan Univ.

This research is a basic study on utilizing artificial intelligence (AI) by applying deep learning to the fields of architecture and urban design.

In recent years, the use of budding technologies, such as deep learning, has increased in the field of architecture and urban design. While this technology has potential in various fields, this study focuses on learning and deduction of sensibility evaluation and impression of design. Needless to say, the relationship of design with sensibility and impression is important, and the design heightens sensibility and impressions. However, the causal relationship of quantitative representation (feature value) and feature value of design with impression is complex and is characteristically difficult to deduce. Such a characteristic is a property similar to fields where deep learning has been successfully used. It is, therefore, thought that AI using deep learning could be applicable. As mentioned, this research examines the budding properties of AI that deduce “street names and desire to visit” based on city landscapes. Specifically, the “desire/no-desire to visit (classification)” and “degree of desire to visit” are deduced, and as constituents of image consciousness, street names are also classified (21 classes).

The object of the study and the city landscapes were prepared from 21 streets selected from a large city and sightseeing information. The images for city landscapes were obtained from street view on Google Earth to ensure that these images were not of any one building, ground, or sky. A total of 2,100 images, 100 for each street, were considered.

Deduction AI with high precision was first successfully developed to deduce “classification of street names (21 classes)”. Its precision was approximately 86% for the F-value with a K-coefficient of 0.8508 (p-value = 1.6e-15). Next, for the classification, deduction AI with high conformity with desire/no-desire to visit criteria of test subjects was successfully prepared. Its precision had a K-coefficient of 0.8920 (p-value = 2.2e-15). Further, for deducing degree of desire, there was little difference in the degree of desire to visit between test subjects, and AI permitting deduction with high correlation was successfully developed. For its precision, the effect size of Wilcoxon’s signed rank test (test of paired nonparametric data) was 0.18, and Spearman’s rank correlation was 0.7564 (p-value = 0.0005742). Finally, to generalize the methodology of AI using deep learning, the 95% confidence interval that considered 100 kinds of AI developed using this method was confirmed to be small. Specifically, the effect size did not exceed 0.2 (a threshold value indicating small effect size) and did not fall below 0.6 (a threshold value indicating high correlation). Under the experimental conditions of this study, the AI developed using deep learning can be described as a method that presents generality in the degree of precision.

From the perspective of the experimental conditions of the study and usage, a successful impression deduction AI for city landscapes with good precision is developed. This provides the first step in systematically organizing and investigating the hitherto unstudied budding potential of deep learning in the fields of architecture and urban design.