

深層学習を用いた伝統的建造物群保存地区の街路画像の分類

Classification of Street Images of Preservation Districts for Groups of Traditional Buildings by Deep Learning

越智広樹¹・山田悟史²

Hiroki Ochi and Satoshi Yamada

¹立命館大学大学院 理工学研究科環境都市専攻 博士課程前期課程 (〒525-8577 滋賀県草津市野路東1-1-1)

Graduate Student, Graduate School of Science and Engineering, Ritsumeikan University

²立命館大学准教授 理工学部環境都市工学科 (〒525-8577 滋賀県草津市野路東1-1-1)

Associate Professor, Department of Civil and Environmental Engineering, College of Science and Engineering, Ritsumeikan University

In Japan, preservation of Japanese historical sites have been progressing. However, regarding a judgment as to whether the preservation plan is suitable or not, some people have pointed difficulties out. Therefore, the authors attempt quantitative approach to the problem by AI technology based on Deep Learning. In this study, to verify the possibility of application of the technology, the authors create street image classification AI on scene of preservation districts for groups of traditional buildings. Moreover, the authors visualize where AI looking in the image when AI classify by XAI technology. As a result, classification AI reaches high precision and the judgment is systematic by the class, so the result show AI has the possibility of application.

Keywords: street scene, landscape preservation, Deep Learning, XAI

1. はじめに

昭和50年に「文化財保護法」の改正により伝統的建造物群保存地区（以降、伝健地区と表記）制度が定められて以来、該当地域では歴史的まちなみと伝統文化や周辺環境が一体となった歴史的風致の保存活用が進められてきた¹⁾。一方で、規制の範囲外では歴史的価値が認められる建造物が空き家となって解体される事例も見られ²⁾、中間領域と称される周辺市街地において、商業化や生活様式の変化に合わせた建築物の変化等の新たな計画との調整が課題となっている。この課題については国としても強く認識されており、「地域における歴史的風致の維持及び向上に関する法律（愛称：歴史まちづくり法）」が制定され、歴史的風致維持向上計画によって重要伝統的建造物群保存地区（以降、重伝健地区と表記）等の文化財を中心とした一帯の包括的な保全が試みられている³⁾。しかし課題も報告されている。それは、保全上の基準が文章で示されており人によって解釈が変わり得ることである。修理・修景を例にすれば、視覚的要因のみならず、地区ごとの歴史や文化などの多様な条件も考慮する旨が文章表現で示されている。事実、文化庁は伝建地区制度の条例において定めるべき項目の1つとして「伝統的建造物群を構成している建築物等（以下「伝統的建造物」という。）の増築若しくは改築又は修繕、模様替え若しくは色彩の変更でその外観を変更することとなるものについては、それらの行為後の伝統的建造物の位置、規模、形態、意匠又は色彩が当該伝統的建造物群の特性を維持していると認められるものであること」（伝統的建造物群保存地区制度の実務の手引き、p.37, 2021）と記している。このような文章表現は包括的かつ柔軟な運用を可能にするという利点がある。一方で、定量的でない言語表現ゆえの曖昧さは、保全計画そのものや個々の新たな計画の適正さの判断を困難にする側面もあることが指摘されている⁴⁾。

上述の課題に対するアプローチとして、街路の組成や性質に対する定量的な研究がある。本研究もそれらに属するものであり、特にDeep Learning（深層学習）を用いたAIによる画像分類に着目するものである。この手法は画像の特徴を自ら学習するため、人間の代替的な分類、及び人間では困難な分類において高い精度が報告されている。またDeep Learningによる分類が注目を集めた当初に指摘された根拠の読解についても、AIが把握した特徴量を人間が理解可能な形で説明する技術であるXAI（説明可能なAI）技術が発展している⁹⁾。その代表例にはAIが根拠とした画像内の領域の可視化が挙げられる。これらのDeep Learningを基盤とするAI技術には、前述した課題に対するアプローチとして萌芽性がある。一例には、新たな計画の適正診断AIの作成が挙げられる。これは、計画が対象の歴史都市の特徴に適しているか否かの判断をAIが定量的に支援するという、人とAIの共創の可能性の一つである。このときAIは都市の画像群からその総合的な特徴を学習している。そのため、前述した「伝統的建造物群の特性の維持」が為されているか否かなどの項目に対して、その学習内容を元に定量的な観点で推定が可能となる。但し、これはAIの学習内容が正確な場合に限られる。仮にAIの学習内容が不正確な場合は推定も同様に不正確なものとなり、本来ならば許諾され得ない計画によって歴史都市の特徴が著しく損なわれる可能性すら存在する。つまり、こうした深層学習を用いた画像分類の応用技術は、前提となる分類において高い精度が確認されて初めて実現可能性が担保されるものである。そのため、歴史都市の保全に関して慎重を期すためにも精度の追求は必要不可欠である。

以上のように、Deep Learningを基盤とするAI技術は歴史都市の保全において発展的な萌芽性を持つが、発展に必要な適用可能性を検証した基礎研究は少ない。特に重要かつ原始的な基礎となる重伝建地区内と近傍、及び重伝建地区同士の分類可能性を検証し、その根拠となる特徴量を把握した研究は、筆者が知る限り見当たらない。そこで本研究は、前述した萌芽性の発展に向けた基礎研究として、重伝建地区を対象に地区内と近傍、及び重伝建地区同士の分類に対するDeep Learningを基盤とするAI技術の適用可能性をXAI技術と合わせて検証するものである。具体的には、分類精度と学習の際にAIが獲得する特徴量^{注1)}を把握する。本研究の対象に対してこのような結果を示した研究は筆者が知る限り見当たらず、本研究の新規性に該当する。

2. 概要

(1) 研究の流れ

本研究では、2種類の分類問題を通して街路画像の分類（以降、街路分類と表記）における深層学習の適用可能性を検証する。第一に、単一の重伝建地区を対象に、その内部と周辺地域の街路の分類（以降、内外分類と表記）を試みる。第二に、複数の重伝建地区を対象に、内外分類で用いた街路全ての分類（以降、全街路分類と表記）を試みる。内外分類は分類問題において最もシンプルな形式の分類であり、多クラス分類となる全街路の前提として高精度である必要があるため実施する。全街路分類は、広範な対象を取るより困難な分類問題であり、この分類問題に対して高精度の分類が可能であるかの検証が研究の主目的である。また内外分類と同等の精度を保つことができるか否かの把握も研究目的であるため、前提となる内外分類も原稿に示す。検証においては高精度のAIを目指した学習パラメータの調整を行う。この際には、単純な分類精度の追求だけでなく、それぞれの分類モデルにXAI技術を適用して分類に寄与した画像中の領域（以降、判断領域と表記）の可視化を併用する。これにより各モデルが系統だった判断を行っているか否か確認する。

(2) 対象街路

京都市内の祇園新橋重伝建地区・上賀茂重伝建地区・産寧坂重伝建地区・嵯峨鳥居本重伝建地区の4カ所の重伝建地区から全15カ所の街路を対象とした。図1に対象の重伝建地区及び周辺地域の景観計画区域図と対象街路を示す。以降、各街路は図1内の記号で呼称する。

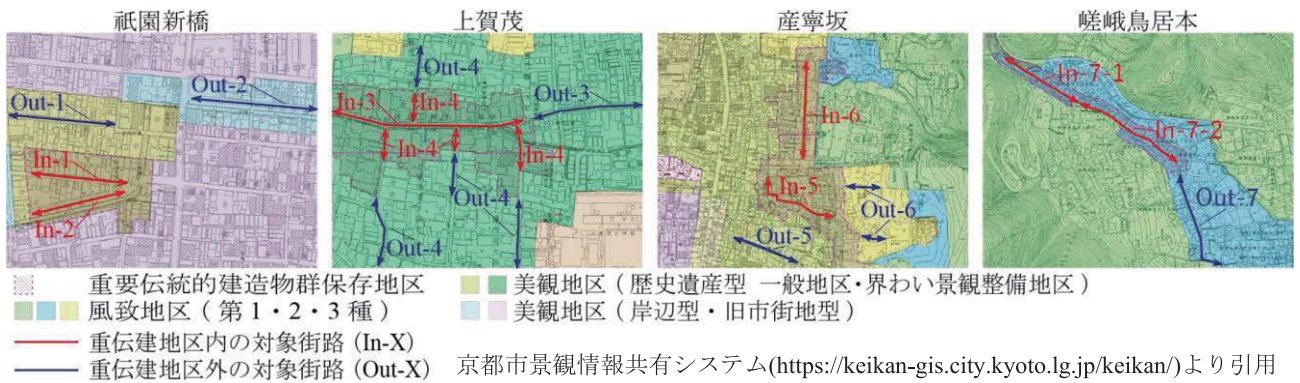


図1 対象重伝建地区の景観計画区域図と対象街路

(3) 各分類のクラス定義

クラスとは学習対象がそれぞれ属する集団（本研究では前述の各街路が該当）であり、対象をそのクラスに適切に分類することが深層学習による画像分類の目的である。内外分類では、前述した重伝建地区内外の15の街路を組み合わせ、2または3クラス分類を7種実施した。この組み合わせをパターン（A～G）と称す。その詳細を表1に示す。例えば、パターンAは画像がIn-1かOut-1のどちらのクラスに属するか分類するものである。なお、この組み合わせは内外以外の条件を一致させるために、美観地区や風致地区等の伝建地区制度以前に指定された法令に従って同種の景観計画に属するように設定した。全街路分類では、内外分類で利用した15種の街路全てを用いて15クラス分類を実施した。この際のクラスと街路の対応関係を表2に示す。

表1 内外分類の組み合わせ及び対象街路の詳細

| パターン | 対象街路 | 重伝建地区 | 景観計画区域 | クラス数 | |
|------|--------|-------|----------------------|------|---|
| A | In-1 | 祇園新橋 | 歴史遺産型美観地区(一般地区) | 2 | |
| | Out-1 | | | | |
| B | In-2 | 上賀茂 | 岸部型美観地区 | | |
| | Out-2 | | | | |
| C | In-3 | 上賀茂 | 歴史遺産型美観地区(界わい景観整備地区) | | |
| | Out-3 | | | | |
| D | In-4 | 上賀茂 | 歴史遺産型美観地区(界わい景観整備地区) | | |
| | Out-4 | | | | |
| E | In-5 | 産寧坂 | 歴史遺産型美観地区(一般地区) | | |
| | Out-5 | | | | |
| F | In-6 | 産寧坂 | 風致地区(第3種地域) | | |
| | Out-6 | | | | |
| G | In-7-1 | 嵯峨鳥居本 | 風致地区(第2種地域) | | 3 |
| | In-7-2 | | | | |
| | Out-7 | | | | |

表2 全街路分類のクラス-街路対応表

| クラス | 対象街路 |
|-----|--------|
| 0 | In-1 |
| 1 | Out-1 |
| 2 | In-2 |
| 3 | Out-2 |
| 4 | In-3 |
| 5 | Out-3 |
| 6 | In-4 |
| 7 | Out-4 |
| 8 | In-5 |
| 9 | Out-5 |
| 10 | In-6 |
| 11 | Out-6 |
| 12 | In-7-1 |
| 13 | In-7-2 |
| 14 | Out-7 |

(4) データセットの作成方法

データセットは、各街路を徒歩で往復しながらビデオカメラで進行方向を撮影した動画をフレームレート毎に静止画を切り出して生成した。このとき撮影時の条件として、画質を4K、フレームレートを30fps、地面からの高さを1.5mと定めた。また、時刻や周辺環境の様子でカメラの設定が変化しないようにマニュアル操作に固定した。加えて、可能な限り動画に人が映り込まないように早朝から昼にかけて撮影した。ただし、この状態では街路の長さによって切り出し後の画像の枚数が異なるため、最短街路のIn-1を基準としてフレームレートを変更し、1秒ごとに切り出すフレーム数を街路ごとに調整した。これにより大まかな枚数を揃えた後に、街路の両端から超過分の画像を取り除いて各街路750枚の画像を用意した。この操作内容は表3に示す。また、切り出した画像の例は図2に示す。この画像群を訓練用データ：検証用データ：テスト用データ^{注2)}に7：2：1の比率で無作為に分割した。そして、後述の学習モデルに合わせて画像にzero paddingとresizeを施して、画像のサイズを224×224pixelに変更した。

表3 枚数調整時の内訳

| 街路 | 撮影時間(秒) | 調整前fps | 調整前の枚数 | 調整後fps | 調整後の枚数 | 最終枚数 |
|--------|---------|--------|--------|--------|--------|-------|
| In-1 | 161 | 30 | 4830 | 5.0 | 805 | 750 |
| Out-1 | 166 | | 4980 | 5.0 | 830 | |
| In-2 | 314 | | 9420 | 3.3 | 1036 | |
| Out-2 | 196 | | 5880 | 5.0 | 980 | |
| In-3 | 377 | | 11310 | 2.5 | 942 | |
| Out-3 | 380 | | 11400 | 2.5 | 950 | |
| In-4 | 446 | | 13380 | 2.5 | 1115 | |
| Out-4 | 481 | | 14430 | 2.0 | 962 | |
| In-5 | 160 | | 4800 | 5.0 | 800 | |
| Out-5 | 220 | | 6600 | 5.0 | 1100 | |
| In-6 | 373 | | 11190 | 2.5 | 932 | |
| Out-6 | 396 | | 11880 | 2.5 | 990 | |
| In-7-1 | 304 | | 9120 | 3.3 | 1003 | |
| In-7-2 | 297 | | 8910 | 3.3 | 980 | |
| Out-7 | 434 | 13020 | 2.0 | 868 | | |
| 全体 | | | 141150 | | 14293 | 11250 |



図2 対象街路（一例）

(5) 学習モデルの構造と学習時の設定

本研究では、学習モデルにKarenらのVGG16⁶⁾を使用した。この際、実施する分類に合わせて最終層の出力数を変更した。図3にその構造を示す。また、学習時のハイパーパラメータ^{注3)}を表4に示す。batch sizeは1step内で並列処理するデータ数を指しており、本研究では96枚の画像を同時に計算させた。損失関数は、入力データが本来属するクラス（以降、正解クラスと表記）と、入力データに対してAIが属すると予想したクラス（以降、予想クラスと表記）の不一致度合い（以降、lossと表記）を示す計算をする関数である。最適化関数はlossが最小になるように学習時の重みを更新する関数である。schedulerは重みの更新具合を制御する学習率の調整を行う関数であり、本研究ではpytorchに用意されているReduceLROnPlateau⁷⁾を使用した。これは一定epoch^{注4)}経過時にモデルの精度が向上していなかった場合に学習率を減衰させるよう定義されており、本実験では10epoch経過時にlossが低下しなければ学習率を0.1倍するよう設定した。

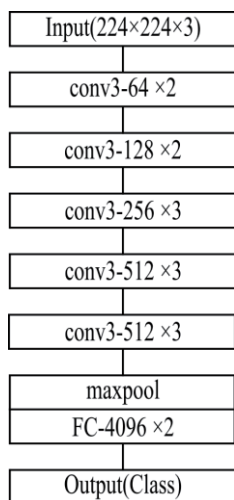


図3 VGG16の構造

表4 学習時のハイパーパラメータ

| | |
|------------|-------------------|
| Batch size | 96 |
| 損失関数 | 交差エントロピー誤差 |
| 最適化関数 | SGD |
| Scheduler | ReduceLROnPlateau |
| 初期学習率 | 0.0001 |
| epoch数 | 100 |

(6) 判断領域の可視化

判断領域の可視化にはMarcoらのLIME⁸⁾を使用した。画像分類において、LIMEは説明対象画像の分割された領域ごとに、分類に対する寄与度を解釈可能な形で提示するXAI技術である。同様のXAI技術は他に複数存在するが、領域単位での寄与度を提示する点で他の技術と比較して着目対象の識別が容易になることからLIMEを採用した。この適用例を図4に示す。図4左のように、画像をエッジやコントラストに沿って分割することで、画像内で寄与度が大きい部分を視覚的に確認することができる。図4中央は寄与度上位5つの領域を一括して示しており、その内訳を下の5つの画像が左から順に示している。

なお本研究では図5に示すように、全データを撮影方向に並べたときに重なる画素としての位置にある値を合算し、データ数で割って平均化するすることで、街路全体の総合的な寄与度の分布を示せると考えた。こうして各街路ごとに生成した判断領域が特定のエリアに集中する性質（以降、系統性と表記）を持つか否かを確認した。なお確認の際には図4右のような抽象図への照らし合わせも試行した。

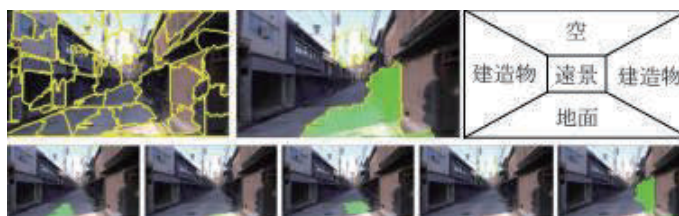


図4 本研究（内外分類）におけるLIMEの適用例

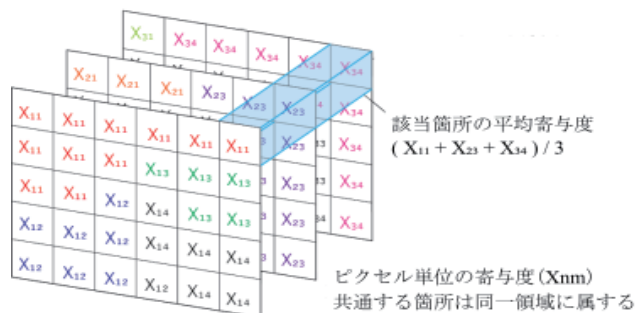


図5 平均寄与度の算出方法

3. 結果

(1) 内外分類

a) 分類精度

表5に各パターンにおける分類精度を示す。いずれのパターンにおいても最高の正答率^{注5)}（以降、best accuracyと表記）が99%以上を記録している。その一方で、CやDなど一部のパターンでは達成時epochが小さいことから、学習において非常に早い段階で安定して高い精度を記録し続けていたことが読み取れる。この点から学習モデルが正常に動作していない可能性が考えられた。そのため入力データや学習過程、精度の算出手法に関して検証を行ったところ、いずれにおいても異常は確認されなかった。これらのことから、表3の結果は内外分類の難易度に起因するものであり、学習自体は適切に行われたと判断した。

表5 内外分類の各パターンにおける分類精度

| パターン | Best Accuracy | 達成時epoch | Loss | 達成時epoch |
|------|---------------|----------|---------|----------|
| A | 1.000 | 41 | 0.00290 | 93 |
| B | 1.000 | 42 | 0.00468 | 100 |
| C | 1.000 | 13 | 0.00103 | 98 |
| D | 0.970 | 95 | 0.08360 | 100 |
| E | 0.997 | 25 | 0.00462 | 100 |
| F | 1.000 | 63 | 0.00325 | 97 |
| G | 0.993 | 65 | 0.04170 | 99 |

b) 判断領域

内外分類における各街路の判断領域を図6に示す。多くのヒートマップ^{注6)}で、系統だった判断を示唆する画像内の特定の部分への高い寄与度のまとまりが確認できる。その場所については、空や地面などに集中している傾向が視覚的には読み取れる。

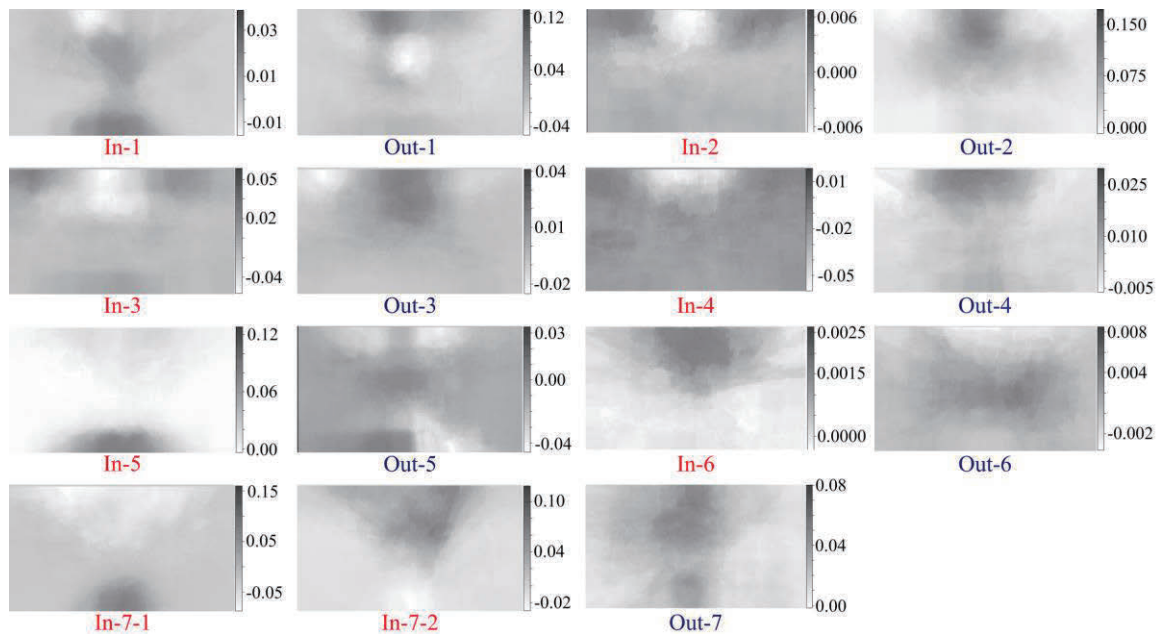


図6 内外分類における各街路の判断領域

(2) 全街路分類

a) 分類精度

全街路分類時の学習の推移を図7に示す。この時のbest accuracyは77epoch目の0.998, 最小のlossは86epoch目の0.0103である。

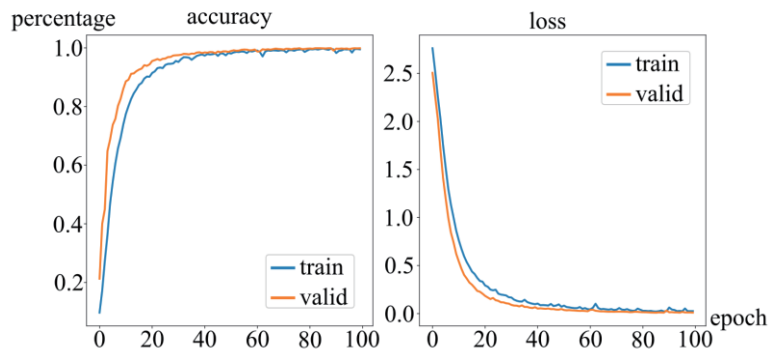


図7 全街路分類における学習の推移

b) 判断領域

全街路分類における各街路の判断領域を図8に示す。内外分類と同様に、多くのヒートマップで寄与度の高い箇所が画像内の特定の部分に集中しており、系統だった判断を示唆する結果が読み取れる。その場所についても着目すると、内外分類から変化して遠景に集中している場合が多いことが視覚的に読み取れる。

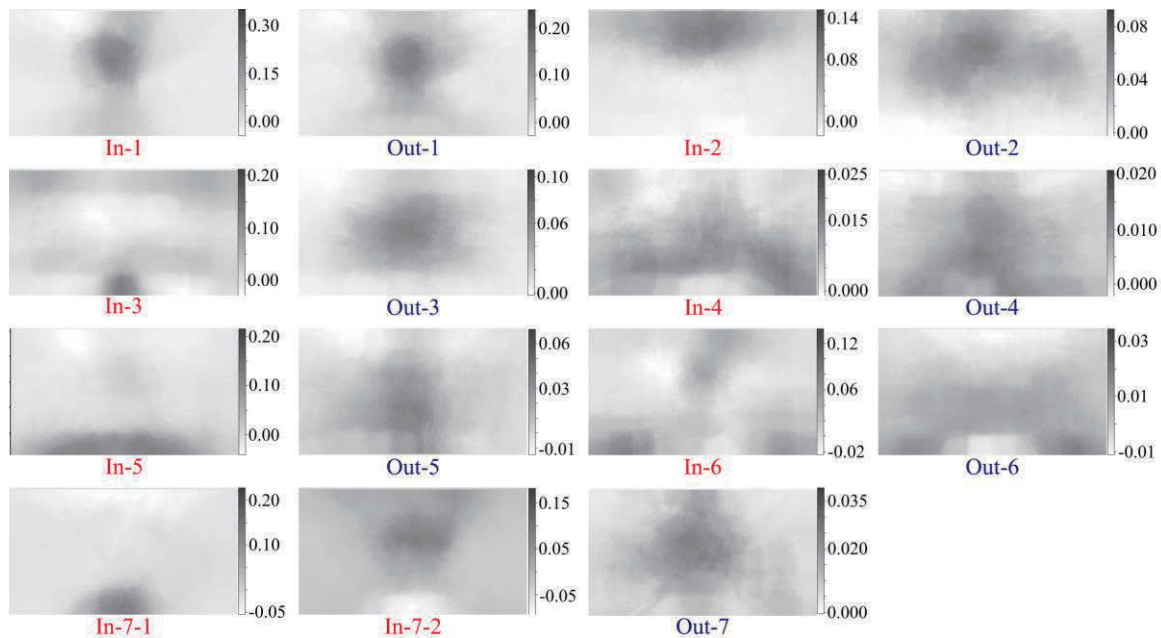


図8 全街路分類における判断領域

4. 考察・まとめ

以上のように本研究ではDeep Learningを基盤とするAI技術の街路画像に対する適用可能性として、重伝建地区を対象に地区内と近傍、及び重伝建地区同士の分類に対する分類可能性を検証した。その結果、内外分類（2・3クラス分類）の平均精度0.994、全街路分類（15クラス分類）の精度0.998という、AI技術を用いた画像分類に関する既往研究において高い分類精度と言及可能な精度をもつAIの開発に成功した。加えて判断領域において、多くの街路で画像内の特定の部分に寄与度の高い箇所が集中する結果となった。このことから本研究が開発したAIは各街路の特徴を把握して系統だった判断を下していると考えられる。また、その箇所の多くは建造物ではなく空や地面、遠景である。これは、街路の特性が従来の歴史都市の保全において重要視されている建造物の意匠に限らず発現することを示唆している。これ踏まえると、空や遠景からは建造物のボリュームが、地面からはそのテクスチャが街路の特性として反映されていると考えられる。

以上のような本研究の基礎的かつ原始的な成果は、発展にむけた適用可能性の示唆として有用であると考えられる。なぜなら1章に記した萌芽性の発展は高精度な分類可能性が前提となるからである。例えば、本研究で採用したDeep Learningを基盤とするAI技術は、対象画像（本研究における街路の画像）がクラスに分類される確率的な高さを把握することができる。この技術的な背景と本研究の成果をふまえれば、周辺地域の街路の変化という計画に対する適切さを重伝建地区との視覚的な類似性として判断する場面において、誘導結果をシミュレートした画像の類似性を定量的に示すという発展性があり得る。具体的には、保存地区内の建造物の改築などの現状変更に対して、それが対象保存地区の特性を維持しているか否か市町村の職員が判断する際に参照することが挙げられる。但しこのような発展においては、本研究が開発に成功した精度が、さらに多様な街路においても保たれる必要がある。そのため学習データセットの拡充が直近の課題である。また、保全計画は視覚的な要因のみで立案される訳ではないことから、画素情報以外もふまえたマルチモーダルな学習^{注7)}も課題である。更には、精度を測る上で分類の判定基準の許容範囲もXAI技術による可視化などを踏まえて取り組む必要がある。これらの精度にまつわる課題に加えて、重伝建地区を対象に、現行の指定条件に基づいた人による判定と深層学習による判定の比較などを通して実務における有効性を提示することも今後の課題として挙げられる。

注

- 注1) 深層学習において学習対象の特徴を定量的に表現し、判断の基準に用いられる次元のこと。
- 注2) 訓練用データはAIの学習に用いるデータ、検証用データは学習内容の確認に用いるデータ、テスト用データはAIの汎用性の確認に用いるデータのこと。
- 注3) 学習の進み具合など、深層学習において学習開始時の設定する条件のこと。
- 注4) 学習の繰り返し回数のこと。本研究では訓練・検証用の全データでの学習を終えた際に1epochとカウントする。
- 注5) 全て入力データに対して、AIがそれらのデータが属するクラスを正しく予想できた割合のこと。
- 注6) 2次元行列の値（本研究では寄与度を指す）の数値上の関係性を色の濃淡などで表現して可視化したもの。
- 注7) 複数種のデータを用いる深層学習のこと。画像に加え、その撮影地点も情報として入力するものなどが挙げられる。

参考文献

- 1) 文化庁文化財第二課 伝統的建造物群部門：伝統的建造物群保存地区制度の実務の手引き [https://www.bunka.go.jp/seisaku/bunkazai/shokai/hozonchiku/pdf/93312701_03.pdf] , pp. 1-2, 2021.
- 2) 特定非営利活動法人 全国街並み保存連盟, 北島力：重伝建地区周辺の伝統的建物の取り壊しが後を絶たない、制度を駆使して保存の手を打とう [https://onl.la/KsryLSj] , 2020.
- 3) 日本：地域における歴史的風致の維持及び向上に関する法律（平成二十年法律第四十号）, 2021.
- 4) 文化庁文化財第二課 伝統的建造物群部門：前掲(1) , pp. 86-89, 2021.
- 5) 大坪直樹・中江俊博・深沢裕太・豊岡祥・坂元哲平・佐藤誠・五十嵐健太・市原大暉・堀内新吾：XAI（説明可能なAI）そのとき人工知能はどう考えたのか？, 株式会社リックテレコム, 2021.
- 6) Karen, S., Andrew, Z.: Very Deep Convolutional Networks for Large-Scale Image Recognition, International Conference on Learning Representations, 2015.
- 7) pytorch : REDUCELRONPLATEAU [https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLRonPlateau.html#torch.optim.lr_scheduler.ReduceLRonPlateau], 2019.
- 8) Marco, T. R., Sameer, S. and Carlos, G.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier, KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1135-1144, 2016.
- 9) pytorch : SGD [https://pytorch.org/docs/stable/generated/torch.optim.SGD.html#torch.optim.SGD], 2019.
- 10) Vinod, N., Geoffrey, E, H.: Rectified linear units improve restricted boltzmann machines, ICML'10: Proceedings of the 27th International Conference on International Conference on Machine Learning, pp.807-814, 2010.
- 11) Nitish, S., Geoffrey, E, H., Alex, K., Ilya, S., Ruslan, S.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, Journal of Machine Learning Research, Vol.15, pp.1929-1958, 2014.
- 12) Ilya, S., James, M., George, D., Geoffrey, E, H.: On the importance of initialization and momentum in deep learning, Proceedings of the 30th International Conference on Machine Learning, PMLR, Vol.28, No.3, pp.1139-1147, 2013.