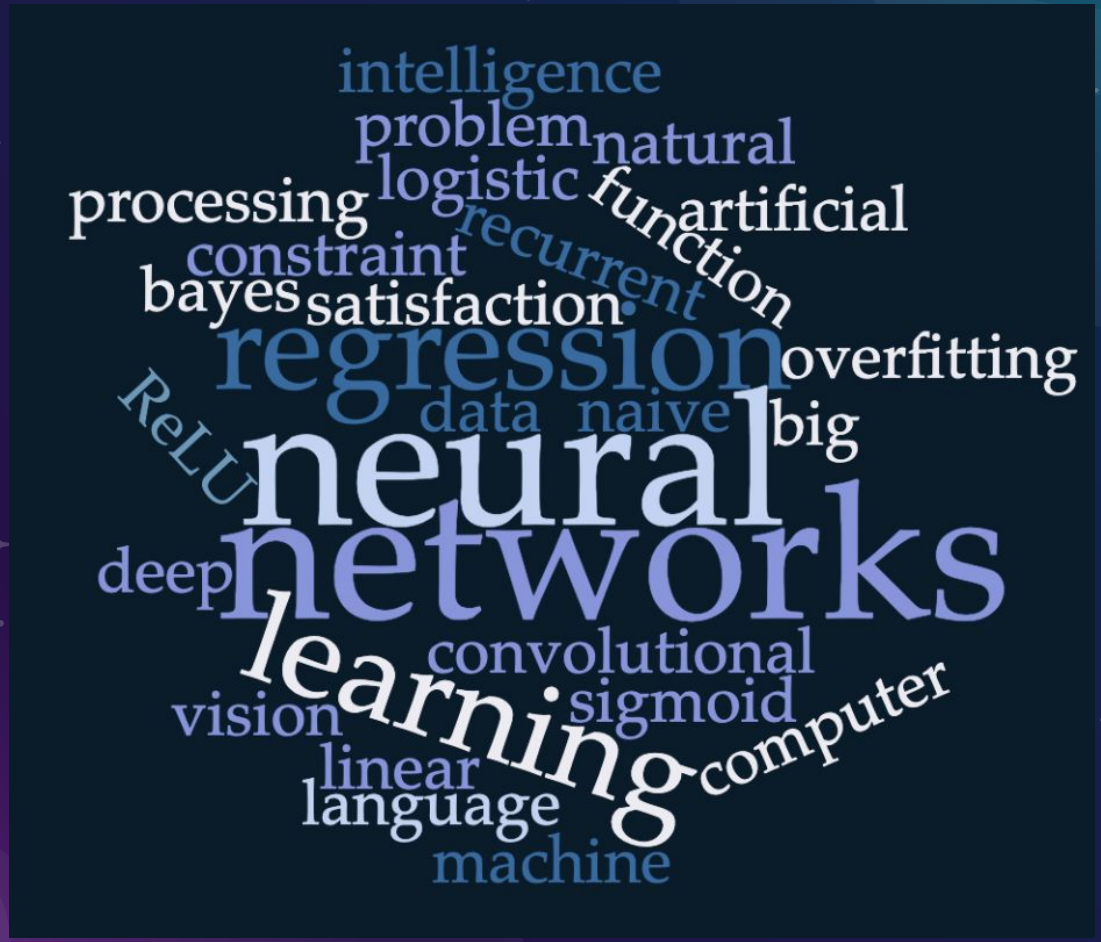
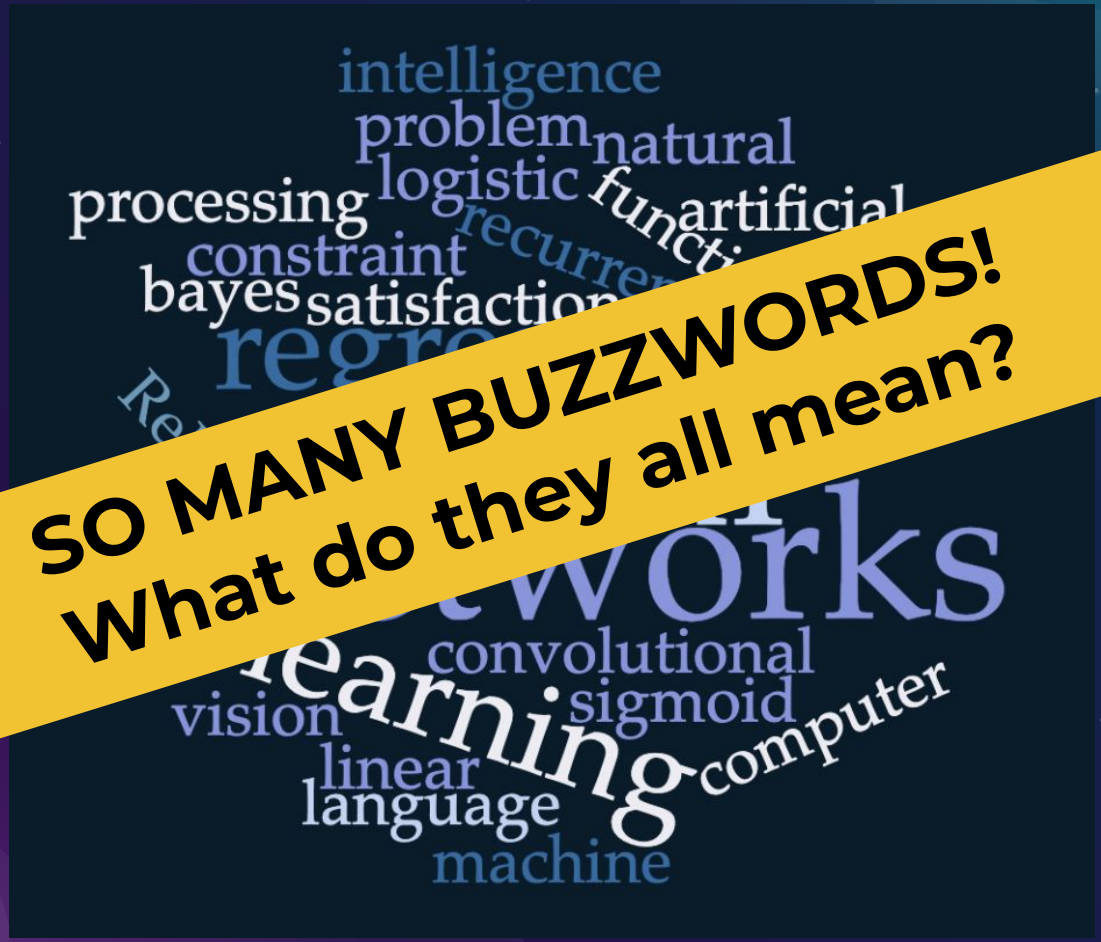
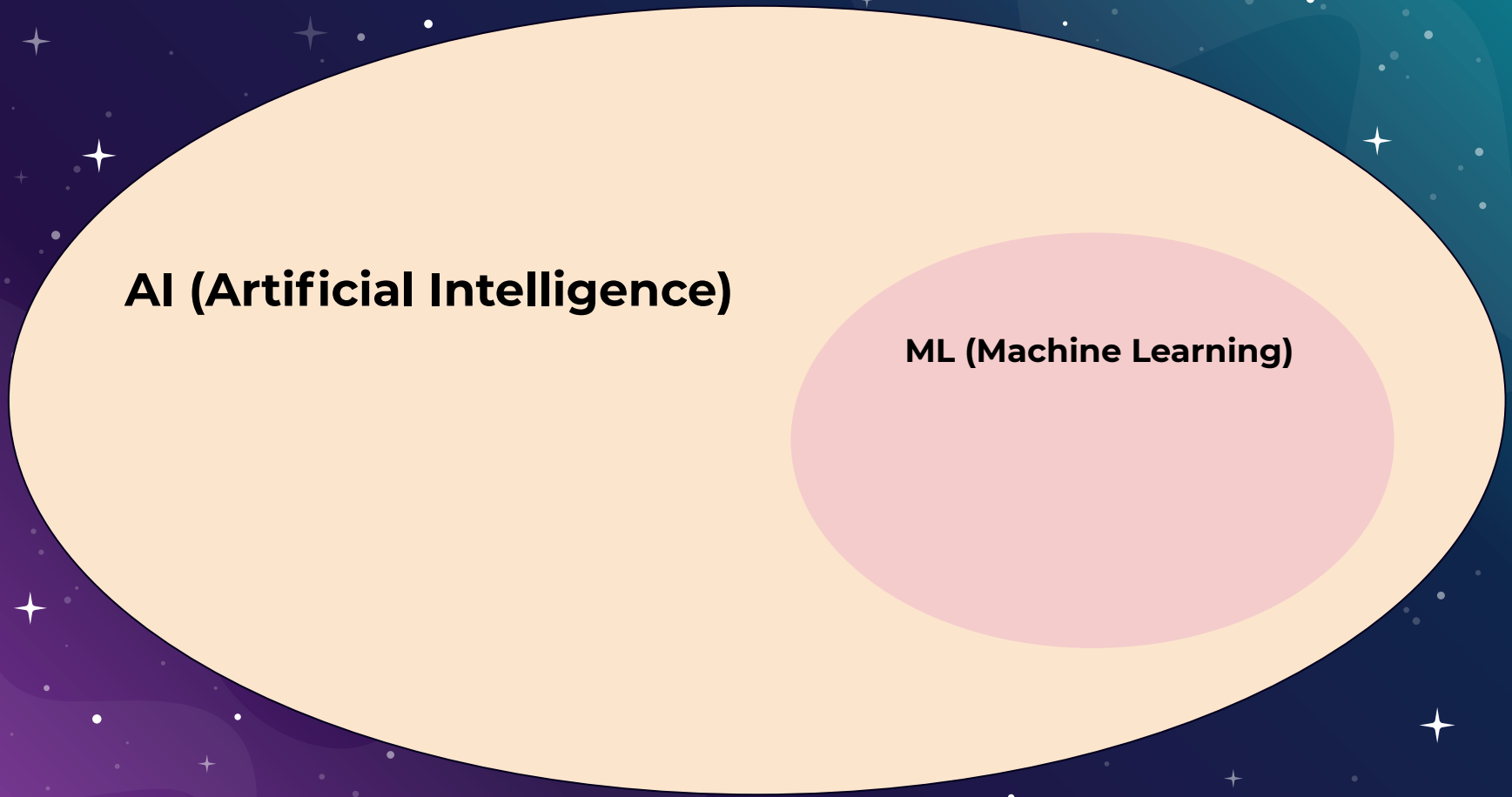
Code For China
Lesson 9

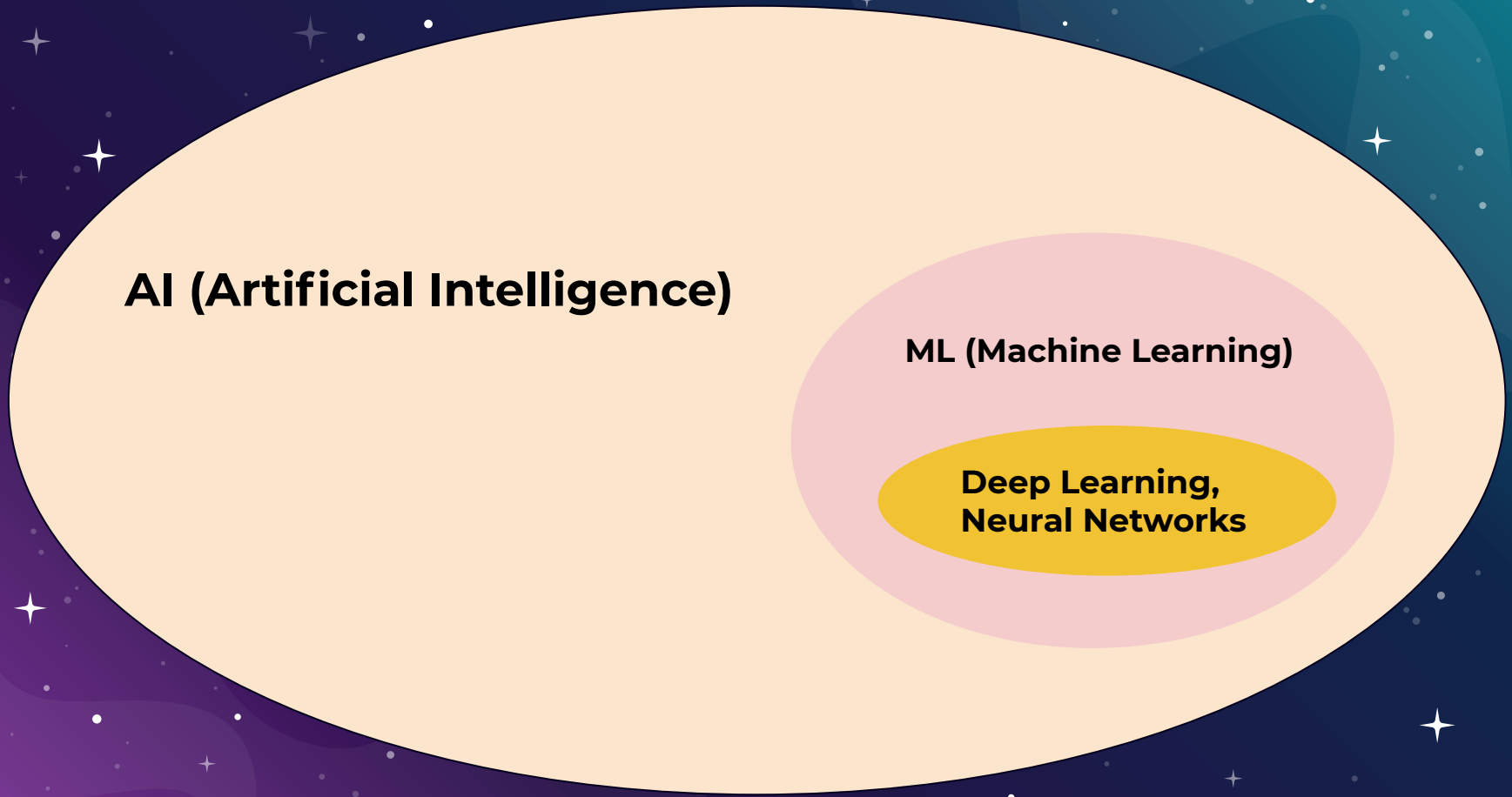

intelligence
problem
natural
logistic
processing
function
artificial
recurrent
constraint
bayes
satisfaction
regression
overfitting
ReLU
data
naive
big
neural
networks
deep
convolutional
learning
vision
sigmoid
computer
linear
language
machine



SO MANY BUZZWORDS!
What do they all mean?

AI (Artificial Intelligence)


AI (Artificial Intelligence)
ML (Machine Learning)


AI (Artificial Intelligence)
ML (Machine Learning)
Deep Learning,
Neural Networks

What is Machine Learning?

Arthur Samuel (1959):
Machine Learning is the field of study that gives the computer the **ability to learn without being explicitly programmed.**

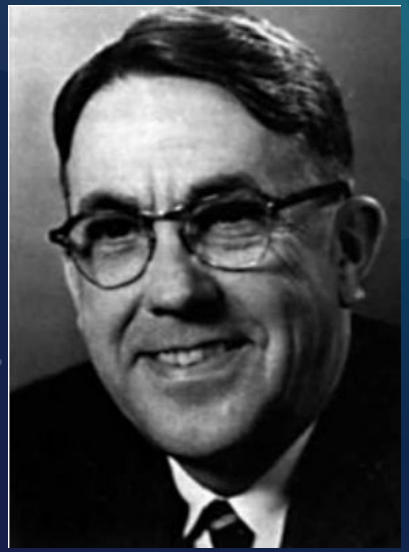
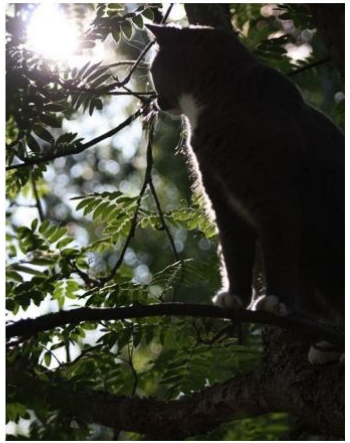


Image Classification: Detecting Cats



How would you determine if something is a cat?

Old Method: Feature Detectors

- Does it have ears?
- Does it have fur?
- Does it have 4 legs?

Old Method: Feature Detectors

- Does it have ears?
- Does it have fur?
- Does it have 4 legs?

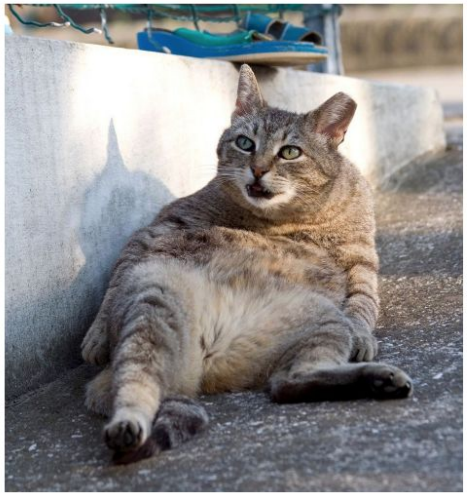
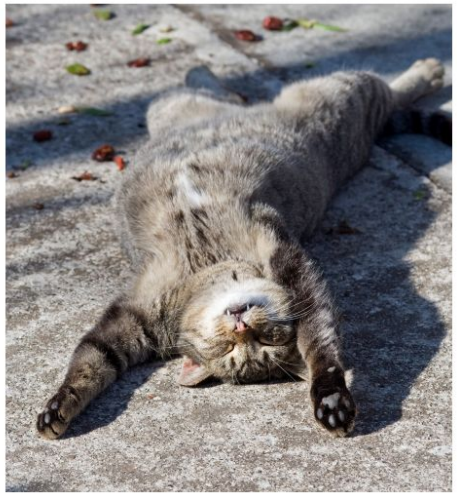
Cons of Feature Detectors

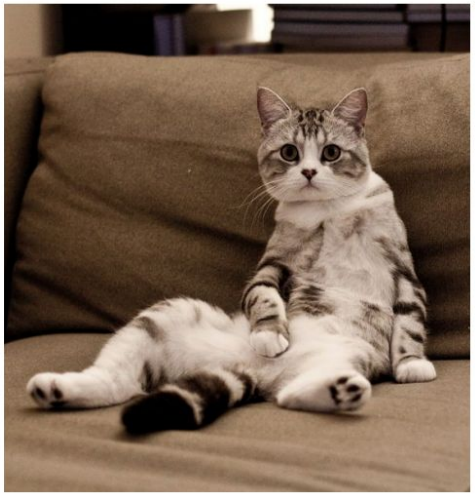
- For each task, every detector must be manually designed - this is **inefficient** (millions of tasks, millions of detectors...)

- **Not scalable** - each one requires human design

- **Not adaptable** - different images may not work with the feature detectors

Challenge: Deformation
(when they are in unusual positions)



Challenge: Occlusion
(when parts are hidden)



Challenge: Inter-Species Variation
(different types of cats look different)

Solution: Machine Learning

Neural Network

YES this is cat

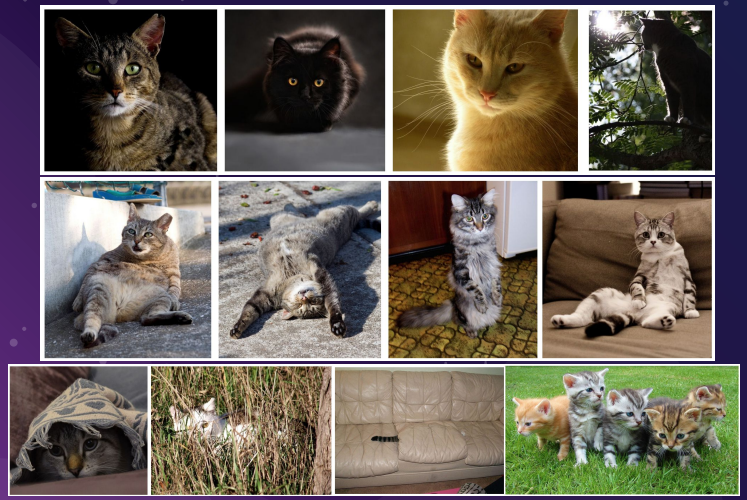
Solution: Machine Learning

Neural Network

NO not a cat

How to make a good neural network?

Answer: Lots of training examples!



Neural Network

Babies learn by repetition:

seeing a “spoon” many times and learning what it is

Requires ~20-50 times

Neural Network

Requires ~500-1000 times
(needs a lot more data!)

How does it all work?

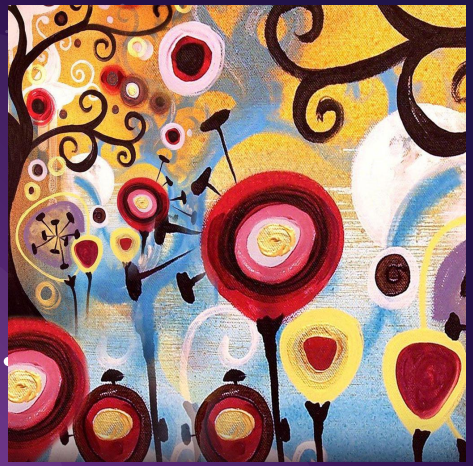
- We want to minimize our prediction error
  - Minimize when we say YES when it's not a cat and NO when it is a cat
- To do so, we:
  - Define a **cost function** that penalizes errors
  - Minimize the cost function using **stochastic gradient descent (SGD)**
- Everything in ML is ruled by stochastic gradient descent (SGD)

Why are neural networks so popular now?

- Requires a lot of memory - before, very expensive, but now computing costs are going down
- One main idea that can be adapted in millions of ways
- Very Effective - neural networks are used for voice recognition (like Siri), face recognition, handwriting recognition,

Applications of Machine Learning

Art: Neural Style Transfer



Candy Style

+

Stanford University Campus

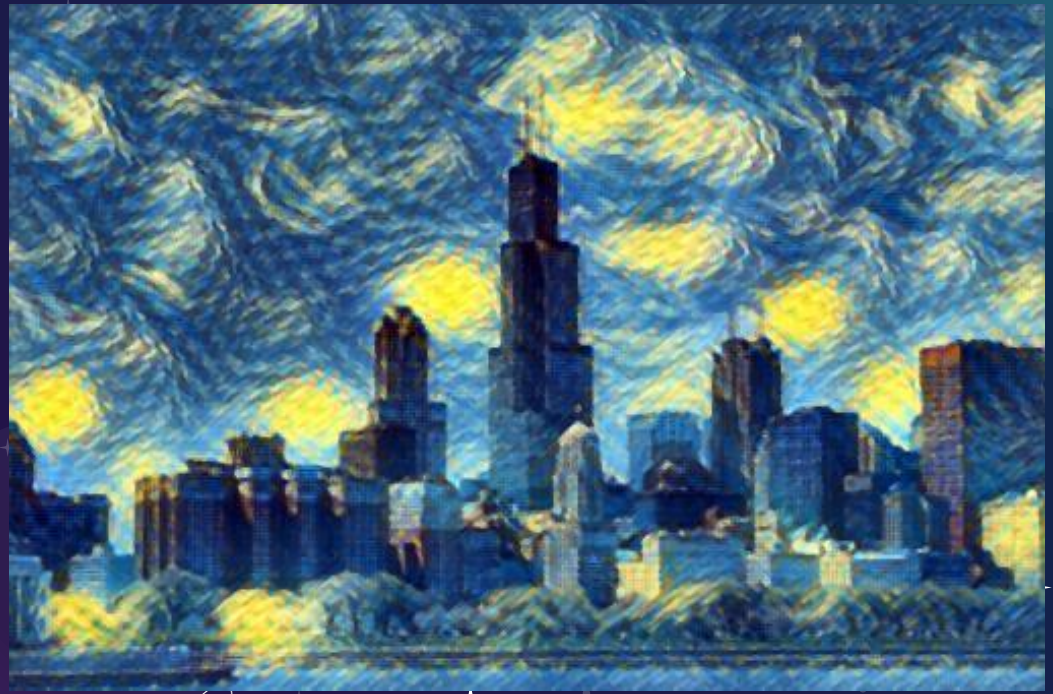
=

+

=

+

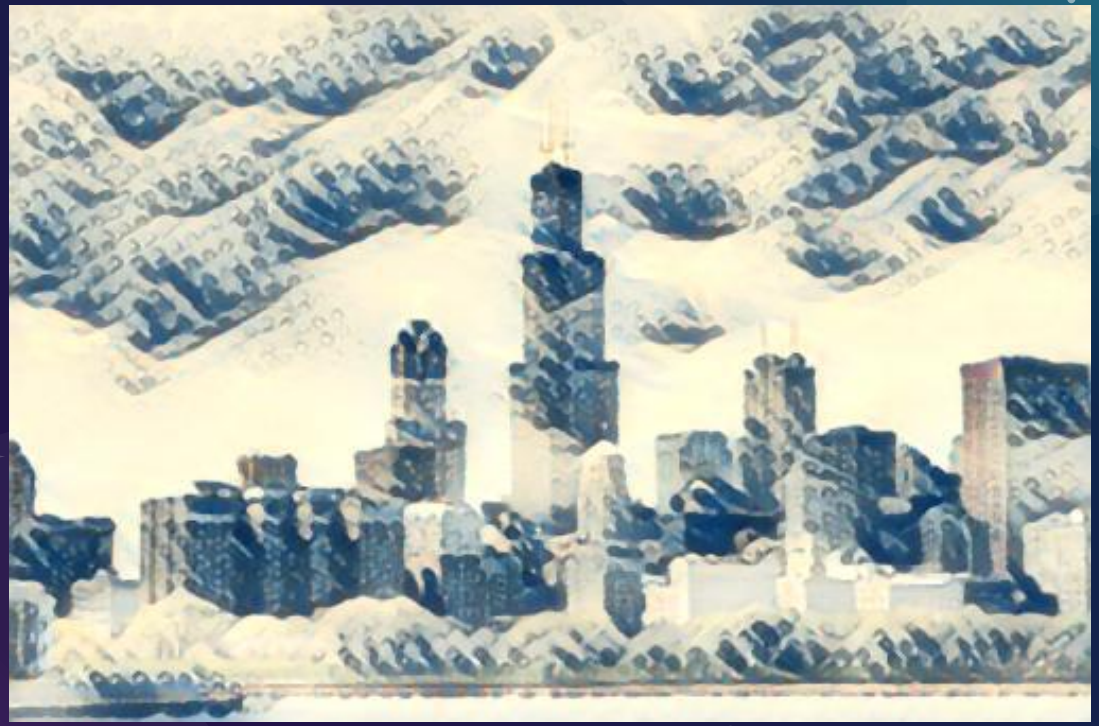
=



In the style of:
Vincent van Gogh's "Starry Night" (1889)

+

=



In the style of:
Hokusai's "Great Wave Off Kanagawa" (~1830)

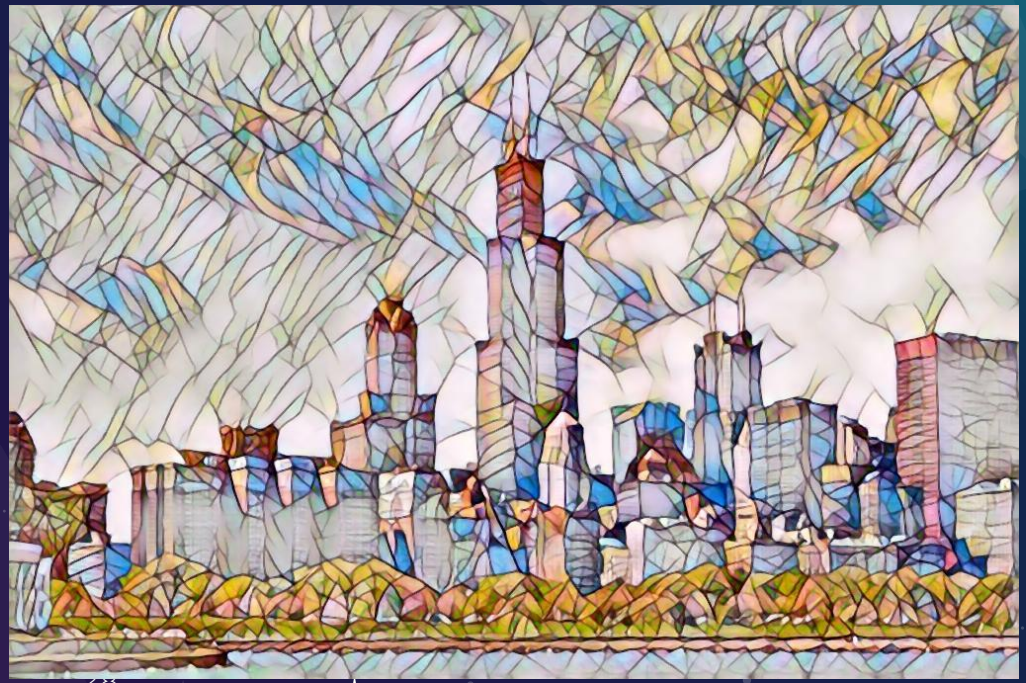
+

=

In the style of:
Edvard Munch's "The Scream" (1893)

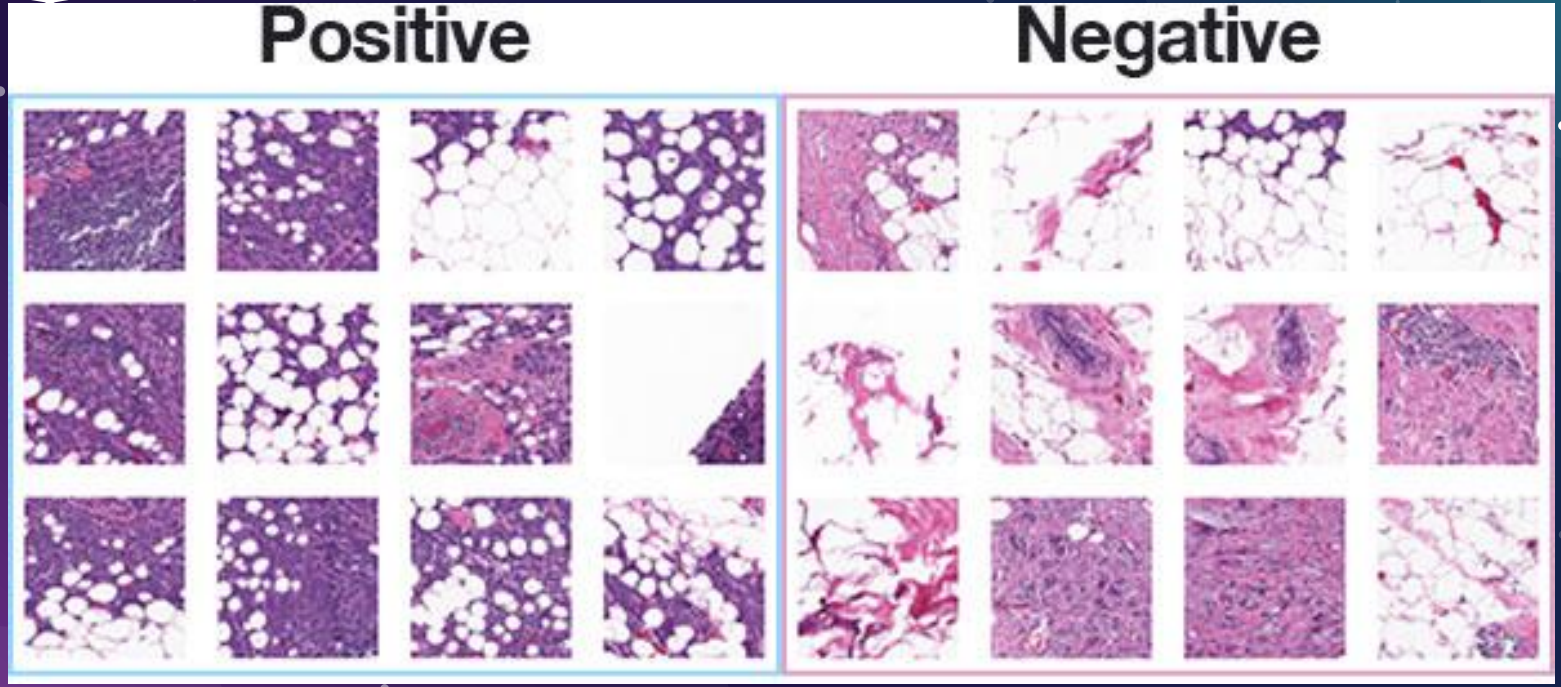
+

=



In the style of:
Stained Glass Mosaic

Medicine: Detecting Cancerous Cells



Economics: Poverty Prediction from Satellite Imagery

Which neighborhood has higher poverty?

Economics: Poverty Prediction from Satellite Imagery

Which neighborhood has higher poverty?

High Poverty (More Poor)

Low Poverty (Less Poor)

YOLOv3 for Household Poverty Prediction

Stanford | ENGINEERING
Computer Science
Jessica Zhao, jesszhao@cs.stanford.edu

TASK

We seek to measure household poverty in impoverished regions. Given a household's geographic coordinates, we use corresponding satellite imagery to predict its poverty level.

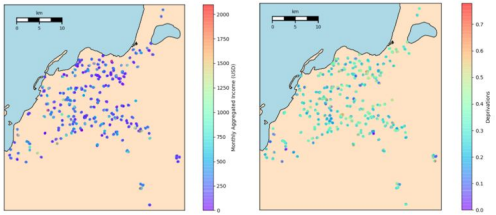
Input *Can you tell which is low & high poverty?*

Outputs

- y1 = Income (aggregated per household)
- y2 = Deprivation (a measure between 0-1, where 0 is better-off and 1 is worse-off)

DATA

1107 household surveys from Luanda, Angola randomly split into 80% Train, 10% Val, 10% Test



Income Distribution | Deprivation Distribution

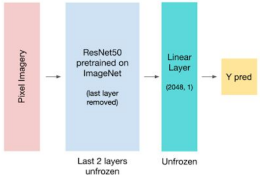
FEATURES & MODELS

In all models, we minimize the Mean Squared Error (MSE) between the ground-truth Y_i and predicted outcome $\hat{Y}_i$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

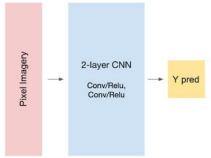
Digital Globe Imagery as Features

- 3x224x224 at 0.3m resolution
- RGB-separated histogram (bins=10)



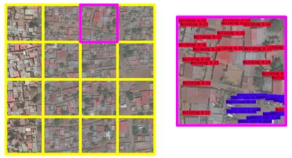
VIIRS Nightlights Imagery as Features

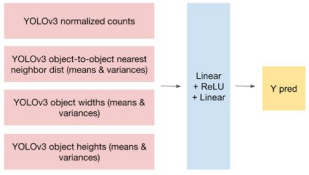
- 225x225 at 30m resolution
- Histogram (bins=10, 20)



YOLOv3 Outcomes as Features

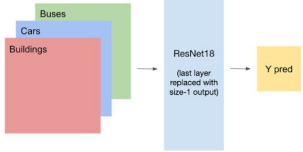
- 1000x1000 image → 16 250x250 tiles
- Extract objects and confidences from YOLOv3 pre-trained on xView





YOLOv3 Detection Masks as Features

- 3-top object categories (Buildings, Cars, Buses) become the 3 channels in a 3x1000x1000 binary image



RESULTS & DISCUSSION

Model	Features	Income r^2	Deprivation r^2
Baselines			
Linear Regression	DG Imagery	3e-4	8e-4
Linear Regression	DG RGB-Separated Histogram	**0.07**	2e-4
Linear Regression	NL Imagery	1e-3	4e-5
Linear Regression	NL Histogram (10 bins)	6e-4	5e-3
Linear Regression	NL Histogram (20 bins)	1e-4	6e-4
kNN	DG Imagery	0.06	0.03
kNN	DG RGB-Separated Histogram	0.03	0.07
kNN	NL Imagery	0.03	0.01
kNN	NL Histogram	7e-3	7e-3
kNN	Latitude, Longitude	0.05	0.09
CNN			
2-layer CNN	NL Imagery	0.07	0.04
Fine-tuned ResNet50	DG Imagery	0.04	0.02
YOLO			
YOLO-FC-60obj	Full	**0.12**	**0.11**
YOLO-FC-5obj	Full	0.09	0.07
YOLO-FC-1000dim	Full	0.09	0.09
YOLO-FC-500dim	Full	0.09	0.07
YOLO-ResNet18	Building, Car, Bus	0.05	0.05
YOLO-ResNet18	Building, Car, Truck	0.04	0.02

Top performing YOLOv3 model on both Income and Deprivation considers all 60 object categories, achieving r^2 scores of **0.12** and **0.11**

Ablation Analysis: YOLOv3 Outcomes

Features	Income r^2	Deprivation r^2
Full	0.12	0.11
- Counts	0.069	0.072
- Neighbors	0.0084	0.0078
- Widths and Heights	0.059	0.063

- Using raw pixel data directly does not yield the best predictive signal at the household-level, but **extracting YOLO features** holds promise, especially nearest neighbors features as seen in Ablation Analysis
- **Future approach:** YOLO detection masks over different objects & confidences, varying confidences by object class – may capture neighbor densities and spatial information without feature engineering

ACKNOWLEDGEMENTS

Many thanks to Burak Uzkent, Marshall Burke, David Lobell, and Stefano Ermon for their feedback; Otto Xu for his pre-trained YOLOv3 model; Justin Johnson for brainstorming different models.

Machine Learning Tools

`pytorch` (made by Facebook)

`tensorflow` (made by Google)

Currently the 2 leading frameworks used in Silicon Valley, as of 2019