

# 深層学習による画像変換技術を用いた都市街路景観の インスタンスセグメンテーションのための合成データ自動生成法

環境エネルギー工学専攻 環境工学コース

環境設計情報学領域 矢吹研究室 菊池 拓哉

## 1. はじめに

建物を取り壊した後の将来景観を可視化したり、大量のストリートビュー画像から建物などの都市景観を分析したりするために、深層学習を用いた方法が提案されている<sup>[1,2]</sup>。しかし、建物を1棟ずつ抽出できるデータセットが公開されていないため、高度なシミュレーションや分析が極めて困難であった。また、深層学習を用いて、画像中に含まれる物体を検出するためには、学習用画像と画像中に含まれる物体の種類と位置を示すデータ（アノテーションデータ）のペアを事前に用意しなければならない。高い精度で検出するためには大規模なデータセット、つまり、大量のデータペアが必要である。しかし、手作業で大規模データセットを作成するためには大量のコスト（時間、労力）がかかる。そのため、データセットの自動生成方法は様々な分野で研究されている。都市景観分野でも提案されているが、これまでの方法では実在する都市の詳細なテクスチャを含む3次元都市モデルやストリートビュー画像が必要だった<sup>[3,4]</sup>。

そこで、実在しない3次元都市モデルを自動生成するプロシージャラルモデリングと画像変換（Image-to-Image、I2I）技術を用いて、建物を対象とするインスタンスセグメンテーション<sup>\*1</sup>用合成データ<sup>\*2</sup>セットを自動的に生成する方法の開発を本研究の目的とした。

## 2. 提案方法

提案方法は4つのステップで構成される（図1）。Step 1では3次元都市モデルをプロシージャラルモデリングにより自動生成する。なお、生成する都市モデルには窓や外壁などの詳細なテクスチャは不要である。Step 2では生成した3次元都市モデルからアノテーションデータの生成とセグメンテーション画像（種類ごとに領域を塗り分けた画像）を生成する。この際、データペアを作成する地点はランダムに選択される。そして、Step 3でI2I技術を用いてセグメンテーション画像からフォトリアルなストリートビュー画像を生成する。そして、Step 4でアノテーションデータとストリートビュー画像を関連付けて、データセットとする。提案方法は、3次元都市モデルの生成をランダムにし、さらに、データペアの作成地点を3次元都市モデル内からランダムに選択している。これにより、生成されるデータセットの種類を大幅に増加させることが可能である。

## 3. 検証実験・結果

提案方法の有用性を示すために、検証実験ではデータセット生成にかかる時間とストリートビュー画像の生成精度を確認した。また、生成したデータセットで学習したインスタンスセグメンテーションモデルが現実の建物を検出することができるか確認した。

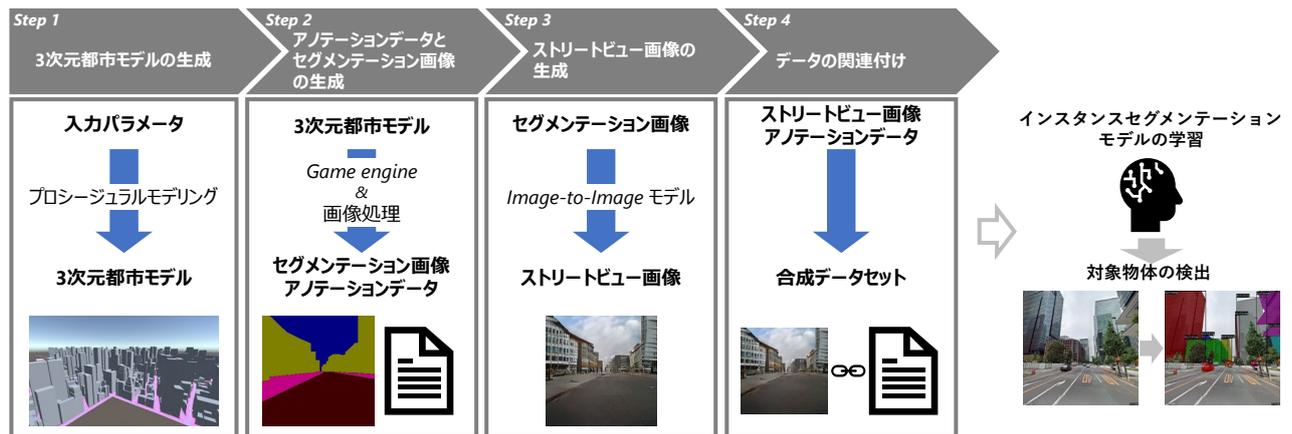


図1 非実在都市の3次元モデルを用いた合成データセット自動生成方法の概要

検証実験の結果、現実都市の街並みの特徴を捉えたストリートビュー画像を1ペアあたり約4.45秒で生成できることを確認した。また、生成したデータセットを用いて学習したモデルで現実の建物を検出できるかを確認した結果、建物同士の間隔が空いている、もしくは、隣り合う建物のファサードが大きく異なる場合に正しく検出できることが確認された(図2)。

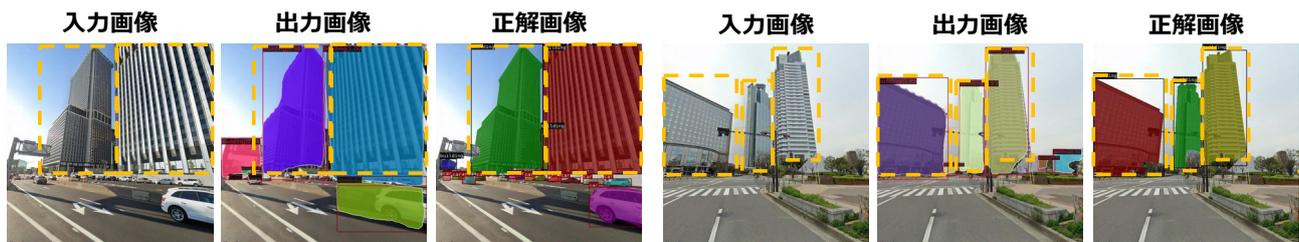


図2 生成したデータセットを用いて学習したモデルによる検出結果(一部)

#### 4. まとめ

本研究では、建物を対象としたインスタンスセグメンテーション用合成データセットを自動的に生成する方法を開発し、有用性を評価した。本研究で得られた結果は、少ないコストでデータセットを作成可能である。今後、I2I技術によるストリートビュー画像の生成精度を改善するなどして、合成データセットで学習した深層学習モデルの検出精度を向上させることが必要である。

※1 インスタンスセグメンテーション：物体を個別に検出し、種類を判別する深層学習モデル

※2 合成データ：コンピュータシミュレーションや現実に基づくアルゴリズムを用いて、もしくはゼロから人工的に生成されたデータ

- [1] Kido, D., Fukuda, T., & Yabuki, N. (2021). Assessing future landscapes using enhanced mixed reality with semantic segmentation by deep learning. *Advanced Engineering Informatics*, 48, 101281. <https://doi.org/10.1016/j.aei.2021.101281>
- [2] Wang, M., He, Y., Meng, H., Zhang, Y., Zhu, B., Mango, J., & Li, X. (2022). Assessing Street Space Quality Using Street View Imagery and Function-Driven Method: The Case of Xiamen, China. *ISPRS International Journal of Geo-Information*, 11(5), 282. <https://doi.org/10.3390/ijgi11050282>
- [3] Abu Alhaija, H., Mustikovela, S. K., Mescheder, L., Geiger, A., & Rother, C. (2018). Augmented Reality Meets Computer Vision: Efficient Data Generation for Urban Driving Scenes. *International Journal of Computer Vision*, 126(9), 961-972. <https://doi.org/10.1007/s11263-018-1070-x>
- [4] Zhang, J., Fukuda, T., & Yabuki, N. (2022). Automatic generation of synthetic datasets from a city digital twin for use in the instance segmentation of building facades. *Journal of Computational Design and Engineering*, 9(5), 1737-1755. <https://doi.org/10.1093/jcde/qwac086>