Using machine learning to examine neighbourhood characteristics associated with physical function

Dr Jerome Rachele
Project team

Jerome Rachele  
MSPGH

Mark Stevenson  
MSD

Bec Bentley  
MSPGH

Haifeng Zhao  
MSD

Jasper Wijnands  
MSD

Gavin Turrell  
Deakin

Jingcheng Wang  
MSPGH
Background

• We normally use GIS to examine associations between neighbourhood design and health, but this has limitations:
  • Data often not available (footpaths, tree coverage, building height etc)
  • Models with multiple neighbourhood design characteristics can become complex
  • Measurement error when measuring the built environment

• Therefore, let’s explore another approach
Aim

• To explore the potential of Generative Adversarial Networks (GAN) – machine learning – to understand the association between neighbourhood design and streetscape characteristics and physical function
GAN – What is it?

• A class of machine learning, in technical terms...
  *Unsupervised Machine Learning*, Neuron network, Game theory, Discriminator/Generator

• An analogy

  Generator = A **child** who is learning how to draw a house
  Discriminator = A **teacher** that knows what a house looks like, tells the child how good their drawing is
Methods

• Image-to-image transfer
  • We train the discriminator (teacher) using example images
    • Neighbourhood of high-functioning people vs low functioning people
  • The generator (child) provides example images
    • The generator makes changes to a “high functioning” neighbourhood, to make it look
      more like a “low functioning” neighbourhood

• Used UniMelb Spartan High Performance Computing
Data sources

• N=5,300 HABITAT participants living in Brisbane + their addresses + their physical function measures
• Images from Google StreetView, Google Maps, NearMap
Street View images are panoramic
Google maps (colours are meaningful)
NearMap
Methods

• Download and sort images
  • Script written in Python to download and save images for each XY coordinate
  • Images classified as high or low function depending on participant
  • Task undertaken in SPARTAN (UniMelb HPC Services)
• [https://github.com/mingyuliutw/UNIT](https://github.com/mingyuliutw/UNIT) UNIT
• We use an existing open-source model and tailor it to our needs.
Issue 1

• HABITAT’s original sample was clustered at the ABS 2001 Census Collectors District level

• A meaningful area to examine for health and behaviours outcomes is around 1.6km

• Clusters ended up being too close together
Issue 2

• The more similar images are, the more you need to train a model.
Issue 2

• The more similar images are, the more you need to train a model.
Google maps
Nearmap
Example from cycling crash sites (Zhao, 2019)
Methods version 2

• HABITAT is contains 200 neighbourhood clusters (baseline)
• We aggregated physical function to the neighbourhood level and used the top 10 and bottom 10 neighbourhoods, then took multiple streetview images (around 5000) from within those neighbourhoods
Finding 1 – Greenery
Finding 2 – Dwelling structure
We’re extending the model

• Feasibility of:
  • Walking in the neighbourhood
  • Subjective wellbeing
  • Psychological distress
  • Transport mode
Future applications of method

• Survey data - simple random sample
• Ecological data with multiple cities

• Drafting manuscript for health journal