01

Urban Hotspots

by Esther Xie

PROJECT DESCRIPTION

Abstract

The project investigates the correlation between morphology and human activities in Helsingborg. It aims to examine the assumption that to attract more visitors and activities in the city, designers must create spatial qualities that are unique and outstanding. By analyzing data of visits and spatial qualities using Helsingborg as a sample site, the project aims to understand what physical characteristics of places relate to popularity and attractiveness. The project primarily uses Twitter data as an indicator of activities, and then employs visual AI to analyze Google street view images to understand the spatial qualities. The finding implies that the hotspots in the city generally have bigger footprint, and have more complex façades. By examining the physical characteristics of successful urban design projects, the project aims to provide reference for designers on applicable qualities when designing for urban hotspots.

Introduction

When designing urban projects, urban designers usually do morphology study of existing site and comparable projects as reference. With morphology study, designers can get a broad sense of the spatial forms that could be effective on the site. However, the common morphology study is generally subjective and speculative, as the designers will decide which projects are interesting enough for the study by themselves, and what spatial qualities are important for the success of these projects. Therefore, this project derives from the motivation to improve such morphology study process with data analysis. By mapping social media data and visual AI, the project aims to provide some data legitimacy for the urban design process.

The morphology study method is inspired by literature and design practice of Professor Joan Busquets. For example, his book Urban Grids: Handbook for Regular City Design (2019) examines innovative urban projects in numerous study, and study their physical characteristics such as street offsets, building heights, floor-area ratio (ratio), planting areas etc. The handbook is valuable as a guideline for urban designers.

Numerous projects exist in the field of using visual AI to analyze the correlation between spatial qualities and human conceptions. One instance is the [In]distinct Cities conducted by the Senseable City Lab. The project uses more than 2 million social media images to develop a model that quantifies the distinctiveness of 18 cities around the world and identifies their visual similarities. Thereby, it reveals the visual identity of each city and the commonplace between different cities. Correspondingly, my project is also interested in understanding the distinct and indistinct qualities of urban places while focusing on only one city, Helsingborg.

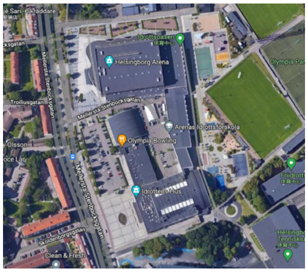
The project uses Helsingborg as a sample site. Helsingborg is a scenic coastal city, with many traditional buildings in the historic center. It is also celebrated for its innovative-driven spirit, and has taken efforts to provide accessible open data. It is the ninth largest city in Sweden and has a population of 113,816 in 2021. The project primarily choses Helsingborg because of its coherent urban morphologies with diverse building styles.

Right: Four samples of activity hotspots in Helsingborg

Next Page: Buildings in Helsingborg with tweet counts, and three buildings with the highest tweet counts

1. Idrottens Hus

Sports complex/ concert hall
tweet count: 1238

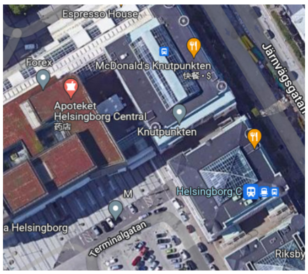


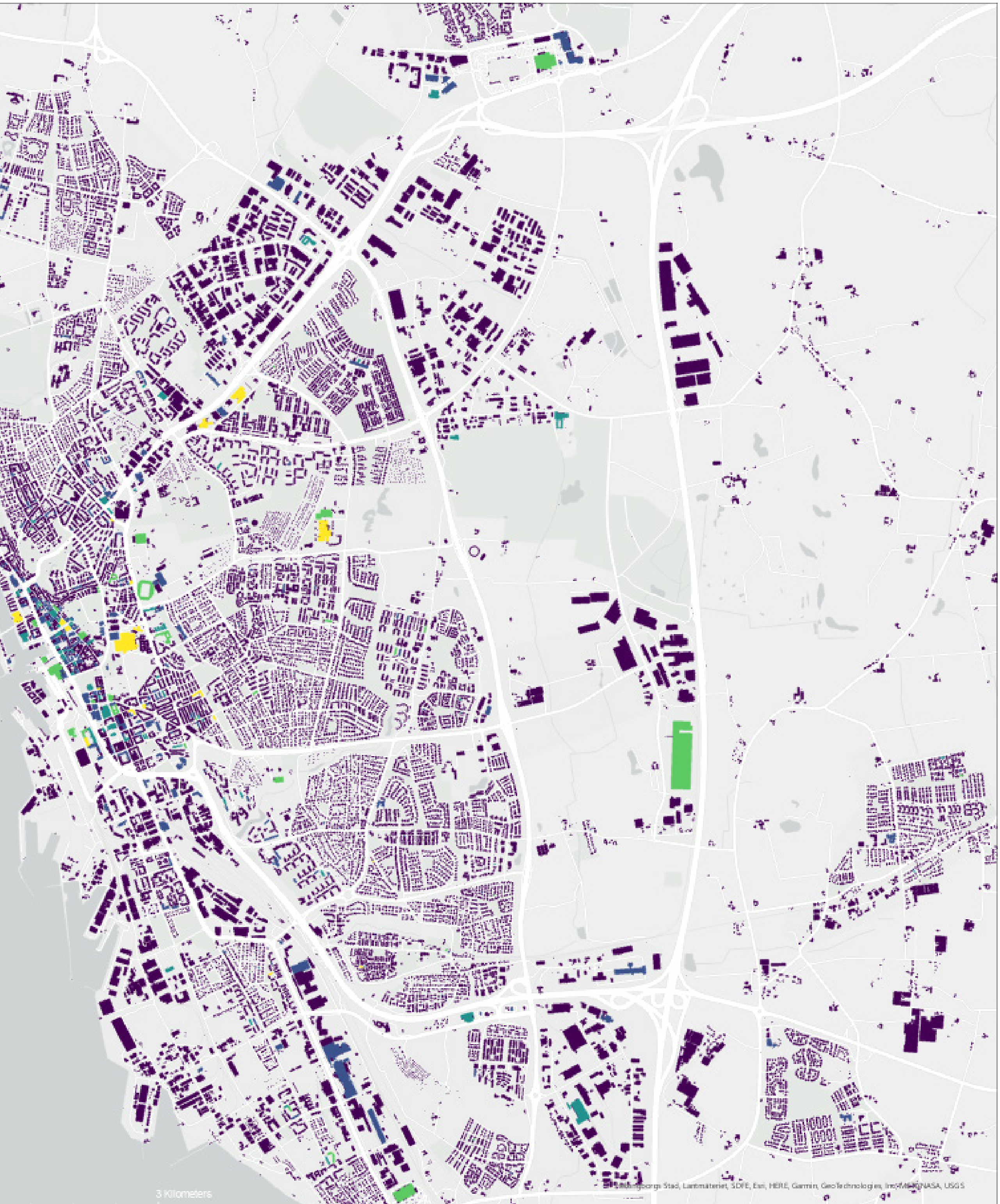
2. Väla Centrum

Shopping mall
tweet count: 963

3. Knutpunkten

Train/ bus station
tweet count: 852



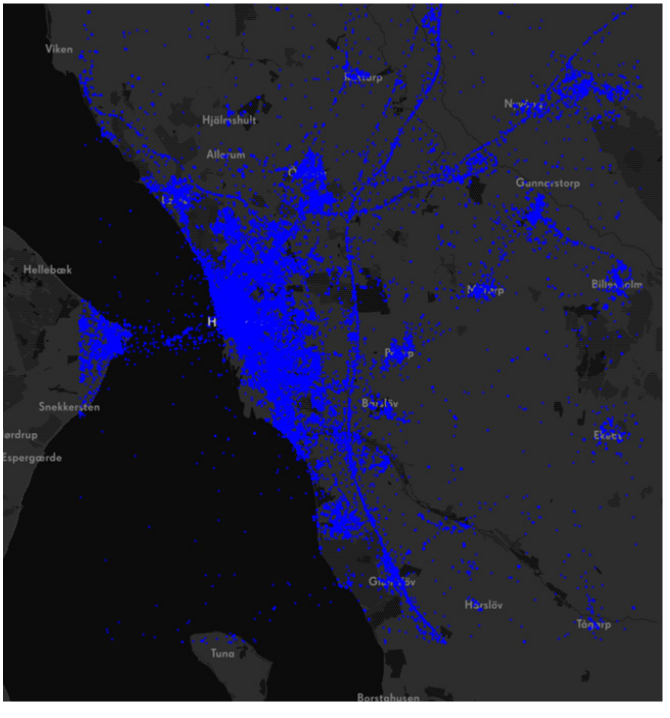

3 Kilometers
Helsingborgs Stad, Lantmäteriet, SDFE, Esri, HERE, Garmin, GeoTechnologies, Inc, METI/NASA, USGS

DATA AND METHODOLOGY

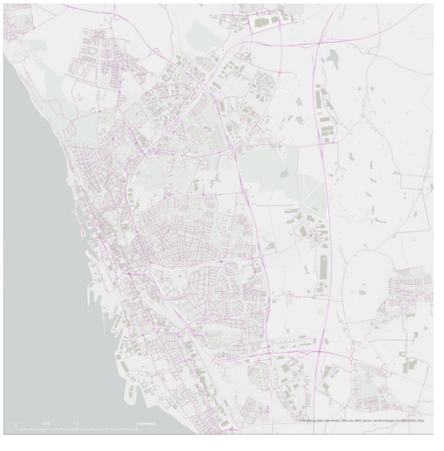
The project mainly uses three data sources: Twitter data, Google street view images, and GIS data from Helsingborg open data portal. The overarching logic is to use tweet number as an indicator of interests, and then use visual AI to analyze the street view images of spots with high interests. Because the project is interested in the visitation and activities taking place in each location, tweet numbers are an appropriate representation of how many activities take place in one location.

The first step is to scrape twitter data in Helsingborg from Twitter API. The project scraped 34,635 tweets with user id, date, content, and location etc. The next step is to map the locations of Twitter data using the coordinates contained in each tweet. Then, using the building footprint data downloaded from Helsingborg open data portal, I was able to join the tweet data to each building with R. In this process, tweets outside the building footprints are left out. As a result, more than 10,000 tweets are selected and aligned with buildings in Helsingborg. It is worth mentioning that the location information contained with Twitter data may not be precise enough to reflect if it is inside or outside the building. Yet, this experiment assumes that the posts in locations close enough to the building are related to the activities in the building. Therefore, building footprints with tweet counts were generated to indicate the activity intensity.

The next step is to scrape Google street view images of the activity hotspots. I select buildings with more than one hundred tweet counts, which sum up to 60 buildings. To scrape street view images from Google API, longitudes and latitudes of the street view location are required. I intersect the

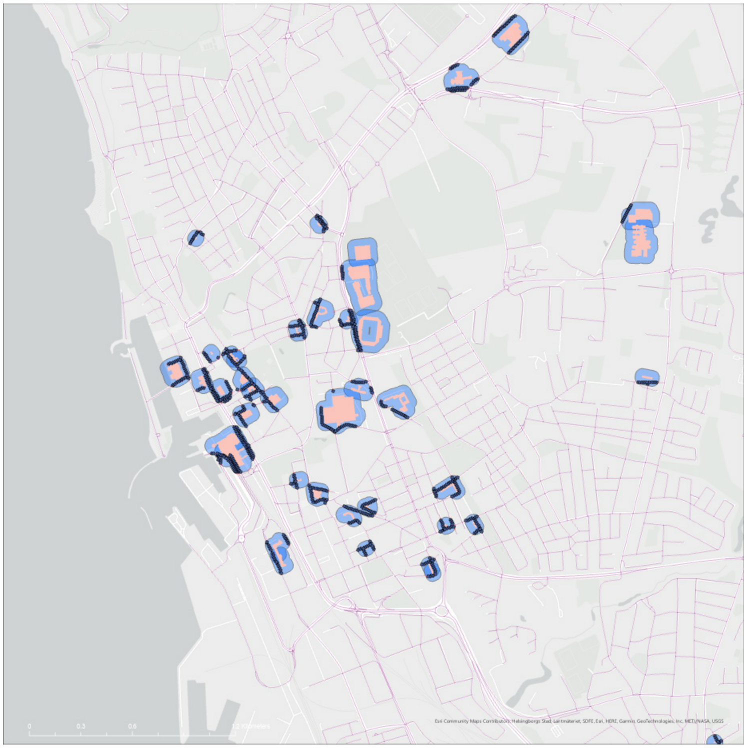



```
# A tibble: 34,635 × 90
   user_id   status_id           created_at          screen_name     text   source
   <chr>     <chr>               <dttm>              <chr>           <chr>  <chr>
 1 282726205 184258677702791168  2012-03-26 12:41:06 juliafruerlund  Aaaaa… Twitt…
 2 282726205 156338620371554304  2012-01-09 11:36:48 juliafruerlund  Jag t… Twitt…
 3 282726205 102007638186917888  2011-08-12 13:24:50 juliafruerlund  @pier… Twitt…
 4 282726205 184675396669800448  2012-03-27 16:17:00 juliafruerlund  Vart … Twitt…
 5 282726205 101655951072968704  2011-08-11 14:07:22 juliafruerlund  @Pier… Twitt…
 6 282726205 133001723018084352  2011-11-06 02:04:16 juliafruerlund  Ojojo… Twitt…
 7 282726205 161516623558230016  2012-01-23 18:32:18 juliafruerlund  Axlar… Twitt…
 8 282726205 160802437140193280  2012-01-21 19:14:23 juliafruerlund  Ser t… Twitt…
 9 282726205 160111930265190400  2012-01-19 21:30:33 juliafruerlund  Åh he… Twitt…
10 282726205 225504831614365696  2012-07-18 08:18:36 juliafruerlund  Utfly… Twitt…
# … with 34,625 more rows, and 84 more variables: display_text_width <dbl>,
#   reply_to_status_id <chr>, reply_to_user_id <chr>,
#   reply_to_screen_name <chr>, is_quote <lgl>, is_retweet <lgl>,
#   favorite_count <int>, retweet_count <int>, quote_count <int>,
#   reply_count <int>, hashtags <list>, symbols <list>, urls_url <list>,
#   urls_t.co <list>, urls_expanded_url <list>, media_url <list>,
#   media_t.co <list>, media_expanded_url <list>, media_type <list>, …
[1] "34635 tweets scraped!"
```



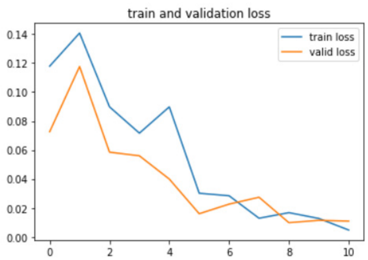
street network and the building with a 30-meter buffer, to find the front street of activity hotspots. Then, a list of points with coordinates are generated with 10-meter intervals on the intersection. As a result, around 1,000 points are selected, generating around 1,000 Google street view images.

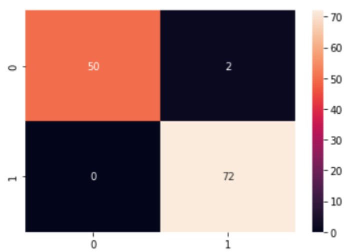
The final part of the project uses visual AI to analyze the acquired street view images. The project uses the method of image classification to see if the machine learning model can be trained to differentiate the images of activity hotspots and others based on their spatial qualities. To train the model, I scraped another 1,500 street view images from Google API with random locations in Helsingborg. The images are divided into train data (700 images from the activity hotspots and 1,100 images from random locations), valuation data (250 images from the hotspots and 300 from random locations), and test data (50 images from the hotspots and 100 from random locations). The model used is pretrained ResNet-18.



index	date	lat	lon	panoID
0	2010-01	56.0433396	12.6931393	CAoSLEFGMVFpcE1WdFY4TFEtYzdXMWV1cVhrcTJSaWdmRXNvZnVqSTNZSVdBaUkx
1	2018-12	56.0432515	12.6931139	CAoSLEFGMVFpcE9BOFdMcGpzbHYwMWZfSVlPem56aWNQa2sxWU1CNnFSNUdLbFI2
2	2022-03	56.02005175973961	12.7220088302747	OsoipIGHR4JzmKkvlck09Q
3	2022-03	56.01996263136059	12.72200119079991	Nx_96K4PQtjNEorMaGW6Xg
4	2022-03	56.0198730787413	12.72199347474167	LoX2YZ4mdfZSToznljcl9w
5	2022-03	56.01978291142786	12.72198584952129	Ug23zCY1o4xygvEWjsWY8Q
6	2022-03	56.01969309921057	12.72197813705556	qDhKDwuZv3uOIDO2Xo021Q
7	2022-03	56.01960328708342	12.72197042460916	sfVmPvAuk7ULl3wWod6kDg
8	2022-03	56.01951279094175	12.72196263913455	QpGhf0UUPkL4gSGdMc0_kQ
9	2022-03	56.01942169306259	12.72195482070998	8cwhNDk30_UVfT_H3d6MxQ
10	2022-03	56.01971267188758	12.72429486309011	BY7fwM9VRuUWkLxW4BilDg
11	2022-03	56.01962476076255	12.7242871592209	RfcAqRA_5n6MpTTVxbbTQg
12	2022-03	56.01944970514693	12.7242718647247	C16H27PUFYVyDqhGV_479g
13	2022-03	56.01935948220107	12.72426394515137	mTZ4hlErC5RVmwSnc82i9w
14	2021-08	56.06544563242271	12.71809856386464	NNJ-xON_rNoid8C2QNfhlg
15	2021-08	56.06552325178424	12.7182432160085	OzhVrElxvuaLcrd56YWr_Q
16	2021-08	56.06560049446065	12.71838898839566	6lPwBzqBpG6MauoPowWTDA
17	2021-08	56.06560049446065	12.71838898839566	6lPwBzqBpG6MauoPowWTDA
18	2021-08	56.06567863792645	12.71853441905611	Ypyg7-NHFwEDO6uR73-e9g
19	2021-08	56.06575946653759	12.71867816273801	QsZU9V0EltD_YKDGimxSPg

```
# plot a test batch of images
show_batch(train_iterator)
```





Opposite (from left to right): Tweet scraped from Twitter API

Building footprint and street network acquired from Helsingborg open data portal

Right (from top to down): Points generated to scrape Google street view images

Data scraped from Google street view API

Train and validation loss during 15 epochs

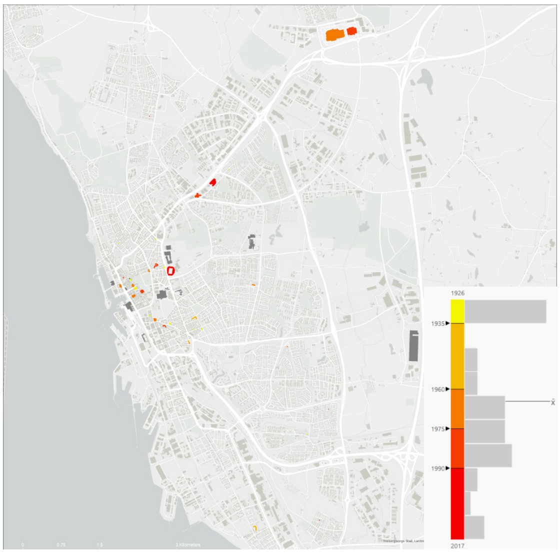
Confusion matrix of the trained model

RESULTS AND DISCUSSION

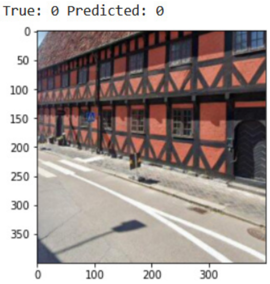
Among the 60 buildings with more than 100 tweet counts, the construction years distribute between 1926 and 2017, with the mean of 1967. The mean footprint of the activity hotspots is 5,693 m^2, far exceeding the average building footprint of 137 m^2 in Helsingborg.

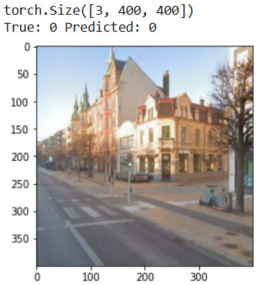
Using visual AI, the model is able to distinguish the street view of urban hotspots and general locations with 98.4% precision, which means that there are distinctive differences between the spatial qualities. When comparing four randomly selected samples from the urban hotspots and four samples from the general places, we can see that urban hotspots generally have more complex facades, including bigger window-to-wall ratio and more variations in the materials. Furthermore, the urban hotspots are often well integrated into the pedestrian network, and have minimum setbacks from the sidewalk. By the contrary, the general locations are often surrounded by green space. Besides, only one image of the general location is confused for the urban hotspots, due to the lack of greenery and predominance of façade.

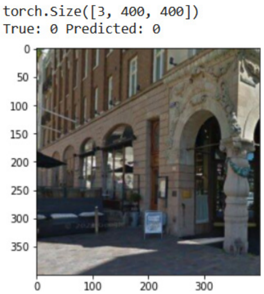
However, the current method has several limitations. The randomly scraped images of general locations are mostly from the suburban area, which are different from the urban center in spatial qualities in many perspectives. The further study would select general locations close to the urban hotspots as comparison. Given the time restrictions, further analysis on the spatial qualities is not conducted. Using image segmentation or object identification in the future will be interesting to identify specific spatial qualities that contribute to the popularity of a place.

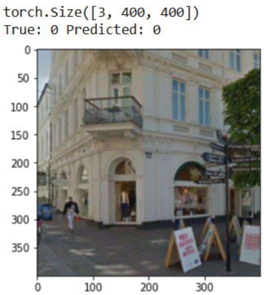


Class A, urban hotspots

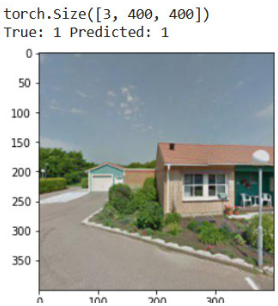


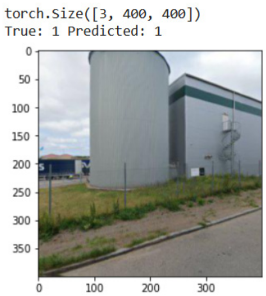


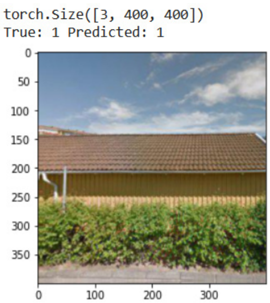


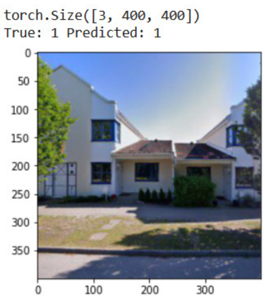


Class B, general places

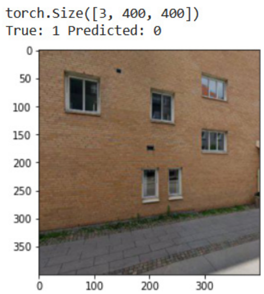








Class B misidentified as class A



Opposite: (up) Activity hotspots with construction year

(down) Activity hotspots with building area

This page: Returned images by the classifier

CONCLUSION

The project shows that there are distinctive differences in spatial qualities between urban hotspots and general locations. However, to understand specific physical characteristics that contribute to higher usage, further analysis of street view images using computer vision is needed. It will also be interesting to apply the trained model to other cities, to understand if the distinction between urban hotspots and general locations is universal. Another potential direction is to scale down the temporal range, to understand what physical characteristics are important among buildings of the same style.

The project shows an example of the preliminary quantitative morphology study for urban designers during the design process. Although design is a subjective process where designers should have their own agencies, conclusions from quantitative analysis can be used as a valuable assistance to determine the morphology of certain projects. On the other hand, the project speculates the visual characteristics attractive to residents in Helsingborg. It will be interesting to validate whether the result reflects the true tastes of people using qualitative study in the future.

In conclusion, distinctive spatial qualities contribute to the high visitation and diverse activities in the space. The project takes a first step to understand these distinctive spatial qualities in the quantitative way, and encourages for further data-driven morphology study to supplement urban design process in the future.