Signs of Gentrification

Analyzing Amsterdam's Storefront Signage with Machine Learning and Street View Imagery

submitted in partial fulfillment for the degree of master of science

Anh Tran
12770698

Master Information Studies
data science
Faculty of Science
University of Amsterdam

Submitted on 30.06.2023

	UvA Supervisor
Title, Name	Tim Alpherts
Affiliation	University of Amsterdam
Email	t.o.l.alpherts@uva.nl

Signs of Gentrification

Analyzing Amsterdam's Storefront Signage with Machine Learning and Street View Imagery

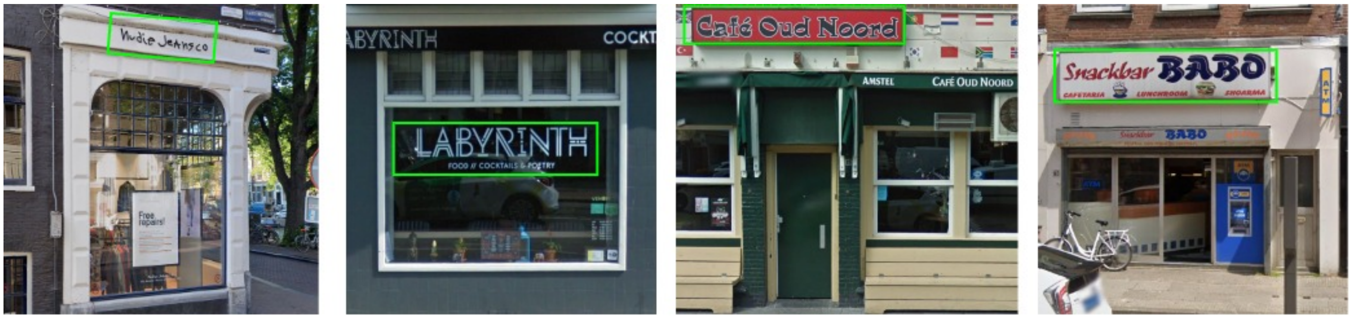


Figure 1: Amsterdam's gentrified (2 left-most images) and non-gentrified (2 right-most images) facades with signage highlighted. It is visible how signage differ in their designs in light of gentrification.

ABSTRACT

Gentrification refers to the process of a neighborhood changing as wealthier residents move in, bringing improvements but displacing existing residents due to rising prices and changing cultures. Studies have pointed out how storefront signage mirrors gentrification: as a communication medium with intentional designs, there are clear differences in signage on gentrified and non-gentrified facades. These qualitative studies, however, were labor-intensive and limited in generalizability, as they were conducted on a few selected neighborhoods, not city-wide. This paper explores the use of computer vision to overcome these limitations, and simultaneously detect gentrification in signage from unseen parts of the city. Trained on street view imagery of Amsterdam, the model learned a city-wide pattern of gentrification, with characteristics similar to what was described in past qualitative research. The model is able to distinguish between gentrified and non-gentrified signage with an F1-score of 0.69. Moreover, the model's output identifies cases where signs did not follow typical patterns - a nuance previous studies did not conclude on. Lastly, the model has the ability to detect the same aesthetics in other areas of the city. While the detection of (non-)gentrification from signage alone are not always accurate, being able to recognize the most prevailing characteristics provides an indication for further investigation.

KEYWORDS

gentrification, storefront signage, street view imagery, computer vision, scene-text detection, classification

GITHUB REPOSITORY

https://github.com/atran13/MSc-Thesis-Signs-of-Gentrification

1 INTRODUCTION

Within urban studies, gentrification is a phenomenon widely discussed. First coined by British sociologist Ruth Glass in 1964 in her work about the inner city of London [1], the term refers to a neighborhood changing as a result of wealthier residents moving in, gradually displacing existing residents as local housing and service prices increase, and cultures homogenized or replaced. Gentrification thus involves an economic and demographic shift, as well as changes in the aesthetics of the built environment. For its negative effects on marginalized communities, it is worthwhile to understand and detect gentrification. This study took a focus on the visual indicators of gentrification, as visual elements are arguably the most telling factor of a neighborhood's cultural identity, demographic and economic characteristics.

Gentrification is a multi-dimensional and multi-step process. Döring and Ulbricht [2] defined gentrification as having 4 aspects: functional (establishment of businesses and cultural institutions), architectural (upgrade to the built environment), social (marginalization, displacement and replacement of existing residents), and symbolic (communication of a new image of the neighborhood to the wider public). As has been noted by Feiereisen and Sassin [3], the functional, architectural, and symbolic aspects can be seen as constituting the visual indicators of gentrification. While these aspects take shape in multiple characteristics of the built environment, storefront signage (hereafter: signage) is a rich communication medium that embodies all three. It is through signage that businesses directly establish their presence, communicate their commercial purposes and values, and distinguish themselves via curated aesthetics [4]. Furthermore, businesses understand the socio-cultural values and identity of the neighborhood, and thus design their appearance to best attract and serve this audience:

> "Shop signs are public texts that communicate what stores sell, who is perceived to be on the street and what their commercial desires are thought to be. [...] Similar to spoken utterances and all written texts, signs are designed for particular audiences [...]. Well-crafted stories are place-making tools inasmuch as they maintain and reproduce prevailing cultural standards and values." [5]

It can thus be expected that once there is a change in the neighborhood - demographically and economically (i.e. gentrification) - signage would mirror this change. Analyzing signage can help understand gentrification, and here lies the interest of this study.

Existing research into the visual indicators of gentrification often analyzes improvements in the neighborhood's physical appearance, and changes in architecture style [6–9]. Comparisons have been drawn in terms of old versus new features, openness of the properties (e.g. boarded up windows, fences), greenery, colors,...; but not as much attention has been paid to signage as a standalone feature. Most research that has been done in this regards are in the context of the US, in which clear distinctions were noted between gentrified and non-gentrified signage [4, 5, 10, 11]. The vast majority of these studies on signage employ qualitative methods, whereby the researchers conduct observational data collection (i.e. manually photographing facades) and summarize what is present in their samples. While their findings provide invaluable insights, their methodologies are undoubtedly labor-intensive. Furthermore, as has been pointed out by Reades et al. [12] and Barton [13], such selection of neighborhoods and facades per neighborhood often comes with limitations in terms of generalizability on a city-wide scale. Conclusions were made about the most differentiating characteristics, but to which extent are these characteristics present in the neighborhood, and in the city? Can it really be assumed that all signage from a gentrified neighborhood look the same? If not - i.e. if there exist non-gentrified storefronts in a gentrified neighborhood - what can be said about the actual state of gentrification, such as in terms of the neighborhood demographic makeup? These are some nuances that existing research has not discussed, presumably due to their selection of data and limiting methodologies.

On the other hand, the availability of street view data and machine learning techniques has led to developments in Urban Visual Intelligence [14], whereby cities' built environments are understood in conjunction with socio-economic circumstances and residents' activities on a large scale. Besides predictive modelling for gentrification [12, 15], work has been done to visually measure gentrification via documenting changes [7], detect [6] and deep-map to reveal gentrifying areas [9]. A small number of machine learning research took a focus on signage, but in terms of linguistic landscape [16, 17], typeface (font type) [18], or to classify points of interest [19, 20], instead of to understand the aesthetics of signage. Systematic literature reviews on Urban Visual Intelligence [14, 21] have noted its importance in generating insights and decision-making, while pointing future research to analyze written languages in images, as well as between-place inference (applying a machine learning model trained with one area to another). This study positions itself in this research gap, where it aims to understand signage aesthetics as a mirror of gentrification in Amsterdam, while utilizing computer vision and street view imagery to overcome limitations of existing urban ethnographic research.

The dataset at the center of this research was from the StreetSwipe project. Using crowdsourcing, StreetSwipe [22] lets people decide whether facades in Amsterdam appear gentrified. With this data, the study drew conclusions based on the subjective and common perception of a diverse group of people - arguably a necessity when it comes to a multi-faceted phenomenon such as gentrification. Moreover, with the images taken from all over the city, the results were not constrained per neighborhood. Understanding the visual state of gentrification would help detect potentially gentrifying areas. As a pilot case, the main goal of the study was to see to which extent a computer vision model can learn the visual perception and, correspondingly, classify signage as perceptually gentrified or non-gentrified. Subsequently, with data from areas of the city that was not covered in StreetSwipe, the model was tested to quantify the extent to which the perception hold against the actual state of the neighborhoods (i.e. given all signage from a gentrified neighborhood (as per census data), how much of the signage would be visually perceived as gentrified?). Lastly, via inspecting the model's classifications, insights were provided into the characteristics of gentrified and non-gentrified signage in Amsterdam, as well as cases where the distinction was less clear.

The research question of this study was stated as follow:
To which extent can a computer vision model learn the differences between perceptually gentrified and non-gentrified storefront signage? How do the learned characteristics generalize to other neighborhoods of the city? Lastly, what are the characteristics that distinguish between gentrified and non-gentrified signage, and which characteristics make the distinction less clear?

(1) Sub-RQ 1: To which extent can the scene text detection model CRAFT identify signage from street view images?
(2) Sub-RQ 2: To which extent can fine-tuned ResNets correctly classify gentrified and non-gentrified StreetSwipe signage?
(3) Sub-RQ 3: How does the model trained on visual perception (StreetSwipe) perform on a test set of data from other neighborhoods, labeled per census data (the extended dataset)?
(4) Sub-RQ 4: What are the characteristics of correctly classified StreetSwipe signage per class with classification probability of 80% and above?
(5) Sub-RQ 5: What are the characteristics of misclassified StreetSwipe signage with high (≥ 80%) and low (50-70%) classification probability?
(6) Sub-RQ 6: What are the characteristics of signage in the extended dataset with classification probability of 80% and above (irregardless of the ground truth)? Compared to the corresponding facades, to which extent are the classifications plausible?

The paper continues with a description of related work, the study's methodology, results, discussion, and conclusion.

2 RELATED WORK

This section first outlines the findings of existing urban ethnographic studies on signage and gentrification, whereby an expectation was formed on the distinguishing characteristics that should be learned by the model. This is followed by a brief discussion on scene text attribute learning as an initial research direction. Next, the state-of-the-art of scene text detection are given for the task of extracting signage from facades. Finally, the computer vision model used in the study is discussed.

2.1 Signage and gentrification

The current body of research about storefront signage aesthetics in relation to gentrification exists largely in the context of the US. While some studies considered multiple features, namely font type (or typeface), colors, text density, language, and meaning, others analyzed elements individually.

Findings from Brooklyn, New York[5, 10, 11] showed that non-gentrified signage typically was text-dense and had larger typeface; names that refer to the location, the owner's name, the type of business, products or services; languages other than English; complementary symbols or images; reference to religion, ethnicity, country of origin, and race. On the contrary, gentrified signage had shorter texts, written in smaller font sizes, lower case letters; more cryptic or ambiguous names; languages other than English that shows sophistication and worldliness. In parallel, a study from Cincinnati, Ohio [4] found that signage became more homogeneous as neighborhoods gentrified (less variation in colors and typeface).

Next to this, there are studies that focused on individual attributes of signage. A popular topic in this regard is the linguistic landscape of a city or neighborhood. Examples could be found in Seoul, South Korea [16]: in the neighborhood where the majority of the Chinese population in Seoul reside, Korean signage had been gradually replaced by Chinese signage, mirroring both the demographic composition and the socio-economic standing of these groups. Similarly, more English signage had appeared in a commercial neighborhood in Phnom Penh, Cambodia, seemingly displacing French as a second language - a trend observed alongside globalization and gentrification [23]. Besides languages, typeface had also been found to be highly correlated with average household income of the corresponding neighborhoods in London [18].

All in all, findings showed clear differences between the characteristics of gentrified and non-gentrified signage. It was therefore hypothesized that the same pattern can be found in Amsterdam's signage, and a computer vision model can be used to distinguish and detect gentrification based on signage.

2.1.1 Scene text attribute learning. Taking inspiration from the aforementioned studies, where differences were found per feature (font, color, semantic, etc.), an exploration in this direction was done for this study. The goal was to learn these specific attributes and quantify their variations among gentrified and non-gentrified signage. However, given the state of the data at hand - namely, the lack of annotations other than the gentrification label - as well as resources available, this research direction was deemed unsuitable. The state-of-the-art approaches are described below.

(1) There exist models that output multiple attributes of interest. Examples are NeurTEx [24] (outputs text bounding box, transcript, and typeface), Fontnet [25] (color and typeface), and TaCo [26] (color and typeface). However, these models are meant for text in graphic design and documents. They are not as reliable when it comes to scene text because of the added noise and distortions of natural images.
(2) Another way is to use individual models for each element:
    - Font: State-of-the-art models include DeepFont [27], HENet [28]. These models perform well on synthetic data (e.g. AdobeVFR [27], VFR-2420 [29]), but are not as robust when it comes to natural images.
    - Text transcripts and semantics: State-of-the-art scene text recognition models include MORAN [30], CRNN [31], and PARSeq [32]. For learning semantics, word embedding models such as FastText [33] or Word2Vec [34] could be used on the transcribed text. However, as was found from text detection (discussed in Section 4.1), some instances were returned as single words or fragments of words, while others were phrases. Transcribing the text and learning semantics from this data would not give consistent results.
    - Color: Text colors could be learnt via creating a color histogram [35]. Instead, by using a convolution neural network (discussed in Section 4.4), colors could be learned, among many other features.

To account for the lack of ground truth labels, a synthetic scene text dataset could have been created for training, before inference was done on StreetSwipe. Gupta et al. [36] offer a method for this that takes into account image segmentation and depth in generating text, with annotations on bounding box, typeface, color, and text transcript as needed. This approach is also non-viable: It requires either mining street view images that have suitable empty spaces to be synthesized on, or using pre-generated images from the authors, which largely include many types of background other than facades. Either method proves impractical - the former is time-consuming, and the latter is a mismatch compared to StreetSwipe.

Not having found a viable approach to learning signage attribute-by-attribute, a more appropriate research direction was devised that involved learning the general visual representation of signage with a convolutional neural network (CNN). Signs would first be extracted using a scene text detection model, and an image classification model would be trained and tested on these signage.

2.2 Scene text detection

Scene text detection involves localizing text in natural images. The challenge in this task stems from the complexity of natural images, with text surrounded by objects, varying in sizes, perspectives, orientations, sometimes curved, obstructed, or blurry.

Benchmark datasets for this task include ICDAR 2013 [37] and 2015 [38], and TotalText [39]. The most widely implemented models are EAST [40] and CRAFT [41]. CRAFT was the best performing model on ICDAR 2013, and surpassed EAST on ICDAR 2015 and TotalText. By calculating character region scores (localizing characters) and affinity scores between characters (grouping characters into sequences), the model returns bounding box coordinates. CRAFT has better accuracy on curved, long, and non-horizontal text compared to EAST. The model is also multilingual - a necessary feature considering signage in the data at hand are (at least) in Dutch, English, Chinese, Arabic, and Korean. CRAFT was thus the model used in this study for extracting signage from facades.

2.3 Image classification

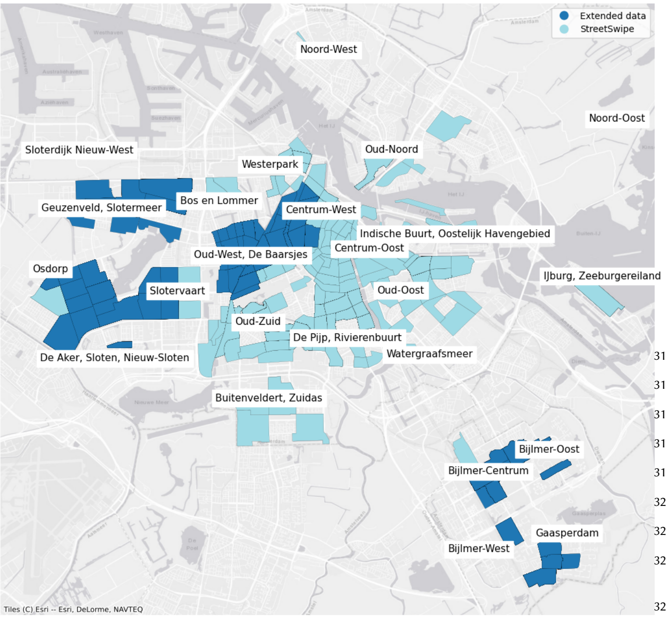
Using an image classification model helps learning features and detecting gentrification in signage. It also enables qualitative analysis into signage aesthetics as they mirror gentrification, via inspecting what the model correctly identifies to be (non-)gentrified (correct classifications), and which characteristics lead to the model failing to differentiate between classes (misclassifications).

The Residual Network (ResNet) [42], with the use of skip connections in its architecture, had enabled deeper networks to learn more efficiently, without the problem of vanishing or exploding gradients. Fine-tuned ResNet18 and ResNet50 were chosen as candidate models for classifying signage, initialized with pre-trained weights from ImageNet. After achieving satisfactory performance, the model's classification outputs were analyzed.

3 METHODOLOGY

3.1 Data

This study used two datasets: the main dataset from StreetSwipe [22], and an extended dataset of other areas in Amsterdam. An illustration of the areas covered by each dataset can be seen in Figure 2.



Tiles (C) Esri -- Esri, DeLorme, NAVTEQ

Figure 2: Areas of Amsterdam present in StreetSwipe and the extended dataset. StreetSwipe had images mainly from the city center (Centrum-Oost and -West), Oost, and Zuid. The extended dataset had images taken from Centrum-West (Jordaan), Oud-West, De Baarsjes, Nieuw-West, and Zuidoost.

3.1.1 *StreetSwipe.*

StreetSwipe's images were originally from Google Street View, which were cropped to focus on facades (examples are in Figure 3. Using crowd-sourcing, the project lets people decide whether each facade appear gentrified, by voting "Yes" or "No" on the images. The official *Gentrified* and *Non-gentrified* label per facade are based on what the majority voted for. Thus, this dataset represents common visual perception of gentrification.

Since there were two versions of StreetSwipe, the data retrieved existed in two sets: 1912 higher resolution images from the older version, and 529 lower resolution images from the new one. The dataset thus had 2441 images in total, each with its numbers of "Yes" and "No" votes, and metadata on the facade's coordinates and street name. The new version's images also had more detailed address, name, and type of business/services. There were also more votes in the new version than in the old one. The images from the old version were available directly, while the new ones were provided via URLs to a Google APIs bucket, and thus were scraped.

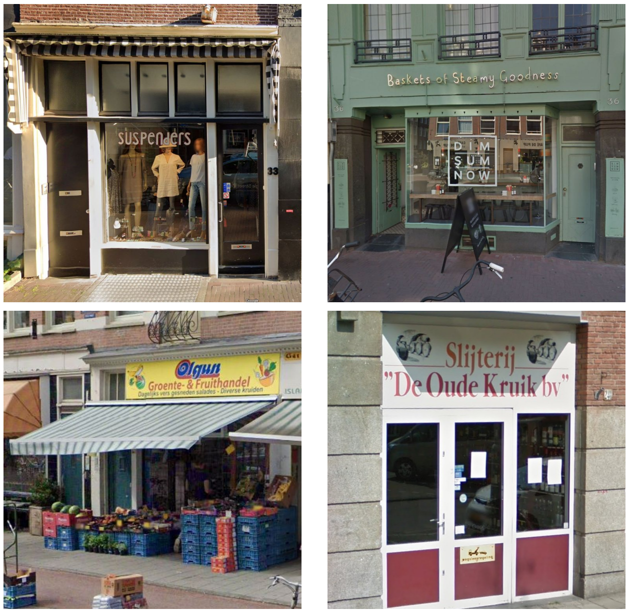


Figure 3: Examples of StreetSwipe images.

Feature engineering was done to create the gentrified/ non-gentrified labels, by taking the vote higher in volume. The images were then re-grouped per their corresponding labels. There was class imbalance in the data, with 71% of the images being non-gentrified (1731 images; versus 710 gentrified images). The images had quite consistent aspect ratios of approximately 1:1; however, they varied in size, ranging from around 300x300 to 1700x1700, with one outlier of size 2500x1300, approximately.

3.1.2 *Extended data.*

Data from areas not present in StreetSwipe was used to test the model's generalizability. Given literature on gentrification in Amsterdam, gentrified and non-gentrified neighborhoods were selected, and street view images from these areas were retrieved from a dataset made available by the Civic AI Lab.

Using Dutch resident register data, Hochstenbach and Van Gent [43] identified gentrification in Amsterdam's neighborhoods via residential mobility (characteristics of neighborhood's in-movers and out-movers), social mobility (change in current residents' socio-economic situation), and neighborhood's demographic changes (age, income) over the 2004-11 period. Following these indicators, downgrading neighborhoods (i.e. declined median income) were found to be in Zuidoost and Nieuw-West; and no trend of social mobility and displacement was found in these areas. Upgrading neighborhoods (i.e. increased median income) were largely in Oud-West, De Baarsjes and Centrum-West - particularly De Jordaan; and these areas underwent social mobility and displacement (see Appendix A for visualizations of these findings). Thus, Zuidoost and Nieuw-West were selected as non-gentrified areas; and De Jordaan, Oud-West, and De Baarsjes were selected as gentrified areas. Images taken from these neighborhoods were labeled accordingly. This dataset thus represents gentrification according to census data (regardless of how facades might be visually perceived).

The dataset consisted of street-view panoramas taken at certain coordinates along the streets of Amsterdam. The panoramas had

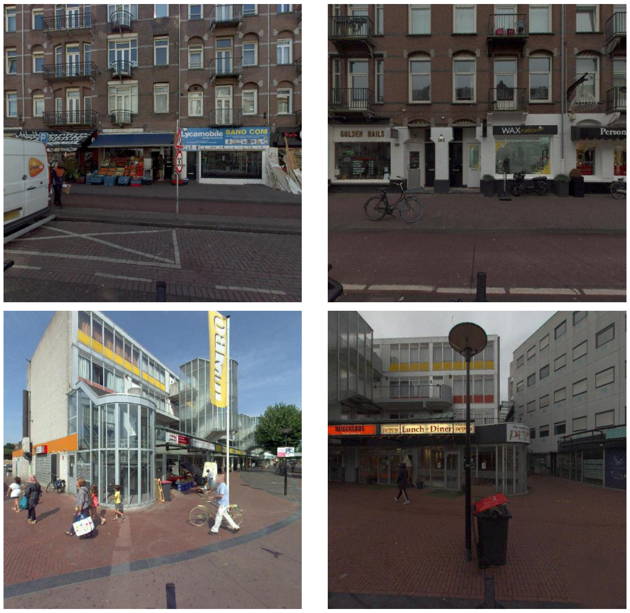


Figure 4: Examples of left- and right-view images from the extended dataset.

corresponding front, back, left, and right views of the vehicle already extracted (Appendix B shows an example of the full shape of the data). Per location, panoramas were taken annually; however, with the scope of the current study, only the latest ones were retrieved. Furthermore, to ensure that the data contained mainly facades with signage, coordinates of shops, restaurants, cultural venues, etc. in the selected neighborhoods were queried from OpenStreetMap, and panoramas within a small radius from these points of interest were retrieved.

Experiments were done with the text detection model to determine which version of the images would return signage with the least amount of noise. It was found that passing the panoramas through the model returned the most noise (e.g. rows of windows, traffic signs), while the left and right views of the vehicle returned the least noise. Therefore, the left- and right-view images are used in the study. Figure 4 shows some example images. In total, this dataset contains 9340 images, with 5374 gentrified images (58%) and 3966 non-gentrified (42%). All images have size 512x512.

3.2 Experimental setup

The data pipeline is visualized in Figure 5. This section outlines the implementations of CRAFT for scene text detection and the ResNets for classification.

3.2.1 CRAFT. The first task was to detect and extract signage in the images of both datasets, using the pre-trained CRAFT model via the EasyOCR Python package [44]. More specifically, the *detect* method was used, with text confidence threshold *(text_threshold)* set to 0.75, bounding box extension *(add_margin)* set to 0, and all other parameters set to their default values. Text instances were cropped out by their bounding boxes and grouped per class.

StreetSwipe returned 10079 instances in total, with 2610 gentrified instances (26%) and 7469 non-gentrified instances (74%). The instances varied in size, typically with width larger than height. The widths ranged from 8 to 1576, and heights ranged from 6 to 795 (see Appendix C - Figure 15 for a visualization of the distributions). These text instances were used to train and optimize the classifier.

The extended dataset returned 2473 instances in total, with 1633 gentrified instances (66.03%) and 840 non-gentrified instances (33.97%). The instances varied in size, typically with width larger than height. The widths ranged from 8 to 510, and heights ranged from 4 to 191 (see Appendix C - Figure 16 for a visualization of the distributions). These text instances were used to test the classifier for generalizability.

3.2.2 ResNet. StreetSwipe text instances were split into training, validation and test sets with 80:10:10 ratio. The data was randomly shuffled prior to splitting to maintain class distribution per split. Table 1 shows the size of each split as well as class distributions.

	Total	Gentrified	Non-gentrified
Train	8063	2092 (25.95%)	5971 (74.05%)
Val	1008	238 (23.61%)	770 (76.39%)
Test	1008	280 (27.78%)	728 (72.22%)

Table 1: Sample size per data split, per class

Images in the training set were randomly cropped and resized to 224x224, plus a random horizontal flip. Validation and test set images were resized and center-cropped to 224x224. All data was normalized with ImageNet's mean and standard deviation.

Model training was done in PyTorch. Both the ResNet18 and ResNet50 were initialized with pre-trained ImageNet1K-V1 weights. As a baseline, all the weights in the model were frozen and only the final layer was optimized, and no action was taken to account for class imbalance. Learning rate was set to 0.001, batch size was 32, and the model was trained for 50 epochs.

Next, the models were fine-tuned. PyTorch's WeightedRandomSampling was applied, with class weights calculated as 1/(class size). The loss function used was the cross entropy loss. The optimizer was stochastic gradient descent with a 0.9 momentum. StepLR learning rate decay scheduler was also implemented with a step size of 7 and gamma of 0.1. Hyperparameters tuning was done on the learning rate, batch size and number of training epochs.

3.2.3 Visual inspection of model's outputs. Random samples were taken from each of the following subsets of the output:

- StreetSwipe's correctly classified signage per class with classification probability of 80% and above. This showed the most distinguishing characteristics of gentrified and non-gentrified signage. In comparison to past research (Section 2.1), conclusions were drawn on the extent to which the model could achieve similar results as qualitative analyses.
- StreetSwipe's misclassified signage, categorized by high (≥ 80%) and low (50-70%) classification probability. Misclassifications showed characteristics of cases that the model failed to distinguish. This demonstrated the extent to which signage alone could fully determine (non-)gentrification to a computer vision model. It was expected that the varying degree of certainty would correspond to different visual

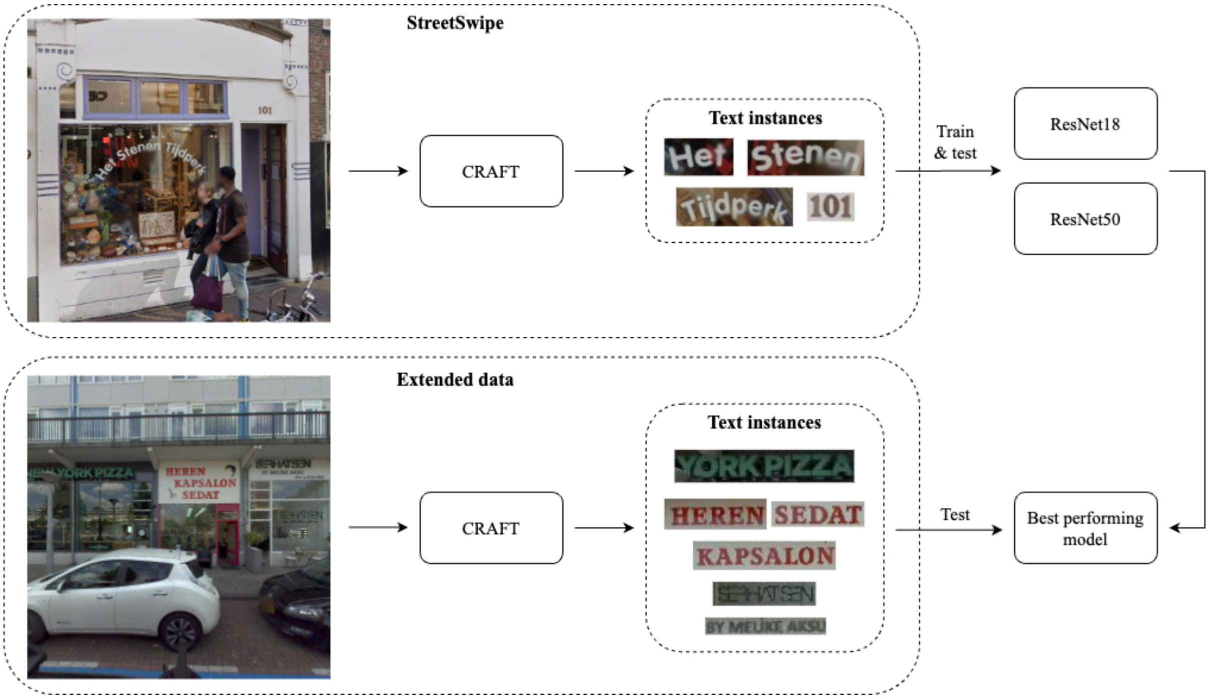


Figure 5: Pipeline: Images of facades were first passed through the text detection model CRAFT to extract text instances. The ResNets were then trained and tested on instances from StreetSwipe. Subsequently, the best performing model was tested on instances from the extended data.

patterns: misclassifications with high certainty would most likely have the same characteristics as the class they were assigned to, and misclassifications with low certainty would have less typical characteristics.

- Extended data's signage per class with classification probability of 80% and above (irrespective of ground truth labels). It was expected that classifications follow the same characteristics in StreetSwipe, and by inspecting the corresponding facades we can get an idea of how well the model can detect perceptual gentrification via signage in unseen data.

3.3 Evaluation

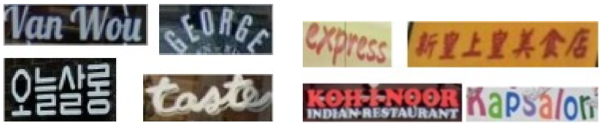
3.3.1 Scene text detection. Since there was no bounding box annotations of the signage, evaluation was done by visual inspection.

3.3.2 Classification. Accuracy, precision, recall, and F1 score were calculated for the classifier. The macro-averaged metrics were used to compare models, and additionally metrics per class were calculated to examine the effect of class imbalance. The metrics were implemented using torchmetrics' MulticlassAccuracy, MulticlassPrecision, MulticlassRecall, and MulticlassF1Score. torchmetrics' ClasswiseWrapper was used to obtain metrics per class.

4 RESULTS

4.1 Scene text detection

Some example signage instances per class can be seen in Figure 6.



(a) Gentrified **(b) Non-gentrified**

Figure 6: Signage instances examples.

It was found that the text detection model returned almost all signage instances present in the original street view images, including signage with non-horizontal and curved text. Cases where the model failed include very small, and therefore illegible texts, especially in lower resolution images. Additionally, the model also returned some noise, namely texts on street signs (e.g. traffic signs, street names), and watermarks ("©Google", as the images in StreetSwipe were originally from Google Street View) - these instances were manually removed.

4.2 Classification

4.2.1 Baseline. Test set results of the baseline model are shown in Table 2.

4.2.2 Fine-tuned. The best performing model with ResNet18 architecture was found with learning rate set to 0.001, batch size 32, and 60 training epochs. The best performing model with ResNet50 architecture was found with learning rate set to 0.01, batch size 64, and 60 training epochs. The macro-averaged metrics for these

Metrics	Class	Baseline model	
		Classwise	Average
Accuracy	Gentrified	0.2143	0.5810
	Non-gentrified	0.9478	
Precision	Gentrified	0.6122	0.6852
	Non-gentrified	0.7582	
Recall	Gentrified	0.2143	0.5810
	Non-gentrified	0.9478	
F1-score	Gentrified	0.3175	0.5800
	Non-gentrified	0.8425	

Table 2: Classwise and macro-averaged test set metrics of baseline model (ResNet50, no weighted sampling, no fine-tuning). This model achieved very high performance for non-gentrified signage, but performed poorly on gentrified signage, due to class imbalance.

models on the validation and test sets are shown in Table 3. The fine-tuned ResNet50 had better performance, its classwise metrics are shown in Table 4.

Metrics	ResNet18		ResNet50	
	Val	Test	Val	Test
Accuracy	0.6497	0.6960	0.6506	**0.7033**
Precision	0.6185	0.6715	0.6209	**0.6795**
Recall	0.6497	0.6960	0.6506	**0.7033**
F1-score	0.6222	0.6781	0.6256	**0.6865**

Table 3: Macro-averaged metrics of fine-tuned ResNet18 and ResNet50 on the validation and test set. Between these two models, the fine-tuned ResNet50 performed better.

Metrics	Class	ResNet50	
		Val	Test
Accuracy	Gentrified	0.5714	0.6429
	Non-gentrified	0.7299	0.7637
Precision	Gentrified	0.3953	0.5114
	Non-gentrified	0.8464	0.8476
Recall	Gentrified	0.5714	0.6429
	Non-gentrified	0.7299	0.7637
F1-score	Gentrified	0.4674	**0.5696**
	Non-gentrified	0.7838	**0.8035**

Table 4: Classwise metrics of the best model. Even though there was an improvement in classifying gentrified signage compared to the baseline model, this model still performed better for non-gentrified signage, as shown in the F1-scores of each class.

4.3 Testing on the extended data

The average and classwise metrics of the best model in classifying the extended data are presented in Table 5.

Metrics	Class	ResNet50	
		Classwise	Average
Accuracy	Gentrified	0.5340	0.5807
	Non-gentrified	0.6274	
Precision	Gentrified	0.7359	0.5725
	Non-gentrified	0.4092	
Recall	Gentrified	0.5340	0.5807
	Non-gentrified	0.6274	
F1-score	Gentrified	0.6189	0.5571
	Non-gentrified	0.4953	

Table 5: Macro-averaged and classwise metrics of the best model on the extended data. Compared to the metrics on StreetSwipe's test set (Table 3), there was a decrease of approximately 10% in all average metrics.

4.4 Inspecting model's output

4.4.1 Correct classifications. StreetSwipe's correctly classified signage showed the most typical and distinguishing characteristics of signage per class, which can be seen in Figure 7.

Gentrified signage were more similar in font types (more modern and minimal fonts) and often did not vary in text sizes, while non-gentrified signage used more types of fonts (more classic and decorated fonts), sometimes more than one font on a single sign, and sometimes with varying text sizes and orientations. In addition, gentrified signs mostly had white texts, with minimal variation in background colors. Non-gentrified signs were the opposite: text colors and background colors varied more; plus a notable usage of neon signage. In terms of languages, besides Dutch, gentrified signage had more English text, with very rare appearances of non-Latin languages (e.g. Korean); while non-gentrified ones were largely in Dutch, with appearances of English, Chinese and Arabic. And finally, although out of scope of the study, the model also picked up graffiti as text instances belonging to non-gentrified facades.

4.4.2 Incorrect classifications. StreetSwipe's misclassified signage showed characteristics of cases the model failed to distinguish. The misclassified instances with high certainty showed similar characteristics to correctly classified instances (gentrified: minimal text fonts and colors, less variation in font styles and background colors; non-gentrified: more variation in fonts, text colors and background colors). On the other hand, mis-classified instances with lower certainty showed a more nuanced picture. Signal strength diminished and the model's outputs of the two classes were more or less indistinguishable. There were similar variations in font styles, text colors and background colors, without any characteristic that stood out. Results can be seen in Figure 8.

4.4.3 Classifications on extended data. On the extended data, signage classified by the model followed the same patterns per class as learned from StreetSwipe; however, visual inspection of the corresponding facades showed not all classifications were true to human perception. Results can be seen in Figure 9.

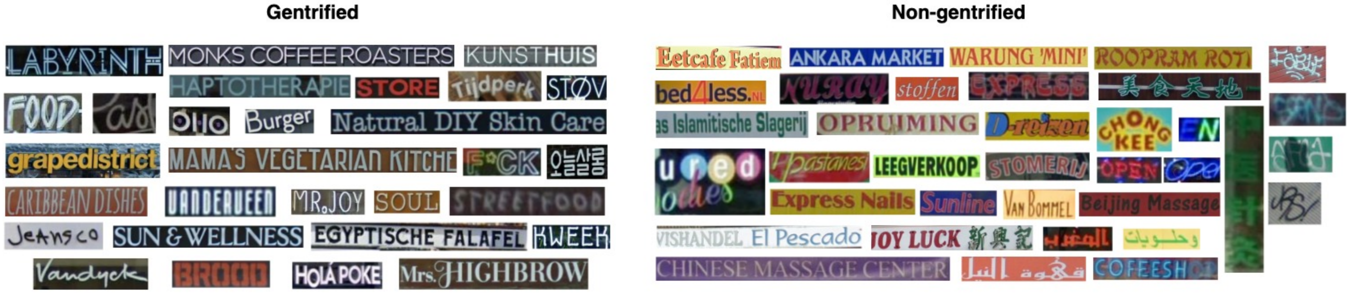


Figure 7: StreetSwipe's correctly classified signage per class with probability of 80% and above. Note how non-gentrified signage varied more in their characteristics (more font types, colors, and languages) while gentrified signage appeared more homogenized.

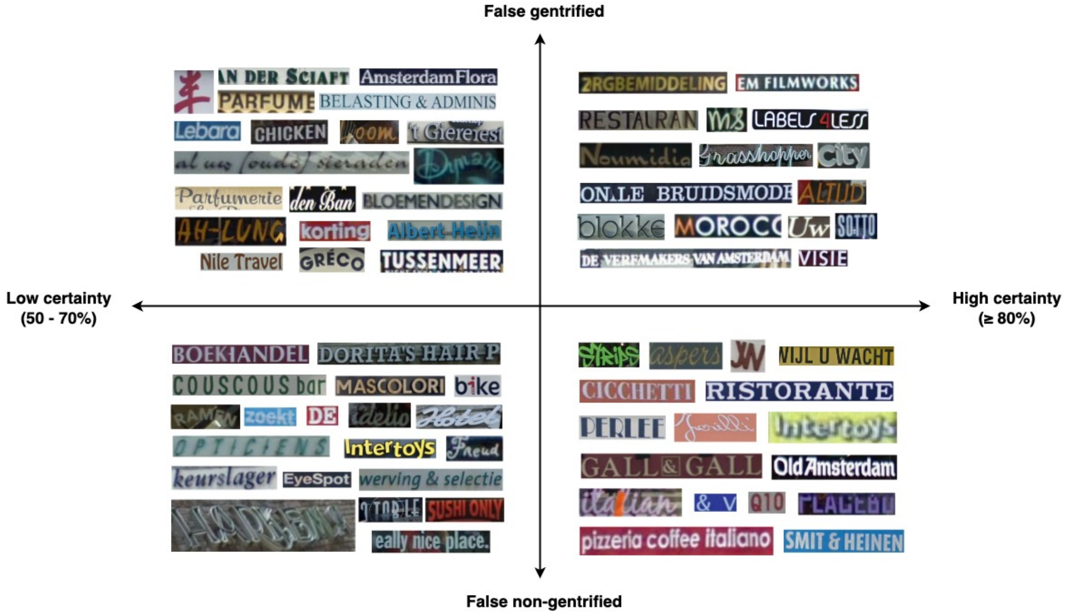


Figure 8: StreetSwipe's misclassified signage per class, grouped by high and low classification certainty. Incorrect classifications with high certainty from both classes generally had the same characteristics as correctly classified instances. As classification certainty decreased, variations in fonts and colors were no longer distinctive across the two classes.

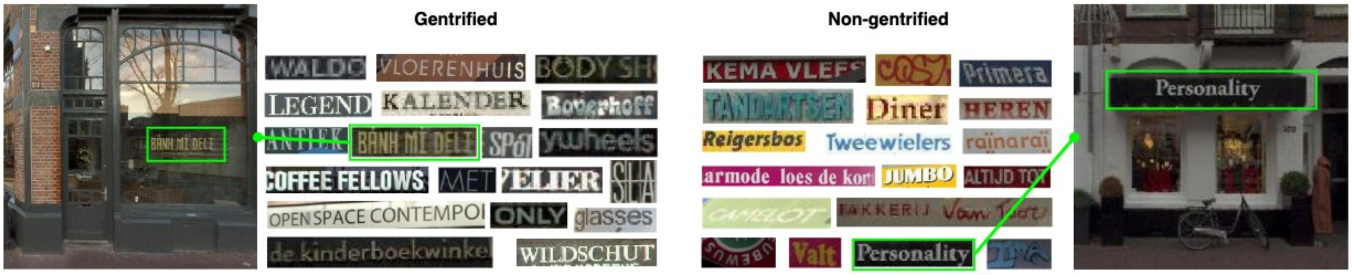


Figure 9: Model's classifications on the extended data's signage with classification probability of 80% and above (disregarding ground truth label). While the model classified signage based on the same typical class characteristics as in StreetSwipe, classifications were not always plausible: Bánh Mì Deli (left) and Personality (right) were gentrified-looking facades, but Personality was misclassified as non-gentrified via its signage.[9]

5 DISCUSSION

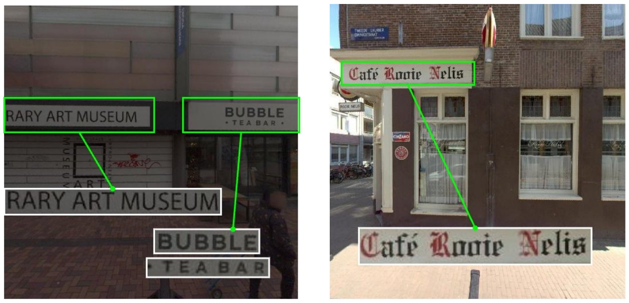
5.1 Summarizing results

This research found that a fine-tuned ResNet50 was able to classify gentrified and non-gentrified signage with an F1-score of 0.69. The model was able to find the same patterns of gentrification in Amsterdam signage characteristics as previous visual qualitative research, in terms of colors and fonts [4, 5, 10, 11], as well as languages [5, 23].

Using an image classification model also enabled the study to make conclusions on fuzzy cases, via inspecting the model's misclassified instances. On one hand, there were signs that had the defining characteristics of the opposite class, and thus were identified as such by the model with high certainty. On the other hand, as classification certainty decreased, we were able to see signage with appearances untypical of any class. Variations in fonts and colors were no longer differentiating the classes. Such output showed more nuances in identifying gentrification via signage. Signage alone could not always tell the full story, as signage of non-gentrified facades can still appear gentrified and vice versa, with consistent visual signals found by the model. Especially for the cases of lower classification certainty, the model failed to clearly tell apart gentrified from non-gentrified signs as signals from both classes appeared simultaneously. While we were able to see what signs looked like beyond the most typical cases of (non-)gentrification (something previous studies did not point out), ultimately, these were the cases that the model failed to distinguish due to added nuances.

Utilizing street view imagery and computer vision also enabled this study to overcome the limitations of past research in terms of generalizability. Past studies, in using labor-intensive manual data collection and qualitative methods, could only make conclusions on a neighborhood scale [12, 13]. Generalizability of the model was tested on a set of extended street view data, label as gentrified and non-gentrified based on census data. Two analyses were done on this set. Firstly, looking at the model's performance, there was a consistent decrease in performance of approximately 10% in all average metrics. This suggests that given a gentrified neighborhood, not all signage in the neighborhood would be visually perceived as gentrified. Secondly, looking at the model's inference on this data, the same visual patterns were found for both classes. This supports the model's ability to identify the most typical cases of visually (non-)gentrified signs from unseen data on a city-wide scale. An example of detecting gentrification can be seen in Figure 10.

Nonetheless, it is also worth noting that not all detections are reasonable per human perception. Keeping in mind the model's misclassifications on StreetSwipe due to added nuances, we can see some examples of incorrect detection in Figure 11. While these are signage of long-established businesses in currently non-gentrified neighborhoods, all but one of them were detected as gentrified.



(a) Gentrified: A contemporary art museum & bubble tea store in the non-gentrified neighborhood of Amsterdamse Poort.

(b) Non-gentrified: A classic and reputable Dutch brown cafe in the gentrified neighborhood of De Jordaan.

Figure 10: Examples of plausible detection.

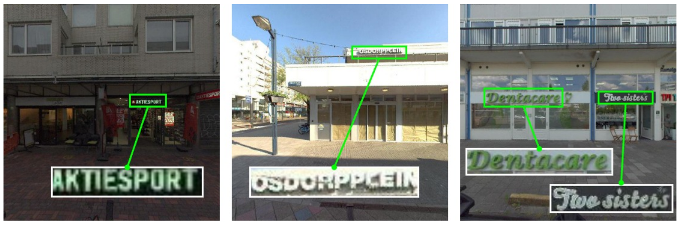


Figure 11: Examples of implausible detection: A sports clothing store in non-gentrified Amsterdamse Poort (left), and a shopping center, dental clinic, and hair salon in non-gentrified Osdorp (middle & right). With the exception of Dentacare (right), all other signage were classified as gentrified.

5.2 Limitations

5.2.1 Data-related limitations. StreetSwipe had the following limitations:

- The number of votes in the pre-july 2020 version of the data was less than in the newer version. This could have lead to lower validity of the results.
- There was strong class imbalance, with 75% of the data belonging to non-gentrified instances. This affected the model's performance in detecting gentrified instances.
- The data had varying dates, with some images dating back to 2009. While we technically could still learn perception, as the images were human-annotated in recent years, the results - especially conclusions made about visual characteristics of signage - should not be interpreted as fully representing the most up-to-date state of gentrification in Amsterdam.

The extended dataset had a limitation in terms of resolution: Text instances generally had much lower resolutions compared to StreetSwipe, and this is because the street view images were taken from a further distance from facades. This could have contributed to model's lower performance on this set.

5.2.2 Methodology-related limitations. The methodology had the following limitations:

- Cropping out text instances meant losing information on the entirety of the signage, such as where the text are placed (windows or above the stores' entrances, or standees, posters etc.), text density on signage, text size, total numbers of font types and colors used, which signage type and whether a combination of types (above entrance, on window, standee, neon, backlit) was used. These elements could have served as extra signals for better classification of signage.
- Cropping out text instances also led to lower reliability for semantic learning. Learning semantic meanings of text could have helped distinguishing between the classes, especially when instances have visual characteristics of the opposite class.

5.3 Future research

Some avenues for future studies are listed below:

- Via object recognition, future studies can include more visual signals by identifying signage in their entirety, and not as (potentially fragmented) text instances. Then, text placement, text density, signage types etc. could be used as features. Further, this would also enable more reliability for semantic learning.
- Other features could be combined to aid model accuracy. Other visual indicators include the appearance of the rest of the buildings, what else appeared in the store facade other than signage (e.g. window displays, outdoor seating). Non-visual indicators include types of business, locations, neighborhood socio-economic indicators such as housing prices, residents' age, education level, income, etc. Similar to semantics, these features can further improve classification, especially when signage visual signals are a misdirection for the model.
- The mismatch in model performance on StreetSwipe and the extended data - or between visual and socio-economic gentrification - points to questions about the neighborhood's demographic makeup. It could be the case that displacement did not happen to old residents, or it did happen but business owners were able to cater to new residents and remained in the gentrified neighborhoods. Indeed, the relationship between gentrification and displacement was found to be more complicated - neighborhood change does not always mean displacement [43]. A research in this direction calls for combining socio-demographic data in combination with street view imagery. A step further would be to incorporate time series data to enable an analysis into the process of neighborhood change, e.g. whether displacement took place over time, and whether/how aesthetics of the built environment reflected change.

6 CONCLUSION

This study set out to study storefront signage aesthetics in relation to gentrification, utilizing street view image data and computer vision for a larger-scale analysis and better generalizability. By fine-tuning ResNets on the StreetSwipe dataset and analyzing the model output, conclusions to the research questions are as follow:

(1) The scene-text detection model CRAFT was able to detect signage with reasonable accuracy and completeness, though noise remains in the form of unwanted text (e.g. traffic signs).
(2) A fine-tuned ResNet50 was able to learn signage features and classify them with a macro-average F1 score of 0.69.
(3) There was a gap of approximately 10% between the model's performance on Street-Swipe and on extended data of other gentrified/non-gentrified neighborhoods. This suggested that "statistically" gentrified did not entirely mean visually gentrified.
(4) Characteristics of signage correctly classified by the model were in-line with findings of previous research: gentrified signage looked more homogenized, and were more likely to be in English, while non-gentrified signs varied more in fonts and colors, and were largely in Dutch, with more appearances of other languages.
(5) Mis-classified signage with high certainty followed the styles of the class they were assigned to. As classification certainty decreased, the distinction diminished.
(6) Signage classifications in the extended dataset follow the same characteristics learned from StreetSwipe. However, inspection of corresponding facades showed that the model was not always giving a plausible detection.

In general, it was found that a computer vision model could do a reasonable job at identifying typical characteristics of gentrified and non-gentrified signage. Further, the model is able to detect the same characteristics on new data, and point out the nuance between the visual and socio-economic states of gentrification. Nonetheless, there were cases where the model failed to detect gentrification based on signage alone. When signage adopted characteristics of the opposite class, perception of gentrification most likely depended on other visual and non-visual features of the facade. While this study was able to show the potential of computer vision in studying signage gentrification, further work is needed in this regard. Expanding the feature space beyond visual signals can be promising in aiding computer vision models to understand the nuances of human perception.

7 ACKNOWLEDGEMENTS

A special thanks to Tim Alpherts for supervising this thesis, whose endless insights and encouragements made for a fruitful and rewarding project, a very fun time, and a tremendous growth in my confidence. I would like to also thank my family and friends, especially Angelo and Thao, for always being my biggest supporters.

REFERENCES

[1] Ruth Glass. *London: Aspects of change.* MacGibbon amp; Kee, 1964.
[2] Christian Döring and Klaus Ulbricht. *Gentrification Hotspots and Displacement in Berlin A Quantitative Analysis*, page 9–35. Springer VS, 2018.
[3] Florence Feiereisen and Erin Sassin. Sounding Out the Symptoms of Gentrification in Berlin. *Resonance*, 2(1):27–51, March 2021.
[4] Muhammad Nafisur Rahman and Vikas Mehta. Signage Form and Character: a window to neighborhood visual identity. *Interdisciplinary Journal of Signage and Wayfinding*, 4(1), 2020.
[5] Shonna Trinch and Edward Snajdr. What the signs say: Gentrification and the disappearance of capitalism without distinction in Brooklyn. *Journal of Sociolinguistics*, 21(1):64–89, 2017. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/josl.12212.
[6] Tianyuan Huang, Timothy Dai, Zhecheng Wang, Hesu Yoon, Hao Sheng, Andrew Y. Ng, Ram Rajagopal, and Jackelyn Hwang. Detecting Neighborhood

Gentrification at Scale via Street-level Visual Data. *2022 IEEE International Conference on Big Data (Big Data)*, December 2022. Conference Name: 2022 IEEE International Conference on Big Data (Big Data) ISBN: 9781665480451 Place: Osaka, Japan Publisher: IEEE.

[7] Evelyn Dawn Ravuri. A Google Street View analysis of gentrification: a case study of one census tract in Northside, Cincinnati, USA. *GeoJournal*, 87(4):3043–3063, August 2022.

[8] Nikhil Naik, Scott Duke Kominers, Ramesh Raskar, Edward L. Glaeser, and César A. Hidalgo. Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences of the United States of America*, 114(29):7571–7576, July 2017.

[9] Lazar Ilic, M. Sawada, and Amaury Zarzelli. Deep mapping gentrification in a large Canadian city using deep learning and Google Street View. *PLOS ONE*, 14(3):e0212814, 2019. Publisher: Public Library of Science.

[10] Edward Snajdr and Shonna Trinch. Old School Rules: Generative Openness in the Texts of Historical Brooklyn Retail Signage. *Interdisciplinary Journal of Signage and Wayfinding*, 2(2):12–29, July 2018. Number: 2.

[11] Edward Snajdr and Shonna Trinch. To preserve and to protect vanishing signs: activism through art, ethnography, and linguistics in a gentrifying city. *Social Semiotics*, 32(4):502–524, August 2022. Publisher: Routledge _eprint: https://doi.org/10.1080/10350330.2022.2114728.

[12] Jonathan Reades, Jordan De Souza, and Phil Hubbard. Understanding urban gentrification through machine learning. *Urban Studies*, 56(5):922–942, April 2019. Publisher: SAGE Publications Ltd.

[13] Michael Barton. An exploration of the importance of the strategy used to identify gentrification. *Urban Studies*, 53(1):92–111, January 2016. Publisher: SAGE Publications Ltd.

[14] Fan Zhanga, Arianna Salazar Mirandaa, Fábio Duarte, Lawrence Vale, Gary Hack, Yu Liu, Michael Batty, and Carlo Ratti. Urban Visual Intelligence: Studying Cities with AI and Street-level Imagery. 2023. Publisher: arXiv Version Number: 1.

[15] William Thackway, Matthew Kok Ming Ng, Chyi Lin Lee, and Christopher Pettit. Building a predictive machine learning model of gentrification in Sydney. December 2021.

[16] Seong-Yun Hong. Linguistic Landscapes on Street-Level Images. *ISPRS International Journal of Geo-Information*, 9(1):57, January 2020. Number: 1 Publisher: Multidisciplinary Digital Publishing Institute.

[17] You Xuan Thung, Tom Benson, and Nikita Klimenko. Detecting languages in streetscapes using deep convolutional neural networks. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1711–1718, October 2022.

[18] Ruixian Ma, Wei Wang, Fan Zhang, Kyuha Shim, and Carlo Ratti. Typeface Reveals Spatial Economical Patterns. *Scientific Reports*, 9(1):15946, November 2019. Number: 1 Publisher: Nature Publishing Group.

[19] Shahin Sharifi Noorian, Sihang Qiu, Achilleas Psyllidis, Alessandro Bozzon, and Geert-Jan Houben. Detecting, Classifying, and Mapping Retail Storefronts Using Street-level Imagery. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*, pages 495–501, Dublin Ireland, June 2020. ACM.

[20] Shahin Sharifi Noorian, Achilleas Psyllidis, and Alessandro Bozzon. ST-Sem: A Multimodal Method for Points-of-Interest Classification Using Street-Level Imagery. In Maxim Bakaev, Flavius Frasincar, and In-Young Ko, editors, *Web Engineering*, volume 11496, pages 32–46. Springer International Publishing, Cham, 2019. Series Title: Lecture Notes in Computer Science.

[21] Filip Biljecki and Koichi Ito. Street view imagery in urban analytics and GIS: A review.

[22] StreetSwipe - About.

[23] Luanga Adrien Kasanga. Mapping the linguistic landscape of a commercial neighbourhood in Central Phnom Penh. *Journal of Multilingual and Multicultural Development*, 33(6):553–567, October 2012. Publisher: Routledge _eprint: https://doi.org/10.1080/01434632.2012.683529.

[24] Vinay Aggarwal, Praneetha Vaddamanu, Bhanu Prakash Reddy Guda, Balaji Vasan Srinivasan, Niyati Chhaya, and Vishwa Vinay. NeurTEx: A Neural Framework for Template Extraction from Flat Images. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7, New Orleans LA USA, April 2022. ACM.

[25] Rakshith S, Rishabh Khurana, Vibhav Agarwal, Jayesh Rajkumar Vachhani, and Guggilla Bhanodai. Fontnet: On-Device Font Understanding and Prediction Pipeline. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2155–2159, June 2021. ISSN: 2379-190X.

[26] Chang Nie, Yiqing Hu, Yanqiu Qu, Hao Liu, Deqiang Jiang, and Bo Ren. TaCo: Textual Attribute Recognition via Contrastive Learning, August 2022. arXiv:2208.10180 [cs].

[27] Zhangyang Wang, Jianchao Yang, Hailin Jin, Eli Shechtman, Aseem Agarwala, Jonathan Brandt, and Thomas S. Huang. DeepFont: Identify Your Font from An Image, July 2015. arXiv:1507.03196 [cs] version: 1.

[28] Jingchao Chen, Shiyi Mu, Shugong Xu, and Youdong Ding. HENet: Forcing a Network to Think More for Font Recognition, October 2021. arXiv:2110.10872 [cs] version: 1.

[29] Guang Chen, Jianchao Yang, Hailin Jin, Jonathan Brandt, Eli Shechtman, Aseem Agarwala, and Tony X. Han. Large-Scale Visual Font Recognition. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 3598–3605, USA, June 2014. IEEE Computer Society.

[30] Canjie Luo, Lianwen Jin, and Zenghui Sun. A Multi-Object Rectified Attention Network for Scene Text Recognition, January 2019. arXiv:1901.03003 [cs].

[31] Baoguang Shi, Xiang Bai, and Cong Yao. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition, July 2015. arXiv:1507.05717 [cs] version: 1.

[32] Darwin Bautista and Rowel Atienza. Scene Text Recognition with Permuted Autoregressive Sequence Models, July 2022. arXiv:2207.06966 [cs] version: 1.

[33] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching Word Vectors with Subword Information, June 2017. arXiv:1607.04606 [cs] version: 2.

[34] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space, September 2013. arXiv:1301.3781 [cs] version: 3.

[35] Divya Srivastava, Rajesh Wadhvani, and Manasi Gyanchandani. A Review: Color Feature Extraction Methods for Content Based Image Retrieval. 18(3), 2015.

[36] Ankush Gupta, Andrea Vedaldi, and Andrew Zisserman. Synthetic Data for Text Localisation in Natural Images. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2315–2324, Las Vegas, NV, USA, June 2016. IEEE.

[37] Dimosthenis Karatzas, Faisal Shafait, Seiichi Uchida, Masakazu Iwamura, Lluis Gomez i Bigorda, Sergi Robles Mestre, Joan Mas, David Fernandez Mota, Jon Almazàn Almazàn, and Lluís Pere de las Heras. Icdar 2013 robust reading competition. In *2013 12th International Conference on Document Analysis and Recognition*, pages 1484–1493, 2013.

[38] Dimosthenis Karatzas, Lluis Gomez-Bigorda, Anguelos Nicolaou, Suman Ghosh, Andrew Bagdanov, Masakazu Iwamura, Jiri Matas, Lukas Neumann, Vijay Ramaseshan Chandrasekhar, Shijian Lu, Faisal Shafait, Seiichi Uchida, and Ernest Valveny. Icdar 2015 competition on robust reading. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 1156–1160, 2015.

[39] Chee Kheng Chng and Chee Seng Chan. Total-text: A comprehensive dataset for scene text detection and recognition. *CoRR*, abs/1710.10400, 2017.

[40] Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, and Jiajun Liang. EAST: An Efficient and Accurate Scene Text Detector, July 2017. arXiv:1704.03155 [cs] version: 2.

[41] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, and Hwalsuk Lee. Character Region Awareness for Text Detection, April 2019. arXiv:1904.01941 [cs] version: 1.

[42] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.

[43] Cody Hochstenbach and Wouter Van Gent. An anatomy of gentrification processes: variegating causes of neighbourhood change (Environment and Planning A, 2015, 47(7), pp.1480-1501). Technical report, September 2015.

[44] Jaided AI: EasyOCR demo.

Appendix A HOCHSTENBACH AND VAN GENT'S FINDINGS ON GENTRIFICATION IN AMSTERDAM NEIGHBORHOODS

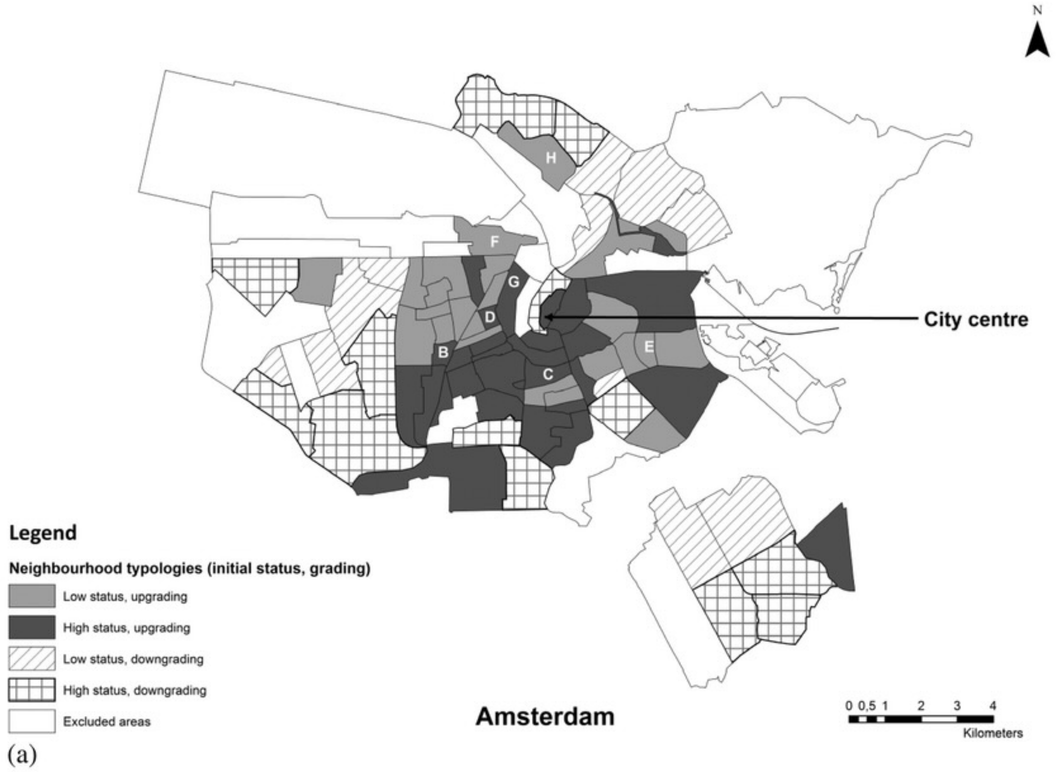


Figure 12: Neighborhood initial status and grading [43]. Neighborhood initial status was defined as above- (high status) or below-average (low status) median income in 2004. Neighborhood grading was defined as an increase (upgrading) or decrease (downgrading) in the median income (corrected for inflation) during the period 2004–11. Note how Zuidoost and Nieuw-West were found to be downgrading, and Oud-West and Centrum-West - particularly De Jordaan - were found to be upgrading.

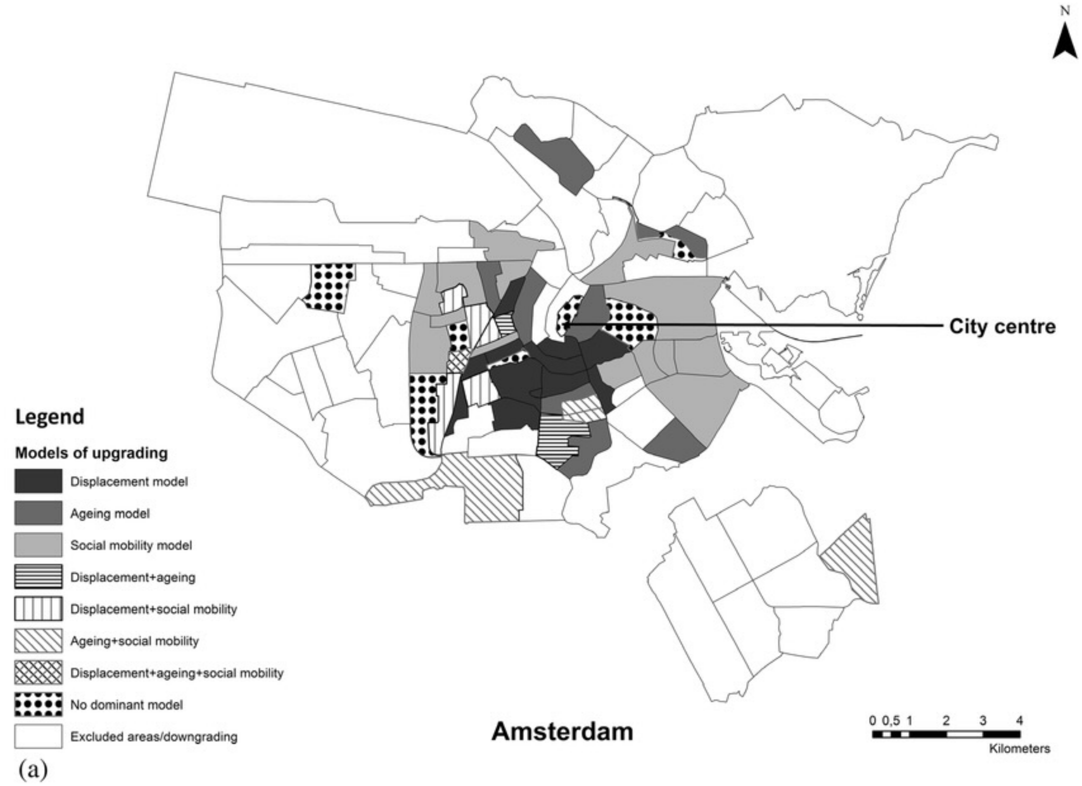
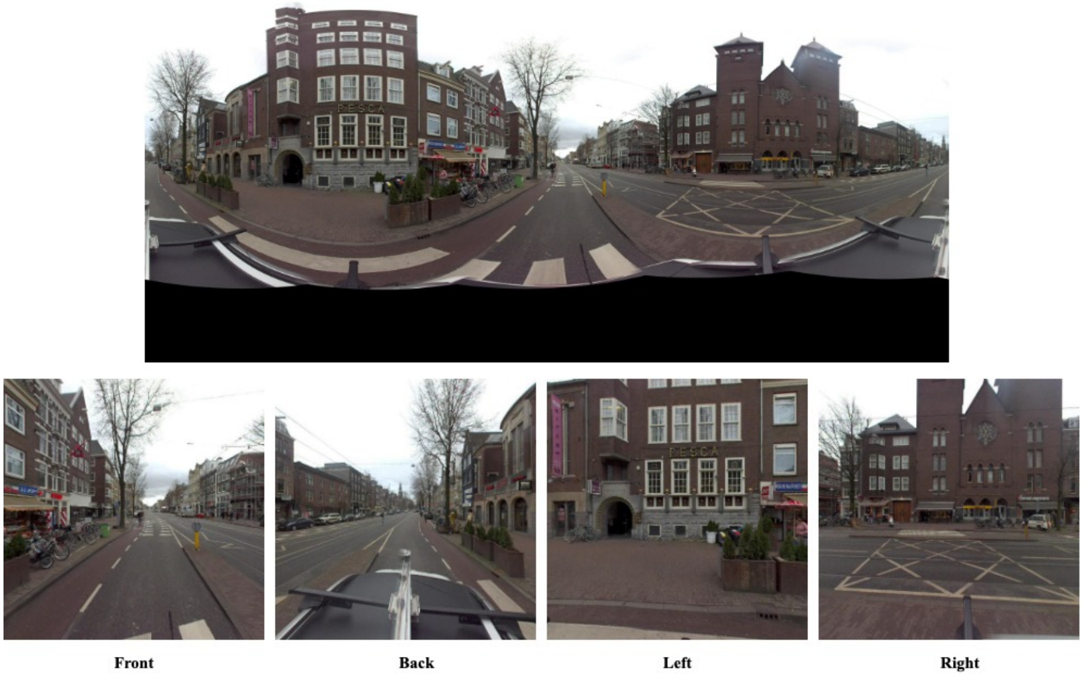


Figure 13: Neighborhood upgrades [43]. Displacement was defined as an above-average share of migration among low-income residents and an above-average share of migration among middle- or high-income residents. Social mobility was defined as an above-average share of low-income residents experiencing upward social mobility (and become middle- or high-income) while staying in the same neighbourhood. Ageing was defined as an above-average share of low-income residents ageing out of the core population. Note how Oud-West and Centrum-West - particularly De Jordaan - were found to have undergone social mobility, displacement, and ageing; while no such trend was found for Zuidoost and Nieuw-West.

Appendix B EXTENDED DATA EXAMPLE IMAGES.



Front Back Left Right

Figure 14: Example of the extended data.

Appendix C TEXT INSTANCES SIZE AND ASPECT RATIO.

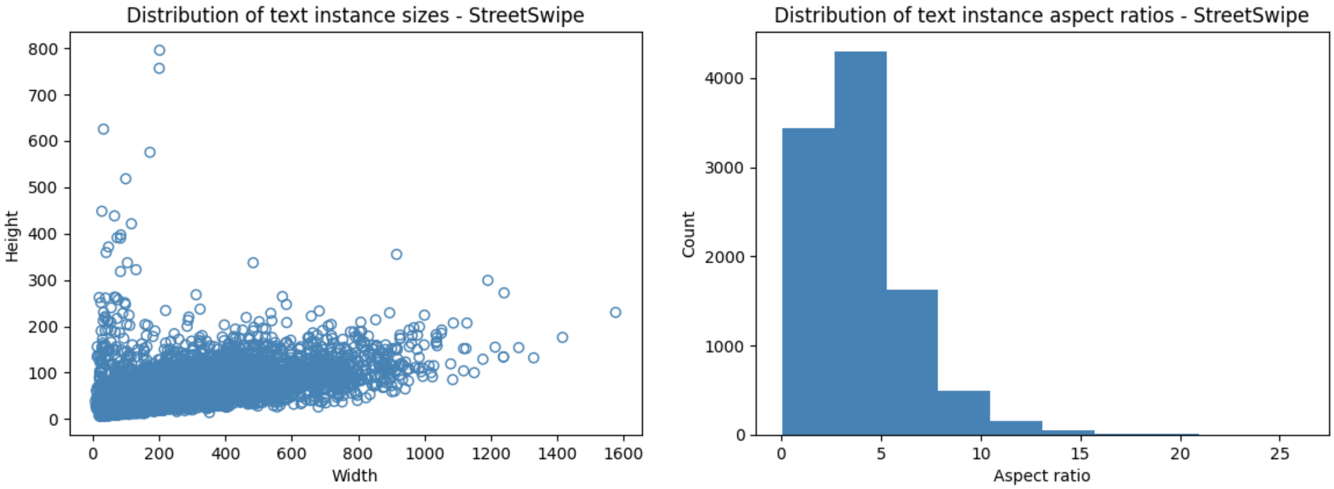


Figure 15: StreetSwipe text instance size and aspect ratio

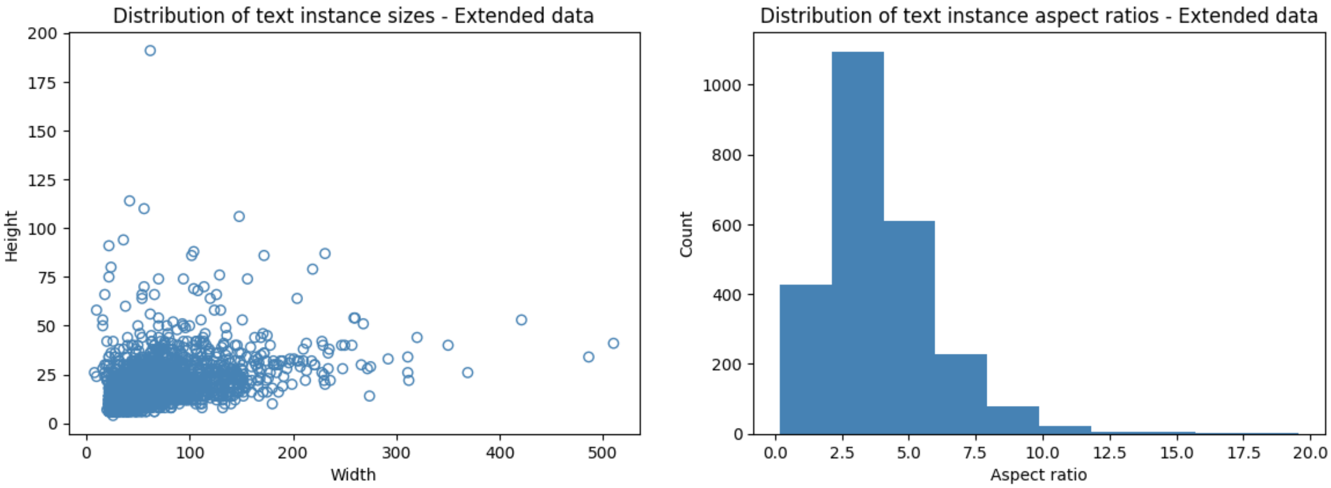


Figure 16: Extended data text instance size and aspect ratio