



UNIVERSIDAD
NACIONAL
DE COLOMBIA


APRENDIZAJE DE MÁQUINAS

Ejemplo - Datos

John Ballesteros

Profesor Asociado
Departamento de Ingeniería Civil
jballes@unal.edu.co

<https://github.com/srobles05/3008422-AprendizajeDeMaquinas>


$$a_{11}X_1 + a_{12}X_2 + \dots + a_{1n}X_n = w_1$$

$$a_{21}X_1 + a_{22}X_2 + \dots + a_{2n}X_n = w_2$$

.

.

.

$$a_{n1}X_1 + a_{n2}X_2 + \dots + a_{nn}X_n = w_n$$

$AX = b$, donde: A, X, b son matrices o tensores, A son los parametros o los pesos, b son los labels

Training Data



Famous Datasets

Mnist: handwritten numbers

Coco: common objects in context

Planet: satellite from amazon rainforest

Camvid: Street datasets

Ways to collect data

- In the field
- In the Internet

Different ways of annotate

- In the file name: it's the most common: ex. dog.jpg, horse.jpg, cat.jpg
- In a separate file: typically in a .csv file
- In a field of a table: target
- In the whole image itself: segmentation

Search and scroll [Google Images](#)

- Be specific, less manual pruning you will have to do.
- The maximum number of images Google Images shows is 700.
- Put things you want to exclude into the search query
- Limit your results to show only photos by clicking on Tools and selecting Photos from the Type dropdown.

What is enough data?

If hyper parameters: architecture, epochs and learning rate are ok but model accuracy is still not good, below 80%

You need more data !!!!

Data Augmentation and sintetic data

- Programming is the solution, there is little workaround
- Python is more than ok: PIL, CV2, and imageio modules do the trick!
- Typical transformations are: rotate=flip, scale (affine), warping

This can multiply actual data by 4 !

- Crappify and blurring is easy.

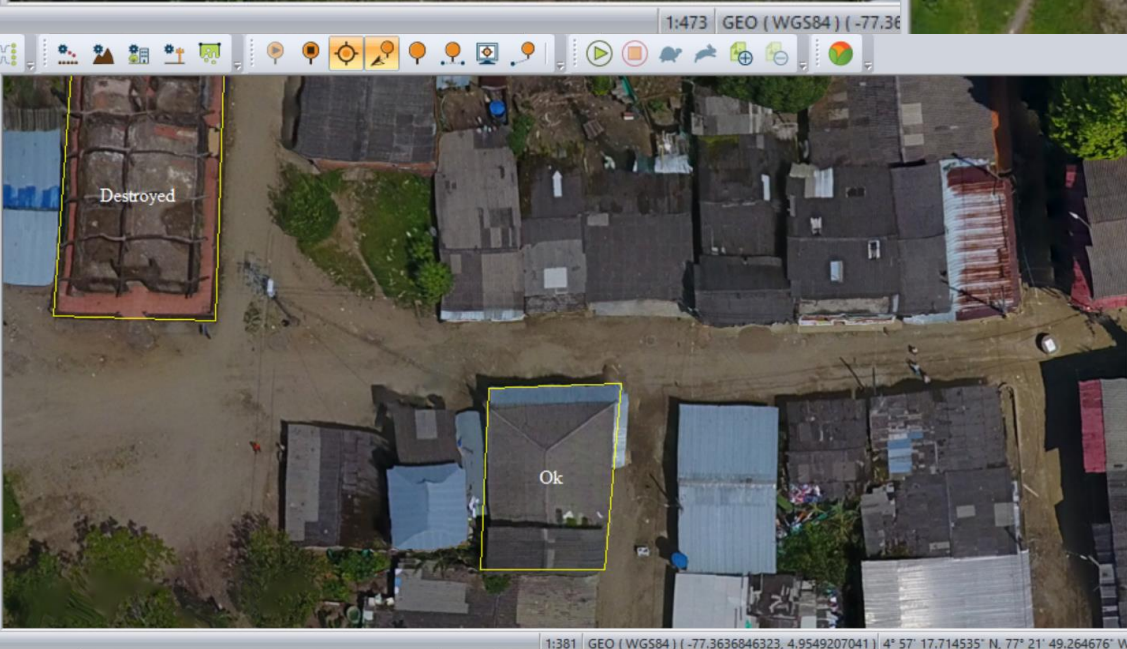
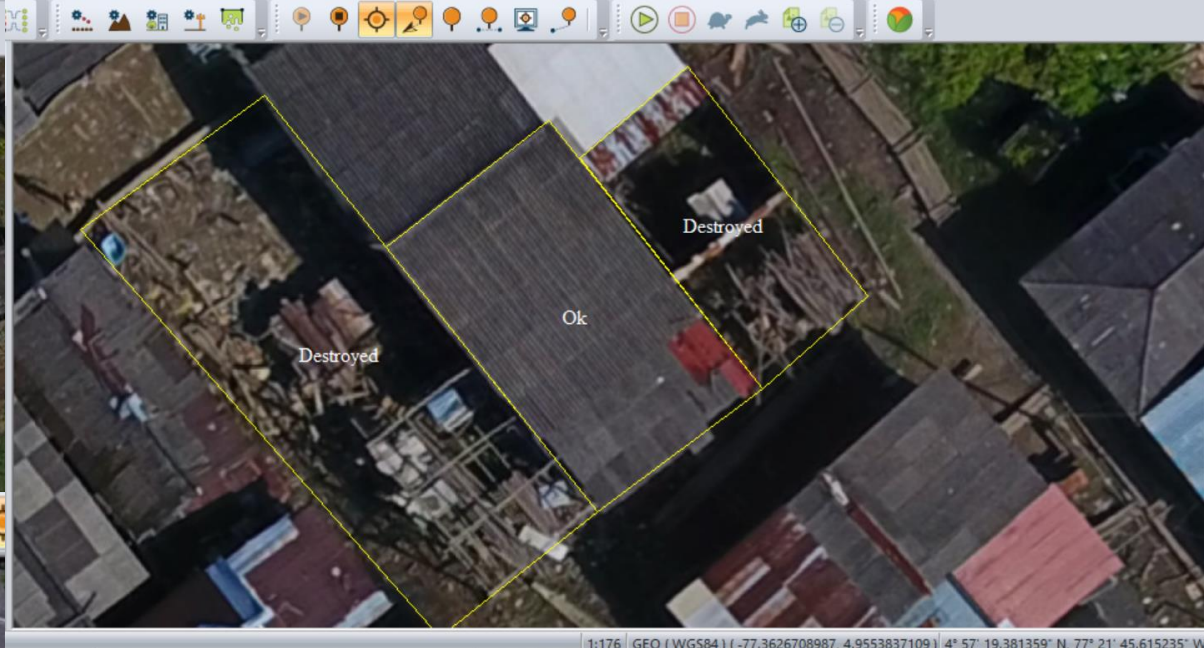
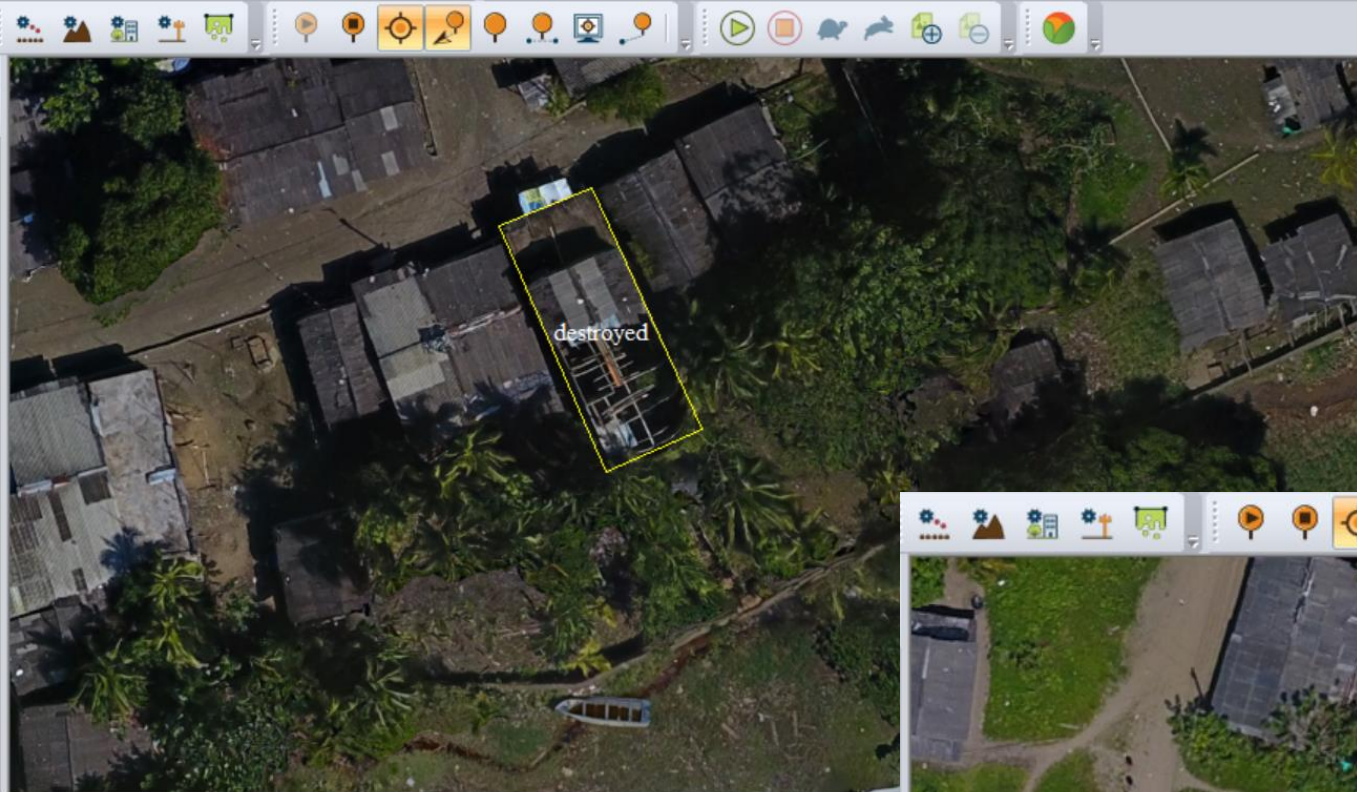
Examples of problems that can be solved

Binary classification: Damaged/No Damaged-----Prob 1/0

Multi classification: Rock classification, Minerals, vipers, trees, outcrops, house assessment-----Prob Treshold above 0.6

Segmentation: each pixel is assigned a value into a category

Cutoff, time, marks, color, resolution (more advanced): GANs



peaton



bus



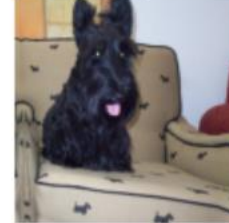
peaton



camion



scottish_terrier



Sphynx



staffordshire_bull_terrier



bus



camion



bus



moto



wheaten_terrier



pug



Siamese



moto



moto



camion



peaton



carbon



rhyolite



basalt



keeshond



British_Shorthair



havanese



moto



moto



moto



moto



andesite



carbon



rhyolite



andesite



sandstone



schist





con acabados, garaje,
ante jardín, varios pisos,
grande, alto valor



con acabados, no garaje,
no ante jardín, varios
pisos, pequeña, medio
valor



sin acabados, no garaje,
no ante jardín, varios
pisos, mediana, bajo valor



con acabados, no garaje,
no ante jardín, 1 piso,
pequeña, bajo valor



alta pendiente, no
cobertura, alto peso, muy
alto riesgo



alta pendiente, no
cobertura, alto peso, muy
alto riesgo



alta pendiente, no
cobertura, bajo peso,
muy bajo riesgo

Camvid: Streets Segmentation Dataset





UNIVERSIDAD
NACIONAL
DE COLOMBIA



Labelling Time (research topic)

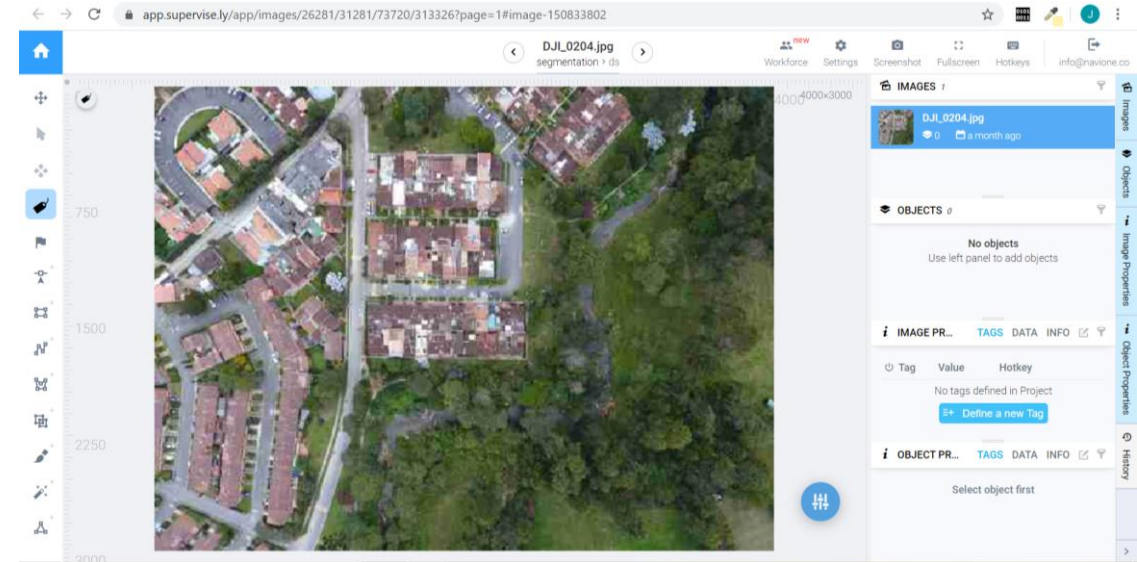


What about text?

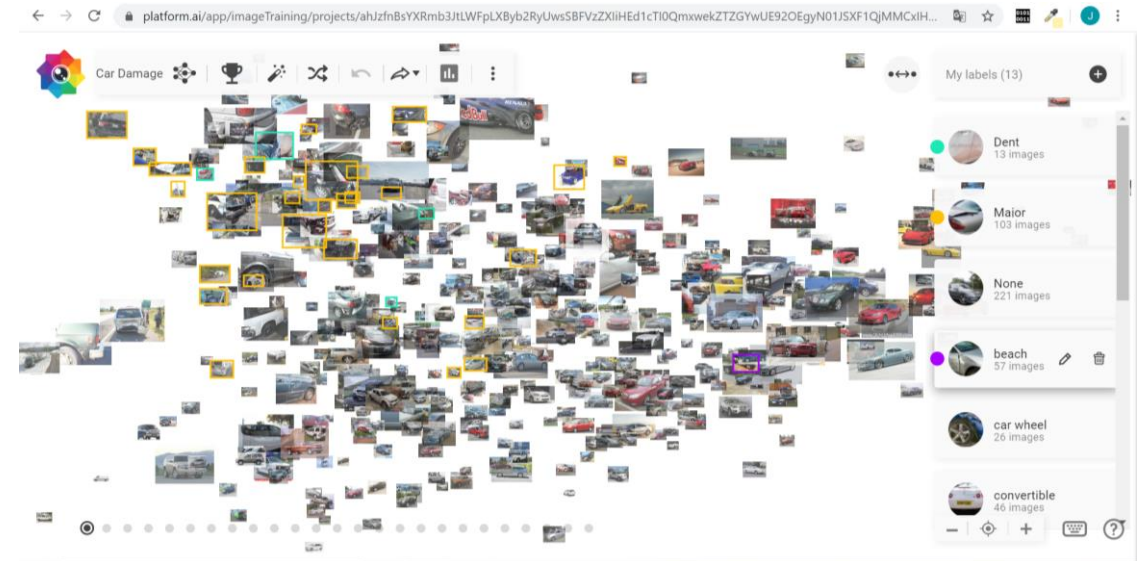
Dir. Inicial	Dirección	Complemento	Barrio / Vereda / I POI	Pto. geográf	Ciudad	Depto	País
CRR 14 #7A 34 SUR BARRIO SAN CARLOS O RETEN SUR	Carrera 14 7A Sur Sur 34		Barrio San Carlos				
CARRERA 26 A # 1A-43 BARRIO SANTA ISABEL	Carrera 26 A 1A 43		Barrio Santa Isabel				
KR 19B1 9 - 25 LOS CORTIJOS	Carrera 19B1 9 25		Los Cortijos				
CALLE 7 NO 11-24	Calle 7 11 24						
APTO 304 T1 EDIF SIEMPRE VERDE CRA 17 #118-32	Carrera 17 118 32	Apartamento 304	Edificio Siempre Verde				
CRA 42 NO 85 A 95 ITAGUI-ITAGUI-ITAGUI-ANTIOQUIA	Carrera 42 85 A 95				Itagui	Antioquia	
CR 7, 10-87,	Carrera 7 10 87						
CALLE 59 # 56-63, NRO. TORRE 6, APTO. 721	Calle 59 56 . 63	Apartamento. 721					
CALLE69SUR 46A 64 CALLE DEL BANCO TORRE ASÍS II APT 701, S.	Calle69Sur 46A 64	Apartamento 701			Sabaneta	Antioquia	Colombia
MANZANA J5 CASA 11 B. VILLA CATALINA	Manzana J5	Casa 11	Villa Catalina				
CRA91#44A39,AMÉRICA NIZA 2PISO MEDELLÍN BARRIO: AMÉRICA	Cra91 44A39		America Niza		Medellin		
CR 34 # 10 581 ACOPI YIMBO	Carrera 34 10 581		Acopi			Yimbo	
CALLE 22 #1-140 AV BOLIVAR FERRETERIA ARGENTINA	Calle 22 1 140		Ferreteria Argentina				
TR 58 BIS # 2 C 60 B CAMELIA	Transversal 58 Bis 2 C 60		B Camelia				
KILOMETRO 3981 ANIKLO VIAL RIO FRIO ZOBA FRANCA SANTANDE	Kilometro 3981 Anillo Vial Rio Frio		Edificio Baiachara		Santander		
CARRERA 4 #2-03 BARRIO CHAPINERO	Carrera 4 2 03		Barrio Chapinero				
CRA 14F 76B 57 SUR INT 1 MARICHUELA USME BOGOTÁ, D.C. BOG	Carrera 14F 76B Sur 57	Interior 1			Bogota D.C.	Bogota	Col
CR 10 A # 11 75 LOCL 103 PASAGE GOMEZ	Carrera 10 A 11 75	Local 103	Pasage Gomez				
CL 24 # 6 67 CENTRO APTO 301 PEREIRA BARRIO: CENTRO	Calle 24 6 67	Apartamento 301	Pereira Barrio Centro	Centro			
CLL 82 # 67 A -51	Calle 82 67 A 51						
CL 21 # 6 36 CENTRO MONTERÍA CÓRDOBA COL	Calle 21 6 36			Centro	Monteria	Cordoba	Col
CL 71 # 6-21 OF 301HIDROCARBUROS DEL CASANARE S.A.S	Calle 71 6 21		Hidrocarburos Del Casanare S.A.S				
CL 20 A # 12 70 METRO SECCION ELECTRO	Calle 20 A 12 70			Metro Seccion Electro			
CARRERA 22B NUMERO 13A 47 BARRIO GUAYAQUIL BARRIO GUA	Carrera 22B 13A 47		Barrio Guayaquil Barrio Guayaquil				
CR 1 # 12 -118 CC PLAZA BOCAGRANDE LC 105	Carrera 1 12 118	Local 105	Cc Plaza Bocagrande				

Software for Labelling

Supervisely: www.app.supervise.ly



Platform.ai: www.platform.ai



PixelAnnotationTool

Working Examples

multilabel.herokuapp.com