

敵対的生成ネットワークを用いた沖縄県伊是名集落の 街路景観の特徴の顕在化

Extracting the Characteristics of a Cultural Landscape in Okinawa by using a Generative Adversarial Network

○北川 拓^{*1}, 藤井 晴行^{*2}, カーロン ユヴァル^{*3}
Taku Kitagawa^{*1}, Haruyuki Fujii^{*2} and Yuval Kahlon^{*3}

*1 東京工業大学 環境社会理工学院 建築系 修士課程

Master's Course, School of Environment and Society, Tokyo Institute of Technology

*2 東京工業大学 環境社会理工学院 教授・博士 (工学)

Prof., School of Environment and Society, Tokyo Institute of Technology, D.Eng.

*3 東京工業大学 環境社会理工学院 助教・博士 (工学)

Assist. Prof., School of Environment and Society, Tokyo Institute of Technology, D.Eng.

Summary: Extracting and archiving the key characteristics of traditional landscapes is extremely important for various reasons, such as preservation and restoration. However, manual extraction of such characteristics has demerits, as it is both timeconsuming and subjective, to some extent. Automatic extraction of these, on the other hand, can potentially be both time efficient and consistent. Further, it may identify features overlooked by human experts. Motivated by this opportunity, we aim to enable such automatic extraction of characteristic features of cultural landscapes using artificial intelligence. In this study, we generated new cultural landscape images in an Island in Okinawa using a deep learning model (GAN) and then attempted to identify characteristic features of these using auto-generated activation maps (Grad-CAM). The activation maps reflect the regions that were used by the learning model in classifying an image as Izena-like, and are therefore informative regarding characteristic features, especially in generated images which closely resemble the original landscape. As a main result, we confirmed that the features of the streetscape can be extracted, and that the landscape features of streets that may have existed in the past can be made apparent by analyzing the generated images.

キーワード: 敵対的生成ネットワーク; 街路景観; 沖縄; 伊是名; 特徴; 顕在化;

Keywords: Generative Adversarial Network; Cultural Landscape; Characteristics; Okinawa; Izena;

1. はじめに

1.1 目的

沖縄県伊是名村の街路景観を学習させた深層学習モデルを用いて、集落の街路景観の特徴を顕在化させることを目的とする。

沖縄県伊是名村のように伝統民家のある集落に関する研究は、その集落の文化や成り立ちを明らかにするという点で重要である。これらの研究は建築物の仕組みやそこで展開されていた生活を明らかにすることに寄与してきた。ま

た民家の集合を、集落の景観として捉えることで集落の違った側面も見えてくる。例えば伊是名の街路景観の特徴の一つである石垣は集落の人々が協力して作り上げることから、石垣には集落における価値共創のエコシステム(平田ら, 2019)が現れる。街路景観という視点を持つことで集落に残る文化的な特徴がさらに明らかになると考える。

一方でこのような伝統的な外観を残す集落では、集落の高齢化や人口流出による景観の継承者の減少問題もある。伊是名村では後辺¹地域を景観計画重点区域に指定するな

¹ 沖縄本島北方に位置する伊是名島の集落。集落の中央に位置する公民館の南道路を境に北を後辺(くしひん)、南を前辺(めいひん)という。

ど景観の維持に努めているものの、空き家の増加が止まらないのも事実である。伊是名の伝統的な街路景観の再生を可能にするために、既存の街路景観の特徴と合わせて、既に失われてしまった可能性のある特徴を顕在化させることは重要である。

1.2 方法

本研究では伊是名の街路景観を顕在化する方法として、対象の特徴を学習した深層学習モデル（生成器）を用いて偽物の街路景観画像を生成する。そして同様に学習させた深層学習モデル（識別器）が生成画像と本物の伊是名の画像を対象に「伊是名らしさ」の有無を判定し、その根拠となる領域をヒートマップを用いて可視化することで視覚的な説明を得る。最後に生成画像と実際の画像のヒートマップを比較分析することで街路景観の特徴の顕在化を試みる。

2. 先行研究

2.1 街路景観の特徴に着目した既往研究

街路景観の特徴や、人に与える心理的な影響については様々な側面に着目して研究が行われてきた。ここでは街路景観の特徴の抽出や再現に関する研究に注目する。例えば近藤は、「良好な景観」を定量化する方法の提案を通して、その特徴の抽出をおこなった(近藤, 2019)。同じ街路に対してもみる人によってその印象は異なるものの、各々が持つ「美」の感覚には一定の共通性があると仮定し、街路から得られる色彩を指標化し、「良好な景観」を検証した。

街路景観の構成要素毎に人が街路に対して抱く印象への影響を調べることで街路の特徴を調べた研究もある。牧田らは景観を構成する要素について、イメージ評価実験をもとに分析と類型化をおこなった(牧田, 1992)。その結果から、看板量・看板の統一性や、建物壁面構成部材や樹木がイメージ評価へ大きく影響を与えることを明らかにした。一方で、街路景観構成要素の統一性・秩序・混乱の起こしやすさに着目した研究もある。松本らは街路景観を乱雑性・整然性という観点から研究を行い、建物・樹木・歩道の視界における面積率や個数などが、乱雑↔整然に影響する因子であることを指摘した(松本, 1991)。

これらの研究では色彩や構成要素といった特定の側面から街路景観を評価・分析することで、人のイメージに対する影響を明らかにしている。複数の観点を組み合わせることで、単一の指標では説明力が不足する場合でも、複合的な指標による判断で街路景観を評価することも可能である。一方で指標を重ね合わせることによる影響の推定は、景観が複雑になるほどに難しくなる。そこで本研究では、街路景観における視覚的な特徴を、人工知能を用いて顕在化し、その要因を分析していくアプローチを試みる。

2.2 深層学習の街路景観へ応用

近年建築の研究分野でも活用されている技術に深層学習がある。深層学習とは、対象の特徴を階層化して学習する機械学習手法の一つであり、人間の脳神経回路を模したニューラルネットワーク（NN）と呼ばれる枠組みが使われる。深層学習を含む人工知能の使われ方には、任意の属性に着目して分類する識別系と、任意の属性を含む偽物のデータを作成する生成系の二つがある。

建築分野における識別系の活用事例として、印象評価実験の結果を学習したモデルを用いて複数の街並みの画像からそれぞれの訪問意欲を推定した研究がある(山田, 2019)。画像を扱うタスクでは一般に、画像の特徴を取り出して低次元の空間に情報を畳み込んでいく、畳み込みニューラルネットワーク(CNN)が用いられる。CNNを用いることで、着目する特徴量を分析者が用意する必要がなくなり、複数の景観要素同士に非線形な関係が成り立つ場合でも高い精度で分類を行うことができる。

CNN含むNN一般の課題として、出力結果に対する説明力の低さがある。そこでZhouらは、ある入力とその出力結果に対して局所的な説明を与える手法を開発した(Zhou, 2016)。この研究では、CNNにおいて特徴マップの位置情報が保持されることを利用して、識別器の隠れ層から出力結果に影響を与える画像内の領域を可視化することによって出力結果に説得力を持たせた。この研究を受けて、本研究では画像内の任意の領域に対して、NNは重み付けを行うことでタスクの遂行にとって重要な部分を選び出し、出力結果に反映させているという仮定を置く。

一方で建築分野における生成系の活用事例として、生成された画像を用いた景観シミュレーションがある(山中, 2022)。この研究では、街路空間の景観変化による印象評価値への影響を推定しており、変化の具現化手法として生成器が用いられる。この研究から再現が容易ではない景観などに関する研究やシミュレーションがより深くなされ得ることが示唆される。生成系の特徴は、例えば、ある街路空間要素の特徴や関係性に従いながら、過去に存在した可能性のある景観画像の生成や、非線形な関係にある街路景観要素の組み合わせを表現できるところにある。街路景観シミュレーションの研究を行う上では、生成された画像の中で景観の特徴が顕在化されていることが望ましい。本研究ではこうした活用方法を前提とすることから、生成された景観画像と実際の景観画像の両方を用いてその視覚的な特徴の顕在化を試みる。

生成系の深層学習にはGAN(敵対的生成ネットワーク)(Goodfellow, 2014)と呼ばれる訓練手法が安定して高い精度で画像の生成を可能にしている(図1)。GANは、生成器と識別器が互いを凌ぐように学習を進めることで、生成器の性能を向上させていく手法である。GANをベースにしたStyle GAN2(Karras, 2020)は、画像を生成する過程において大域的な特徴(木々の大きさ、石垣の位置など)と局所的な特徴(葉っぱの形状、石垣の個々の石の色など)を

分けて画像に適応させることを可能にした。特にそれぞれの特徴を操作可能なパラメータとして扱うことができる点が特徴的で、景観シミュレーションにおいては景観の特定の部分を操作するという点において優れている。

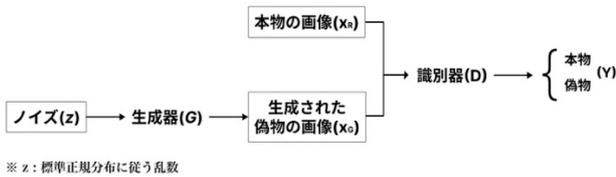


図1. GAN 概要図

3. 実験

3.1 概要

本研究では、伊是名集落の街路景観画像を用いて生成系の深層学習モデルを訓練させる。訓練された深層学習モデルの内部パラメータを取り出し、街路景観画像を判断する際に判別根拠とした領域情報に変換する。

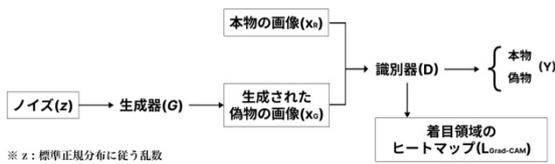


図2. 実験構成図

3.2 データセットの準備と前処理

本研究に用いるデータセットの画像は2022年6月19日～2022年6月25日の間に沖縄県島尻郡伊是名村で筆者が撮影した149枚の街路景観画像である。撮影はCANON社のEOS Kiss X4を用いて行った。街路景観の撮影は、道路の中央に立位の姿勢で、カメラを横向き水平にして、道路の中央とフレームの中央が重なるように行った。撮影時刻は午前9時～午後3時であった。

データセットの前処理として、画像が1024×1024pxの正方形になるように切り抜き・変形を行う。これは今回用いた既存のStyle GAN2モデルの入力形式の制限によるものである。

また、一般的な機械学習タスクと比較してデータセットサイズが小さいのは、学習の対象となる集落に関する資料が少ないケースにおける有効性を検証するためである。

3.3 GANの訓練

本研究では、事前学習済みのモデルを用いた転移学習を行う。転移学習とは、事前に類似する属性データで訓練されたモデルに対して、追加で対象データを用いて訓練することで、最終的な性能の向上や訓練時間を短縮させる手法である(Agrawal, 2014)。事前学習はCityscapes Dataset(Cordtsら, 2016)から3476枚の画像を用いて行なった。Cityscapes Datasetは、ドイツ国内の複数都市の街路を車載カメラで撮影した画像を用いて作成された大規模データセットである。Cityscapes Datasetを事前学習のデータセットに用いたのは、構図の類似性からである。

Cityscapes Datasetの画像は、中央に道路が配され、そこに植栽、建物、信号機、歩行者、サイクリストなどが映り込んでいる構成で撮影されている。本研究に使うデータセットも同様に、中央の道路の脇に伊是名村を特徴付けるフクギ、石垣、民家の外壁、畑そして電柱などが映り込んでいる。このような画像の構図的な類似性から、転移学習によるGANの早期学習と精度向上が期待できると考えた。

3.4 識別領域の可視化モデル

GANの識別器が画像の分類を行う際に根拠とした領域の可視化には、Selvarajuらの提案するGrad-CAM (Gradient weighted Class Active Mapping)と呼ばれる手法を用いる(Selvaraju, 2017)。Grad-CAMを用いることで訓練された識別器が画像の分類を行う際に判断根拠とする領域を抽出することができる。Grad-CAMでは学習済みの畳み込みニューラルネットワーク識別器に画像を順伝播させた後に、出力層に最も近い畳み込み層(最終畳み込み層)まで逆伝播をしてモデルの更新を止め、最終畳み込み層における特徴マップの勾配 A_{ij}^c から判断根拠を導く。勾配の大きさは、識別器に対する更新の大きさであり、同時に識別器の出力結果に対する影響の強さでもある。よって最終畳み込み層における勾配からヒートマップを作成し、入力画像と重ね合わせることで識別器が判別を行う際に重視している領域を視覚的に得ることができる。

実装としてまず、学習済みのモデルの出力結果 y^c に対する特徴マップの勾配を取得し、それにGlobal Average Poolingを適応することで各特徴マップの重要度 α_k^c を得る。 α_k^c は次式

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\delta y^c}{\delta A_{ij}^k}$$

で示される²。さらに、ここで得た各特徴マップの重要度 α_k^c を特徴マップ A^k に重み付けし、 $ReLU(x) = \max(0, x)$ を計算することで $L_{(Grad-CAM)}^c$ を得ることができる。

$L_{(Grad-CAM)}^c$ は次式、

² Selvarajuら[5]より引用

$$L_{(Grad-CAM)}^c = ReLU \left(\sum_k \alpha_c^k A^k \right)$$

で示される³。この $L_{(Grad-CAM)}^c$ を入力画像にマッピングすることで、識別器が判別根拠としている領域を可視化する。ここでは生成された画像と実際の伊是名の画像それぞれ 100 枚に対して Grad-CAM を適応させる。例として Grad-CAM の出力結果を図 3 に示す。

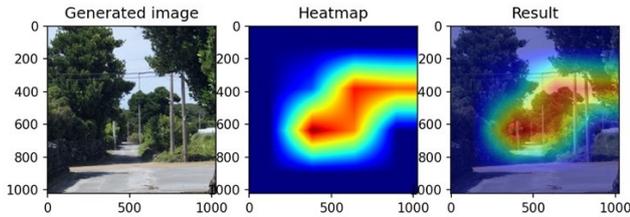


図3. 識別器の判断根拠の可視化の例。左:伊是名の街路景観画像, 中央:勾配のマッピングから生成したヒートマップ, 右:重ね合わせの画像

3.5 手法の評価

本実験で訓練された GAN の評価手法として FID(Frechet Inception Distance)を用いる(Heusel, 2017)。FID とは、生成された画像が既に別で訓練された識別器による識別のしやすさと、生成画像の分布の大きさを指標化した Inception Score を、実画像の分布も考慮して改良された指標である。Karras らは Style GAN2 の性能評価指標としてこの FID を用いており、当該研究で行われた Style GAN2 の FID 値は 10~30 の帯域で収束する傾向にあった。このことから、FID10~30 までは効率的に学習が進行すると考え、本実験に用いる Style GAN も FID 値 30 を下限に訓練を行う。ただし、それ以前に収束する場合は、収束が確認された時点で訓練を完了する。

Grad-CAM を用いた判断根拠の可視化実験においては、出力されたヒートマップにおいて最も着目された画素(勾配値が最も高い)が含まれる景観要素を目視によって取り出し集計する。勾配値が最も高い地点が複数ある場合は全て集計する。ただし、像が崩れているなど目視による街路景観要素を特定できない場合は「特定不可」として集計する。

4. 結果

伊是名集落の街路景観画像を用いた深層学習の訓練時の FID の推移を図 4 に示す。最終的な FID 値は 94.445 であった。また、事前学習済みの深層学習モデルによって生成された街路景観画像と、訓練された GAN が出力する街路景観画像の例を図 5 と図 6 に示す。伊是名の街路景観画像を学習させた深層学習モデルからは、伊是名特有の街路景観要素であるフクギ、石垣、雁行する道路、屋敷囲いが確認

できたほか電柱、ブロック塀、平屋の建物など一般的な街路景観要素の出力が確認できた。フクギについては、道路脇に垂直に生えているパターンや、道路の両側から上から覆いかぶさるように生えるパターン、そして片側から上に覆いかぶさるように生えるパターンなど複数のパターンが生成された。Style GAN2 の生成する画像の空間的な整合性に関しては、約 10 枚に一枚程度の割合で像が歪み、識別不可能な像の生成など、空間的に整合性の保てていない画像が生成された。

次に Grad-CAM による識別器の可視化の一部を図 7 ~ 12 に示す。キャプション後尾についている括弧内の文字は識別器の出力結果ラベルである。生成画像 100 枚を対象に、識別器によって識別の根拠とされた街路景観要素を表 1 に示す。全体的な傾向として、画像中央の街路景観要素が最も着目されることが多かった。画像中央に寄るほど、遠近法によりより多くの街路景観要素が映り込むため、これは妥当な結果となった。

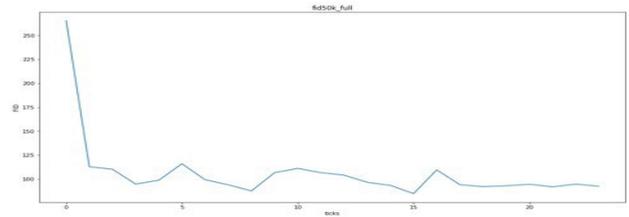


図4. Style GAN2 の FID 値の推移



図5. 伊是名の街路景観画像で訓練した StyleGAN2 による生成画像

表 1. 識別器によって判断根拠とされた街路景観要素

着目された街路景観要素	生成画像	実在画像
フクギ	33	27
空と他の物体との境界	24(2)	13
電柱	18	5
石垣	4	19

³ Selvaraju ら[5]より引用

道路	3	17
道路脇の植生	2	13
民家	0	7
ブロック塀	0	1
特定不可	16	9

※ 括弧内の数字は一枚の画像で最も勾配が高かった景観要素が複数見つけた回数。

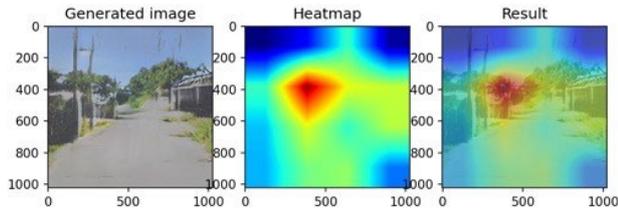


図6. 生成画像に対する Grad-CAM の出力例1(正)

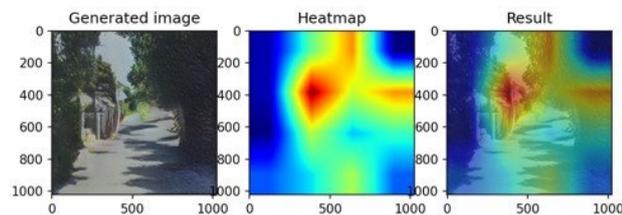


図7. 生成画像に対する Grad-CAM の出力例2(正)

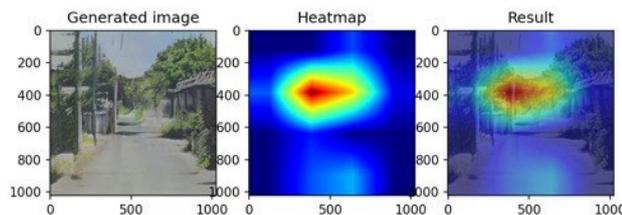


図8. 生成画像に対する Grad-CAM の出力例3(正)

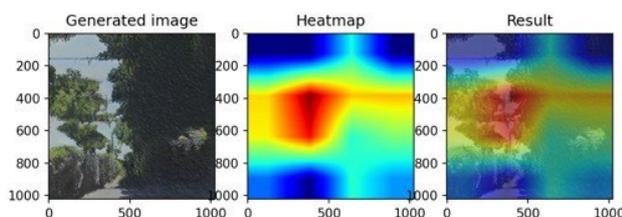


図9. 生成画像に対する Grad-CAM の出力例4(偽)

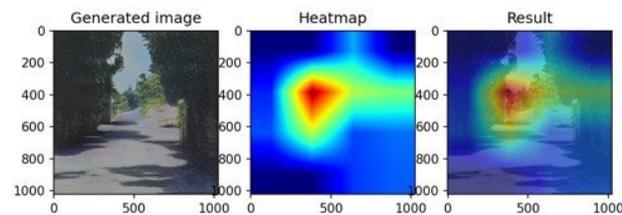


図10. 生成画像に対する Grad-CAM の出力例5(偽)

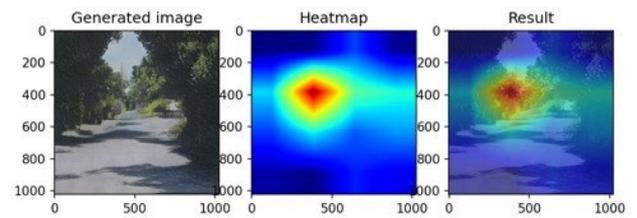


図11. 生成画像に対する Grad-CAM の出力例6(偽)

5. 考察

5.1 GAN の応用可能性

GAN を用いた実験の結果から、伊是名集落の街路景観は部分的に生成・再現された。図7, 図8のように生成画像に出力されたフクギは、位置や大きさも自然であり、その影も太陽の位置などと整合性の取れた自然な位置であった。一方で、全ての画像が必ずしも整合性の取れたものではなかった。例えば、図9の画像の右半分は、フクギが反時計回りに90度回転されて生成されており、また、画像左上でも同様に電柱が90度回転して出力されているように見える。このことから、StyleGAN2は画像を左右上下に半分にした、四分割の領域を一つのまとまった領域として捉えていることがわかる。StyleGAN2は画像の生成過程において、 $2\text{px} \times 2\text{px}$ の画像から $4\text{px} \times 4\text{px} \rightarrow 8\text{px} \times 8\text{px} \rightarrow \dots$ といったように段階的に生成する画像の解像度を大きくすることで、大域的な特徴と局所的な特徴を画像に織り込んでいく。つまり、出力された画像が4分割されて見えるということは、生成の初期段階でマッピングされた大域的な特徴が適切な修正をされないまま、後続の局所的な特徴の反映プロセスが進んだと考えられる。

生成画像から抽出できた特徴は限定的であったものの、生成される街路景観要素にはパターンがあることが分かった。街路景観要素としてフクギに着目すると、生成のされ方には表2のようなパターンが見られた。

表2. フクギの生成パターン

パターン	組み合わせ
左右の配置	両側、片側、なし
高さ	石垣程度、家屋程度、それ以上
植栽位置	敷地内、敷地外
連続性	連続的、断続的、部分的

生成されたフクギの組み合わせの中には、実際には存在しない組み合わせも含まれる。図12のように道路の左右にそれぞれ部分的にフクギが植えられているような組み合わせはデータセットには含まれていなかった。



図12. データセットには含まれていない
フクギの生成の組み合わせ

データセットとして入力していない組み合わせが出現したことによって、実在しない街路景観要素の組み合わせを含む画像が生成できることを確認した。これは本研究が長期的な目標とする「消滅してしまった街路景観の特徴の顕在化」にとって重要な成果である。しかし現段階ではその可能性を示唆するだけにとどまるため、今後さらなる検証手法の確立が必要である。また一方で、本研究で用いた150枚という比較的小さなデータセットから実在しない街路景観要素の組み合わせを生成できたことは、データセットを揃えるのが容易ではない環境下でも景観シミュレーションが可能であることを示す。しかし実在画像では判断根拠とされた要素（民家、ブロック塀）が、生成画像には一度も判断根拠として見つからなかったという結果は、データセットサイズがもたらしたトレードオフであると考えられる。

5.2 Grad-CAMの解釈と改善

本研究では Style GAN2 で生成した画像に対して GradCAM を適応し、街路景観の特徴を可視化することを試みた。その中でも特徴的な結果について考察をする

Grad-CAM の出力結果では、識別器によって着目される領域が十字に表示されてしまう傾向があった。本来であれば放心円に近い形状の出力が期待されるため、改善が必要である。考えられる原因としては、訓練された GAN の識別器が適切に訓練されなかったか、Grad-CAM の勾配計算時に十字が出力されるようなバイアスがかかっていた、の2つが考えられる。前者に関しては、本研究で用いたデータセットとは異なるサイズのデータセットを用意するなどして、訓練の進行度や GradCAM の出力に影響があるかといった精査を行う。後者に関しては、異なる入力画像や識別器を用いて出力を比較することで原因の追求を行っていく。

また、一枚の画像の中で最も勾配が強い地点が二つある出力結果があった。これは複数の景観要素が出力結果に対して同等に強い影響を与えていたことになる。人の場合でも一つの要素を取り上げて、空間の印象を決めていることは考えづらい。今後の課題として、複数の景観要素が出力結果に対してどのような影響を与えるのかも調べていきたい。

6. まとめ

本研究では深層学習を用いて沖縄県伊是名の街路景観画像から特徴の顕在化を試みた。その結果、GAN を使うことで伝統的な集落の街路景観の潜在的な特徴を顕在化できたとともに、GAN を用いた景観シミュレーションの可能性を示すことができた。

[謝辞]

伊是名集落の調査研究は、2014年より東京工業大学藤井晴行研究室と日本大学生産工学部篠崎健一研究室が共同で行なっている。景観の撮影にご協力いただいた伊是名集落の皆さまと調査関係者に謝意を表す。

[参考文献]

- 1) 平田貞代, 藤井晴行, 篠崎健一. (2019). 伝統的集落のまちづくりにおける価値共創のエコシステム-土着建築技術の継承方法の探究その2.
- 2) 近藤桂司. (2019). *Index which Indicates Features of Streetscape Colors*. Transactions of Japan Society of Kansei Engineering, 18(3), 179–186. <https://doi.org/10.5057/jjske.tjske-d-18-00064>
- 3) 牧田和久, 三橋俊雄. (1992). 街路景観構成要素のイメージ評価への影響 - 街路景観のイメージ構造(その1). *BULLETIN OF JSSD*, 94, 75–82. https://doi.org/10.11247/jssdj.1992.75_3
- 4) 松本直司, 西寺敦敏, 仙田満. (1991). 街路景観の乱雑性・整然性要因に関する研究. 日本建築学会計画系論文報告集, 429, 73–82.
- 5) 山田悟史, & 大野耕太郎. (2019). Deep Learning を用いた印象評価推定 AI の作成と検証. *Journal of Architecture and Planning (Transactions of AIJ)*, 84(759), 1323–1331. <https://doi.org/10.3130/aija.84.1323>
- 6) Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). *Learning Deep Features for Discriminative Localization*. <http://cnlocalization.csail.mit.edu>
- 7) 山中理沙, 沖拓弥. (2022). 街路景観画像と印象評価値の変化を統合的に予測可能なシミュレーション手法の開発. 第36回日本人工知能学会
- 8) Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (n.d.). *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. <http://gradcam.cloudev.org>
- 9) Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks*. <http://arxiv.org/abs/1406.2661>
- 10) Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (n.d.). *Training Generative Adversarial Networks with Limited Data*. <https://github.com/NVLabs/stylegan2-ada>
- 11) Agrawal, P., Girshick, R., & Malik, J. (2014). *Analyzing the Performance of Multilayer Neural Networks for Object Recognition*. <http://arxiv.org/abs/1407.1610>
- 12) M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, *The Cityscapes Dataset for Semantic Urban Scene Understanding*, in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- 13) Simonyan, K., & Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. <http://arxiv.org/abs/1409.1556>
- 14) Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). *GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium*. <http://arxiv.org/abs/1706.08500>
- 15) 関千穂子, 渡辺康, 篠崎健一. 沖縄伊是名集落の石垣築造の実地による、琉球の伝統的集落空間整備可能性についての研究.