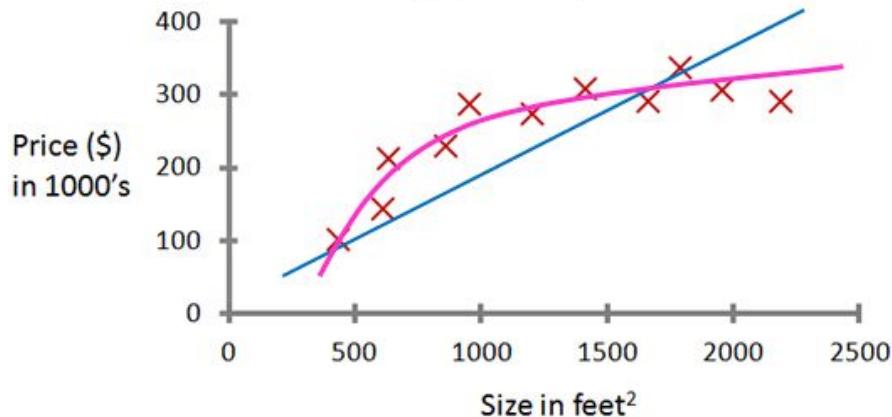


# POKEMON WORLD AND THE INDUS VALLEY CIVILISATION - THE ANALOGY

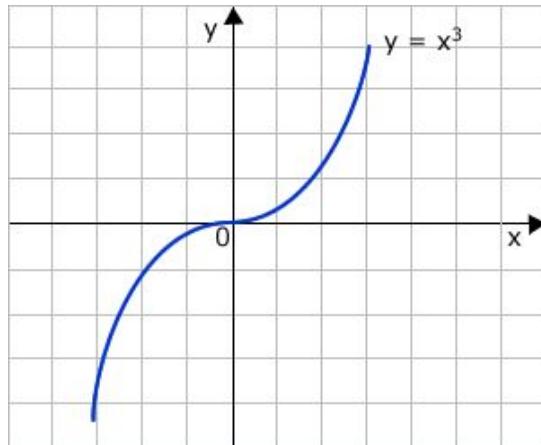
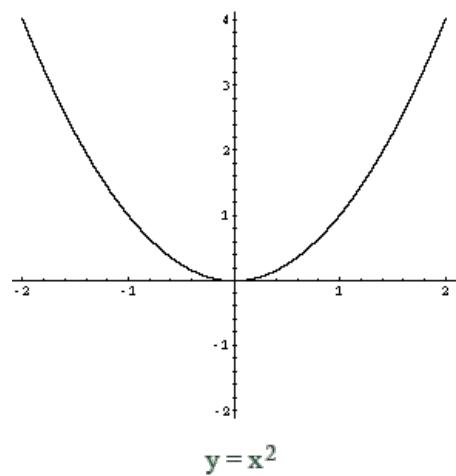
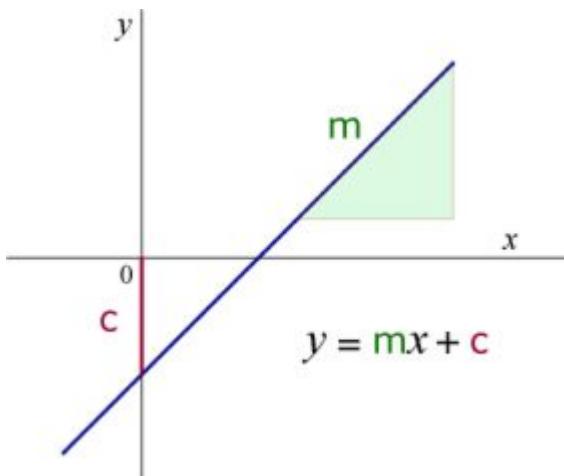
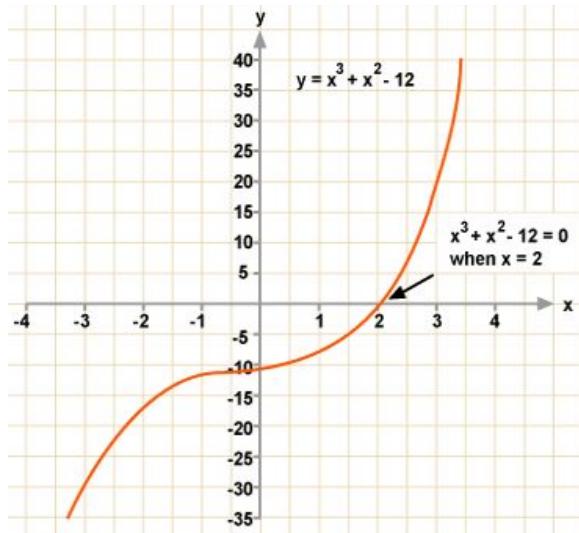
Satish Palaniappan

# WHERE IT ALL STARTED?

## Housing price prediction.

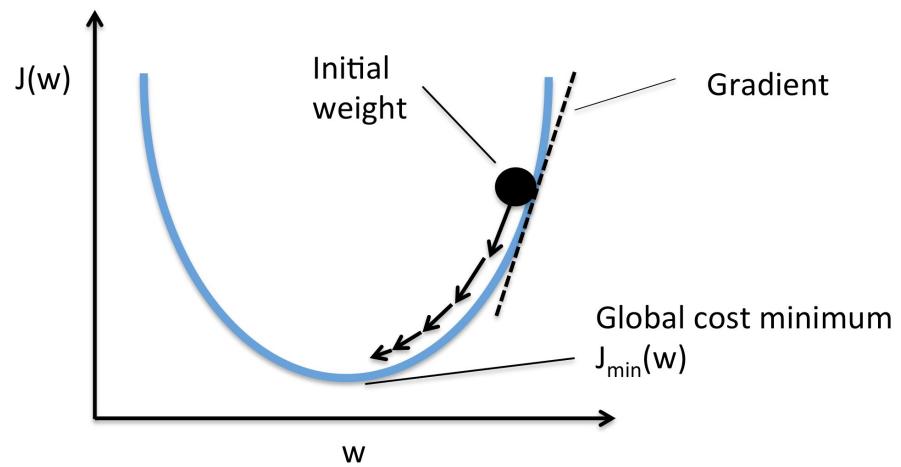
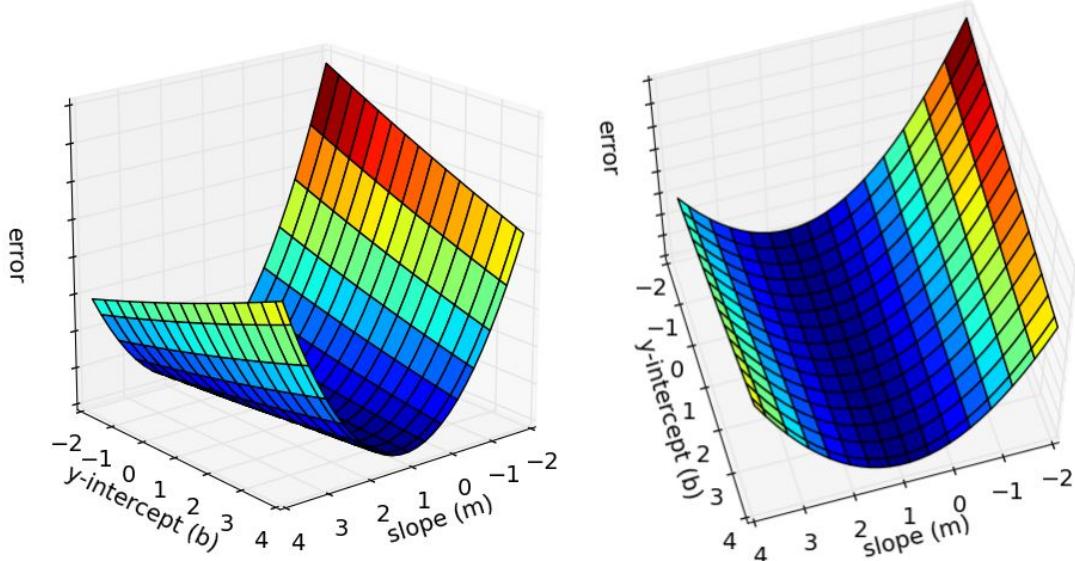
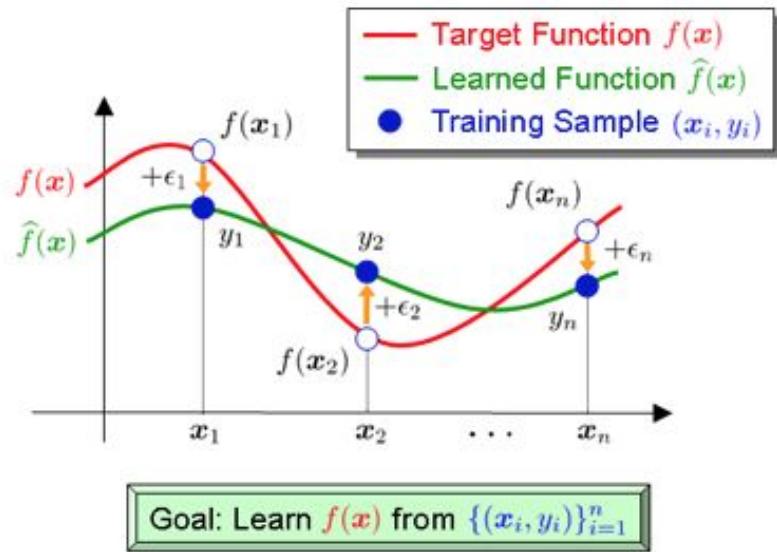


# WHAT IS ML?

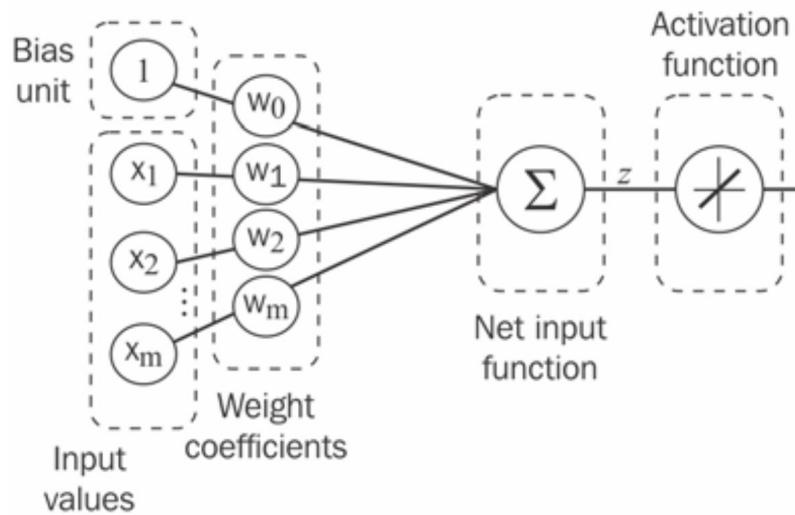


# WHAT IS ML?

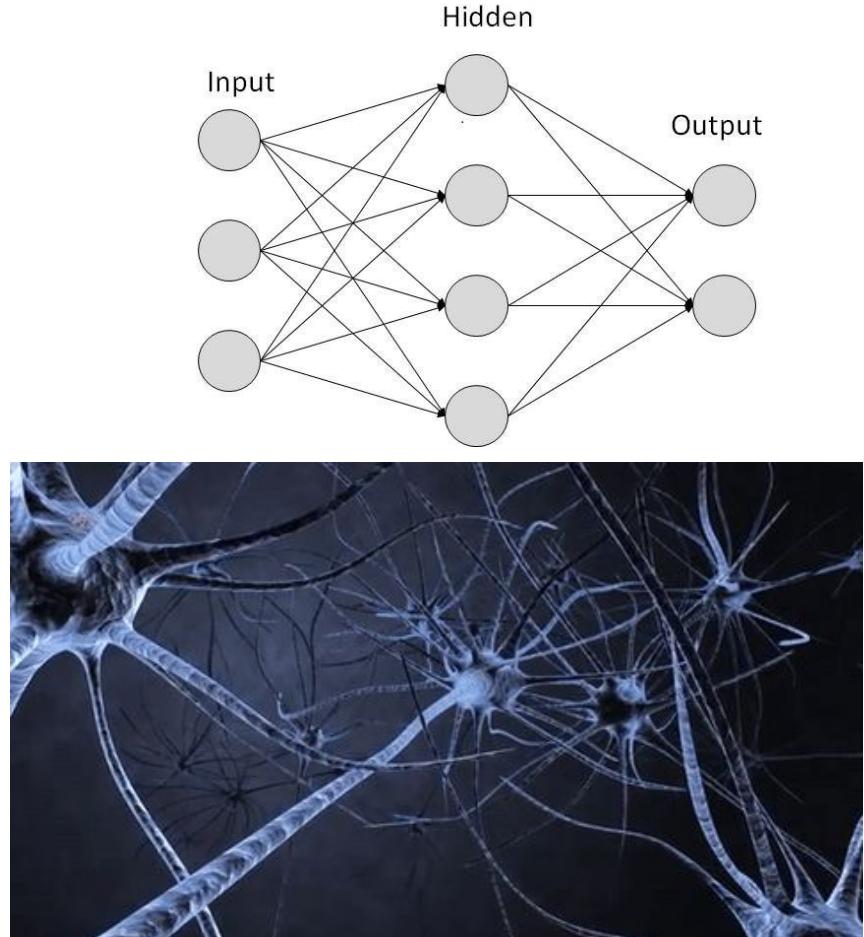
## Supervised Learning as Function Approximation



# ANNS - UNIVERSAL FUNCTION APPROXIMATORS



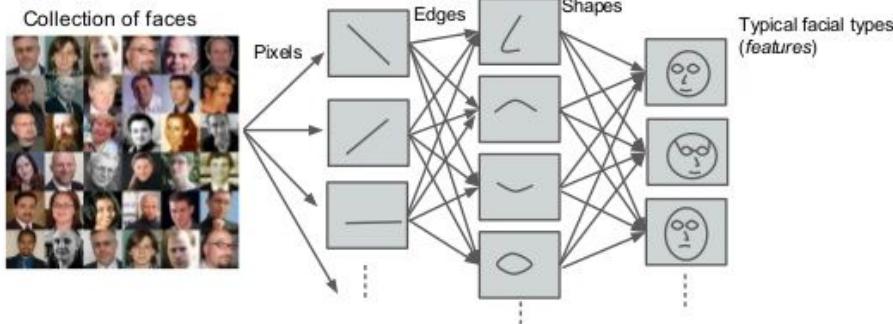
$$f(W_1 \cdot X_1 + W_2 \cdot X_2 + \dots + W_m \cdot X_m + \text{Bias}) \Rightarrow \text{Activation}$$



# WHAT IS DEEP LEARNING?

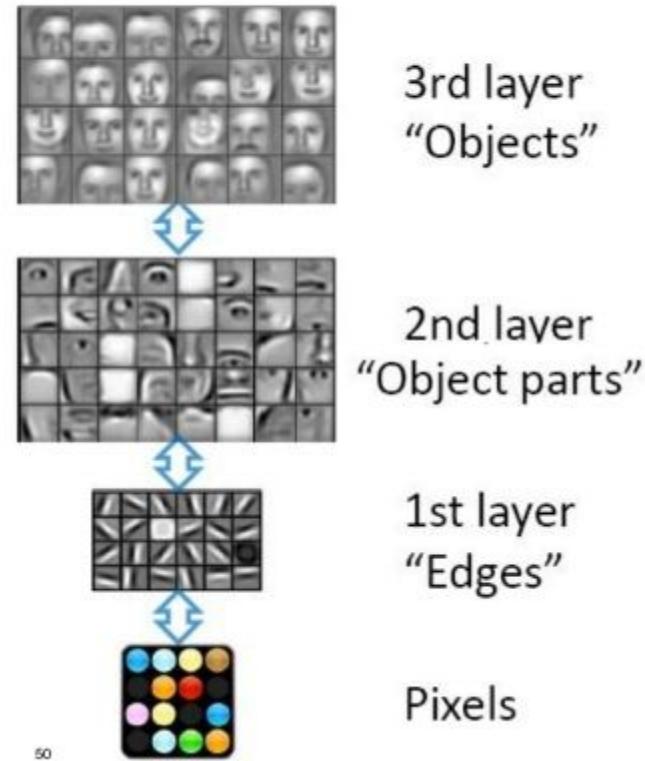
Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. **Handcrafting the features = Nightmare!**

## Distributed Representations



- Human brain uses distributed representations
- We can use deep learning to do the same thing with words (letters -> words -> phrases -> sentences -> ...)

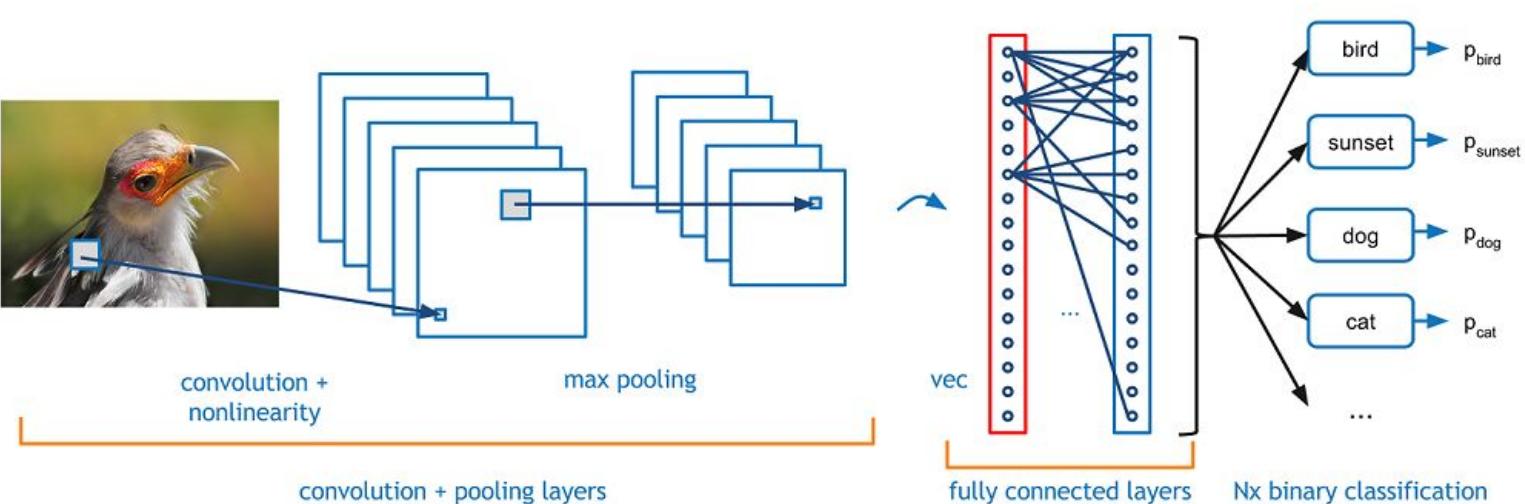
## Feature Representation



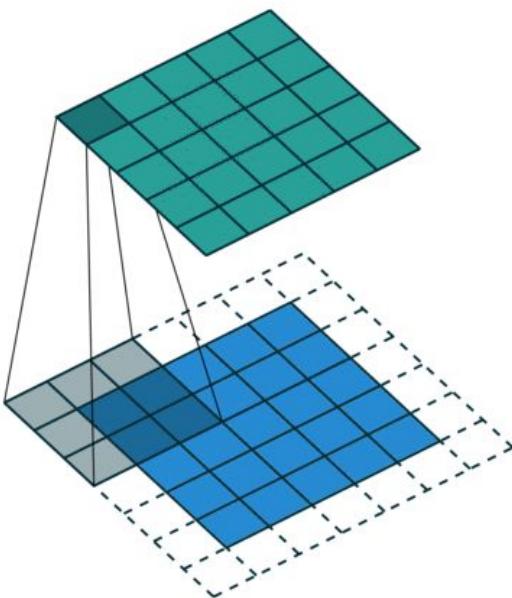
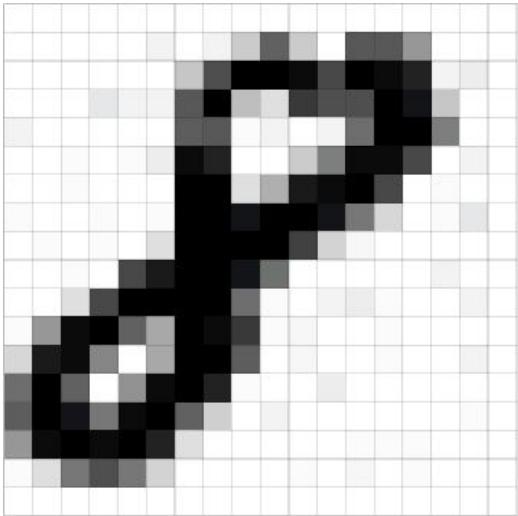
# CONVOLUTIONAL NEURAL NETWORKS

08 02 22 97 36 18 00 40 00 75 06 05 07 78 52 12 50 77 91 08  
49 49 99 40 17 81 18 57 60 87 17 40 98 43 69 48 04 86 62 00  
81 49 31 73 55 79 14 29 93 72 40 67 53 88 20 03 48 13 36 65  
52 70 95 23 04 60 11 42 69 24 68 56 01 32 56 71 37 02 34 95  
22 31 16 71 51 47 63 89 41 92 36 54 22 40 40 28 66 33 13 80  
24 47 32 60 99 03 45 02 44 75 33 53 78 34 84 20 35 17 12 50  
32 98 01 28 66 23 67 10 26 38 60 67 59 34 70 66 16 38 64 70  
67 24 20 68 02 62 12 20 93 63 94 39 63 04 40 91 64 49 94 21  
24 53 58 05 66 73 99 26 97 17 78 78 96 83 14 88 34 89 63 72  
21 34 23 09 75 00 76 44 20 45 35 14 00 63 33 97 34 31 33 95  
78 17 53 28 22 75 31 47 15 94 03 80 04 62 16 14 09 53 56 92  
16 39 05 42 96 35 31 47 55 58 82 24 00 17 54 24 36 29 85 57  
86 56 00 48 35 71 89 07 05 44 46 37 44 60 21 58 51 54 17 58  
19 80 81 68 05 94 47 69 28 73 92 13 86 52 17 77 04 89 55 40  
04 52 08 85 97 35 99 16 07 97 57 32 16 26 26 79 33 27 98 66  
88 34 68 87 57 62 20 72 03 46 33 67 46 55 12 32 63 93 53 49  
04 42 16 73 38 25 38 11 24 94 72 18 08 46 29 32 40 42 76 36  
20 69 36 41 72 30 23 88 34 62 99 69 82 67 59 85 74 04 36 16  
20 73 35 29 78 31 90 01 74 31 49 71 48 86 81 16 29 57 05 54  
01 70 54 71 83 52 58 69 16 92 48 61 43 52 01 89 19 67 48

What Computers See



# CONVOLUTIONS



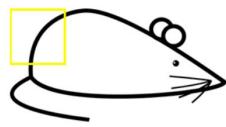
1 <small><math>\times 1</math></small>	1 <small><math>\times 0</math></small>	1 <small><math>\times 1</math></small>	0	0
0 <small><math>\times 0</math></small>	1 <small><math>\times 1</math></small>	1 <small><math>\times 0</math></small>	1	0
0 <small><math>\times 1</math></small>	0 <small><math>\times 0</math></small>	1 <small><math>\times 1</math></small>	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved  
Feature

# CONVOLUTIONS



Visualization of the filter on the image



Visualization of the receptive field

0	0	0	0	0	0	0	30
0	0	0	0	50	50	50	0
0	0	0	20	50	0	0	0
0	0	0	50	50	0	0	0
0	0	0	50	50	0	0	0
0	0	0	50	50	0	0	0
0	0	0	50	50	0	0	0

Pixel representation of the receptive field

\*

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter



Original image



Visualization of the filter on the image

0	0	0	0	0	0	30	0
0	0	0	0	30	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	30	0	0	0	0
0	0	0	0	0	0	0	0

Pixel representation of filter

\*

0	0	0	0	0	0	0	0
0	40	0	0	0	0	0	0
40	0	40	0	0	0	0	0
40	20	0	0	0	0	0	0
0	50	0	0	0	0	0	0
0	0	50	0	0	0	0	0
0	25	0	50	0	0	0	0
0	0	0	0	0	0	0	0

Pixel representation of receptive field

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter

Multiplication and Summation =  $(50*30)+(50*30)+(50*30)+(20*30)+(50*30) = 6600$  (A large number!)

Multiplication and Summation = 0



Visualization of a curve detector filter

# TRANSLATION INVARIANCE

Test  
Image #1



Prediction from  
our network

100% an "8"!

Test  
Image #2



Prediction from  
our network

100% not an "8"!

Test  
Image #1



Prediction from  
our network

No idea!?!?

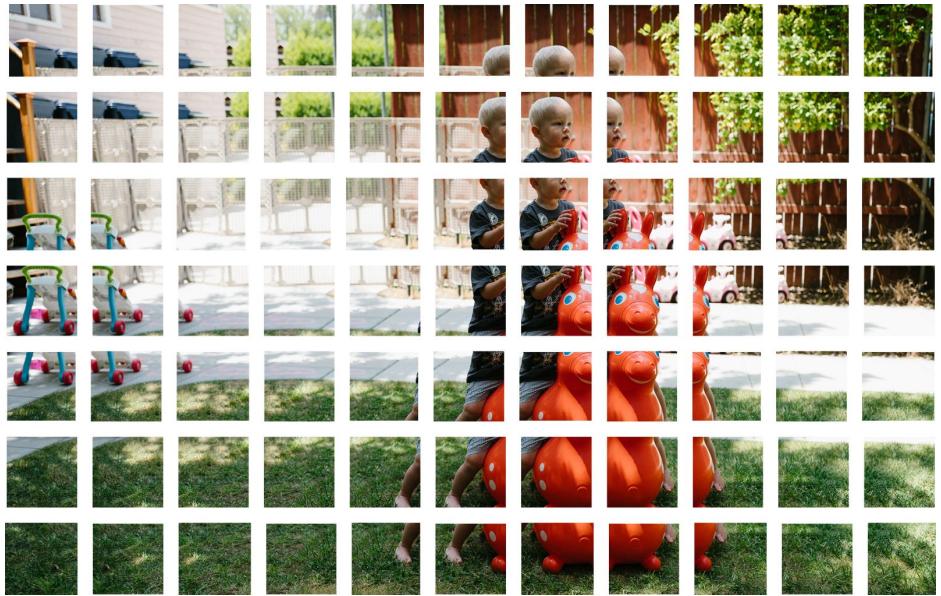
Test  
Image #2



Prediction from  
our network

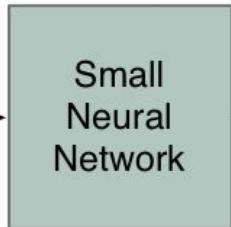
What is this?!@

# TILING AND POOLING

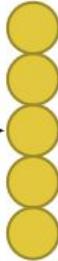


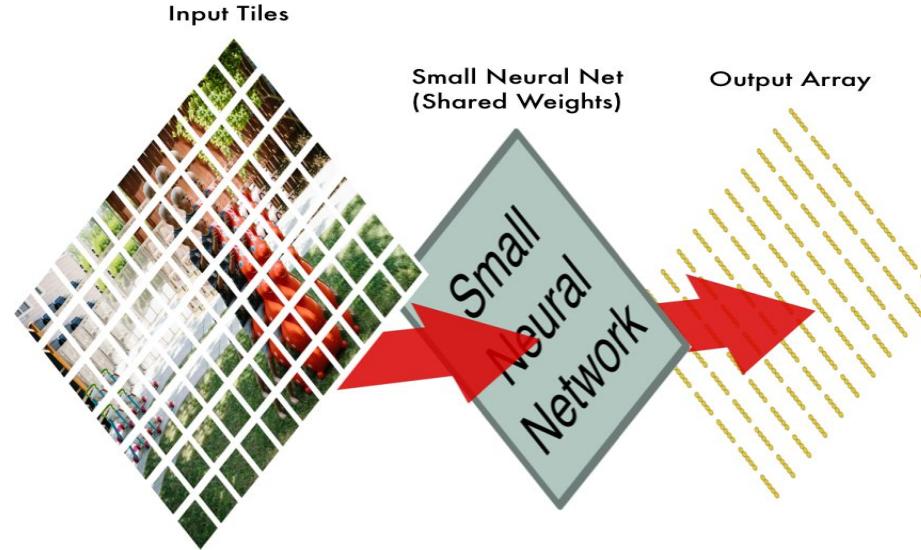
Processing a single tile

Input Tile

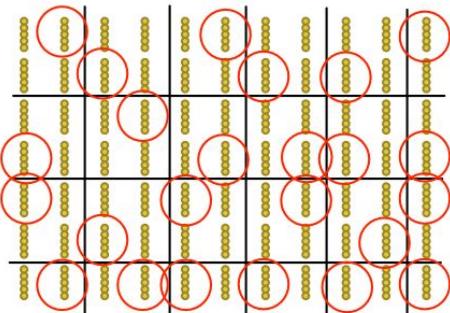


Outputs

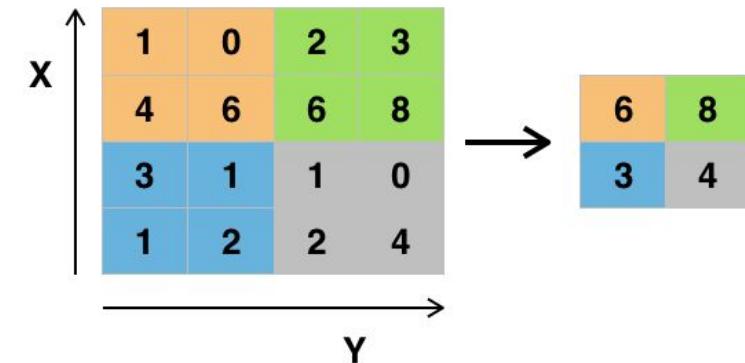




Find the max value in each grid square in our Array

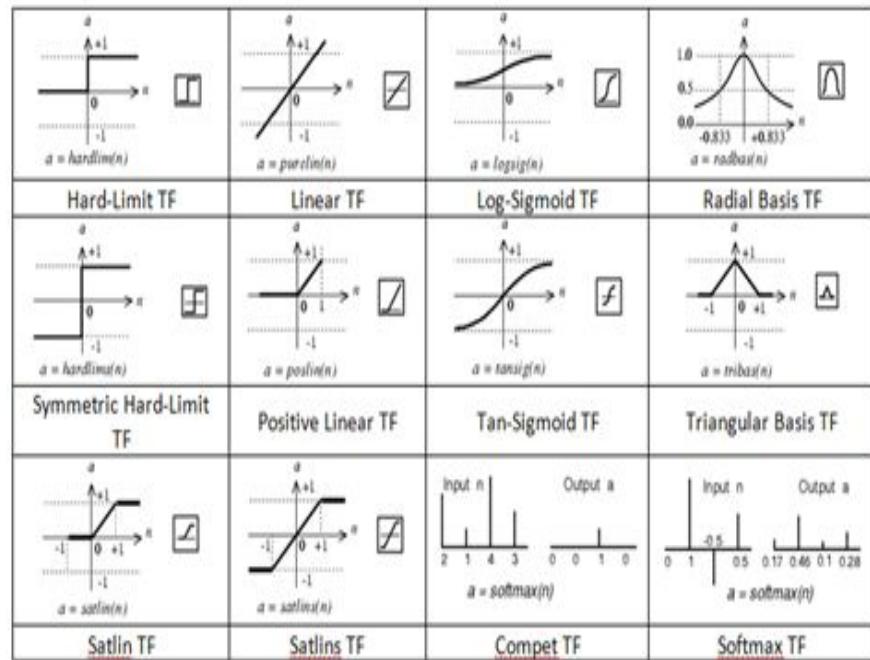
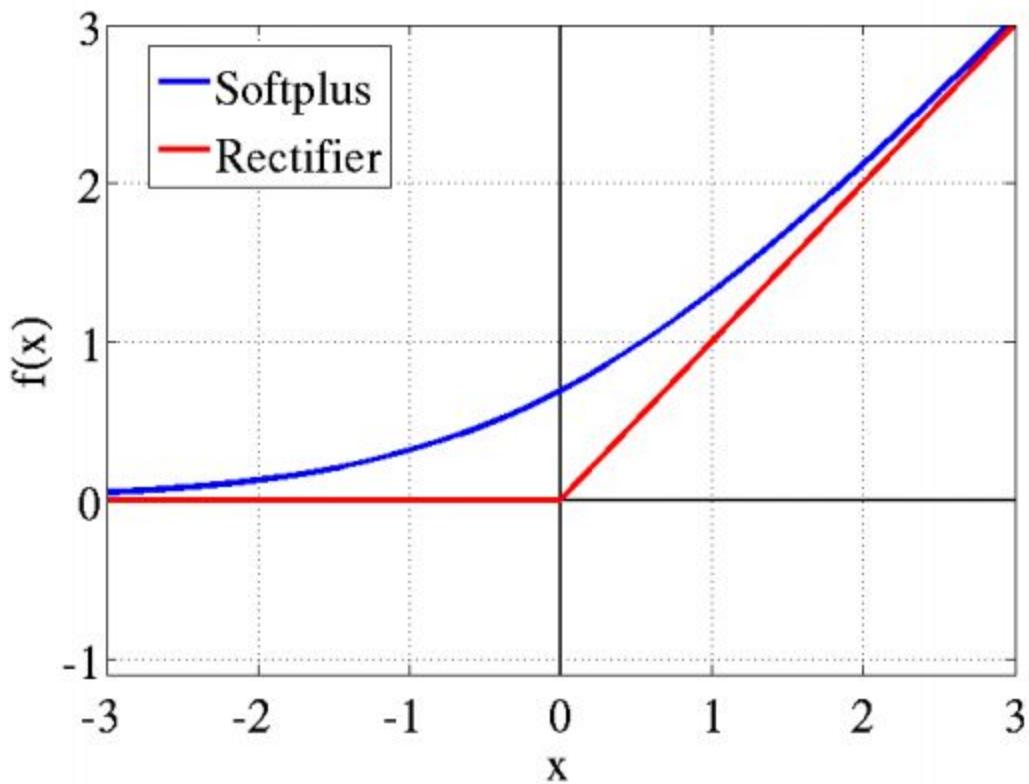


Max-pooled array

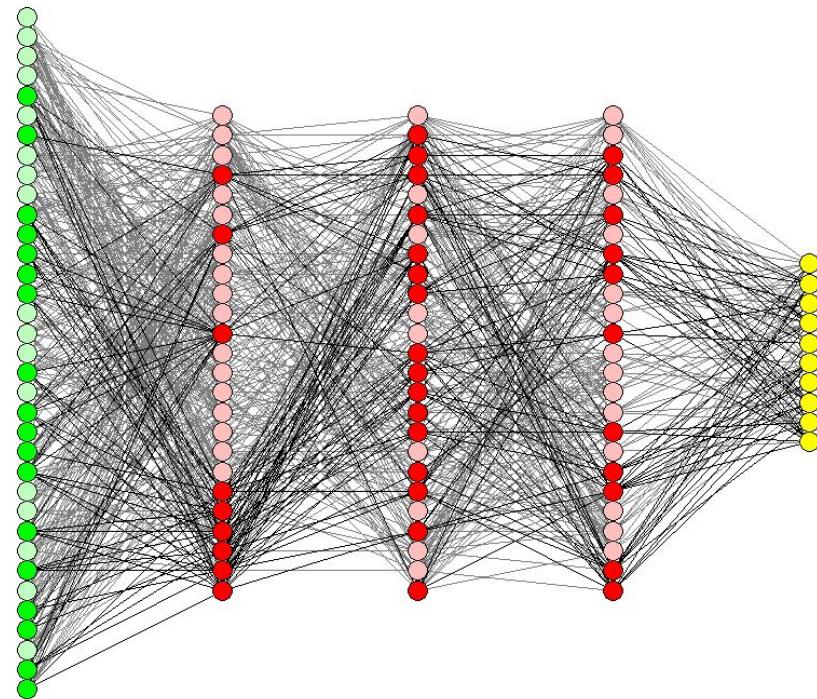
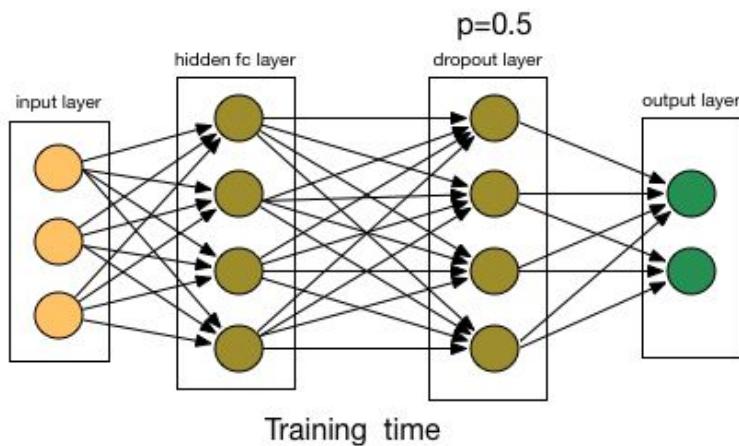


Example of Maxpool with a 2x2 filter and a stride of 2

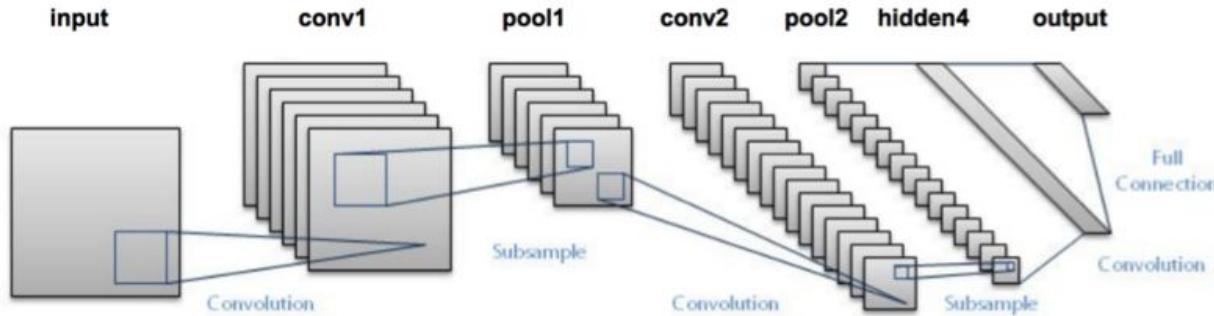
# NON LINEARITY



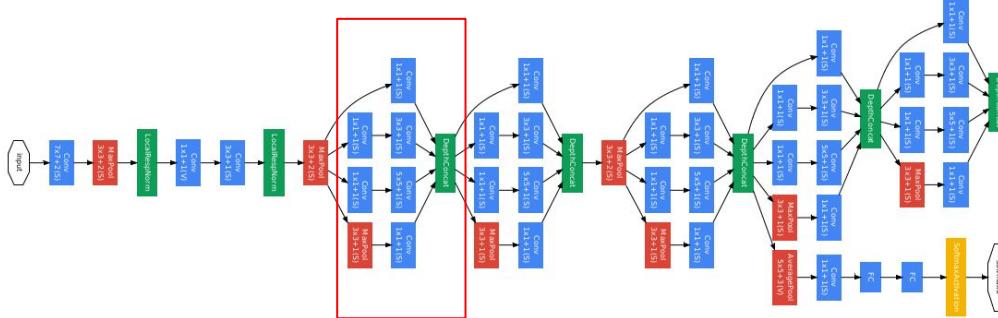
# DROPOUTS



# CNN ARCHITECTURES

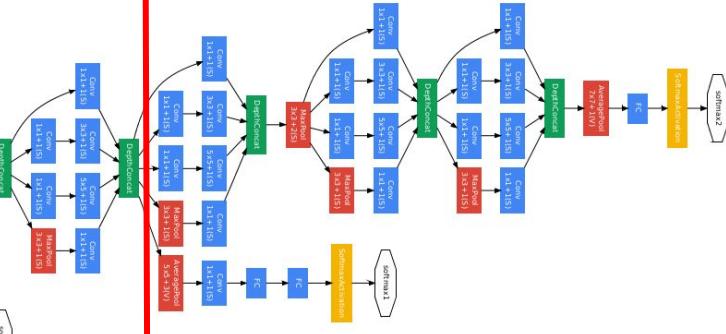


ImageNet weights - Frozen (Transfer Learnt)



Inception Module

ImageNet weights - Initialized

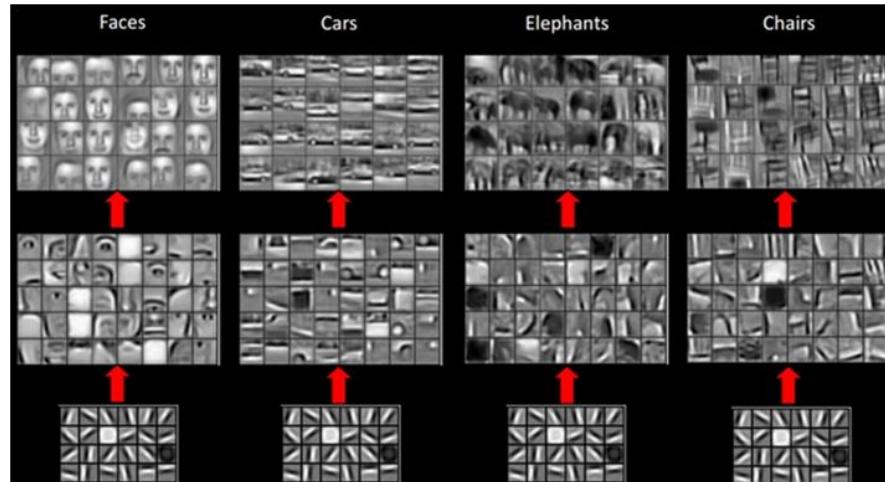
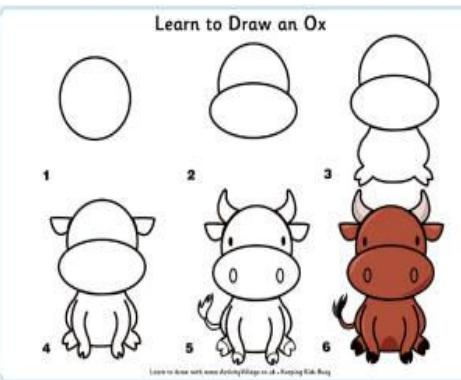
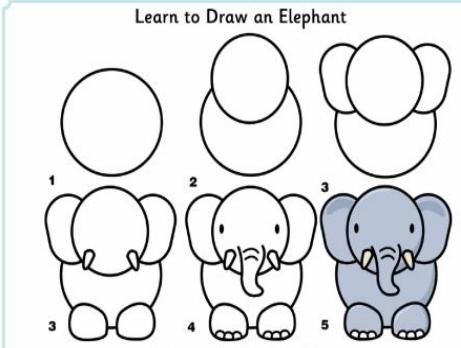
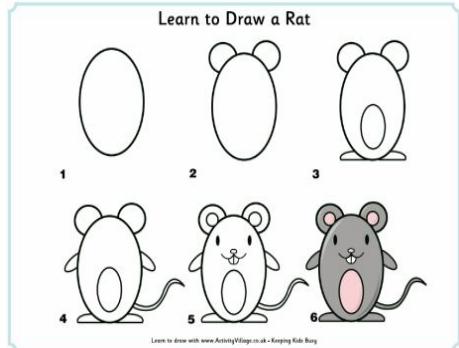


Learning Rate - 2X (Fine-tuned)

# JUST FANCY MATRIX MULTIPLICATION !



# TRANSFER LEARNING AND FINE TUNING



# IMAGENET AND MIT PLACES DATASETS



veterinarians office



elevator door



fishpond



bedroom



cafeteria



watering hole



staircase



bar



field road



conference center



shoe shop



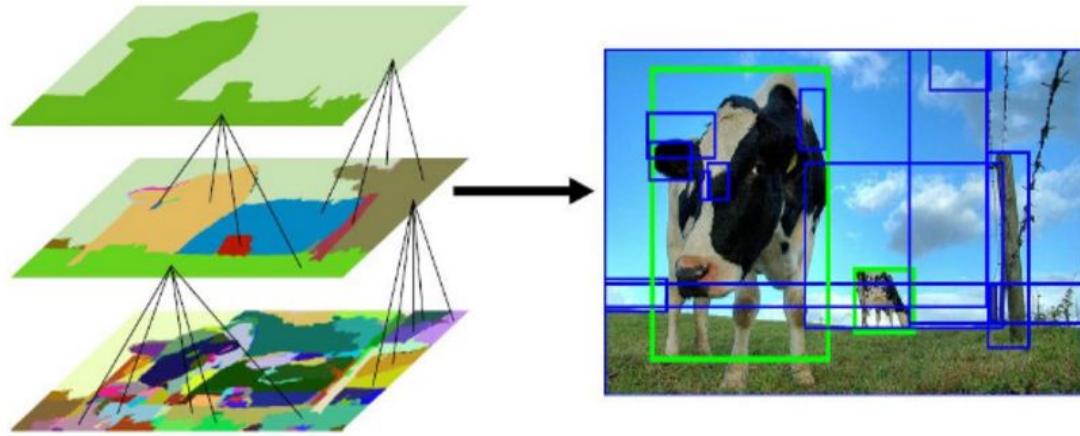
rainforest



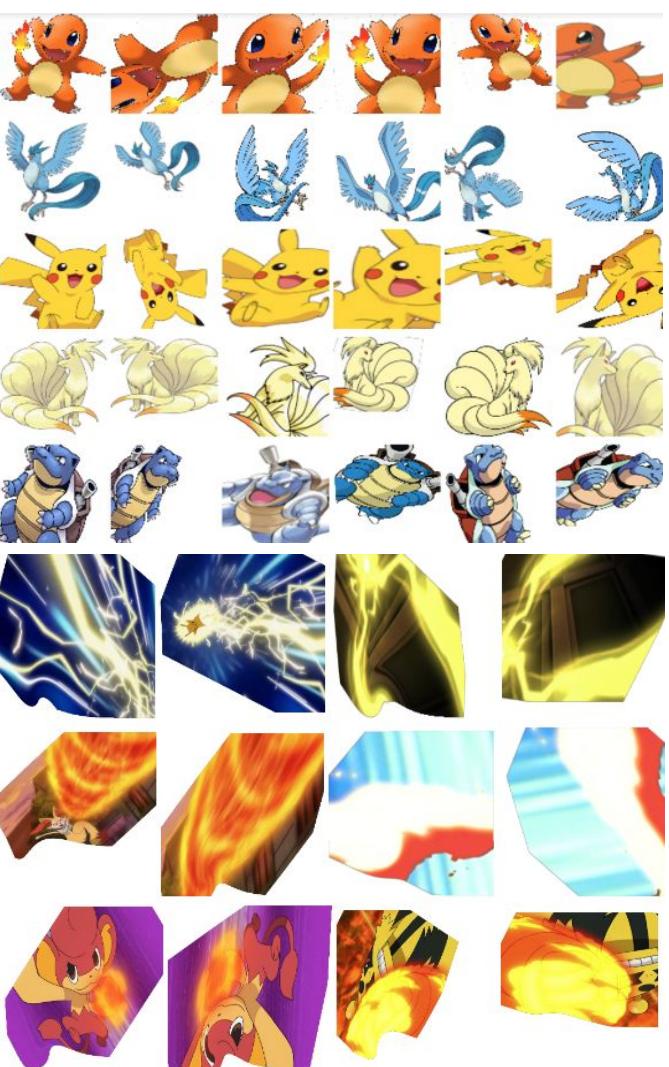
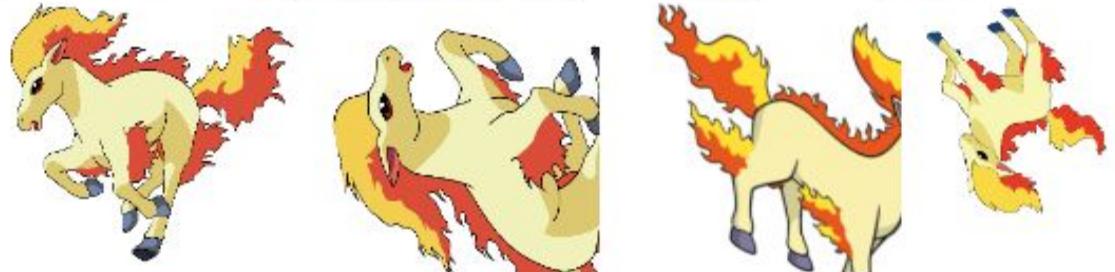
# SELECTIVE SEARCH

Region proposal algorithm - proposes an array of all possible regions, likely to hold objects of interest.

- Combines the advantages of exhaustive search and segmentation
- Hierarchical grouping of the region proposals based on color, texture, size and fills
- Grid search over the four Selective Search parameters to use the Scale, Min Size, Min Area, and Sigma



# DATA AUGMENTATION



ABOUT BOTH THE WORLDS

# POKEMON

What are they?

How many are there?



Pikachu



Squirtle



Bulbasaur

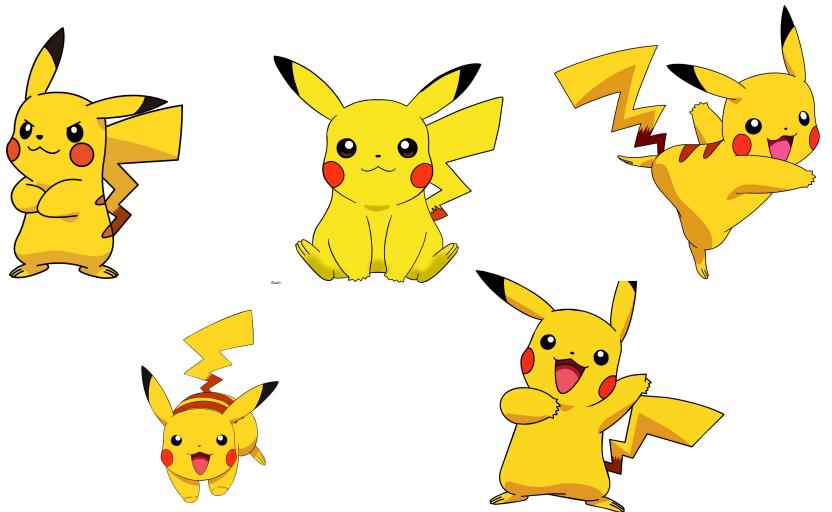


Charmander



Ash

# WHY SOLVE IMAGE CAPTIONING?



Poison Sting



Fire Punch



# THE POKEMON BATTLE IMAGE CAPTIONING PROBLEM



**Sub:** Go Pikachu ! Use Thunder Bolt !

**(Note:** Only one Pokemon and One Action in image)



**Sub:** Pikachu, Use Iron Tail on Roselia !

**(Note:** Two Pokemon and One Action in image)

# INDUS VALLEY CIVILIZATION

- Largest and one of the most ancient civilizations known to mankind
- Major Indus valley sites: (Northwestern regions of India)
  - Harappa, Mohenjo-Daro, Chanhudaro, Lothal, Kalibangan
- Inscription's form factor:
  - Stamp seals and Sealings
  - Amulets, stone tablets, pottery ...
- The script:
  - Around 3700 text inscriptions with 417 known graphemes according to the M77 corpus, with an average of 5 graphemes per text.

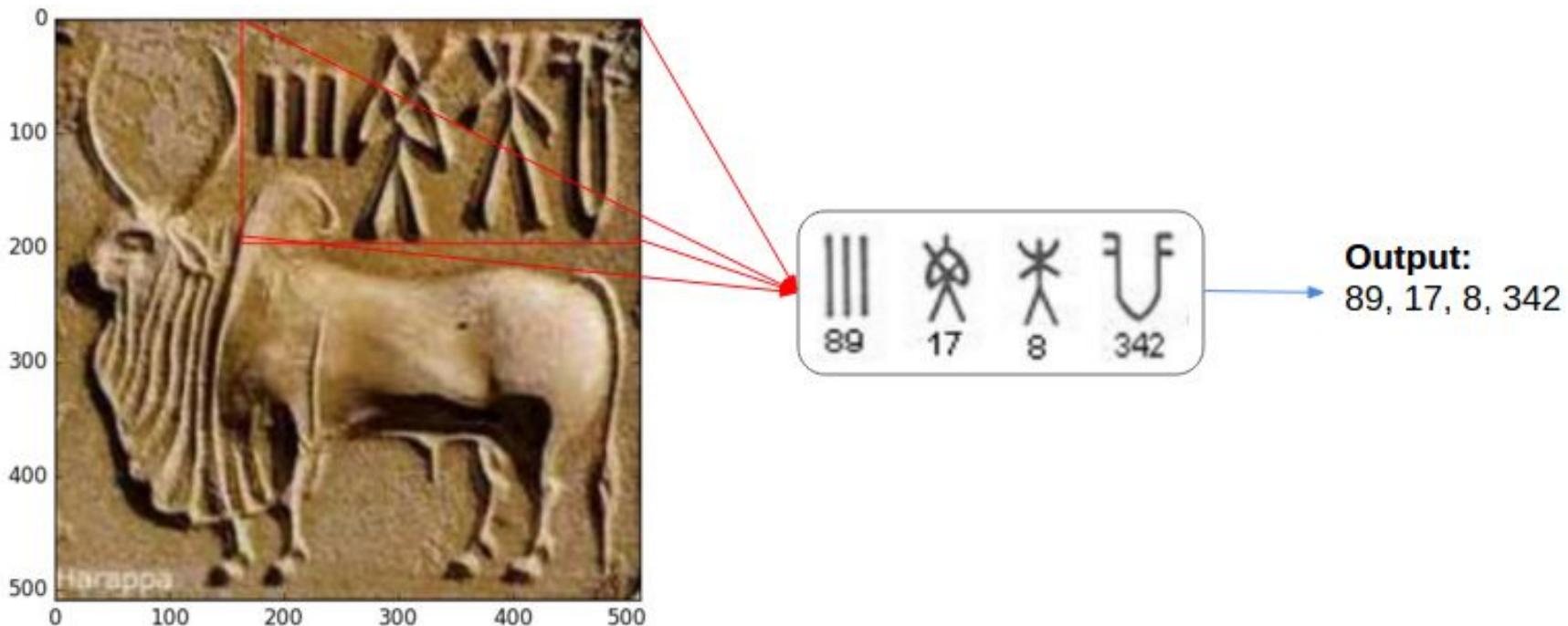
# SAMPLE INDUS SEALS



# WHY SOLVE OCR?

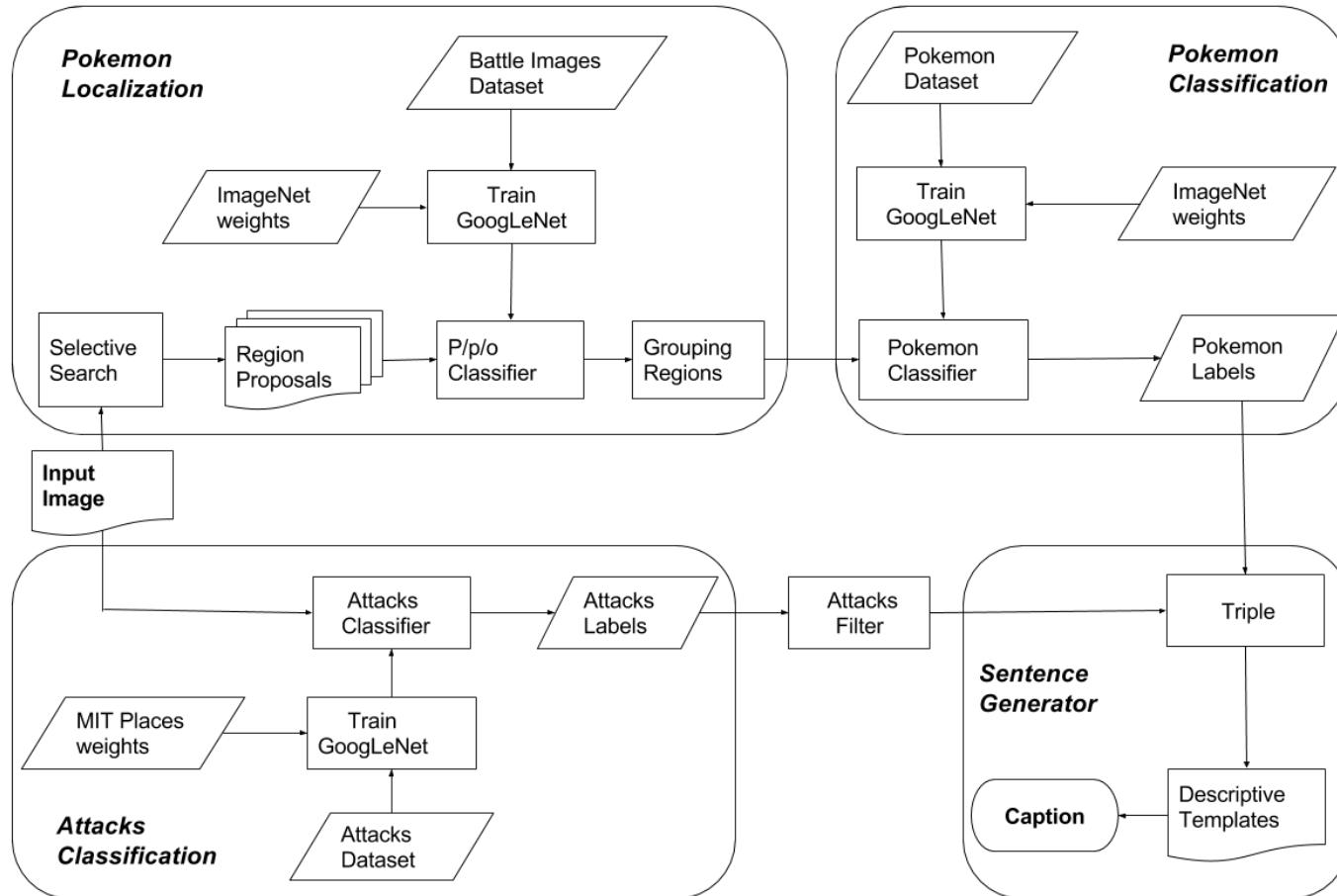
- Why still undeciphered?
  - Paucity of long texts (rarely with 14 graphemes per text)
  - Absence of parallel or bilingual text
  - No definite knowledge about the underlying language
  - The number of graphemes and the very less data (417 symbols with just 3700 texts documented)
- Corpus Formulation:
  - We have been using a 40 years old dataset, with no recent updates
  - Challenges
    - Extremely laborious human (expert) effort
    - Time consuming to standardize for use
    - Other political issues

# THE OCR ENGINE PROBLEM

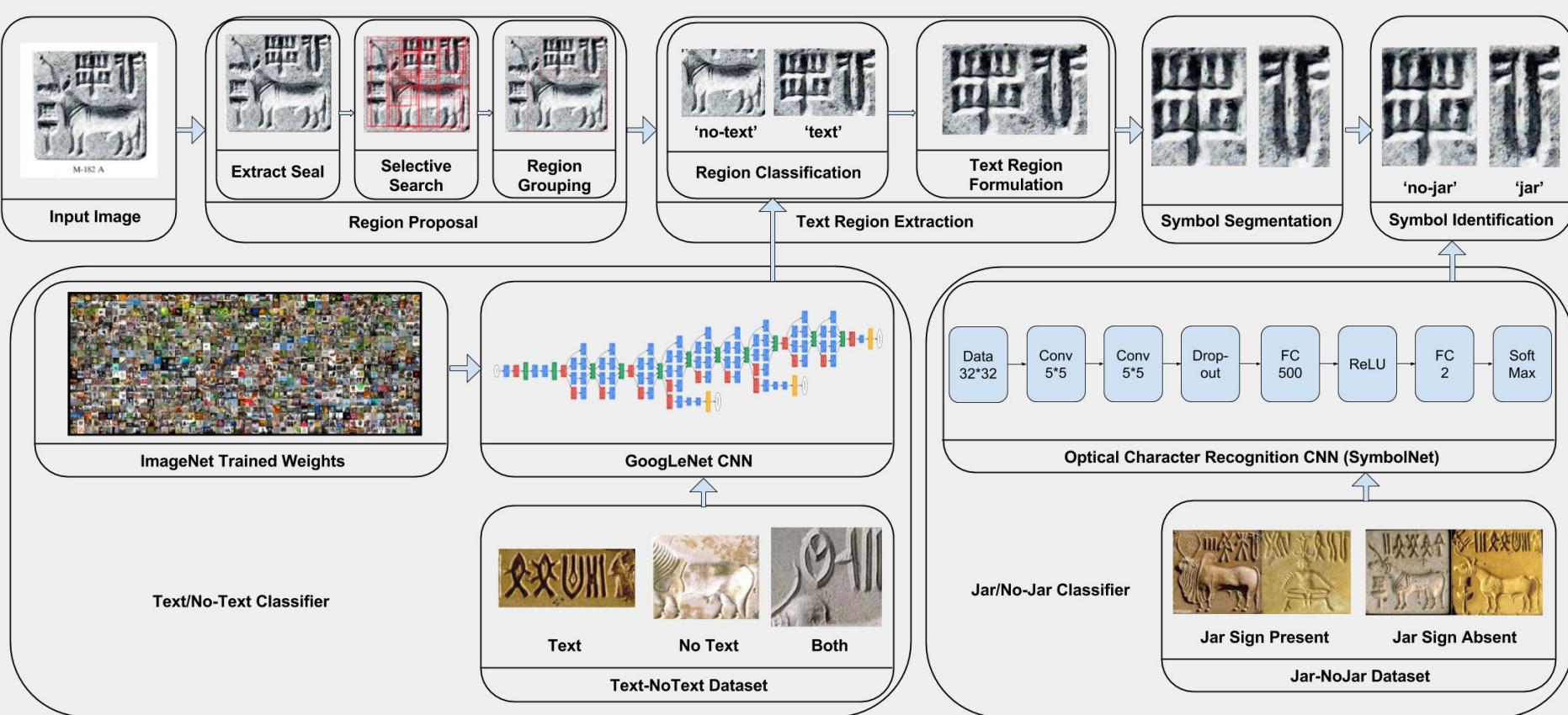


THE PIPELINES

# THE POKEMON BATTLE IMAGE CAPTIONING PIPELINE

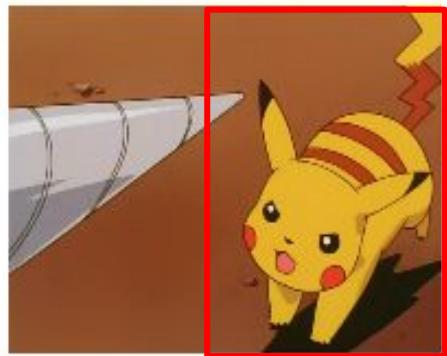
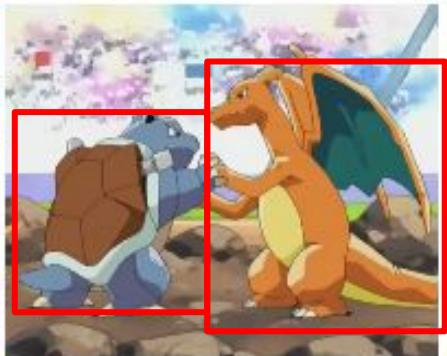


# THE INDUS OCR ENGINE PIPELINE



THE THREE ANALOGIES

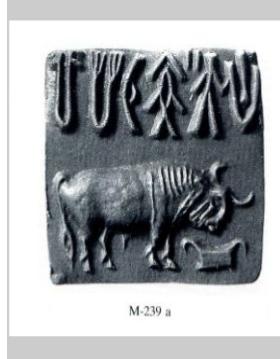
# ANALOGY 1 - OBJECT LOCALIZATION



H-47 a

L-42 a

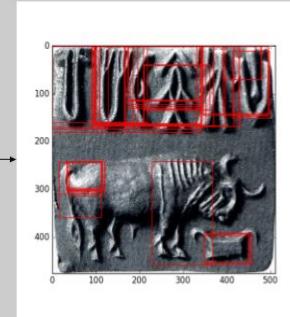
# OBJECT LOCALIZATION - SELECTIVE SEARCH



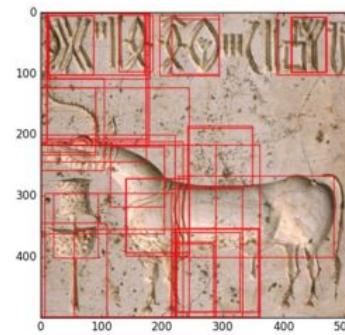
Indus Seal Scan



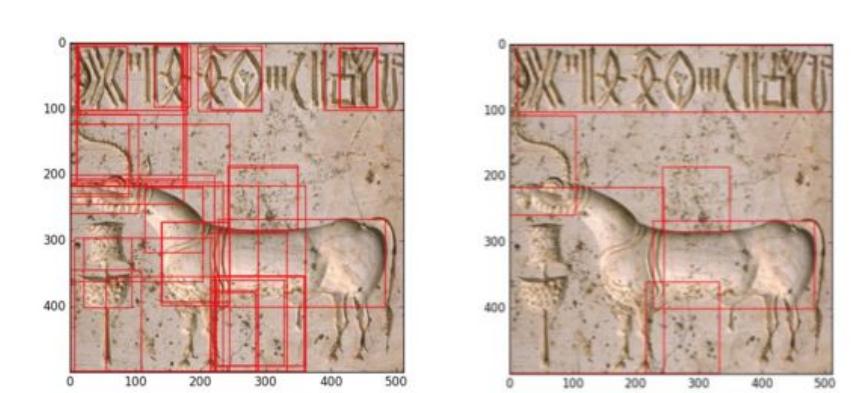
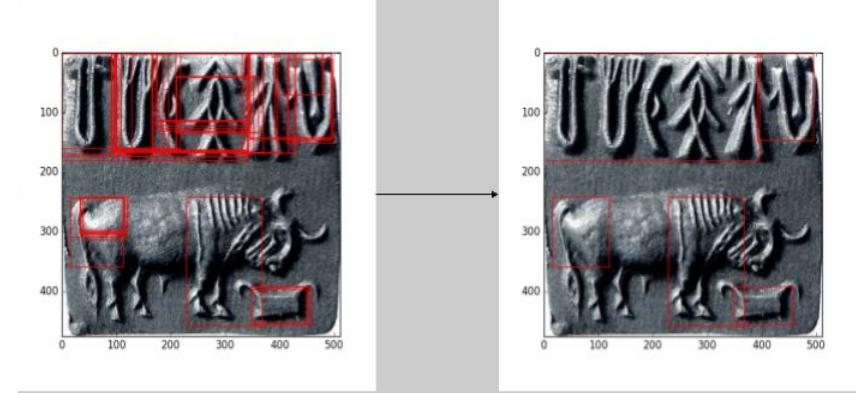
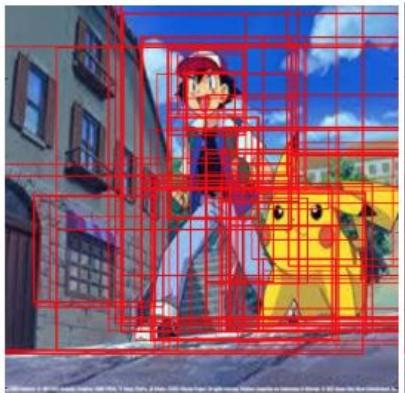
Extract Seal



Selective Search



# OBJECT LOCALIZATION - REGION GROUPING



# OBJECT LOCALIZATION - FULL/PART/OTHER CLASSIFIER

Full (Pokemon)



Part (Partly Pokemon)



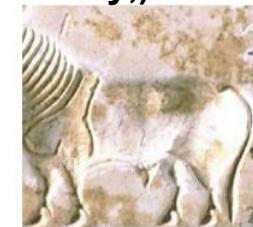
Other (BG, Attack)



Full (Text)



Other  
(Animals,  
Deity,)



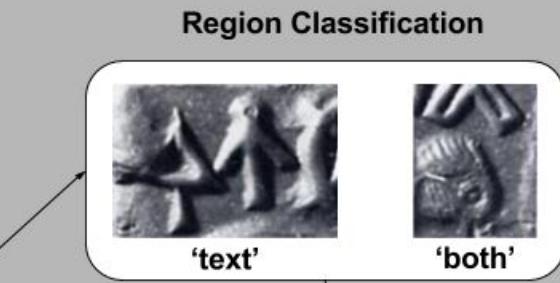
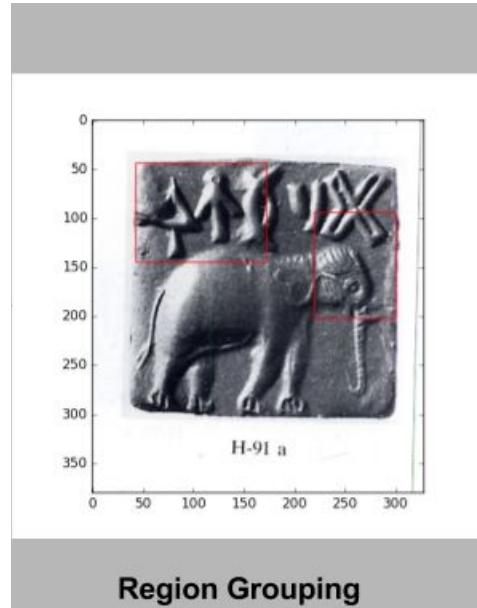
Part (Both /  
Partly Text)



Accuracy: 74.9%

Accuracy: 89.3%

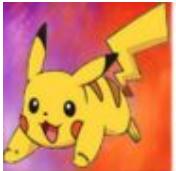
# OBJECT LOCALIZATION - FULL/PART/OTHER GROUPING



Text Region Formulation

# ANALOGY 2 - CLASSIFICATION PROBLEM

Pokemon



Pikachu



Pidgeotto



Rapidash



Snorlax

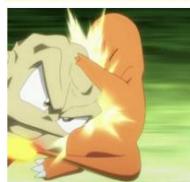
Attack



Blizzard



Water Gun



Tackle



Thunder Wave

Symbol



Sign  
89



Sign  
17



Sign  
8

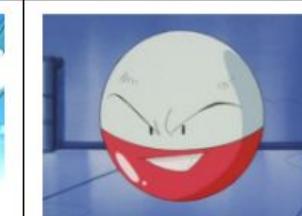


Sign  
342

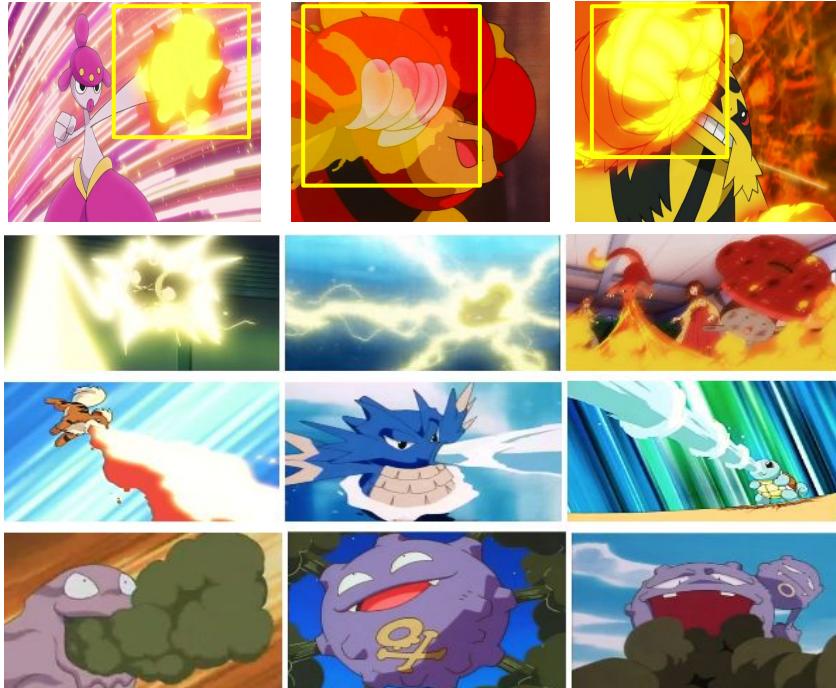
# POKEMON CLASSIFICATION

Levels of testing in the GoogleNet	Augmented Dataset		Backgrounded and Augmented Dataset	
	Top1	Top5	Top1	Top5
Level 1 ( $\frac{1}{3}$ rd of Network)	74.39%	91.8%	79.22%	93.99%
Level 2 ( $\frac{2}{3}$ rd of Network)	76.28%	92.3%	82.06%	95.32%
Level 3 (Full Network)	85.39%	96.43%	91.16%	<b>98.03%</b>

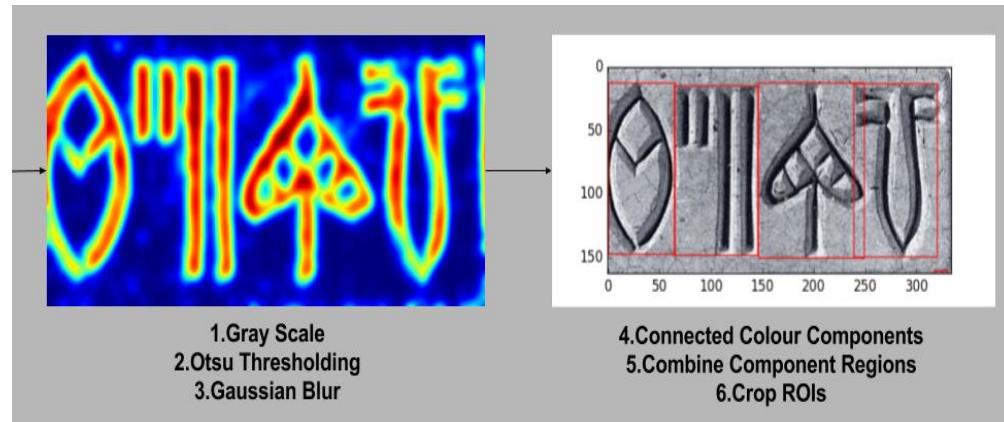
		
'Electrode' 'Voltorb' 'Chansey' 'Horsea' 'Gengar'	'Blastoise' 'Rhydon' 'Rhyhorn' 'Golduck' 'Wartortle'	'Haunter' 'Mankey' 'Nidorino' 'Arcanine' 'Seadra'

		
'Articuno' 'Venomoth' 'Vaporeon' 'Mew' 'Fearow'	'Arcanine' 'Charizard' 'Fearow' 'Charmander' 'Gyarados'	'Voltorb' 'Kakuna' 'Diglett' 'Wigglytuff' 'Exeggute'

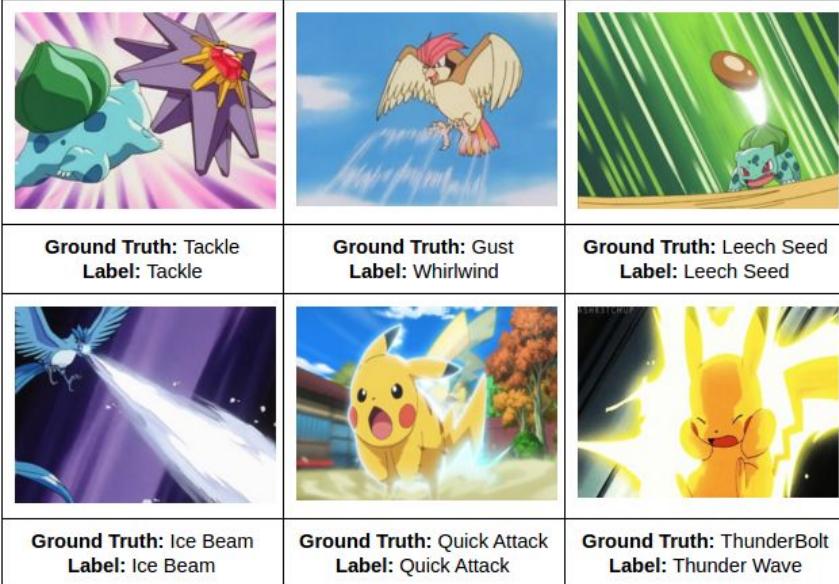
# CLASSIFICATION PROBLEM - ATTACK AND SYMBOL LOCALIZE



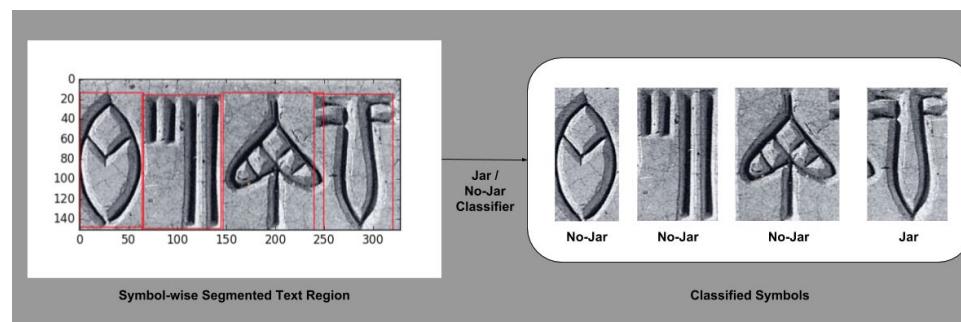
Wait, all these can't be detected 😱



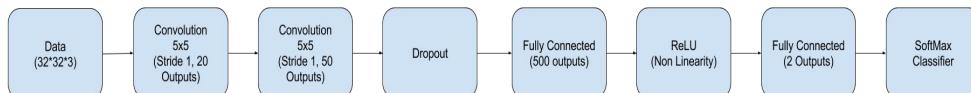
# CLASSIFICATION PROBLEM - ATTACK AND SYMBOL CLASSIFY



Accuracy: 55% (Top 1) & 88% (Top 5)



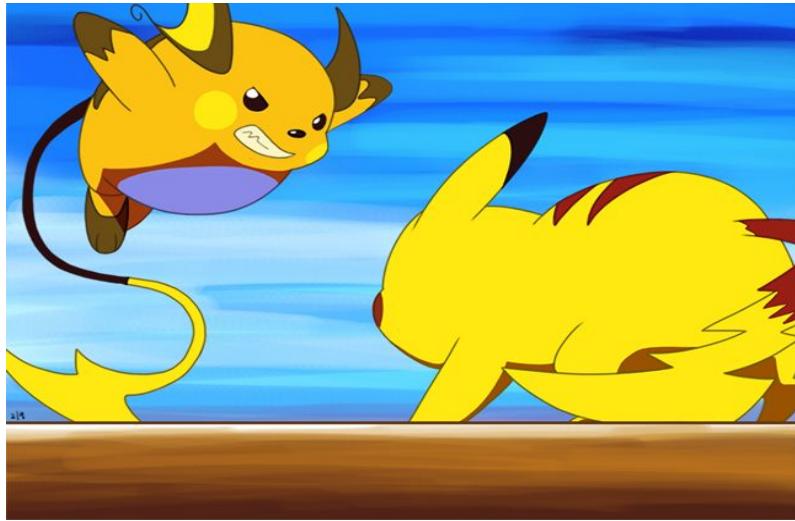
Accuracy: 92% (SymbolNet)



# ATTACK FILTERING (NOT REACHED THIS STAGE IN INDUS OCR)

- Each Pokemon is capable of executing only a small number of attacks, and hence we formed a database of Pokemon-Attacks mapping, and used this in removing attacks which cannot be executed by the Pokemon identified in the image.
  - **Pikachu** - ThunderBolt, ThunderWave, VoltTackle, IronTail, Tackle, Agility, Thunder, TailWhip, Growl, ElectroBall
  - **Charmander** - FlameThrower, Tackle, Ember, TailWhip, Growl, Scratch, SmokeScreen, FlameBurst, FireSpin, Inferno
  - **Squirtle** - WaterGun, Withdraw, Tackle, TailWhip, WaterPulse, HydroPump, IronDefense, SkullBash, Bubble, AquaTail

# ANALOGY 3 - SEQUENCE GENERATION PROBLEM (NAIVE!)



'*<Raichu> executing <body slam> on <Pikachu>*'  
*<Attacker> <Attack> <Target>*

# TEXT CAPTION GENERATION



**Triple:**

<pokemonA, pokemonB, attackC>  
<Attacker, Target, Relationship>

SpokemonA attacks SpokemonB using \$attackC  
SpokemonA attacked SpokemonB using \$attackC  
SpokemonA is attacking SpokemonB using \$attackC  
SpokemonA uses \$attackC on SpokemonB  
SpokemonA used \$attackC on SpokemonB  
SpokemonA is using \$attackC on SpokemonB  
SpokemonB is attacked by SpokemonA using \$attackC  
SpokemonB was attacked by SpokemonA using \$attackC  
SpokemonB is being attacked by SpokemonA using \$attackC  
\$attackC is used by SpokemonA on SpokemonB  
\$attackC was used by SpokemonA on SpokemonB  
\$attackC is being used by SpokemonA on SpokemonB  
SpokemonA has attacked SpokemonB using \$attackC  
SpokemonA has used \$attackC on SpokemonB  
SpokemonB has been attacked by SpokemonA using \$attackC  
\$attackC has been used by SpokemonA on SpokemonB  
SpokemonA attacks SpokemonB utilizing \$attackC  
SpokemonA attacked SpokemonB utilizing \$attackC  
SpokemonA is attacking SpokemonB utilizing \$attackC  
SpokemonA utilizes \$attackC to attack SpokemonB  
SpokemonA utilized \$attackC to attack SpokemonB  
SpokemonA is utilizing \$attackC to attack SpokemonB  
SpokemonB is attacked by SpokemonA utilizing \$attackC  
SpokemonB was attacked by SpokemonA utilizing \$attackC  
SpokemonB is being attacked by SpokemonA utilizing \$attackC  
\$attackC is utilized by SpokemonA to attack SpokemonB

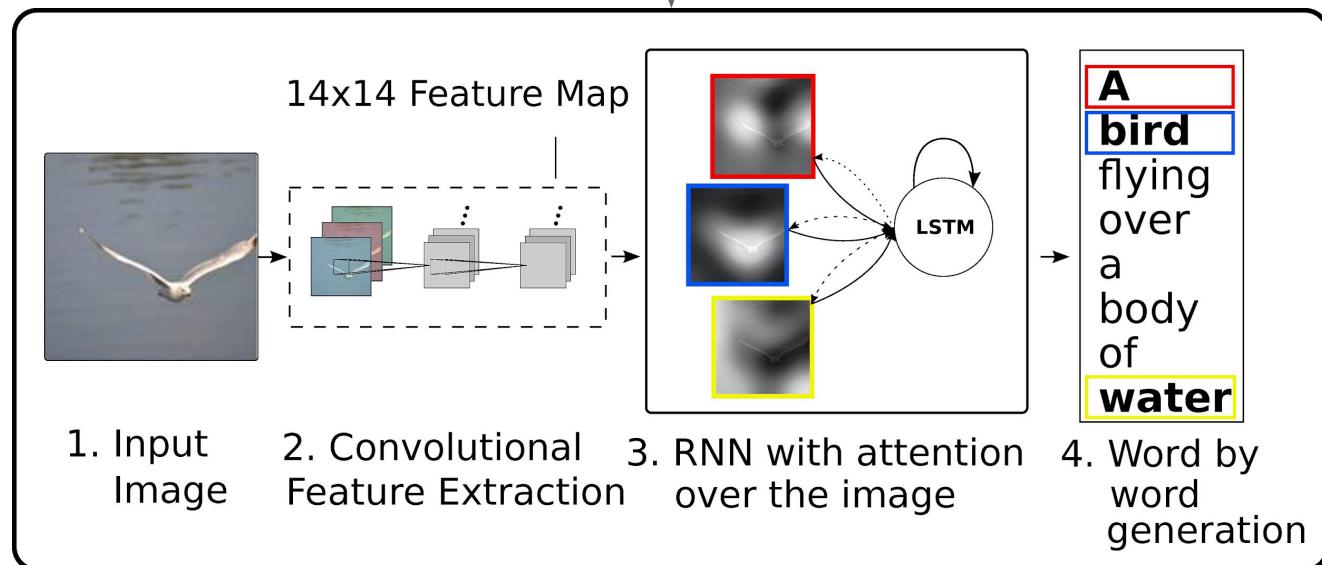
# SYMBOL SEQUENCE GENERATION (PLAN)

Input:



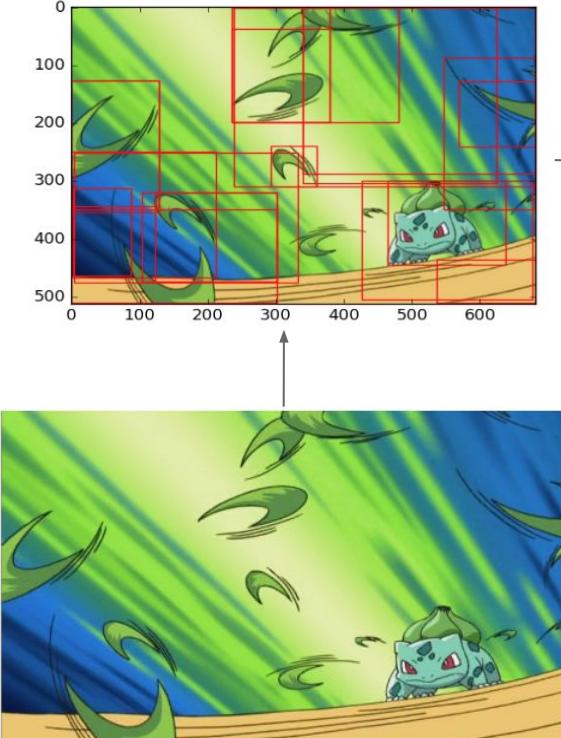
Expected Output:

89, 17, 8, 342

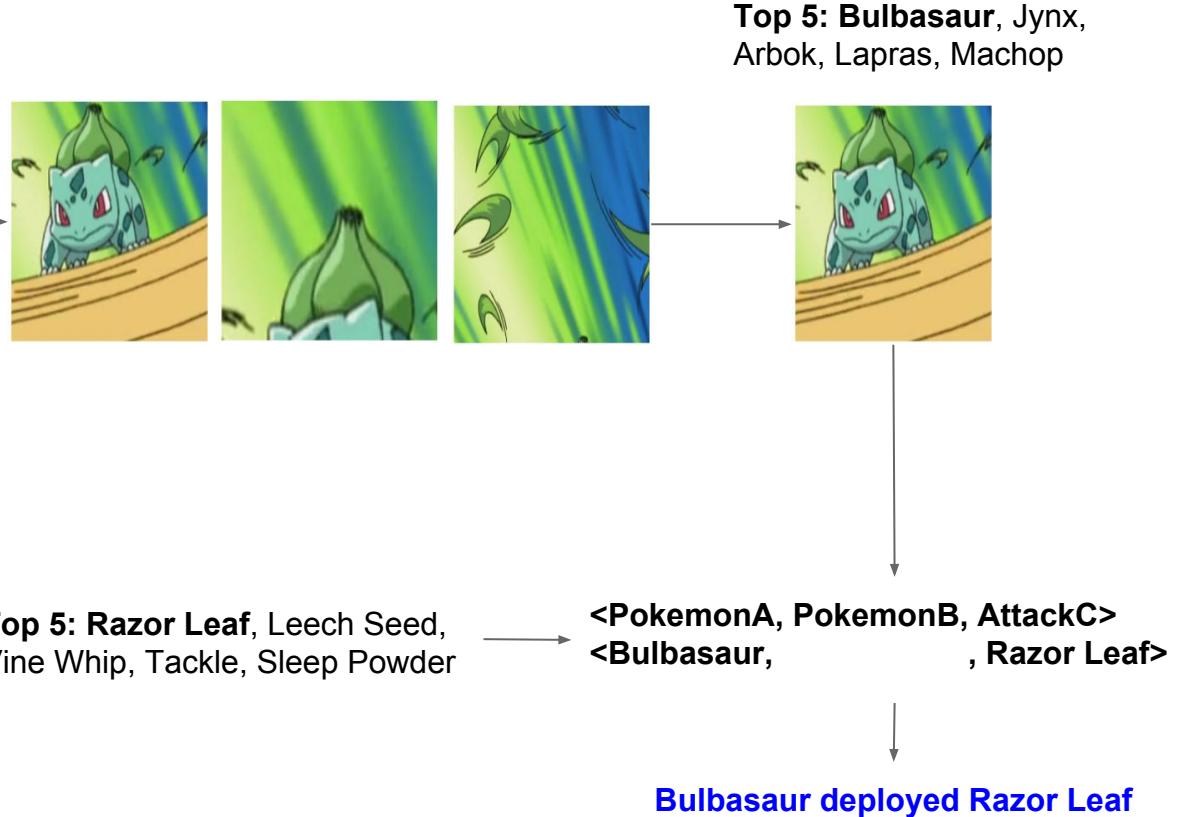


# PIPELINE FLOW RECAP

# IMAGE CAPTIONING FLOW

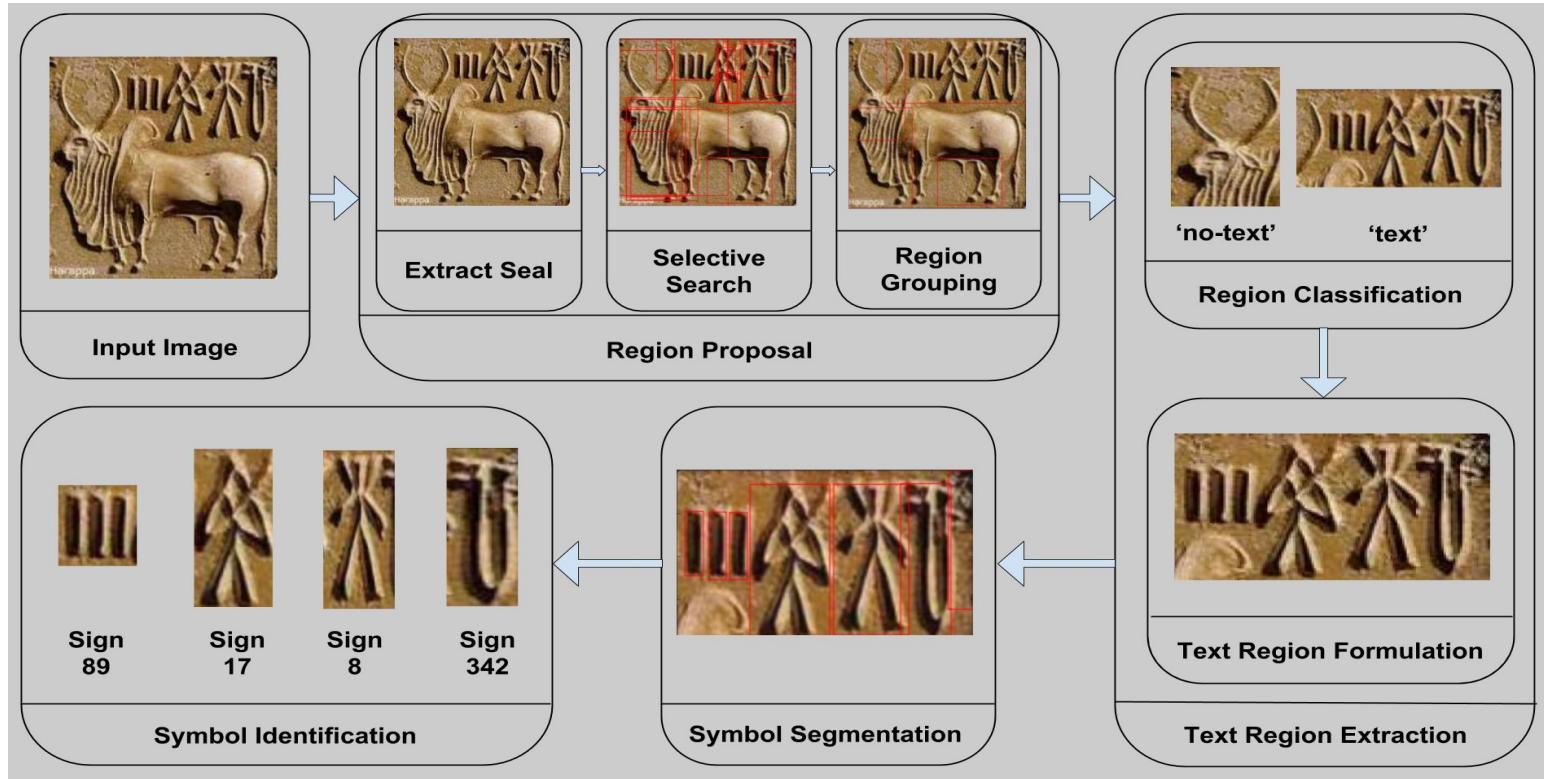


Input



Output

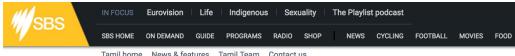
# OCR FLOW



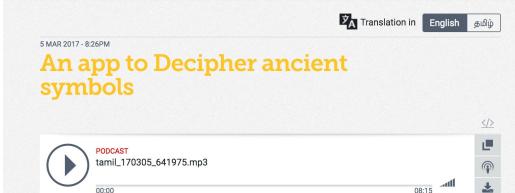
# THE BUZZ ! (NEVER FOR POKEMON !)



Meanwhile, Ronjooy Adhikari, a physics professor at The Institute of Mathematical Sciences in Chennai, India, and his research associate Satish Palaniappan are working on a program that can accurately extract symbols from a photo of an Indus artifact. "If an archaeologist goes to an Indus site and finds a new seal, it takes a lot of time for those seals to actually be mapped and added to a database if it's done manually," says Palaniappan. "In our case the ultimate aim is just with a photograph of a particular seal to be able to extract out the text regions automatically." He and Adhikari are working on building an app that archaeologists can bring to a site on a mobile device that will extract new inscriptions instantly.



Tamil home News & features Tamil Team Contact us



A professor from Institute of Mathematical Sciences (IMSc), Chennai, and an engineer have developed an app that will allow archaeologists and amateur history buffs alike to, say, capture images of seals on pottery and share it online via the app to assist experts devoted to the recognition and transcription of the script. It will also provide an approximate date by recognition of the iconography and its style... Satish Palaniappan, an SSN College of Engineering graduate who worked on the app talks to Kulasegaram Sanchayan about it.

## Researchers Look To Widen Script Database, Solve Mystery New app could help decipher ancient Indus Valley symbols

DAWN OF TIME

The computer application can be used to identify elements belonging to the Indus script

SHUBASHREE DESIKAN

**A**s the Egyptian civilisation flourished, Indian seal engravers documented the rise and fall of one of man's greatest civilisations in the Indus Valley, c.3,500BC-1,300BC, another great civilisation arose in the Indus Valley in the northwest of the Indian subcontinent.

Much less is known about the Indus Valley Civilisation than the Egyptian one, however, about its development, governance, activities, discoveries and daily life – because the Indus script is extremely short on information and yet to fully interpret the script of these ancient symbols.

To the common man, however, the limited corpus of hieroglyphs and other symbols that have been found from the Indus Valley bear an uncanny resemblance to those found along River Nile. Scientists believe that the Indus and Dravidian languages and an early form of Sanskrit but its meaning is still unknown.

Here's a discovery, however, that could help change that.

Artificial intelligence involves the use of self-driving cars, which mimics the functions of the human brain, may now aid researchers to develop a large enough database for Indus script that could eventually help decipher the texts. Scientists are also working on a mobile application for the software.

The technology will allow archaeologists and amateur history buffs alike to, say, capture images of seals on pottery and share them online via the app to assist experts devoted to the recognition and transcription of the script. It will also provide an approximate date by recognition of the iconography and its style. The app will fit the text from the image and identify the presence or absence of individual characters in an existing database. If it is a known symbol, the app will display a number representing each character in the texts in

the database if not, it will include the symbol in the database. The output will be a string of graphemes (characters) and a corresponding number. This will be fed into a neural network-based learning technique for inclusion in a standard corpus.

A professor from Institute of Mathematical Sciences (IMSc), Chennai, and an engineer have developed the app. Satish Palaniappan, an RSSN College of Engineering graduate who worked on the app, said the sequence of numbers may help in the search for similar sequences in the texts. Researchers can use 'deep learning' to develop the technology. "Deep learning is basically

call for a neural network-based learning technique, inspired by large numbers of interconnected neurons in the brain that interact and make decisions based on inputs. "Deep learning has never been used before in epigraphic research."

The app is crucial to make big leaps in epigraphic research. "A researcher has to know the history and sequence of symbols," Palaniappan said. "It takes years to compile texts from artefacts and put them in form. The computer will understand. We want to bridge that gap."

> The image of the seal is scanned on all sides that likely to have the Indus script symbols depicting animals like bulls and unicorns, and deities

> A customised algorithm segments out individual units or letters to identify it

> The results are further classified into text, 'no-text', and non-graphemic unit (or letter), the no-text region consists of non-graphemic elements

> The identified units or letters are classified into one of 417 classes, the known names of Indus graphemes, according to scholar Mahadevan

AS recently excavated structures resembling those at Harappa from Keeladi village, Tamil Nadu. They are believed to be 2,000 years old



two prominent cities, Harrappa and Mohenjo-daro

> Copper, bronze, pottery and terracotta toys, beads (carnelian) and (steatite) seals were retrieved from the sites. Many

> Seals were used for ritualistic, commercial and religious purposes

> These were engraved with animal figures

Though the origin and decline of the civilization remain a mystery, there are theories that links, deterioration or the invasion of Aryans could be among reasons

> The Indus Valley Civilisation is one of the oldest known of the ancient civilisations. It existed from 2500 BC to 1300 BC

> Discovered in the 1920s, it was spread over 1,200 sq km covering parts of modern-day Pakistan, Afghanistan and India

> It included around 1,000 well-planned villages, towns and cities, with

the process consists of three stages: In the first phase, input images are broken into sub-images that contain graphemes only, by trimming out the areas that do not contain the script. The remaining areas are further trimmed into single grapheme pieces. Last, each grapheme is classified to match one of the 417 symbols discovered so far in the Indus script.

**Indus script**

The Indus Valley script is much older than the Brahmi and Tamil-Brahmi scripts. However, unlike the latter two, it has not yet been deciphered because a bilingual text has yet been found.

A bilingual text has in many other cases aided archaeologists in understanding ancient scripts, for example, the Rosetta stone, which was found in the eighteenth century carries inscriptions of a decree issued by Ptolemy V in 190 BCE, in three parts, the first two in ancient Egyptian hieroglyphic and the Demotic script, while the bottom is in Greek script. Since the decipherment of the Rosetta stone, the Indus script remains undeciphered today.

It is a major effort to even build a standard corpus to the language and decode the writing on existing artefacts and map them to this standard corpus. The most widely accepted theory of the Indus script was brought together by the efforts of Iravatham Mahadevan, noted Indus scriptologist, from the 3,700 texts or 417 unique symbols recorded so far.

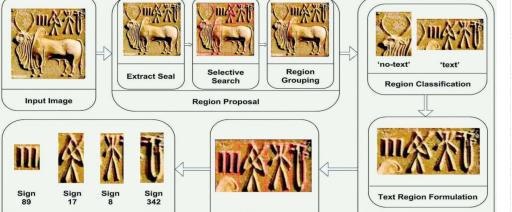
He often asked about the relevance of this work, Dr Mahadevan says, "[t]he [algorithm] represents a significant advance in the computerised study of the Indus script. I wish I had this software 40 years ago when I compiled the Indus concordance."

## Chennai team taps AI to read Indus Script

The algorithm uses 'deep neural networks' which are also used in self-driving cars

SHUBASHREE DESIKAN

The Indus script has long challenged epigraphists because of the difficulty in reading and classifying text and symbols in artifacts. Now, a computer-based team of scientists has built a programme which eases the process.



**Step by step:** Scanning the image, the algorithm smartly processes the data in three steps to place its elements within the standard corpus. SPECIAL ARRANGEMENT

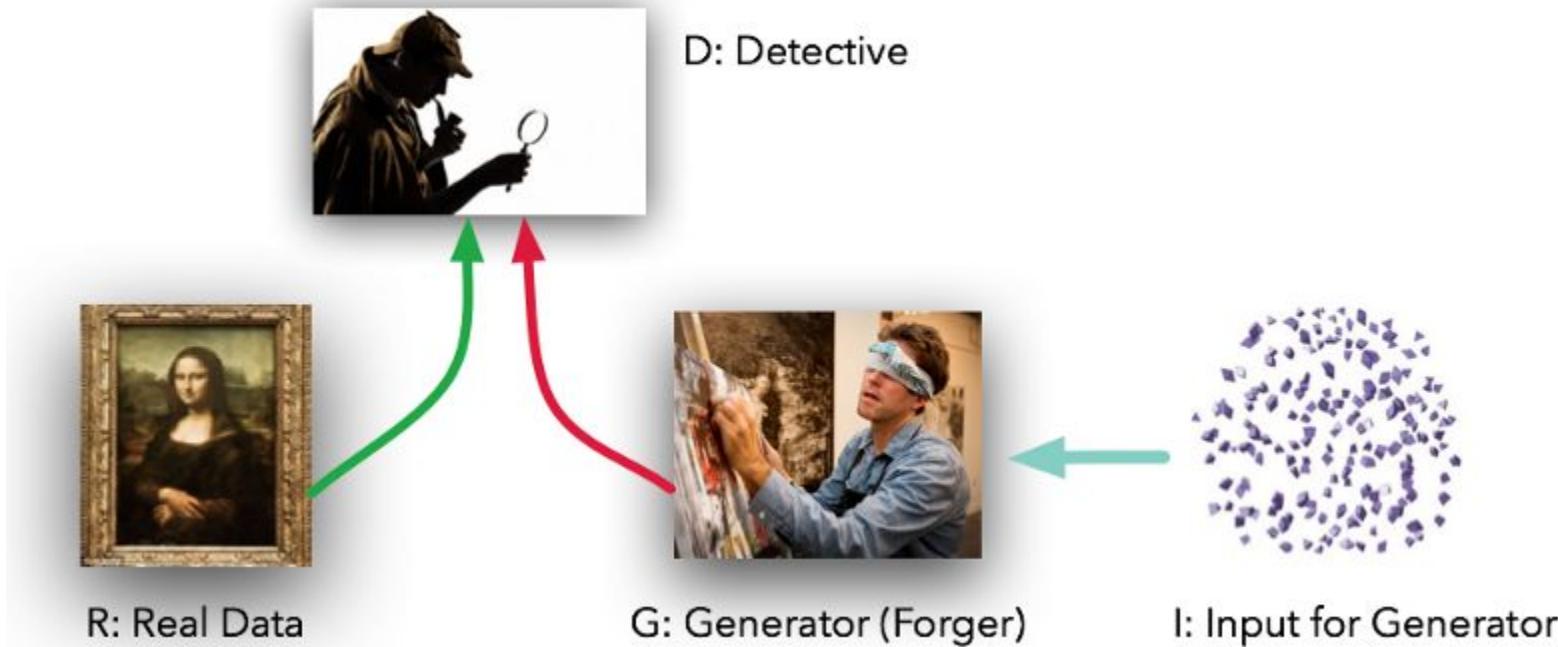
The process consists of three stages: In the first phase, input images are broken into sub-images that contain graphemes only, by trimming out the areas that do not contain the script. The remaining areas are further trimmed into single grapheme pieces. Last, each grapheme is classified to match one of the 417 symbols discovered so far in the Indus script.

The Indus Valley script is much older than the Brahmi and Tamil-Brahmi scripts. However, unlike the latter two, it has not yet been deciphered because a bilingual text has yet been found.

A bilingual text has in many other cases aided archaeologists in understanding ancient scripts, for example, the Rosetta stone, which was found in the eighteenth century carries inscriptions of a decree issued by Ptolemy V in 190 BCE, in three parts, the first two in ancient Egyptian hieroglyphic and the Demotic script, while the bottom is in Greek script. Since the decipherment of the Rosetta stone, the Indus script remains undeciphered today.

**Dr Mahadevan says, "It represents a significant advance in the computerised study of the Indus Script. I wish I had this software 40 years ago when I compiled the Indus concordance."**

# FUN: GENERATIVE ADVERSARIAL NEURAL NETS



# SOURCES!



**Andrew Ng**



**Andrej Karpathy**

## Others:

- [https://github.com/malaikannan/Talks/blob/master/Demystifying\\_AI\\_Keras\\_Intro.ipynb](https://github.com/malaikannan/Talks/blob/master/Demystifying_AI_Keras_Intro.ipynb)
- <https://adeshpande3.github.io/adeshpande3.github.io/>
- <https://github.com/leriomaggio/deep-learning-keras-tensorflow/blob/master/4.%20Convolutional%20Neural%20Networks/4.1%20Convolutional%20Neural%20Networks.ipynb>
- <http://neuralnetworksanddeeplearning.com/>
- <https://pokemondb.net/>
- [https://bulbapedia.bulbagarden.net/wiki/Main\\_Page](https://bulbapedia.bulbagarden.net/wiki/Main_Page)
- <https://arxiv.org/abs/1406.2661>

# Thank You!

Satish Palaniappan

