Machine Learning For City-Scale Building Information Modeling

Development Team: Chaofeng Wang, UC Berkeley | Qian Yu, UC Berkeley | Barbaros Cetiner, UCLA
Advisory Board: Kincho H. Law, Stanford | Stella Yu, UC Berkeley | Ertugrul Taciroglu, UCLA | Frank McKenna, UC Berkeley
Leadership Team: Sanjay Govindjee, UC Berkeley | Gregory G. Deierlein, Stanford | Satish Rao, UC Berkeley | Ahsan Kareem, Notre Dame | Laura Lowes, UW

SimCenter

The Computational Modeling and Simulation Center (SimCenter) is part of the Natural Hazards Engineering Research Infrastructure (NHERI) program funded by the National Science Foundation. The commission of SimCenter is "Transforming the nation's ability to understand and mitigate adverse effects of natural hazards on the built environment through computational simulation."

Framework
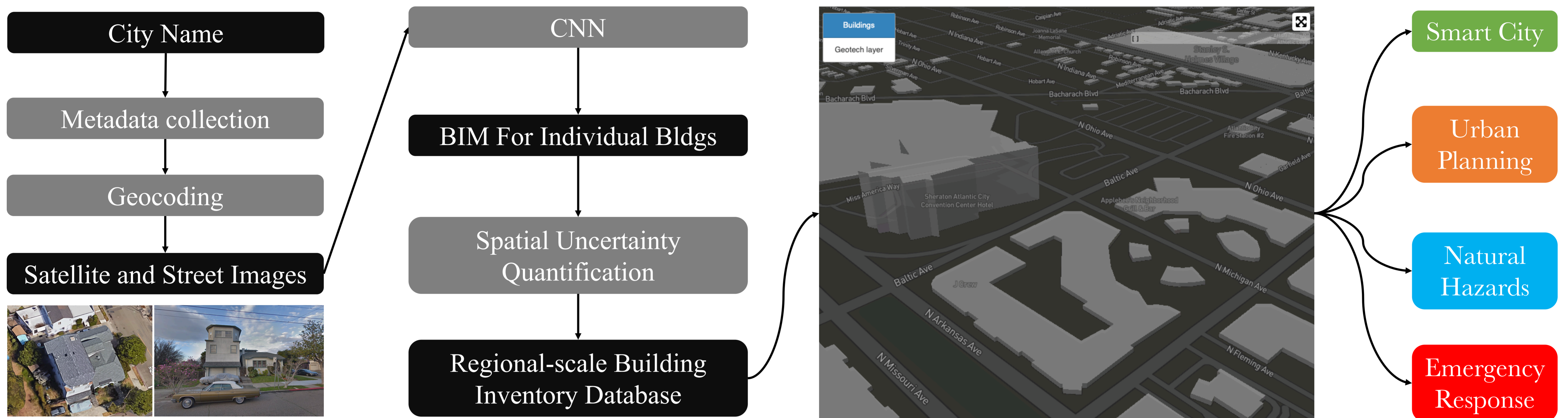
Metadata

When tax assessors value properties within a municipality, information including value and all other descriptive data are recorded. These data are public resources available on various governmental websites. Useful information for describing a building can be extracted from these records. For example, number of stories, exterior structural material, style, year of construction, type of garage, etc. The authors developed a series of automated scripts to scrape these information from public websites and created a preliminary database for each building we can find. This building information model database is called the Metadata-BIM database.

Geocoding

The process of geocoding is to find out the locations of buildings on the earth and tag them with geographic coordinates (latitude and longitude). The reasons why geocoding is needed here are two-fold: First, the coordinate is the 'index' when retrieving images of individual buildings from satellite or street view image databases; Second, in order to enhance the incomplete database, the spatial distribution analysis requires the coordinates of buildings to be known. To geocode each building in the preliminary BIM database, Google Geocoding API is employed to retrieve the latitude and longitude.

Convolutional neural nets

In addition to Metadata-BIM, Convolutional Neural Networks (CNN) are utilized to extract more information from satellite and street view images. We harvest tags describing building properties (height, number of stories, structure type, exterior material, footprint shape, usage, etc) from OpenStreetMap (OSM) and associate them with images downloaded from Google Maps to build a databases for deep learning. The CNNs trained on these database are used to predict building properties when given any images containing unseen and unlabeled buildings. Repeat the above steps for each building property to create a Vision-BIM database.
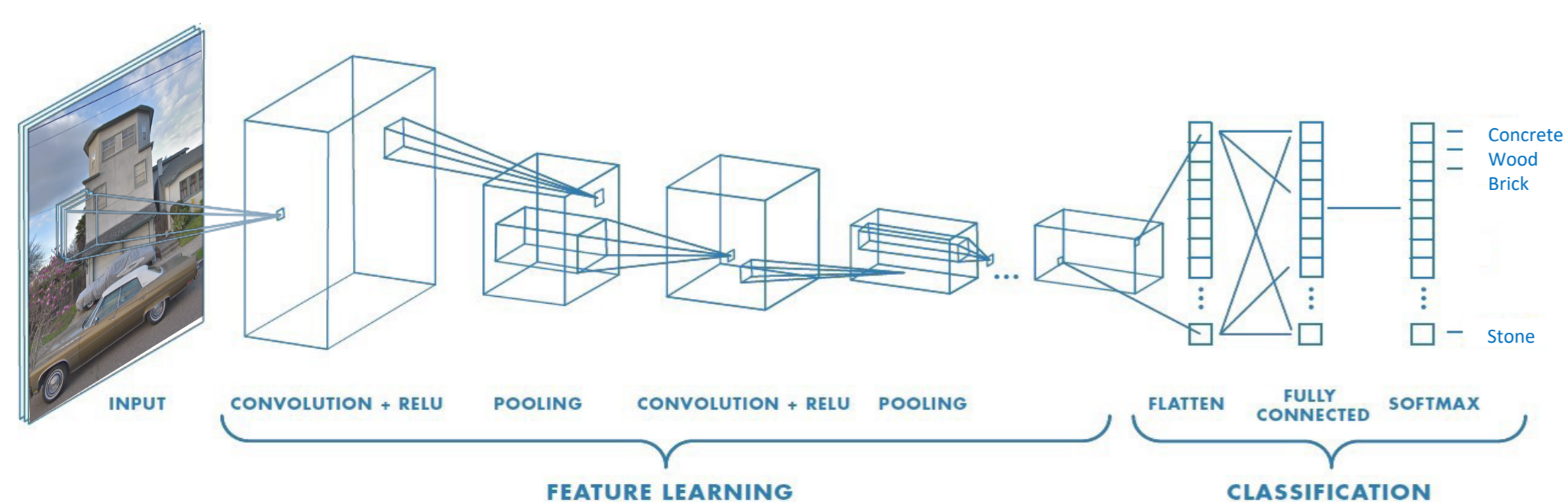
Spatial uncertainty quantification & Data enhancement
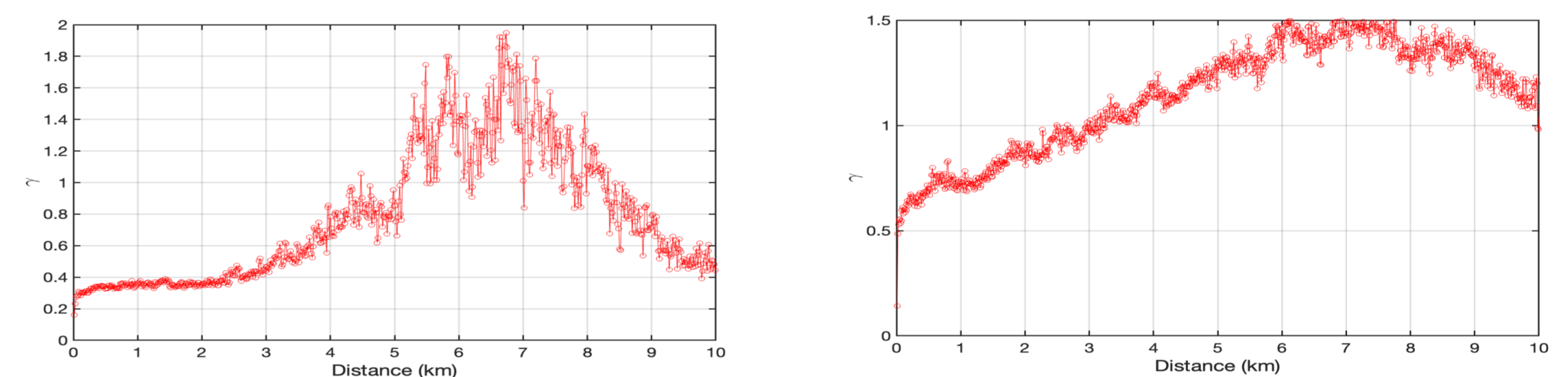
Spatial correlations of building properties.

We investigated the patterns in the spatial distribution of buildings and found that buildings were actually built following certain spatial patterns. Based on this discovery, we developed an AI tool, SURF, to learning the spatial patterns of building information and used it to enhance the merged Metadata-Vision BIM database, where missing values are predicted by neural nets.

Example

The development framework is applied to a coastal region in New Jersey. A BIM database consisting of more than 30,000 is created for regional hurricane simulation and loss estimation.

Buildings identified from this region.

Building information model.

Codes and released BIM database can be found on Github:
https://github.com/charlesxwang/BIM.AI