東京大学　関本研究室 / Sekimoto Lab. IIS, the University of Tokyo.

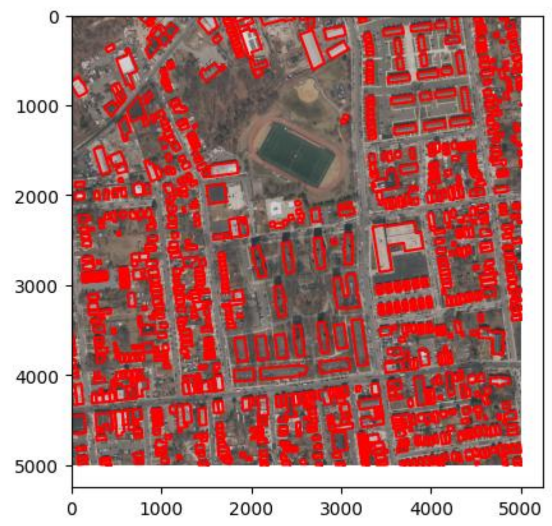
Developing a method for simultaneous extraction of building and road data from satellite imagery using deep learning

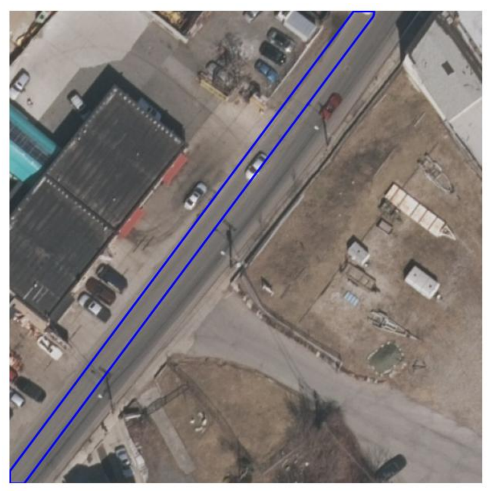
Shinji ITO, Chenbo ZHAO, Yoshihide SEKIMOTO

Background and Objectives

In recent years, the need to monitor urban conditions in real time has been growing, particularly in the contexts of smart city development and disaster response. Traditional approaches have typically extracted buildings and roads separately; however, treating them independently often leads to limitations in both extraction accuracy and processing efficiency. This study proposes a deep learning model that simultaneously extracts buildings and roads by leveraging their spatial correlation, aiming to improve accuracy and reduce inference time.

Dataset

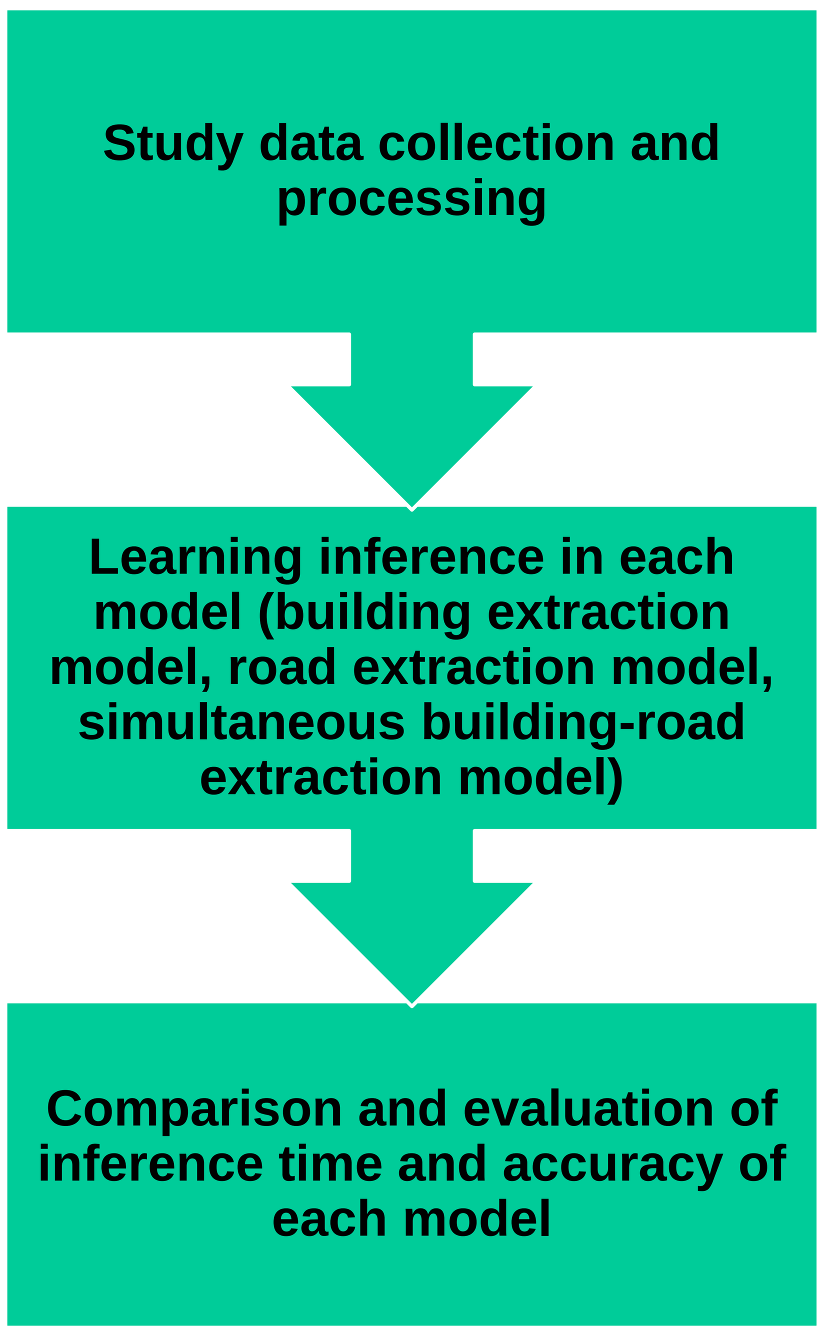




Building annotation (left image)
Road annotation (right image)

The study uses 100 satellite images of New York City (500 x 500 pixel) and corresponding building and road annotation data (961 buildings and 172 roads in total) for training. For validation and testing, 100 satellite images of another area (Nerima Ward, Tokyo) are used to evaluate the generalization performance.

Workflow



- **Study data collection and processing**
  - Pre-processing of satellite images
  Images in TIFF format are divided into 500 x 500 pixels, and the annotation data and coordinate system (CRS) are unified. Each pixel value is normalized from 0 to 1 to ensure learning stability.
  - Data expansion
  Data enhancements such as left-right inversion (probability 50%), scaling (±50%), translation (±10%), and color correction (hue ±1.5%, saturation ±70%, brightness ±40%) are automatically applied by default during training.

- **Learning inference in each model**
  - Fine tuning of models
  YOLO11 pre-trained models (COCO dataset) were used as initial weights. Learning rate scheduling (CosineAnnealing) was introduced and training was performed with default settings of SGD (momentum: 0.937, weight decay: 0.0005) for the optimizer.
  - Design of Loss Functions
  A total loss function combining classification loss (Binary Cross Entropy), location loss (CIoU), and segmentation loss (Binary Cross Entropy) was used for training. The default settings do not weight the loss function per class, so the weighting is done manually.

- **Comparison and evaluation of inference time and accuracy of each model**
  We will perform inference using instance segmentation for buildings and semantic segmentation for roads, and compare and evaluate the accuracy (IoU, mAP) and processing time for each method.

Result

Currently, it has completed to the inference by building extraction model and road extraction model.

Inference by building extraction model

Inference by road extraction model

Training results of building extraction model

Instances	Precision	Recall	mAP0.5	mAP0.5:0.95
216	0.819	0.732	0.778	0.469

Training results of road extraction model

instances	Precision	Recall	mAP0.5	mAP0.5:0.95
38	0.731	0.368	0.459	0.224

Future Prospects

Based on the results of each extraction model, we will develop a simultaneous extraction model using multi-class classification. We will also perform inference using two different classification methods (instance segmentation and semantic segmentation) to verify the differences in extraction accuracy and extraction time. In addition, since the current model is biased toward training data and there is room for improvement in extraction accuracy, we plan to add other training data and conduct further experiments.