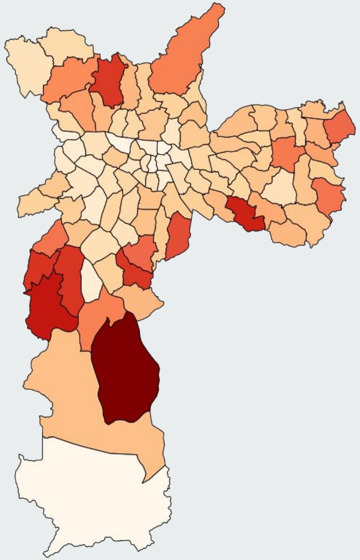
17K-Graffiti: Spatial and Crime Data Assessments in São Paulo City



Bahram Lavi, Eric K. Tokuda, Felipe Moreno-Vera, Luis Gustavo Nonato, Claudio T. Silva, and Jorge Poco

Visual Data Science Lab.,
School of Applied Mathematics,
Getulio Vargas Foundation,
Rio de Janeiro, RJ, Brazil

Email: bahram.lavi@fgv.br

Graffiti and Urbanism

★ Urban elements (city’s appearance) can affect the lives of inhabitants.

★ *Graffiti* is an essential and inseparable social element:

- *to express culture, or*
- *to manifest the vision of a community of people.*

★ *Graffiti* formerly interpreted in the *Broken Window Theory* as a social disordering element.

(this theory plays a significant role in getting police attention to social elements and other offenses).

©**Banksy**, Street Artist.

Motivation

❖ The ultimate goal is to seek relations between:
  ➢ _Graffiti Incidence_ (as an spatial city's element), and
  ➢ _Crime Occurrences_ (as social offences).

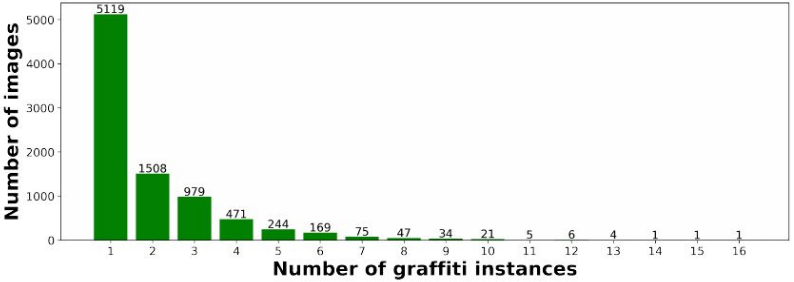
❖ Lack of sufficient Graffiti dataset (only _STORM_ with 1K images).

❖ No robust model to detect and localize Graffiti.

Example images and boundary boxes from _STORM_ dataset

Graffiti Dataset Collection-and-Annotation

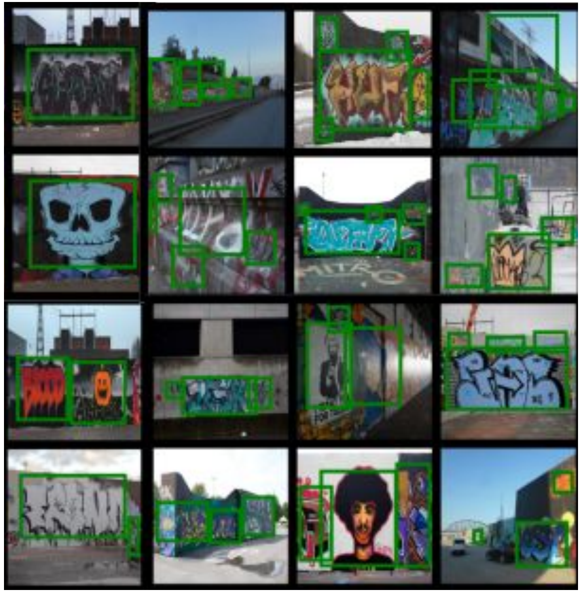
- Graffiti images collected via *Flickr.com* (through an API with hashtag of "graffiti").
- The initial pool was 15K images and we kept only ~9K.
- The boundary box annotation procedure performed manually: yielded ~17K Graffiti instances.
- The dataset was divided as 80% to train and 20% to test.



Set	Images	Boundary box		
		Single-boundary	Multi-boundary	Total
Train	6,956	4,115	9,704	13,819
Test	1,737	1,004	2,008	3,012
Total	8,693	5,119	11,712	**16,831**

Image and Annotation Examples

(a) Graffiti Image Samples

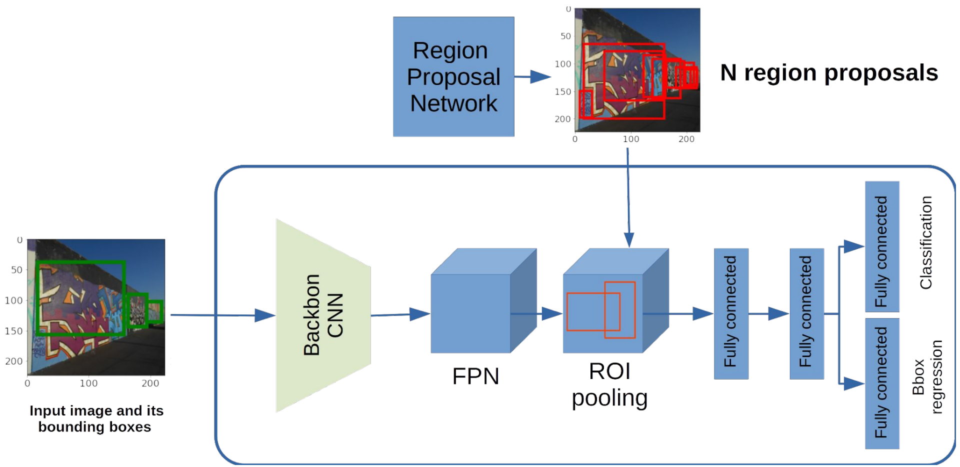


(b) Annotated Boundary Box

Faster R-CNN* Graffiti Detection Model

- ➔ Treated the detector as a binary problem
- ➔ Backbone = ResNet50

  [pretrained weights on MSCOCO]
- ➔ Image size = 224x224x3
- ➔ Batch size of 16
- ➔ Iterated for 27k



* Ren, S., He, K., Girshick, R., and Sun, J. Faster r-cnn: towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence, 2016.

Performance Evaluation and Comparison

- ➔ Mean Average Precision (mAP) over different criteria on IOU.
- ➔ Performance evaluation of the detector with STORM dataset.
- ➔ Comparing the results with a detector developed by Alzate et al*.

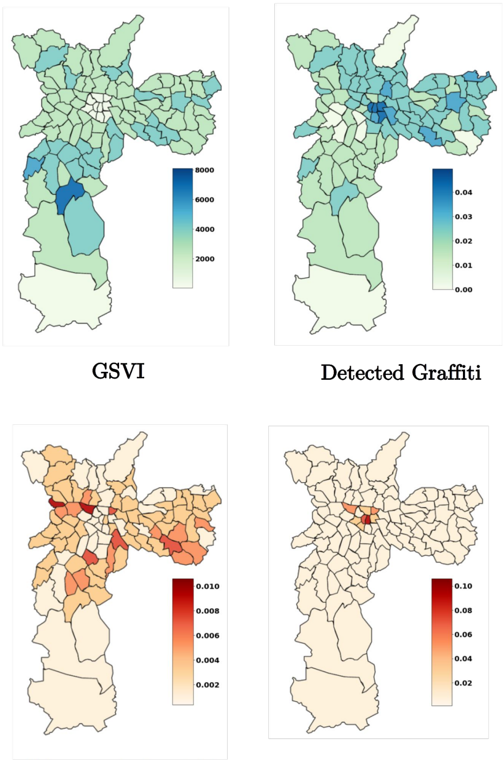
Detector	dataset	mAP		
		@[IOU=0.25]	@[IOU=0.50]	@[IOU=0.75]
(Alzate et al., 2021)	STORM	-	58.30	-
	STORM-Extended	-	69.14	-
Ours	STORM	83.05	71.60	51.53
	17K-Graffiti	**89.13**	**85.20**	**62.64**

* Alzate, J. R., Tabares, M. S., and Vallejo, Graffiti and government in smart cities: a deep learning approach applied to medellı́n city, colombia. In International Conference on Data Science, E-learning and Information Systems 2021.

Graffiti and Crime Data in São Paulo

- Normalization Factors:
  - Graffiti: frequency of GSV images per district
  - Crime: population size per district

Data Type	GSV [Year 2017]		Crime data [Year 2017]	
	Images	*Detected* Graffiti Images	Vehicle	Pedestrian
Frequency	275,349	4,268	31,800	103,945



GSVI — Detected Graffiti

Crime against vehicle — Crime against pedestrian

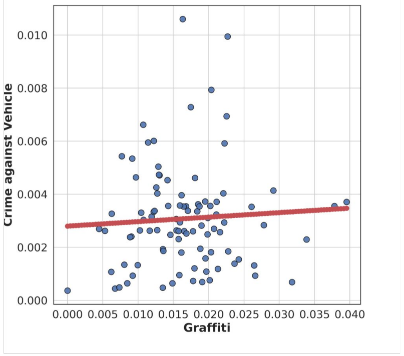
(Top): geographical distributions of downloaded GSV images, and detected graffiti images; **(Bottom):** crime against vehicle, and crime against pedestrian in 96 districts in São Paulo.

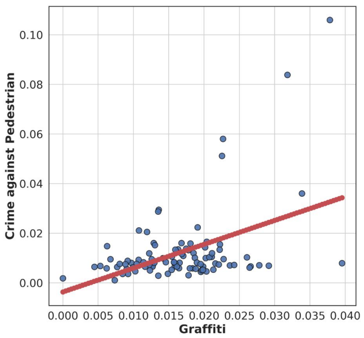
Examples of Graffiti Incidence in São Paulo City

Data Correlation

❖ We report **r** value computed by Pearson Correlation.

Spatial infrastructure	Crime	**r** value
Graffiti vs.	Vehicle	0.06
	Pedestrian	0.44





Conclusion

★ We organized the *17K-Graffiti* dataset and treated it as an spatial city's element.

★ A robust *Graffiti detection model* was developed and performed on the vast number of images from GSV in São Paulo; aiming to detect *Graffiti incidence*.

★ We studied the data correlation between *Graffiti* and two types of *Crime*:

- *Vehicle: No apparent association,*
- *Pedestrian: a relatively high correlation across neighborhoods.*

★ Hypothesized the causes of such effects, mainly related to the factors that favor graffiti production.

Future work

- ❑ Influence of different Graffiti types with crime data
- ❑ Other imagery clues (e.g., trash, garbage bag, container)
- ❑ More infrastructure data such as: health rate, education, tree, bus stop, street light, school, bar
- ❑ Different Crime types and its records over different time periods

Thanks!

For more info, find us through:

- Our Lab. website: www.visualdslab.com
- Our Github repository: https://github.com/visual-ds