Movie Genre Classification Using Deep Multimodal Neural Networks

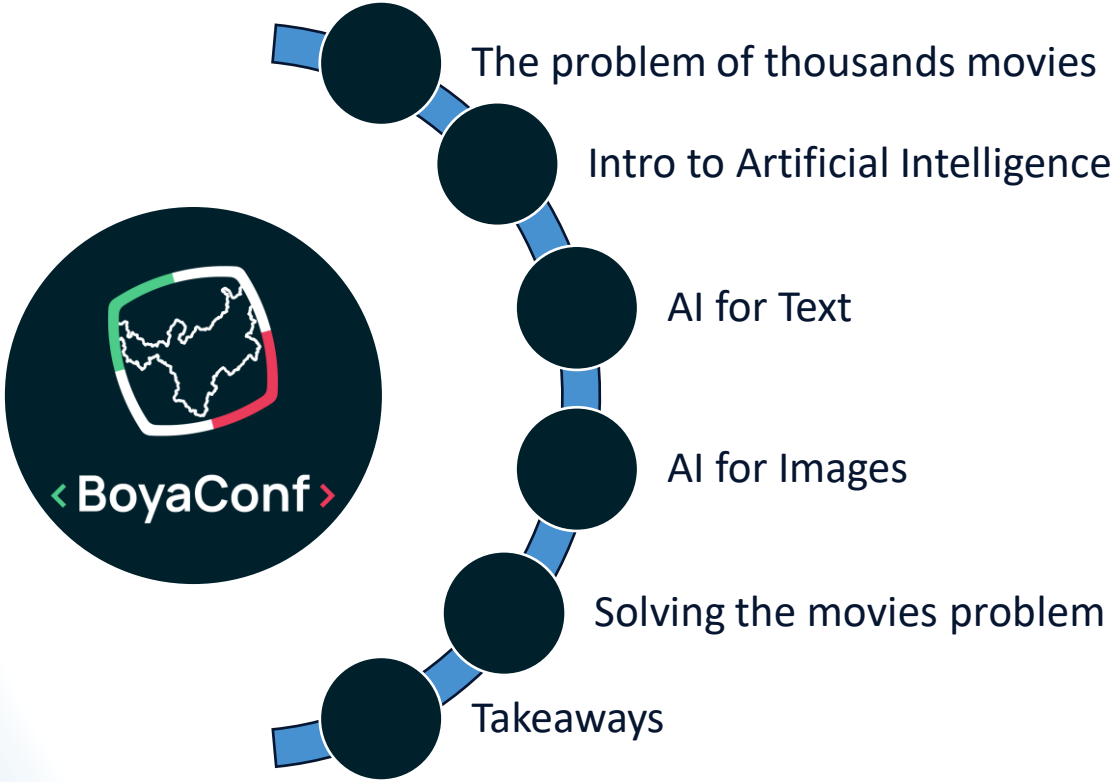
Images and texts analyzed using AI

Jesus Solano

Data Scientist

Cyxtera™

Agenda


The problem of thousands movies
Intro to Artificial Intelligence
AI for Text
AI for Images
Solving the movies problem
Takeaways
BoyaConf

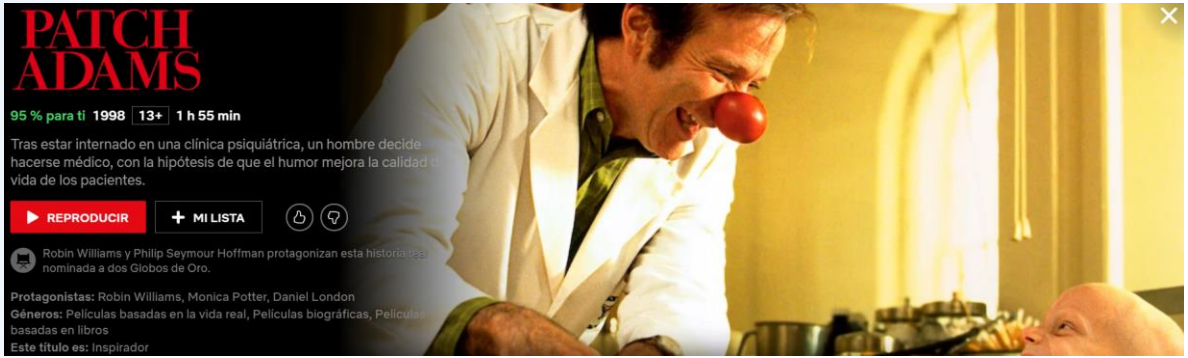
The dilemma…


NETFLIX
Inicio
Programas
Películas
Agregados recientemente
Mi lista
Películas
Géneros
Acción
Anime
Ciencia ficción y fantasía
Clásicas
Colombianos
Comedias
De Hollywood
Deportes
Documentales
Dramas
Fe y espiritualidad
Independientes
Infantiles y familiares
Internacional
Latinoamericanas
Los favoritos de la crítica
Música y musicales
Navidad
Películas policiales
Romances
Terror
Thrillers
NAVIDAD EN ÁFRICA
SIN GRAVEDAD
Dramas basados en la vida real
LA GRAN APUESTA
PATCH ADAMS
HAMBRE DE PODER

The dilemma…


NETFLIX
Inicio
Programas
Películas
Agregados recientemente
Mi lista
Películas
Géneros
Acción
NAVIDAD EN ÁFRICA
SIN GRAVEDAD
Dramas basados
LA GRAN APUESTA
HAMBRE DE PODER
CHOOSING a MOVIE to WATCH
WATCHING THE MOVIE

The dilemma…



How do the humans **abstract the genre** and also the type of movie **with only and poster and a plot**?

LET'S TRY TO CLASSIFY MOVIES USING MACHINE LEARNING!

"The only limit to AI is the human imagination" Chris Duffey

What is artificial intelligence?

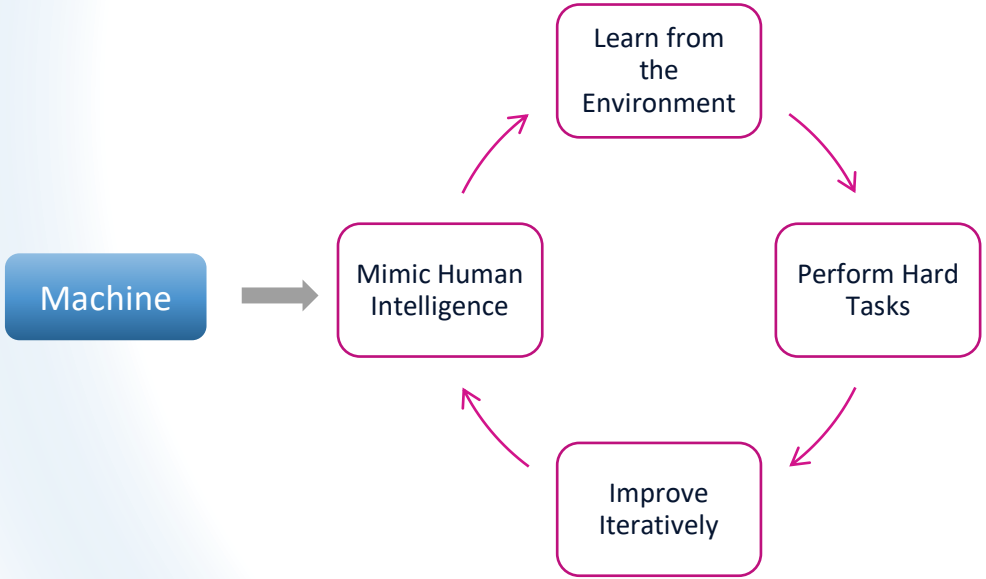
“What is intelligence, anyway? It is only a word that people use to name those unknown processes with which our brains solve problems we call hard. But whenever you learn a skill yourself, you're less impressed or mystified when other people do the same.

This is why the meaning of "intelligence" seems so elusive: it describes not some definite thing but only the momentary horizon of our ignorance about how minds might work. It is hard for scientists who try to understand intelligence to explain precisely what they do, since our working definitions change from year to year”

Marvin Minsky

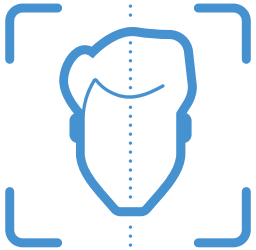
What is artificial intelligence?

Even though we don't yet understand how brains perform many mental skills, we can still work toward making machines that do the same or similar things….



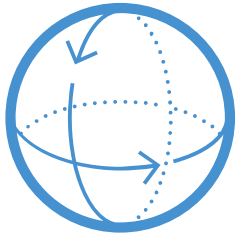
…."Artificial intelligence" is simply the name we give to that research.

Artificial Intelligence



Intentionality

AI is designed to make decisions, often using real-time data



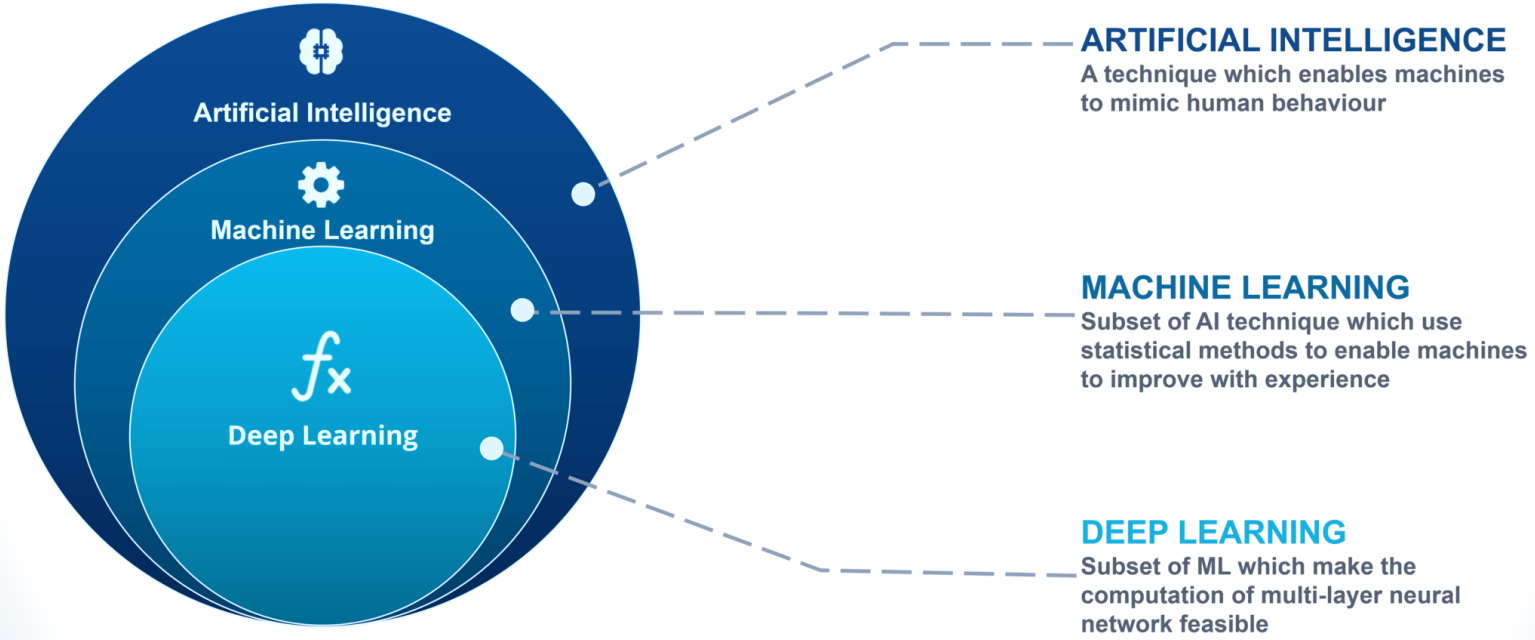
Intelligence

AI systems takes data and looks for underlying trends

Adaptability

Effective AI must adjust as circumstances or conditions shift

Artificial Intelligence



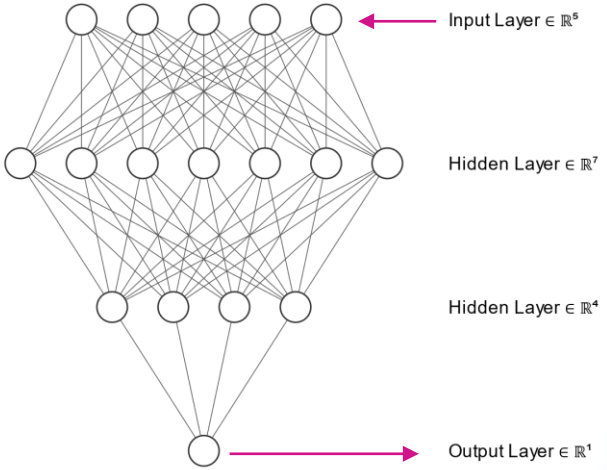
What is Deep Learning?

Algorithms inspired by function of the brain neural networks.

Learn from large amounts of raw data

Sometimes they exceeds human-level performance.

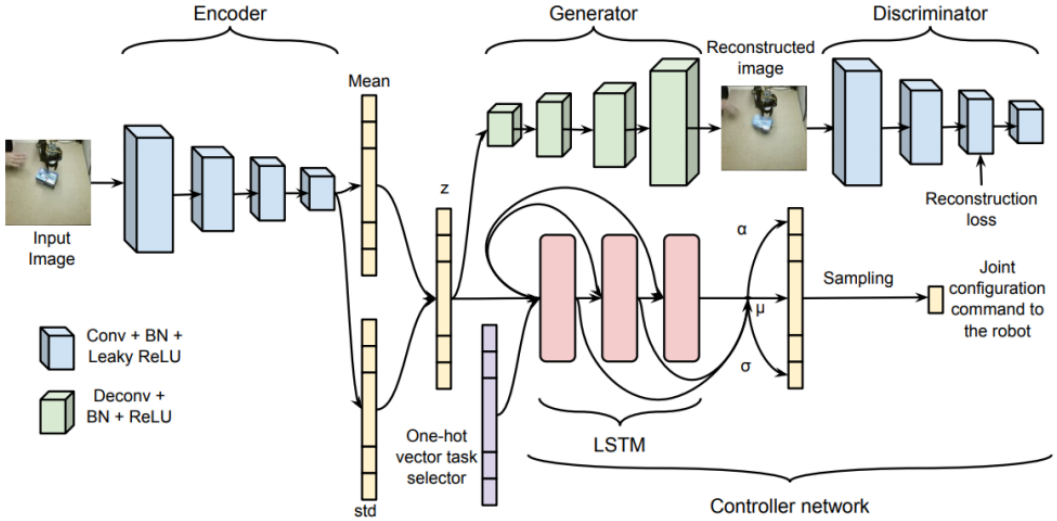
Neural network architectures that contain many layers



Why Deep Learning?
In a word, accuracy.

How today deep learning looks like?

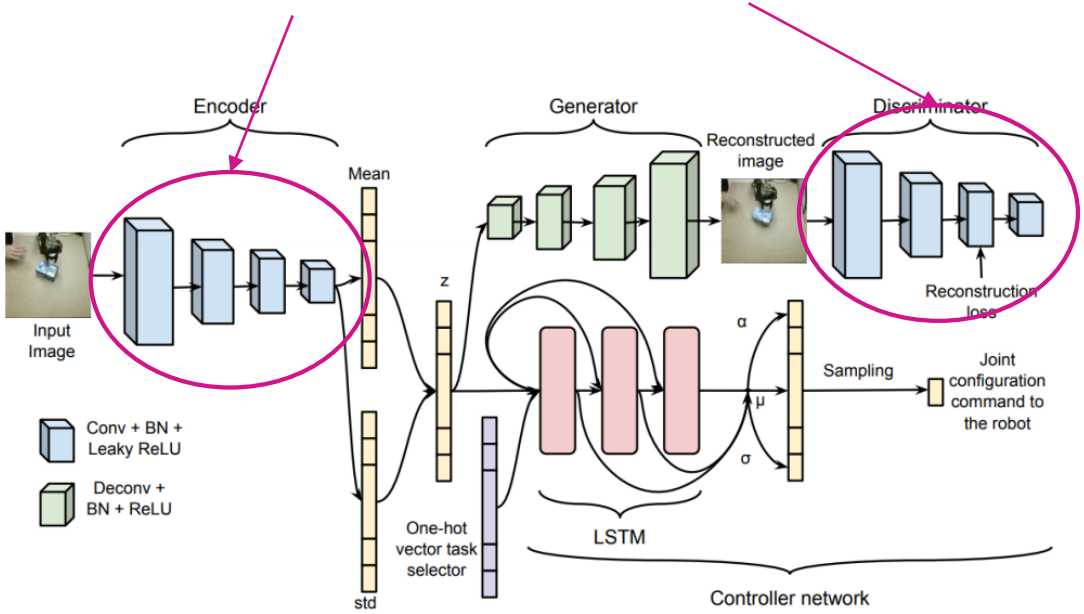
Why if I want to make a robot?
No problem, below is the simple solution.



Francois Fleuret, EE559 Deep Learning, EPFL; Rahmatizadeh et al, 2017, arXiv:1707.02920.

What are we learning today?

Today we are learning a little bit about…



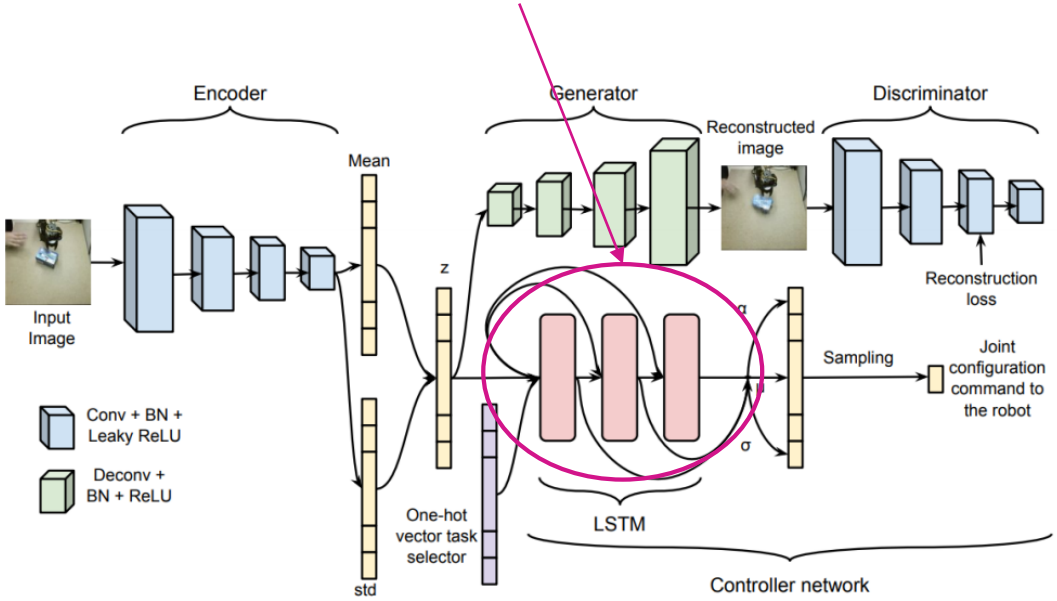
Convolutional Neural Networks.

For computer Vision

Francois Fleuret, EE559 Deep Learning, EPFL; Rahmatizadeh et al, 2017, arXiv:1707.02920.

What are we learning today?

And also a little bit about…



Recurrent Neural Networks.

For Natural Language Processing

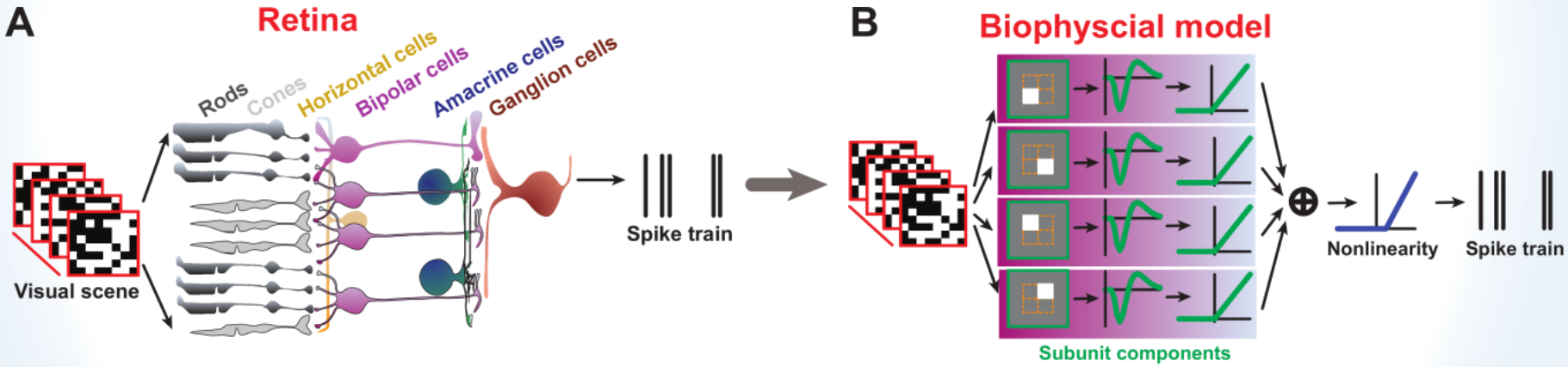
Francois Fleuret, EE559 Deep Learning, EPFL; Rahmatizadeh et al, 2017, arXiv:1707.02920.

AI for Computer Vision

Intro to Convolutional Neural Networks(CNN)

How does a human see?

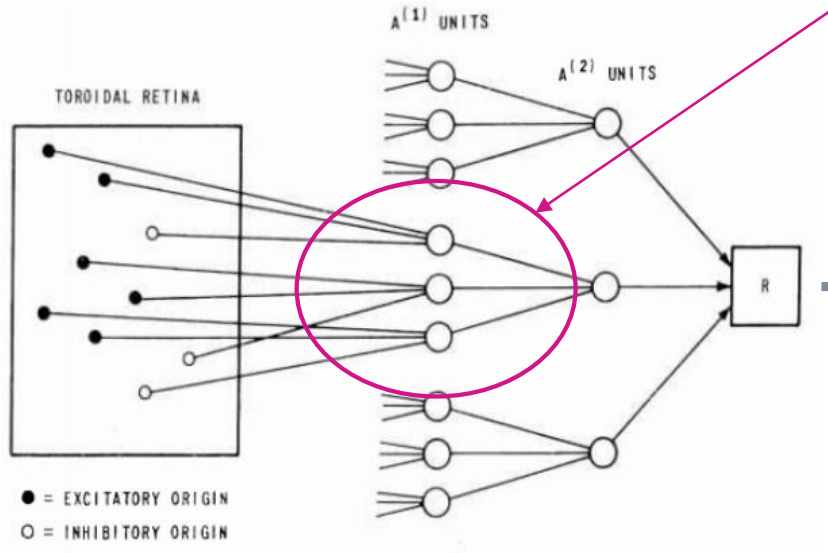
Yan, Q., Zheng, Y., Jia, S., Zhang, Y., Yu, Z., Chen, F., ... & Liu, J. K. (2018). Revealing Fine Structures of the Retinal Receptive Field by Deep Learning Networks. *arXiv preprint arXiv:1811.02290*.



Great! But let's make it simpler…

Intro to Convolutional Neural Networks(CNN)

How does a human see?



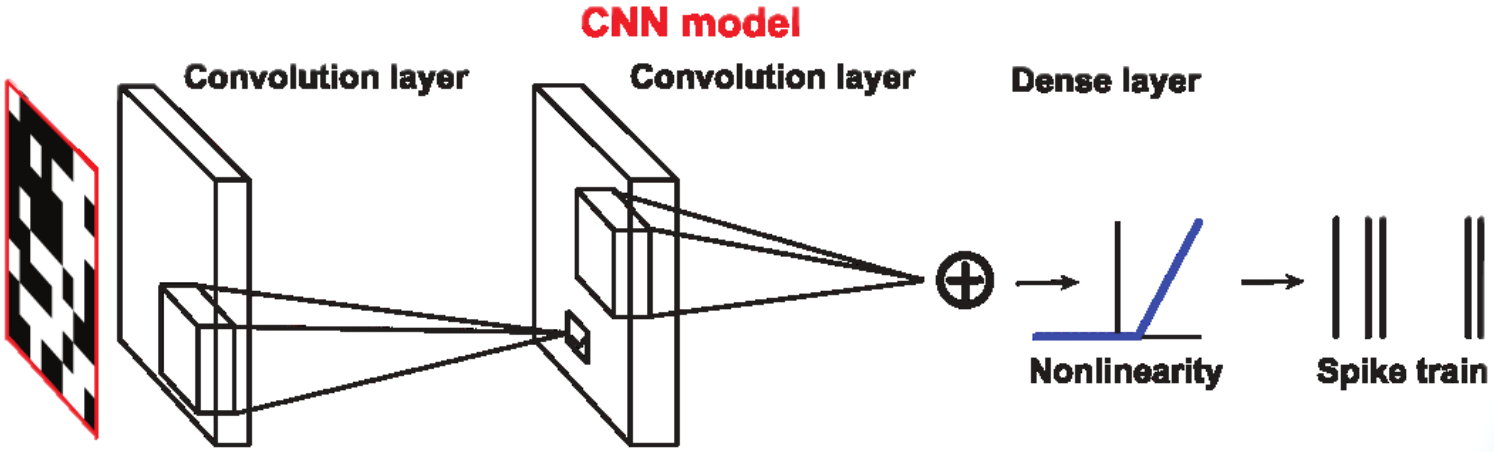
Selective neuron activation !

Human Response

Let's make an analogy with the machine !

Intro to Convolutional Neural Networks(CNN)

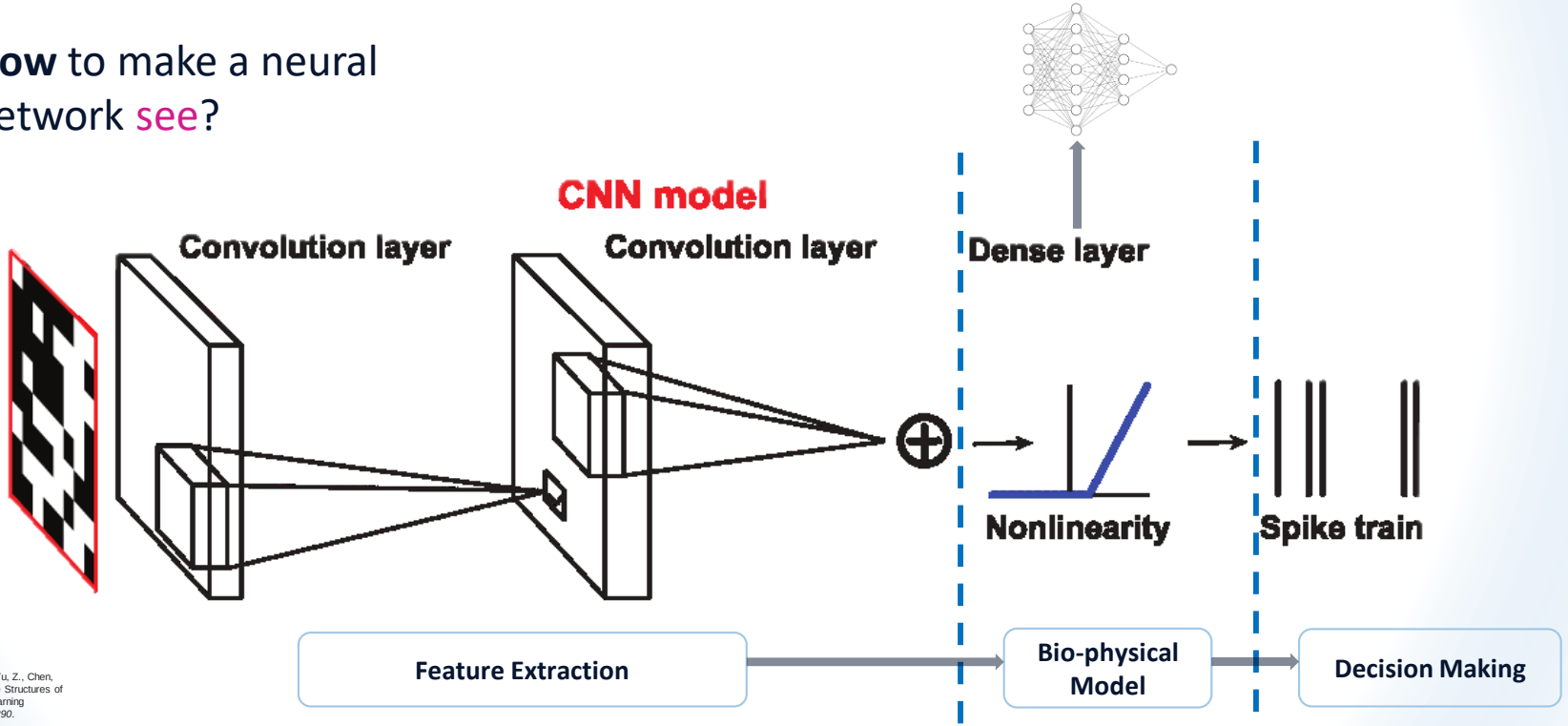
How to make a neural network see?



Yan, Q., Zheng, Y., Jia, S., Zhang, Y., Yu, Z., Chen, F., ... & Liu, J. K. (2018). Revealing Fine Structures of the Retinal Receptive Field by Deep Learning Networks. *arXiv preprint arXiv:1811.02290*.

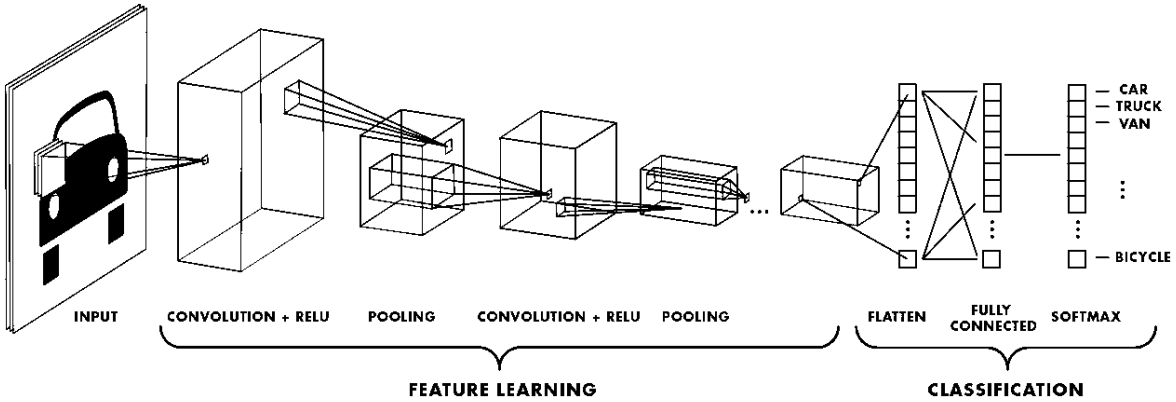
Intro to Convolutional Neural Networks(CNN)

How to make a neural network see?



Yan, Q., Zheng, Y., Jia, S., Zhang, Y., Yu, Z., Chen, F., ... & Liu, J. K. (2018). Revealing Fine Structures of the Retinal Receptive Field by Deep Learning Networks. *arXiv preprint arXiv:1811.02290*.

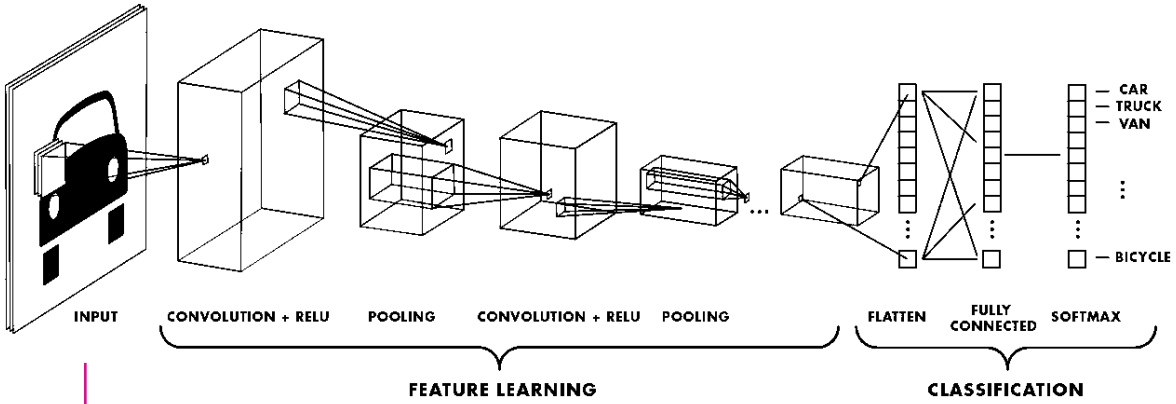
Intro to Convolutional Neural Networks(CNN)



Convolutional Neural Networks.

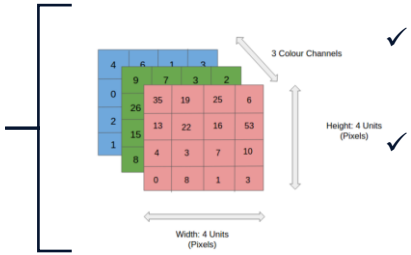
- ✓ ConvNets have the ability to **learn** filters/characteristics **without hard pre-processing**.
- ✓ ConvNets are able to successfully **capture the Spatial and Temporal dependencies** in an image through the application of relevant filters.

Intro to Convolutional Neural Networks(CNN)



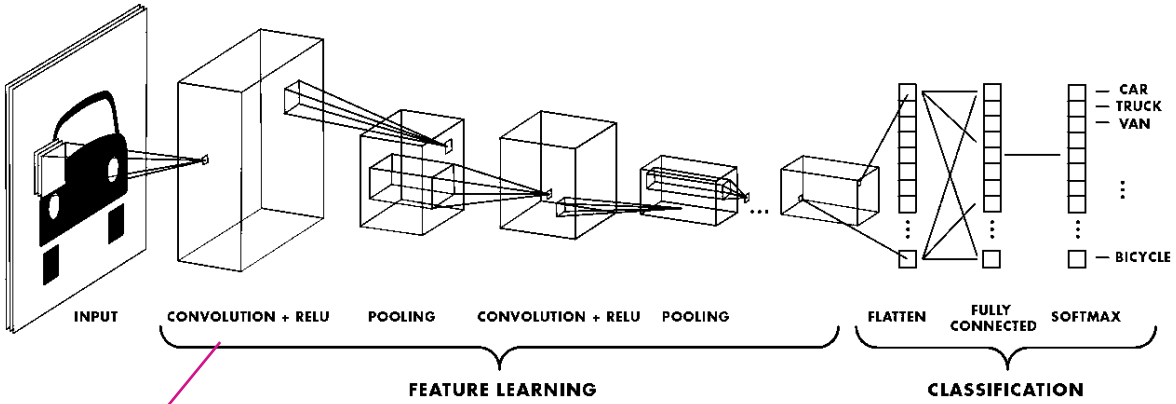
1

Feed Neural Network



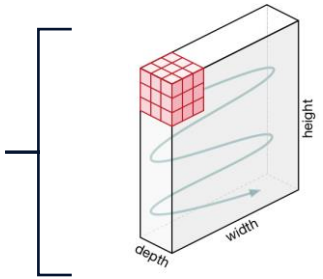
- ✓ ConvNets have the ability to **learn** filters/characteristics **without hard pre-processing**.
- ✓ Why not just flatten the image and feed it to a simple neural network? **Because of Spatial Dependencies**

Intro to Convolutional Neural Networks(CNN)



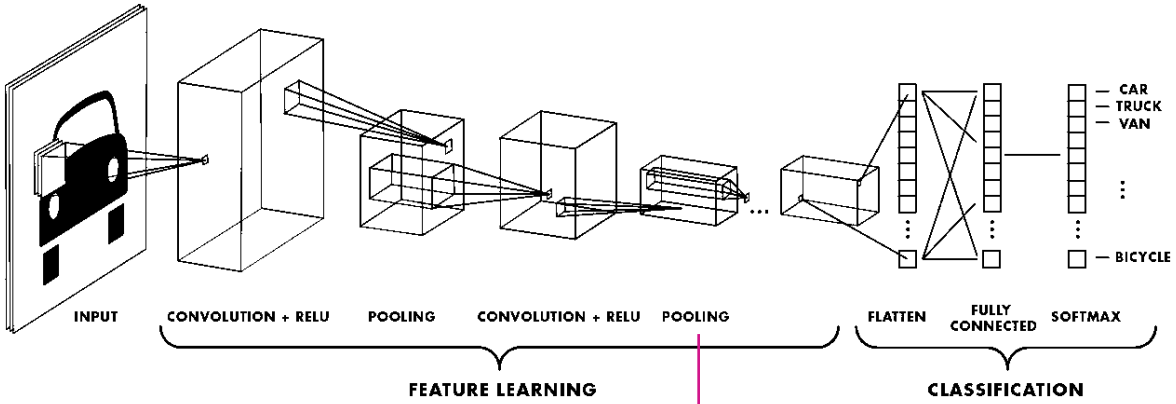
2

Convolutions



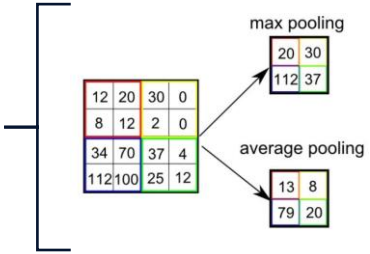
- ✓ The objective of the **Convolution Operation is to extract the high-level features**(such as edges, etc)
- ✓ **Usually more than one ConvLayer** help capturing the Low-Level features such as edges, color, gradient orientation, etc.

Intro to Convolutional Neural Networks(CNN)



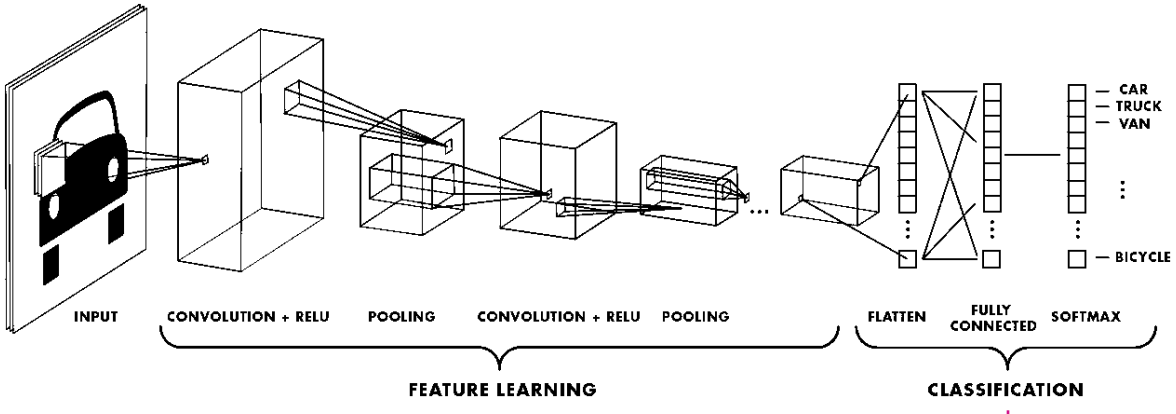
3

Pooling



- ✓ Pooling layer is responsible for **reducing the spatial size** of the Convolved Feature.
- ✓ Decrease the computational power required to process the data.
- ✓ Useful for **extracting dominant features**(rotational and positional invariant)

Intro to Convolutional Neural Networks(CNN)



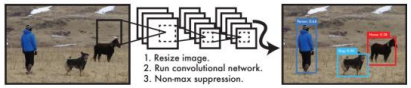
4

Fully connected

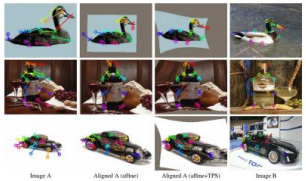
- ✓ Cheap way of **learning non-linear combinations** of the high-level features as represented by Conv output.
- ✓ Fully Connected **distinguish between dominating** and certain low-level **features** in images for the classification problem.

Intro to Convolutional Neural Networks(CNN)

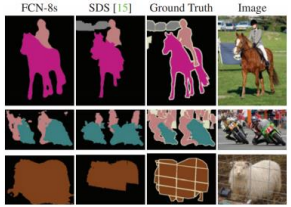
Convolutional neural networks are now used everywhere in computer vision



Object detection
(Redmon et al, 2015)



Geometric matching
(Rocco et al, 2017)

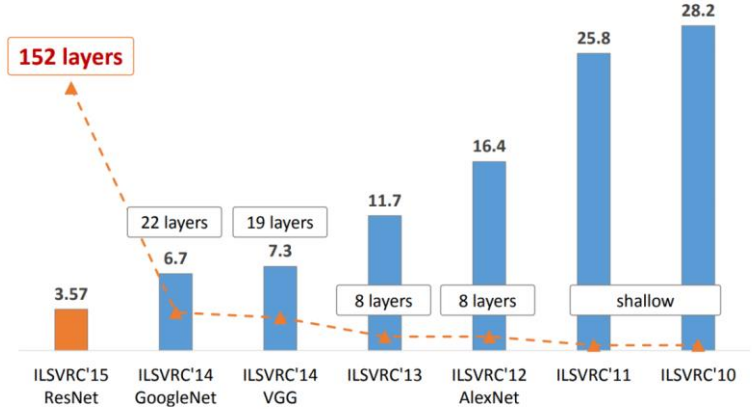


Semantic segmentation
(Long et al, 2015)

Instance segmentation
(He et al, 2017)

Deeper is better !

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)



Canziani et al, An Analysis ofDeep Neural Network Models for Practical Applications, 2016

Beats human-level performance !!

AI for NLP

Intro to Natural Language Processing(NLP)

How to make a neural network read and understand?

First of all !

Machine learning models take vector(arrays of numbers) as input.

So….

How to represent *“This year Boyaconf was amazing”* in a numerical way

Intro to Natural Language Processing(NLP)

How to represent *“This year Boyaconf was amazing”* in a numerical way

Vectorize the text

1. One-hot Encoding
2. Unique Map Encoding
3. Word Embeddings

Intro to Natural Language Processing(NLP)

How to represent *"This year Boyaconf was amazing"* in a numerical way

Vectorize the text

1. **One-hot Encoding**
2. Unique Map Encoding
3. Word Embeddings

"One-hot" encode each word in our vocabulary.

	year	amazing	was	boyaconf	this
This →	0	0	0	0	1
year →	1	0	0	0	0
Boyaconf →	0	0	0	1	0
...			...		

But…This approach is **inefficient**

Intro to Natural Language Processing(NLP)

How to represent *"This year Boyaconf was amazing"* in a numerical way

Vectorize the text

1. One-hot Encoding
2. **Unique Map Encoding**
3. Word Embeddings

Encode each word using a unique number

year → 1
amazing → 2
was → 3
boyaconf → 4
this → 5

This year Boyaconf was amazing"

[5, 1, 4, 3, 2]

But…

(1) it does not capture any relationship between words
(2) An integer-encoding can be challenging for a model to interpret

Intro to Natural Language Processing(NLP)

How to represent *"This year Boyaconf was amazing"* in a numerical way

Vectorize the text

1. One-hot Encoding
2. Unique Map Encoding
3. **Word Embeddings**

Word embeddings give us a way to use an efficient, dense representation in which similar words have a similar encoding

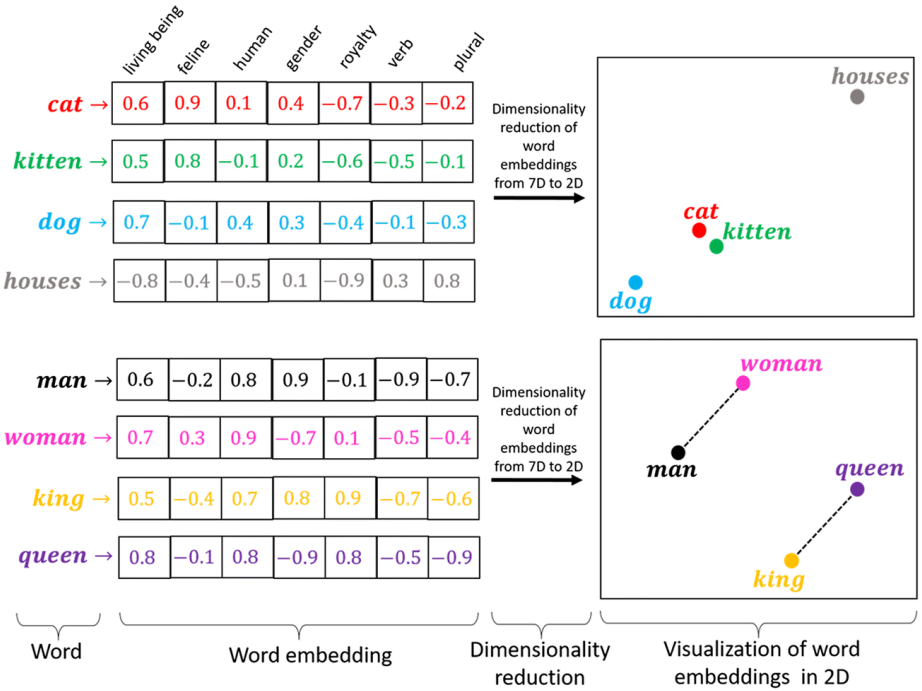
Everything looks great with embeddings !

Intro to Natural Language Processing(NLP)

Word Embeddings

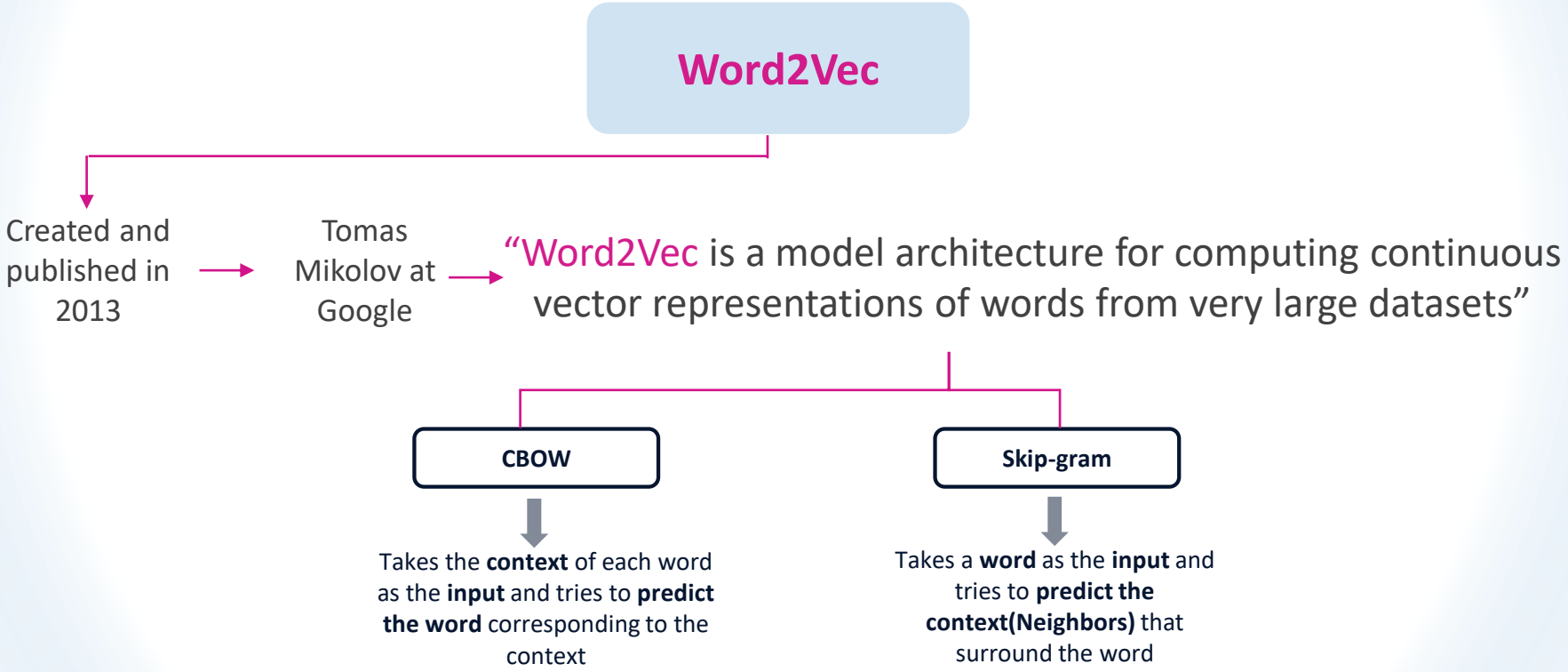
An embedding is a dense vector of floating point values

Higher dimensional embedding can **capture** fine-grained **relationships between words**

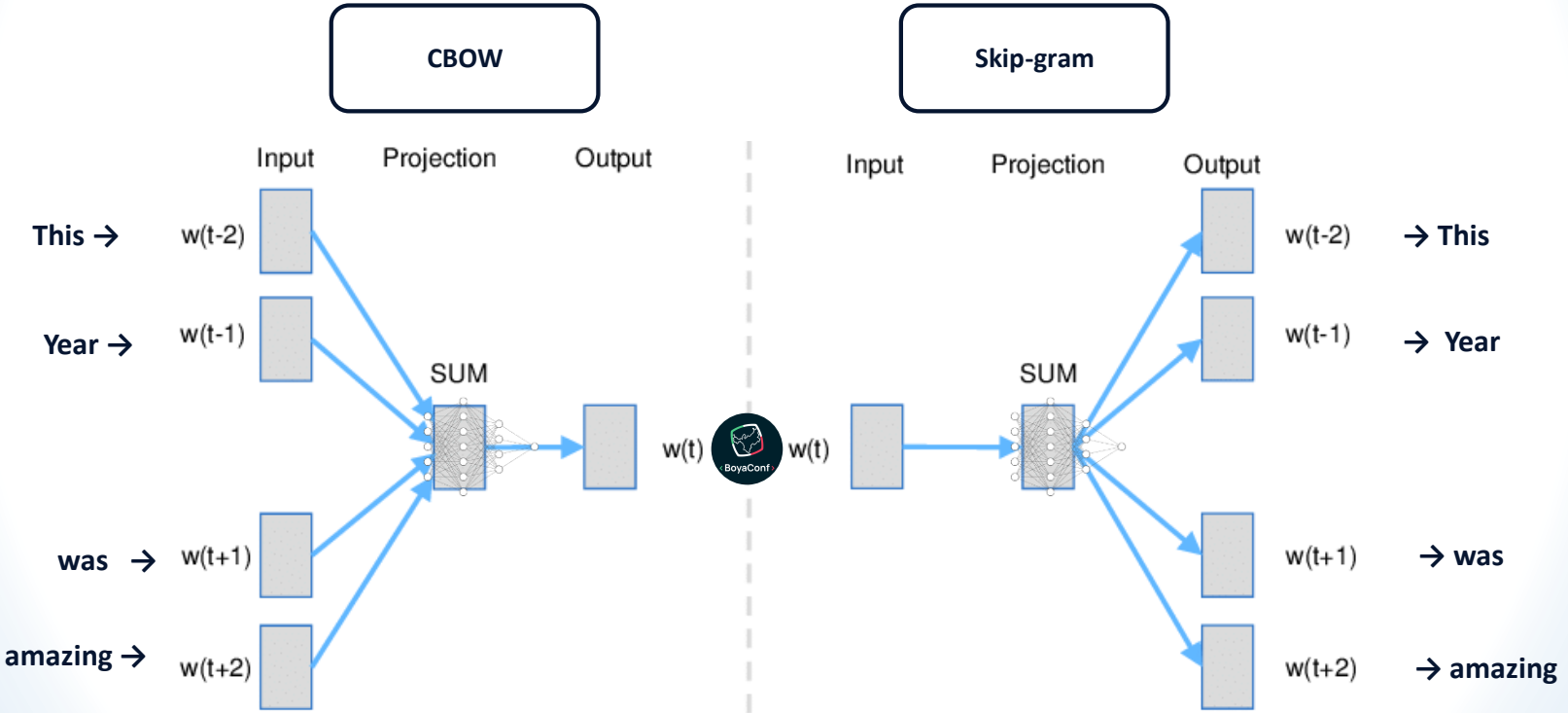


Word2Vec is a method to build such a embedding

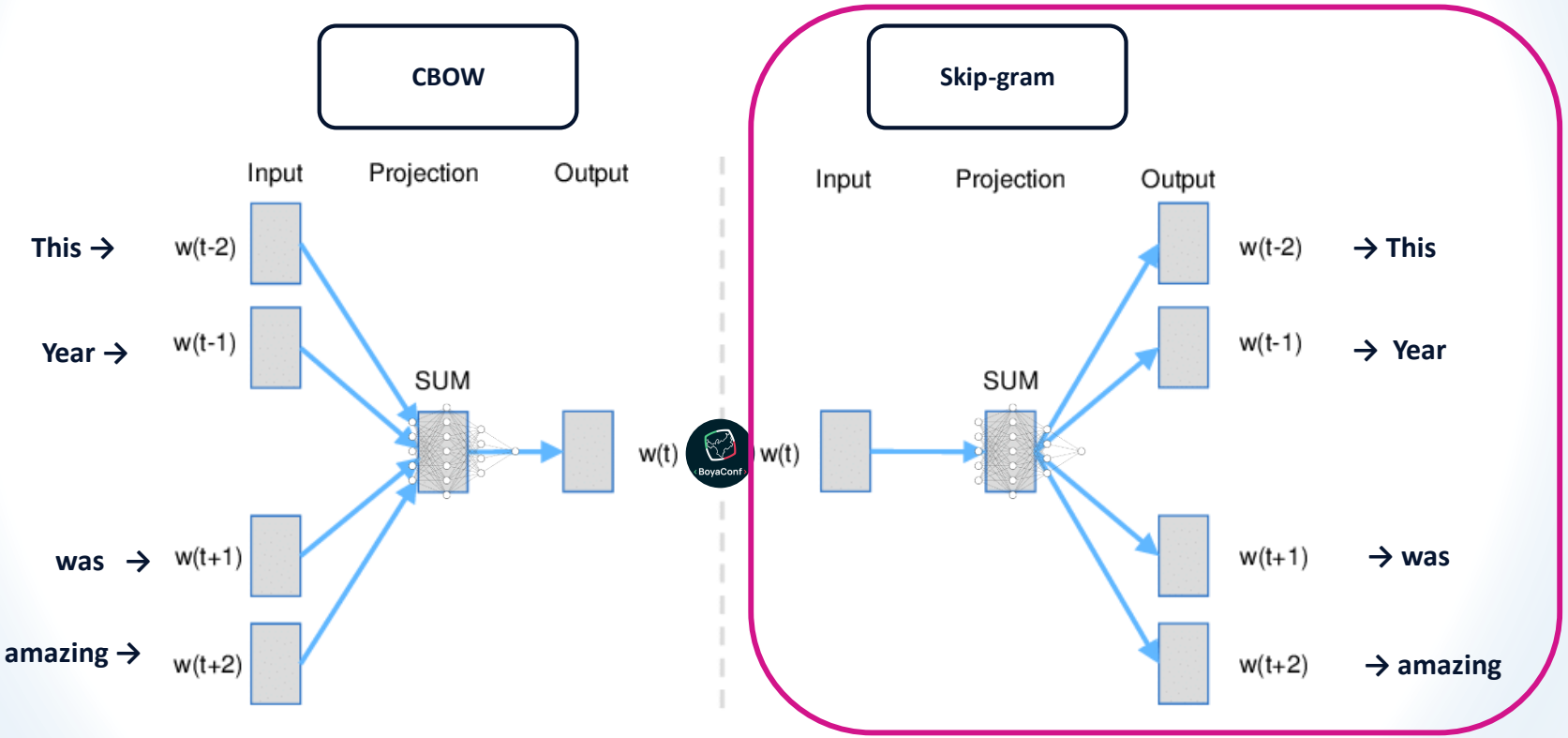
Intro to Natural Language Processing(NLP)



Intro to Natural Language Processing(NLP)



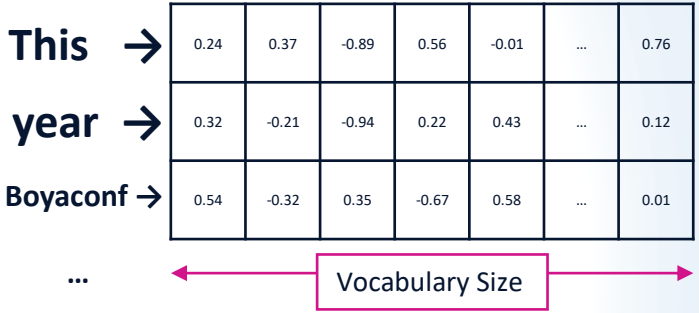
Intro to Natural Language Processing(NLP)



Intro to Natural Language Processing(NLP)

How to represent *"This year Boyaconf was amazing"* in a numerical way ?

Using Word2Vec

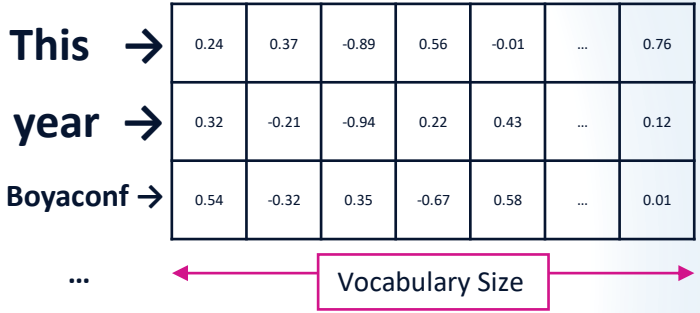


… What's next?

Intro to Natural Language Processing(NLP)

How to represent *"This year Boyaconf was amazing"* in a numerical way ?

Using Word2Vec



Train a Neural Network

Train a Neural Network

… What's next?

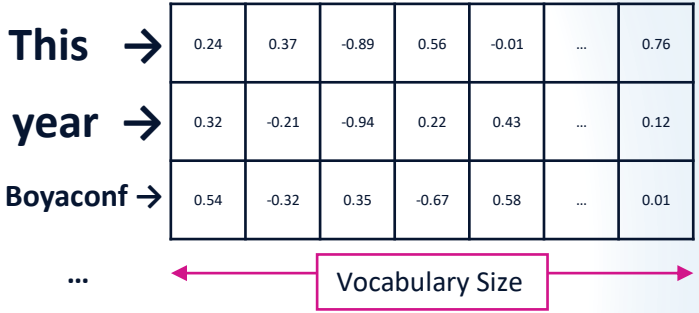
Train a Neural Network

Train a Neural Network

Intro to Natural Language Processing(NLP)

How to represent *"This year Boyaconf was amazing"* in a numerical way ?

Using Word2Vec



What should the network take into account?

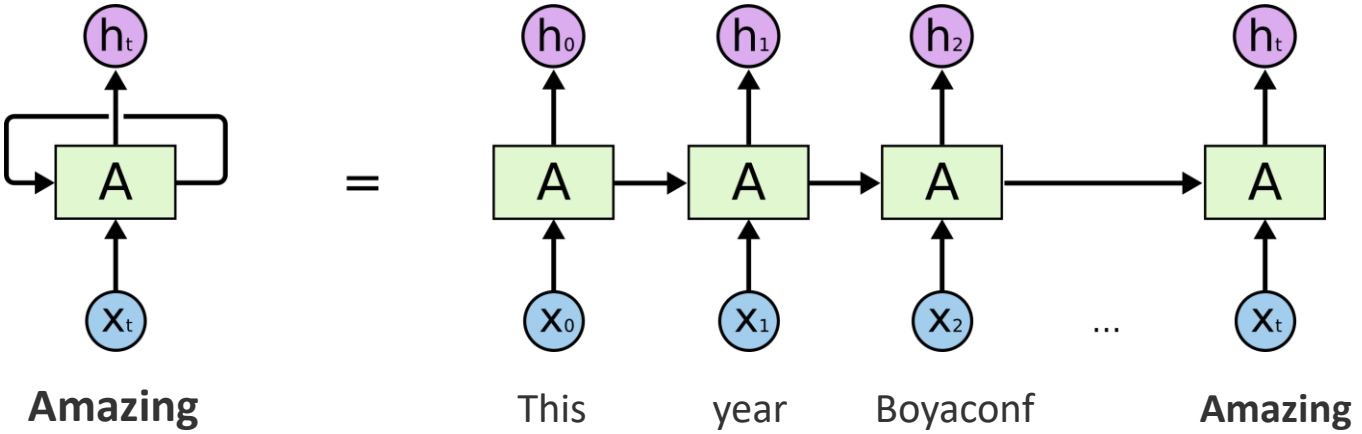
- ✓ The model should **recognize text as a sequence** of words.
- ✓ The neural network should have **memory** -> Words from past define meaning of the future.
- ✓ Computational **efficient**.

→ Recurrent Neural Networks.

Intro to Natural Language Processing(NLP)

Recurrent Neural Networks.

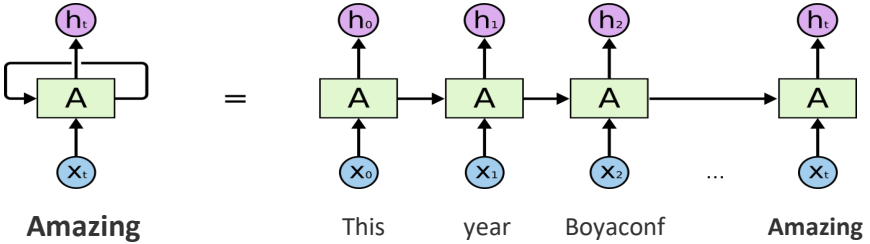
Neural networks with loops in them → Allow information to persist. → **Capture Temporal Correlations !!**

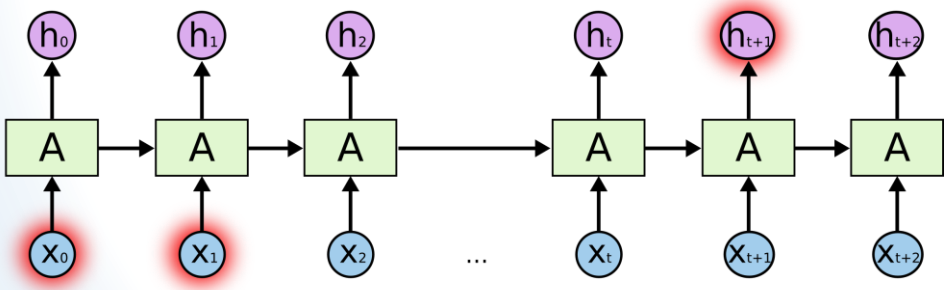


Intro to Natural Language Processing(NLP)

BUT… RNNs usually don't capture **long-term** dependencies…

Recurrent Neural Network



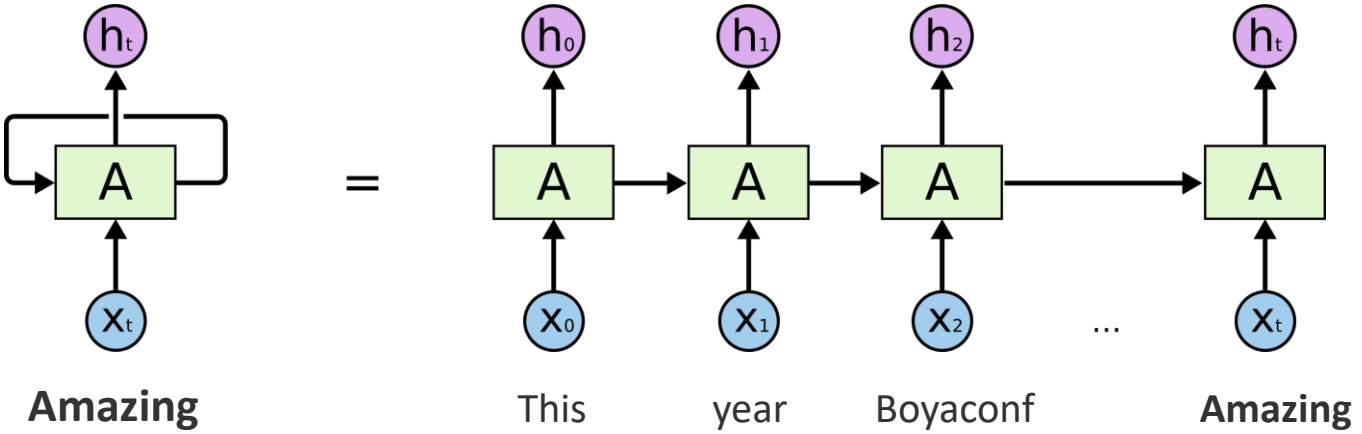


We need long-term because we are analyzing paragraphs !

Intro to Natural Language Processing(NLP)

Long Short Term Memory(LSTM).

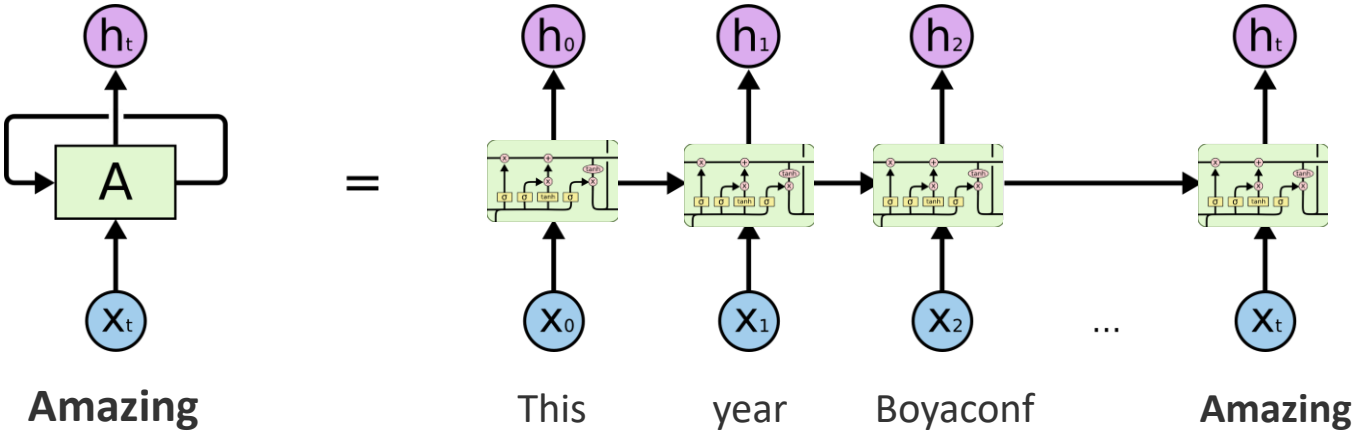
Neural networks with loops in them → Capture Temporal Correlations → **Capable of learning long-term dependencies**



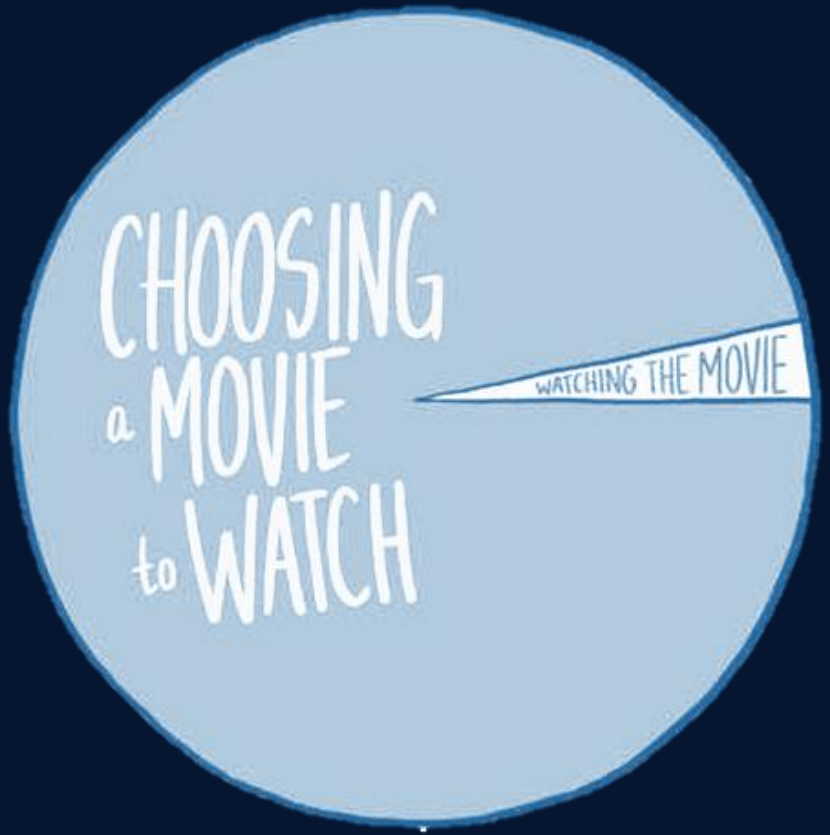
Intro to Natural Language Processing(NLP)

Long Short Term Memory(LSTM).

Neural networks with loops in them → Capture Temporal Correlations → **Capable of learning long-term dependencies**



MEANWILE IN DAILY LIFE ...


CHOOSING
a MOVIE
to WATCH
WATCHING THE MOVIE

Coming back to the movies…

Poster

Plot

In Gotham City, mentally-troubled comedian Arthur Fleck is disregarded and mistreated by society. He then embarks on a downward spiral of revolution and bloody crime. This path brings him face-to-face with his alter-ego: "The Joker".

How can we use deep learning to analyze these information?

AI to predict Movie's Genre

Plot Analysis using Word2Vec

Plot

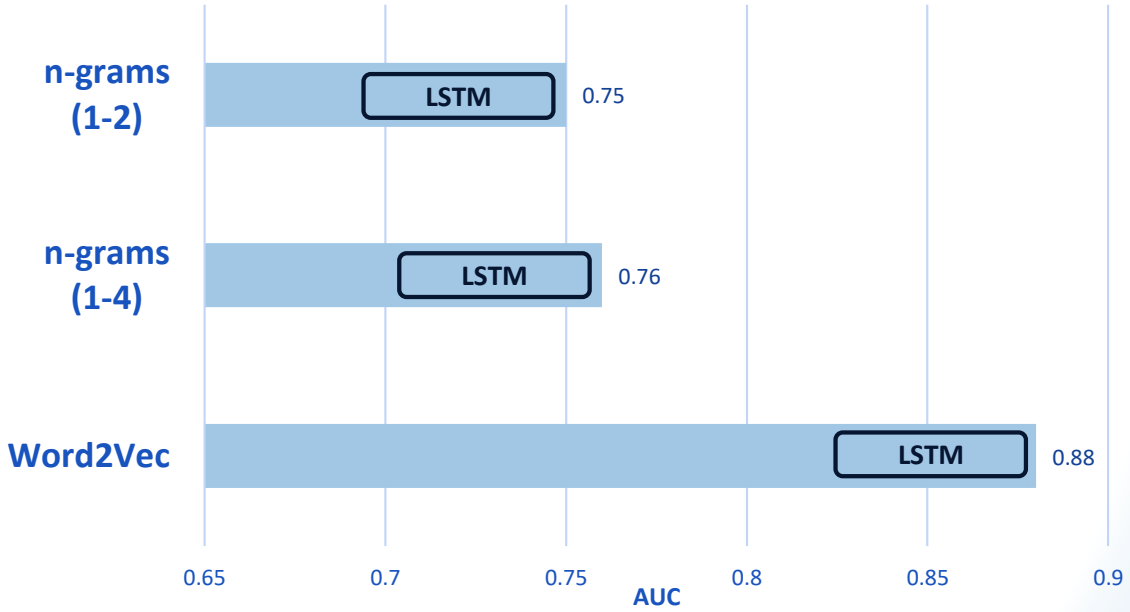
In Gotham City, mentally-troubled comedian Arthur Fleck is disregarded and mistreated by society. He then embarks on a downward spiral of revolution and bloody crime. This path brings him face-to-face with his alter-ego: "The Joker".

- Multimodal IMDb Dataset.
- 7895 plots
- About 100 words per plot

Plot Analysis using Word2Vec

Plot

In Gotham City, mentally-troubled comedian Arthur Fleck is disregarded and mistreated by society. He then embarks on a downward spiral of revolution and bloody crime. This path brings him face-to-face with his alter-ego: "The Joker".



AUC -> Probability of classify a random positive sample higher than a random negative sample

Poster Analysis using CNNs

Poster

- Multimodal IMDb Dataset.
- 7895 images
- 160 x 256 pixels.

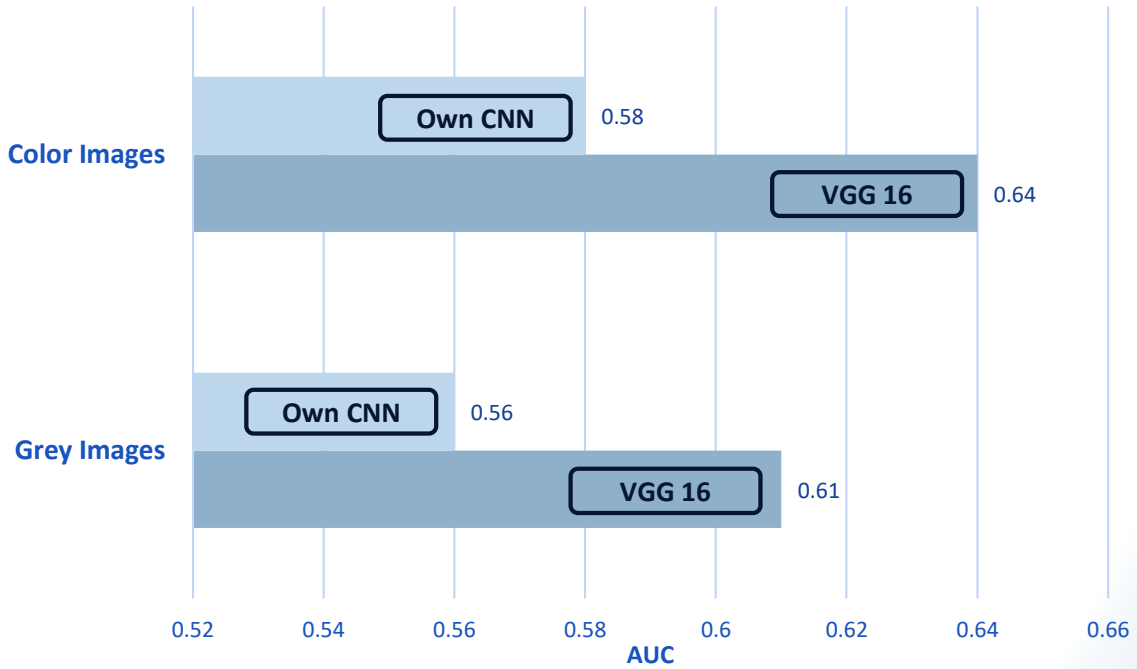
Image Resizing

Genre Fully Connected

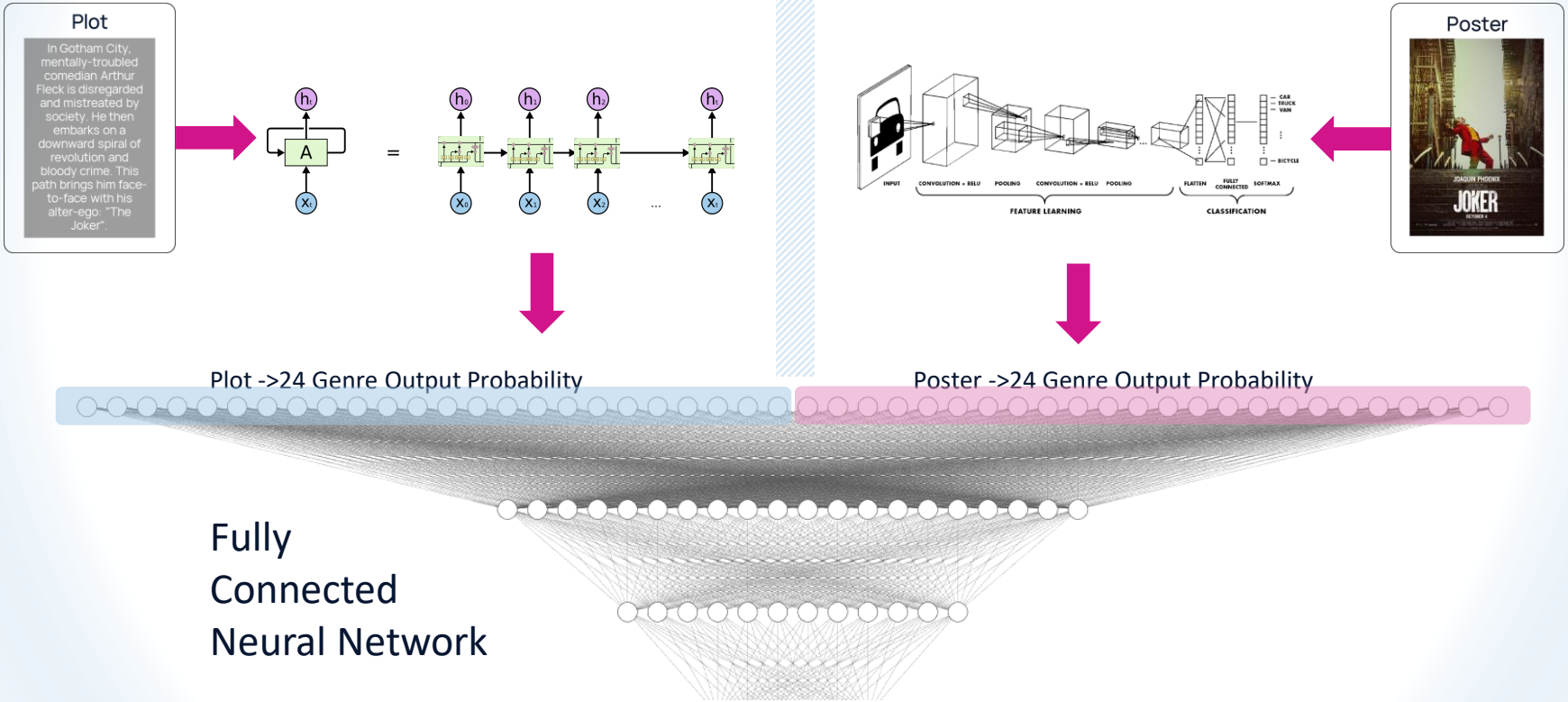
Fine Tuning VGG16

Poster Analysis using CNNs

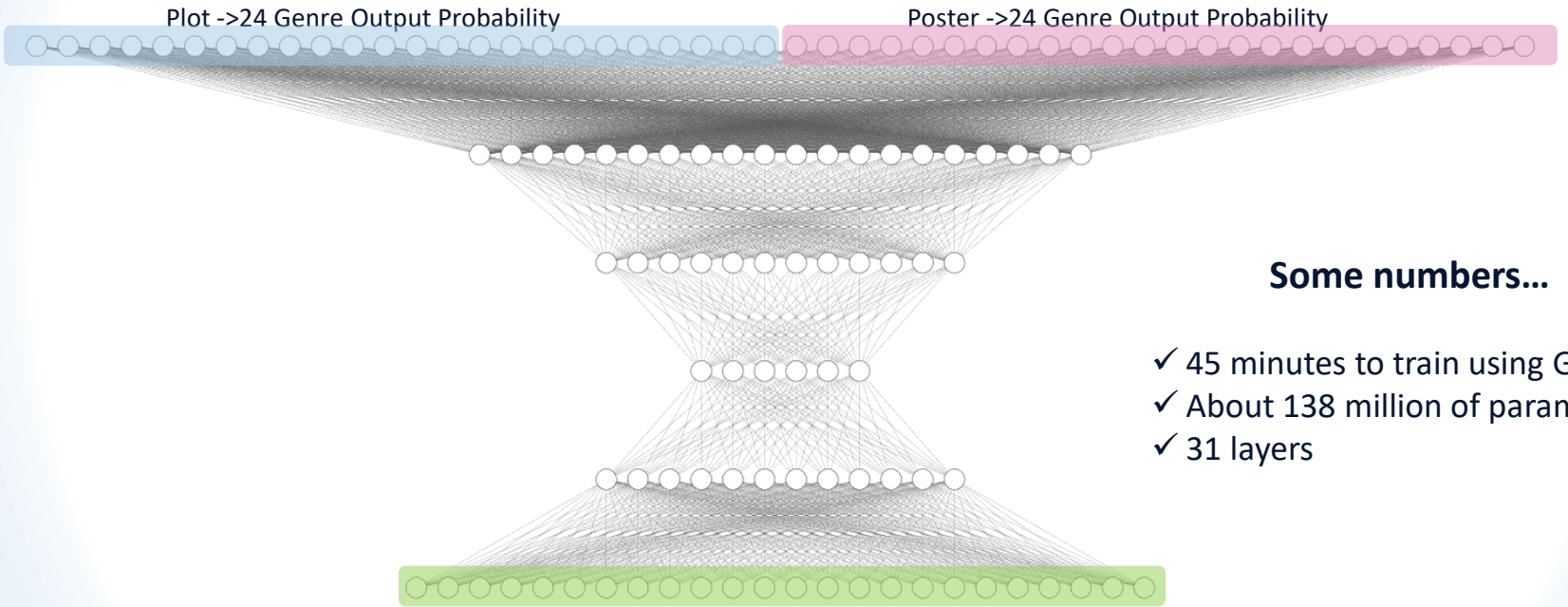
Poster



Multimodal Deep Learning



Multimodal Deep Learning



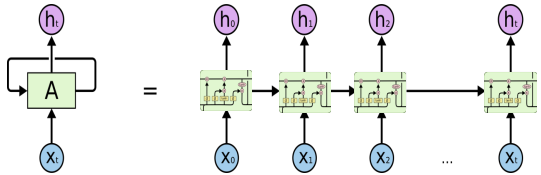
Some numbers…

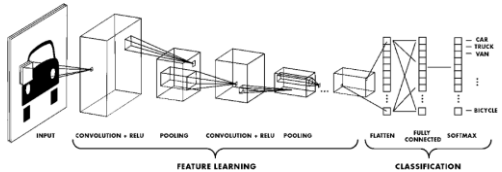
- ✓ 45 minutes to train using GPU
- ✓ About 138 million of parameters
- ✓ 31 layers

Results Multimodal Deep Learning

Plot ->24 Genre Output Probability

Poster ->24 Genre Output Probability





Word2Vec -> LSTM

Color CNN VGG 16

AUC → 0.91

~ 3 % more than best model

Takeaways

- ✓ A good choice Deep Learning model usually improve a lot the performance
- ✓ Deep Learning could be used for almost everything
- ✓ With Deep Learning one can start from raw data. Forget about hard feature engineering.
- ✓ Challenges are out there. Data is out there. Models are out there. What are you waiting for playing with AI?

Questions & Answers

Jesus Solano
Data Scientist

jesus-solano-go
Jesus.solano@cyxtera.com

Cyxtera