**Generating Environmental Measures from Multimodal Data using Machine Learning**

Min Namgung [namgu007@umn.edu](mailto:namgu007@umn.edu)

Yao-Yi Chiang [yaoyi@umn.edu](mailto:yaoyi@umn.edu)

Computer Science and Engineering, University of Minnesota

**1. Goal**:

* Automatically generate various types of environmental measures for India

**2. Target Geographical Areas:**

* We take all respondents’ locations from LASI-DAD Wave 1 (N=4,096) in these cities in India: [Delhi](https://livingatlas-dcdev.opendata.arcgis.com/datasets/087ee998317c42ba8cfdbfa2e5fe13b3/explore?location=28.644039%2C77.093250%2C11.54), [Bengaluru](https://www.openstreetmap.org/relation/2020589), [Greater Mumbai](https://livingatlas-dcdev.opendata.arcgis.com/datasets/67b75ea184424d20a657276251e164ea/explore), and [Kolkata](https://livingatlas-dcdev.opendata.arcgis.com/datasets/22935cc469ee42358cb0512cc0c2b734)
  + We obtain the ward boundary from ESRI India and OpenStreetMap
  + Jan-11-2024 Update:
    - If the ESRI website is blocked, go to [this website](https://geodata.mit.edu/catalog/princeton-gb19f9251).
* [Chennai](https://datahub.shiftcities.org/datasets/5b1072dbce9d4a5780663d61d0b62158/explore?location=12.839182%2C79.359602%2C9.20)

**3. Spatial Analytic Units:**

We generate environmental measures for drivable road intersections in the target geographical areas. The numbers of those road intersections in the target cities are:

* Mumbai: 13,481
* Delhi: 68,409
* Bengaluru: 71,739
* Kolkata: 18,591

For particular data sources and measures, for example, the numbers of road intersections having Google Street View images within a 1,600m buffer (i.e., 20 minutes walking distance[[1]](#footnote-0)) centered at the houses of each respondent are (Figure 1):

* Mumbai: 7,716
* Delhi: 20,228
* Bengaluru: 9,809
* Kolkata: 8,061

**4. Datasets we have generated** **and their download links**

* [**Road intersections**](https://drive.google.com/drive/folders/1HKnDn-QFxcdKyhfngzAaM9g9OXs3fV7g?usp=drive_link)
  + Generated Dataset:
    - mumbai-road.geojson
    - delhi-road.geojson
    - kolkata-road.geojson
    - bengaluru-road.geojson
  + Data schema:
    - osmid: the intersection id
    - center\_lat: the latitude of the intersection point
    - center\_lng: the longitude of the intersection point
    - geometry: a 100mx100m squared buffer for each intersection point
* [**Detected measures/physical objects**](https://drive.google.com/drive/folders/1Glpuu0VDxBAt6bL8Al1moVMNXJdh4ho_?usp=drive_link)
  + Method:
    - We query four (east, west, south, north) Google Street View images at each intersection within a 50m buffer and detect measures/physical objects from the images
      * Mumbai: 11,361 road intersections have Google Street View images
      * Delhi: 52,039
      * Bengaluru: TBA
      * Kolkata: 11,618 road intersections have Google Street View images
  + These measures/physical objects can be derived from general street-view images
    - Greenness indicator[[2]](#footnote-1)
      * Take four images in each location (east, west, south, north) and compute the ratio of green pixels to the total number of pixels
    - The number of person (person\_gt)
      * Person + Riders
    - The number of vehicles (car\_gt)
      * Car + Bus + Truck
    - Motorcycle, Bicycle, Train

→ These objects are included in the [instance-objects folder](https://drive.google.com/drive/folders/1XF00b0DHuekPrtoqe-BDMMM8PMezHooj?usp=drive_link)

* + - Additionally, we can capture the existence or the percentage of
      * Sidewalk
      * Building
      * Sky
      * Wall, Fence
      * Traffic light
      * Poll
      * Traffic sign
      * Road

→ These objects are included in the [segmented-objects folder](https://drive.google.com/drive/folders/1tJ38z8tbC21OC2EJx3vA4wYU0duxQzK_?usp=drive_link)

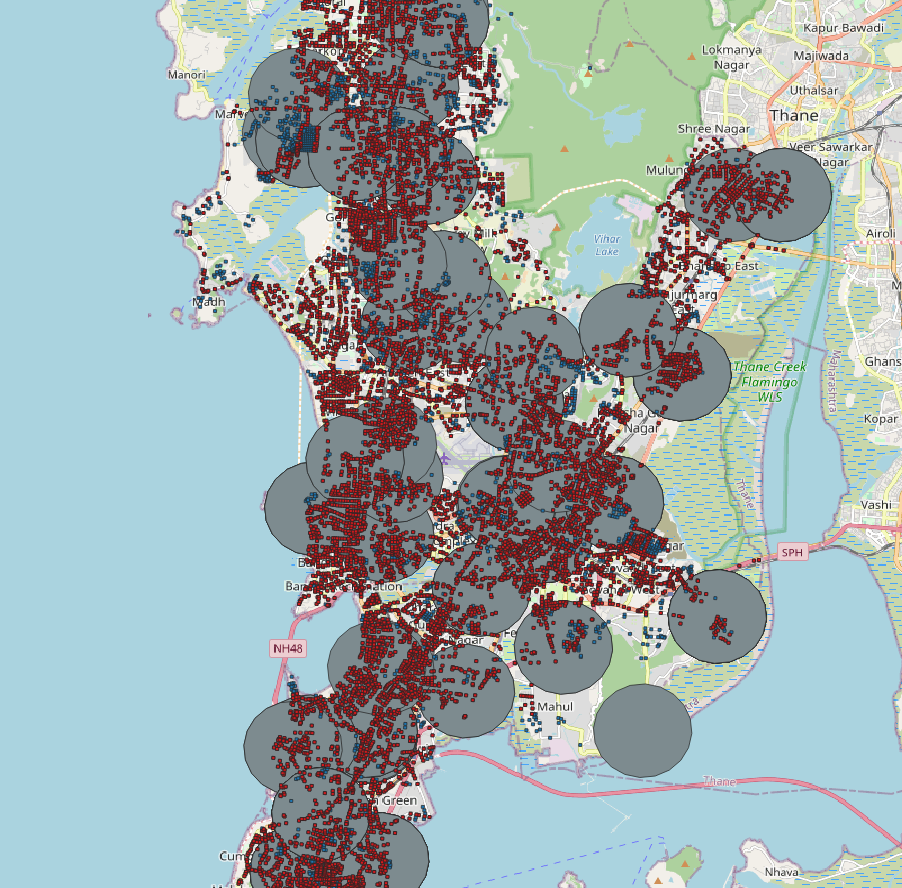
* + Generated Dataset:
    - [Instance-objects:](https://drive.google.com/drive/folders/1XF00b0DHuekPrtoqe-BDMMM8PMezHooj?usp=drive_link)
      * Mumbai-instance-objects.geojson
      * Kolkata-instance-objects.geojson
      * Delhi-instance-objects.geojson
    - [Segmented-objects](https://drive.google.com/drive/folders/1tJ38z8tbC21OC2EJx3vA4wYU0duxQzK_?usp=drive_link):
      * Mumbai-segmentation-objects.geojson
      * Kolkata-segmentation-objects.geojson
      * Delhi-segmentation-objects.geojson
    - [Combined-objects](https://drive.google.com/drive/folders/16GohMz28KDQkm1pSTs3qJP9zPmtpqqjm?usp=drive_link):
      * Mumbai-combined-objects.geojson
      * Kolkata-combined-objects.geojson
      * Delhi-combined-objects.geojson
    - bengaluru- TBA
  + Instance-objects data schema:
    - osmid: the intersection id
    - image\_exist: indicate whether images is existed (Yes or No)
    - objects [car, truck, train, person, etc.]: counts
    - center\_point: the queried street view image’s point location
    - objects [car, person]\_gt: the sum of all associated objects
    - geometry: the 100m x 100m squared buffer polygon geometry
  + Segmentation-objects data schema:
    - osmid: the intersection id
    - image\_exist: indicate whether images is existed (Yes or No)
    - objects [sidewalk\_percent, building\_percent, sky\_percent, etc.]: percentage of the number of pixels from four images (the same method as the gvi)
    - gvi: the calculated greenness view index
    - center\_point: the queried street view image’s point location
    - geometry: the 100m x 100m squared buffer polygon geometry
  + Combined-objects data schema:
    - \*The combination version of instance-objects and segmentation objects
    - Image\_exist\_instance: instance result is existed (Yes or No) derived from image\_exist from instance-objects and image\_exist from segmentation-objects
    - Image\_exist\_segmentation: segmentation result is existed (Yes or No)
    - Note:
      * The number of regions where instance and segmentation exist are different because the classes are complementary to each other
        + For instance, sky or vegetation do not exist in instance since we cannot count the number of sky or vegetation
        + Here are the number of differences for each region:

Segmentation results tend to have more than instances

Mumbai: 14

Kolkata: 360

Delhi: 1,626

**Figure 1.** **Mumbai**. **Circle**: the 1,600m buffers centered at the houses of respondents. **Red Squares**: road intersections where Google Street View images exist. **Blue Squares:** road intersection where street view images don’t exist

**5. Other Data we have collected:**

**5.1 India**

1. Street view image
   1. [Google Street View image (360 degree)](https://developers.google.com/maps/documentation/streetview/?_gl=1*1ditilh*_ga*MTM4MTM5Njc2LjE2OTY5MDc3OTg.*_ga_NRWSTWS78N*MTY5ODA3NTMwMi4xNS4xLjE2OTgwNzUzMDMuMC4wLjA.)
      * Coverage: entire drivable roads in all four cities
      * Price:
        + $7 for every 1000 images
   2. [Indian Driving Dataset with annotation](https://idd.insaan.iiit.ac.in/dataset/details/)
      * Road image (car, road condition, etc.)
      * Price: Free
   3. [Mapiliary street view images](https://www.mapillary.com/datasets)
      * Coverage: Partial coverage for all four cities
      * Price: Free
2. Raster-based dataset
   1. [Planet satellite image](https://www.planet.com/products/hi-res-monitoring/)
      * In the middle of the contacting
      * Spatial resolution: 3m/ pixel
      * Price: TBA
   2. [Slope](https://search.earthdata.nasa.gov/search/granules?p=C1711961296-LPCLOUD&pg%5B0%5D%5Bv%5D=f&pg%5B0%5D%5Bgsk%5D=-start_date&q=NASA%20DEM&circle%5B0%5D=79.73438%2C21.39427%2C3115229&tl=1672181516!3!!&lat=17.129349281721076&long=-114.328125&zoom=1)
      * Spatial resolution: 15m/ pixel
      * Price: Free
   3. [Population](https://www.worldpop.org/datacatalog/)
      * Spatial resolution: 1000m/ pixel
      * Price: Free
   4. [Global Human Settlement](https://ghsl.jrc.ec.europa.eu/download.php) 
      * Spatial resolution: 10m/ pixel
      * Price: Free
   5. [Night lights](https://search.earthdata.nasa.gov/search/granules/collection-details?p=C1624667653-LAADS&q=Day/Night&fi=VIIRS&tl=1679928462!3!!&fst0=Spectral/Engineering&fsm0=Infrared%20Wavelengths&fs10=Infrared%20Radiance&fst1=Spectral/Engineering&fsm1=Infrared%20Wavelengths) 
      * Spatial resolution: 375m/ pixel
      * Price: Free
3. Geographic Data
   1. [Open Street Map](https://download.geofabrik.de/)
      * Polygon
        + Lake, Mountains, Basin, Apartments, School, Gym, Park, etc.
      * Point
        + Restaurants, traffic signals, traffic signs, etc.
      * Line
        + Highway, roads, River, etc.
      * Price: Free
   2. [India Ward Level Data](https://github.com/datameet/Municipal_Spatial_Data)
      * [Greater Mumbai](https://livingatlas-dcdev.opendata.arcgis.com/datasets/67b75ea184424d20a657276251e164ea/explore)
      * [Kolkata](https://livingatlas-dcdev.opendata.arcgis.com/datasets/22935cc469ee42358cb0512cc0c2b734)
      * [Delhi](https://livingatlas-dcdev.opendata.arcgis.com/datasets/087ee998317c42ba8cfdbfa2e5fe13b3/explore?location=28.644039%2C77.093250%2C11.54)
      * [Bengaluru](https://www.openstreetmap.org/relation/2020589)
      * Price: Free
   3. [World Land Use](https://viewer.esa-worldcover.org/worldcover/)
4. Other Data
   1. [SHRUG](https://www.devdatalab.org/shrug)
      * 2011 India census-based survey data
      * Nightlight, population and economic census, rural development
      * Socio-economic and Caste Census
      * Forest coverage
      * Census tract-level survey, which means it is not a fine-grained information
      * Coverage: Village (ward) level
      * Price: Free
   2. [Indian thematic data](https://bhuvan-app1.nrsc.gov.in/thematic/thematic/index.php)
      * Web-based dataset
      * Land Use Land Cover (2005, 2011, 2015)
      * Coverage: Village (ward) level
      * Price: Free
   3. [Weather (Rainfall)](https://www.imdpune.gov.in/cmpg/Griddata/Rainfall_1_NetCDF.html)
      * Yearly min/max rainfall
      * Coverage: Village (ward) level
      * Price: Free

**5.2 Non-India: The purpose of the experiment on data-rich areas with our AI model**

1. Street view image
   1. [Google Street View image (360 degree)](https://developers.google.com/maps/documentation/streetview/?_gl=1*1ditilh*_ga*MTM4MTM5Njc2LjE2OTY5MDc3OTg.*_ga_NRWSTWS78N*MTY5ODA3NTMwMi4xNS4xLjE2OTgwNzUzMDMuMC4wLjA.)
      1. Price: TBA
   2. [Mapiliary street view images](https://www.mapillary.com/datasets)
      1. Price: Free
   3. [Streetlearn](https://sites.google.com/view/streetlearn/)
      1. Pennsylvania (PA), Manhattan (NYC)
      2. Price: Free
2. Raster-based dataset
   1. [Planet satellite image](https://www.planet.com/products/hi-res-monitoring/)
      1. In the middle of the contacting
      2. Spatial resolution: 3m/ pixel
      3. Price: TBA
   2. [NAIP high-resolution satellite image in the U.S.](https://earthexplorer.usgs.gov/) 
      1. Spatial resolution: 0.6m/ pixel
      2. Price: Free
   3. [Slope](https://search.earthdata.nasa.gov/search/granules?p=C1711961296-LPCLOUD&pg%5B0%5D%5Bv%5D=f&pg%5B0%5D%5Bgsk%5D=-start_date&q=NASA%20DEM&circle%5B0%5D=79.73438%2C21.39427%2C3115229&tl=1672181516!3!!&lat=17.129349281721076&long=-114.328125&zoom=1)
      1. Spatial resolution: 15m/ pixel
      2. Price: Free
   4. [Population](https://www.worldpop.org/datacatalog/)
      1. Spatial resolution: 1000m/ pixel
      2. Price: Free
   5. [Global Human Settlement](https://ghsl.jrc.ec.europa.eu/download.php) 
      1. Spatial resolution: 10m/ pixel
      2. Price: Free
   6. [Night lights](https://search.earthdata.nasa.gov/search/granules/collection-details?p=C1624667653-LAADS&q=Day/Night&fi=VIIRS&tl=1679928462!3!!&fst0=Spectral/Engineering&fsm0=Infrared%20Wavelengths&fs10=Infrared%20Radiance&fst1=Spectral/Engineering&fsm1=Infrared%20Wavelengths) 
      1. Spatial resolution: 375m/ pixel
      2. Price: Free
3. Geographic Data
   1. [Open Street Map](https://download.geofabrik.de/)
      1. Polygon
         1. Lake, Mountains, Apartments, School, Gym, Park, etc.
      2. Point
         1. Restaurants, traffic signals, traffic signs, etc.
      3. Line
         1. Highway, roads, River, etc.
      4. Price: Free

**General project priorities:**

1. Low-hanging fruit project on intersections in cities (this data)
   1. Use this for writing retreat paper
2. Expand this to the broader survey
3. Validation or supplementation using pictures taken during Wave 2
4. Road conditions and access - washing out or inaccessibility at various times of the year
5. Agricultural land use - classifying type of land
6. Other signals/indicators important to climate change? Heat islands.

1. Dempsey, S., Lyons, S., & Nolan, A. (2018). Urban green space and obesity in older adults: evidence from Ireland. *SSM-Population Health*, *4*, 206-215. [↑](#footnote-ref-0)
2. Yang, J., Zhao, L., Mcbride, J., & Gong, P. (2009). Can you see green? Assessing the visibility of urban forests in cities. *Landscape and Urban Planning*, *91*(2), 97-104. [↑](#footnote-ref-1)