ABSTRACT

We want to determine future outbreaks of West Nile Virus, based on a fine temporal and spatial scale.

There are many different approaches to this, including statistical modeling and county-level data. However, we wanted individual-level data.

We choose to follow a machine learning image analysis-based approach to determine areas with higher risk of outbreaks via satellite imagery.

Common risk predictors include green space, population density, weather, and housing age. Focusing on housing age, we hypothesized that housing age could be categorized based on roof style and housing shape via satellite imagery.

We discovered that we were able to produce a working image-analysis model using image labeling, housing age validation, and machine learning. This model can highlight houses in satellite imagery and predict housing age semi-accurately.

BACKGROUND

West Nile Virus is a mosquito-borne illness that causes encephalitis in 1 out of 150 infected individuals.

It has no known vaccine or treatment besides testing and isolation.

Previous studies differ on what are the best resources to provide a more in-depth understanding of sources to provide a more in-depth understanding of West Nile Virus, based on a fine temporal and spatial scale.

ABSTRACT

We want to determine future outbreaks of West Nile Virus, based on a fine temporal and spatial scale.

There are many different approaches to this, including statistical modeling and county-level data. However, we wanted individual-level data.

We choose to follow a machine learning image analysis-based approach to determine areas with higher risk of outbreaks via satellite imagery.

Common risk predictors include green space, population density, weather, and housing age. Focusing on housing age, we hypothesized that housing age could be categorized based on roof style and housing shape via satellite imagery.

We discovered that we were able to produce a working image-analysis model using image labeling, housing age validation, and machine learning. This model can highlight houses in satellite imagery and predict housing age semi-accurately.

RESULTS

Using the labeled satellite imagery, validated with the Cook County Assessor’s dataset, we were able to build a model that can reliably detect houses within unlabeled images.

It can also accurately predict the classification of houses some of the time, with the model accurately predicting houses post-1960 about 80% of the time, and the model accurately predicting houses pre-1960 50% of the time.

While the accuracy of classification is not ideal, the model is still able to output newly-labeled images, which can in turn be used in ArcMap to produce a fine-scale map of individual houses, allowing for much closer analysis at a finer scale, instead of generalizations at the county level.

The classifications may be useful as a rough estimate of housing age, as well, and should improve with larger training sets.

CONCLUSIONS

As DuPage county lacks easy housing age validation data, other approaches are required to produce fine-scale housing age maps.

With this tool, we can not only likely produce maps of individual homes, but evaluate which ones are most at risk, based on the age of the housing, using available satellite imagery.

METHODOLOGY

In order to collect the imagery for analysis, we had to find a means of collecting satellite imagery for specific locations; the best way of achieving this is through running a python program to download chunks of images from the Google Maps Static API.

To prepare these images for training, we must annotate them with labels at each house, corresponding to rough housing age. For this, we used labelling. To validate this data, we used data from the Cook County Assessor’s office, and through transformations in ArcMap, were able to create a map of every residential parcel in Cook County. These parcels were classified based on the age listed in the Cook County Assessor’s data, and with this validation we were able to cross-reference the training set with the base dataset, creating an accurate series of images labeled according to housing age.

Once enough images are labeled, we fed the training set into a machine learning model called YOLO, or You Only Look Once, which is a real-time, image-based analysis learning algorithm that can take labeled images, analyze them, and learn to label unlabeled images to fit with the training set.

Running the model takes a great deal of computational resources and time, so the model was trained and run on the HAL cluster.

REFERENCES

• Special thanks to Weihao Ge for mentoring me through the SPIN program
• Special thanks to Tao Zang, for figuring out the Google Maps API and the YOLO modeling
• Fig. 1: Credit: https://www.flickr.com/photos/daleh/3408794695
• https://wiki.ncsa.illinois.edu/display/CPRHD/Report%3A+Housing+Age+as+a
• Determinant+of West+Nile+Virus++risk
• https://wiki.ncsa.illinois.edu/display/CPRHD/Report%3A+Housing+Age+as+a
• Determinant+of West+Nile+Virus++risk
• https://wiki.ncsa.illinois.edu/display/CPRHD/Report%3A+Housing+Age+as+a
• Determinant+of West+Nile+Virus++risk
• https://wiki.ncsa.illinois.edu/display/CPRHD/Report%3A+Housing+Age+as+a
• Determinant+of West+Nile+Virus++risk
• https://wiki.ncsa.illinois.edu/display/CPRHD/Report%3A+Housing+Age+as+a
• Determinant+of West+Nile+Virus++risk
• https://wiki.ncsa.illinois.edu/display/CPRHD/Report%3A+Housing+Age+as+a
• Determinant+of West+Nile+Virus++risk
• https://wiki.ncsa.illinois.edu/display/CPRHD/Report%3A+Housing+Age+as+a
• Determinant+of West+Nile+Virus++risk
• https://wiki.ncsa.illinois.edu/display/CPRHD/Report%3A+Housing+Age+as+a
• Determinant+of West+Nile+Virus++risk
• https://wiki.ncsa.illinois.edu/display/CPRHD/Report%3A+Housing+Age+as+a
• Determinant+of West+Nile+Virus++risk