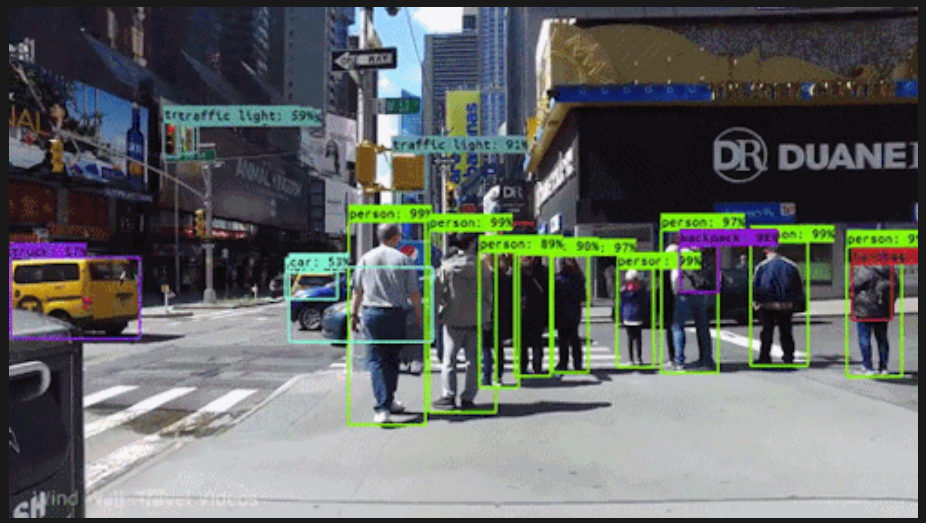
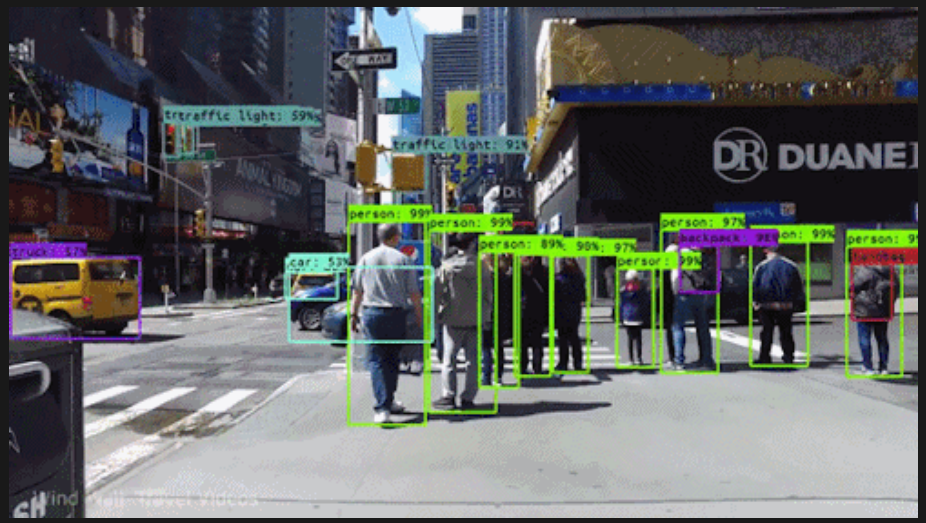
1.00

¿PUEDE UNA MÁQUINA LEER LOS LABIOS?

TRABAJO DE FIN DE GRADO

Roberto Carlos Saavedra Baylón


trtraffic light: 59%
traffic light: 91%
person: 99%
person: 99%
person: 89%
person: 97%
car: 53%
person: 97%
backpack: 91%
person: 99%
person: 99%


DR DUANE
person: 99%
person: 99%
person: 89%
person: 97%
person: 97%
person: 99%
car: 53%

1. Introducción
2. Estado del Arte
3. Sistemas Tradicionales
4. Deep Learning
5. Conclusiones



























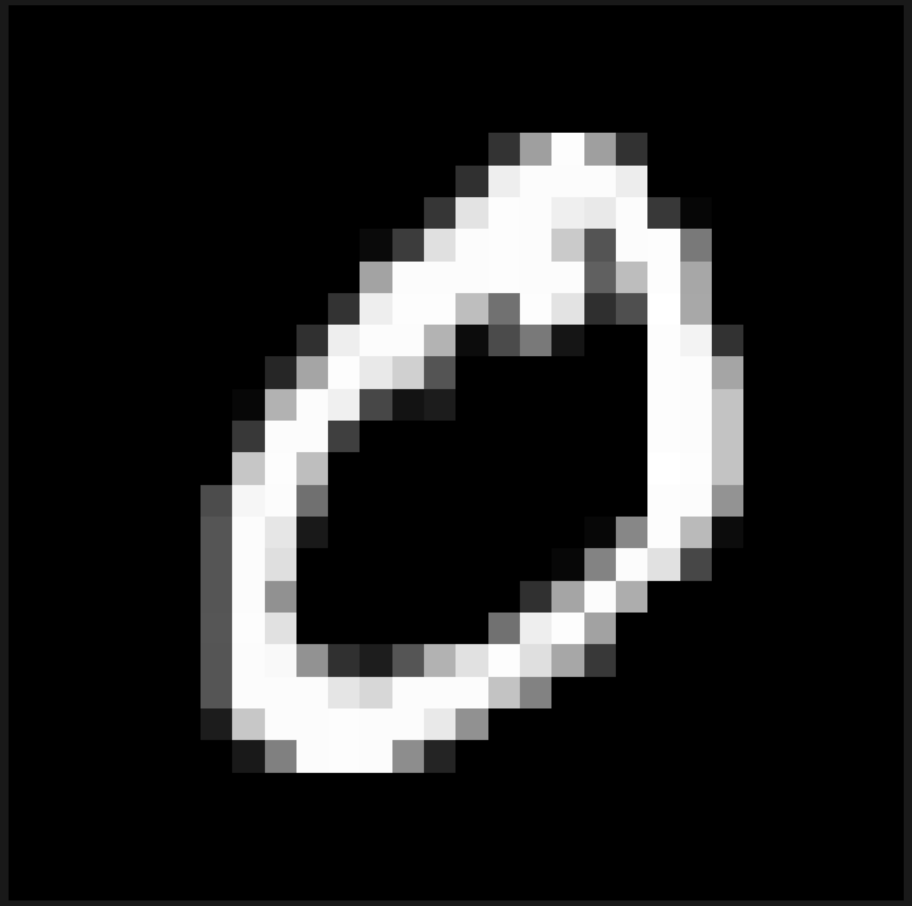


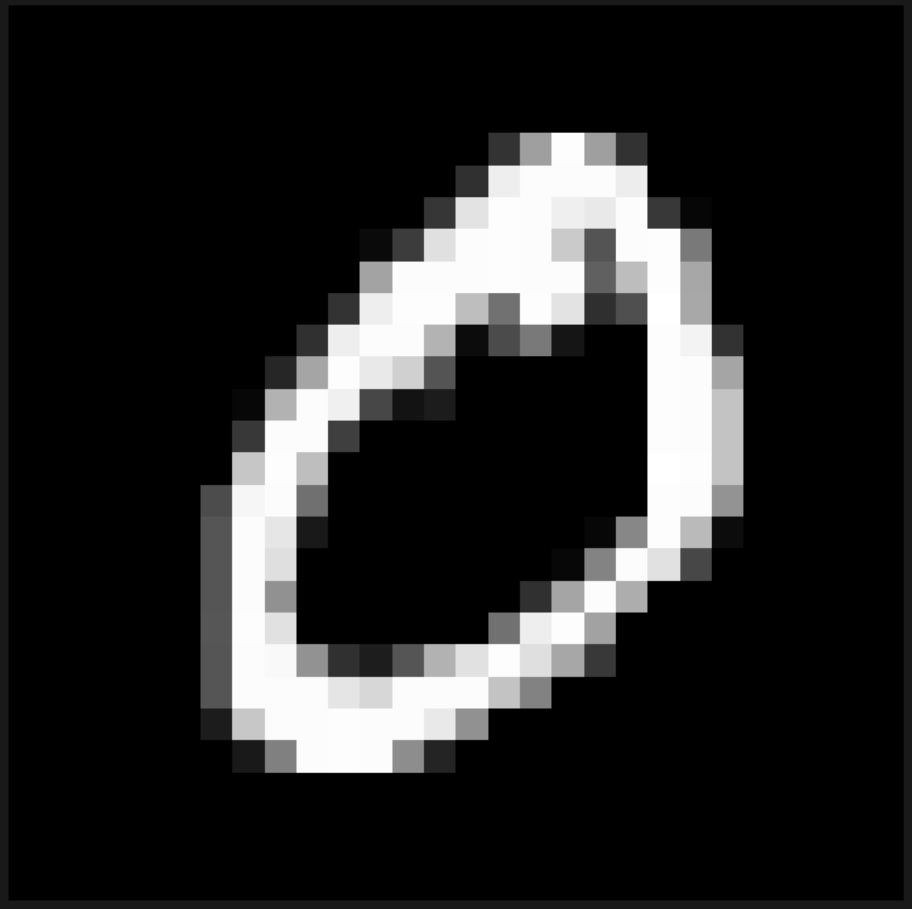




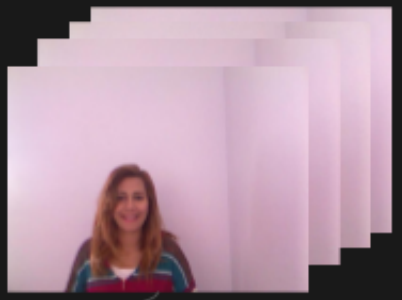








0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	159	253	159	50	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	238	252	252	252	237	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	54	227	253	252	239	233	252	57	6	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	10	60	224	252	253	252	202	84	252	253	122	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	163	252	252	252	253	252	252	96	189	253	167	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	51	238	253	253	190	114	253	228	47	79	255	168	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	48	238	252	252	179	12	75	121	21	0	0	253	243	50	0	0	0	0	0
0	0	0	0	0	0	0	0	38	165	253	233	208	84	0	0	0	0	0	0	253	252	165	0	0	0	0	0
0	0	0	0	0	0	0	7	178	252	240	71	19	28	0	0	0	0	0	0	253	252	195	0	0	0	0	0
0	0	0	0	0	0	0	57	252	252	63	0	0	0	0	0	0	0	0	0	253	252	195	0	0	0	0	0
0	0	0	0	0	0	0	198	253	190	0	0	0	0	0	0	0	0	0	0	255	253	196	0	0	0	0	0
0	0	0	0	0	0	76	246	252	112	0	0	0	0	0	0	0	0	0	0	253	252	148	0	0	0	0	0
0	0	0	0	0	0	85	252	230	25	0	0	0	0	0	0	0	0	7	135	253	186	12	0	0	0	0	0
0	0	0	0	0	0	85	252	223	0	0	0	0	0	0	0	0	7	131	252	225	71	0	0	0	0	0	0
0	0	0	0	0	0	85	252	145	0	0	0	0	0	0	0	48	165	252	173	0	0	0	0	0	0	0	0
0	0	0	0	0	0	86	253	225	0	0	0	0	0	0	114	238	253	162	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	85	252	249	146	48	29	85	178	225	253	223	167	56	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	85	252	252	252	229	215	252	252	252	196	130	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	28	199	252	252	253	252	252	233	145	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	25	128	252	253	252	141	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0



Localización de los labios

Extracción de características

Clasificación

ID	Palabras	ID	Frases
1	Begin	1	Stop Navigation
2	Choose	2	Excuse me
3	Connection	3	I am sorry
4	Navigation	4	Thank you
5	Next	5	Good bye
6	Previous	6	I love this game
7	Start	7	Nice to meet you
8	Stop	8	You are welcome
9	Hello	9	How are you?
10	Web	10	Have a good time

$$WRR = \frac{\#PALABRAS\ RECONOCIDAS\ CORRECTAMENTE}{\#PALABRAS\ TOTALES}$$

Estado del Arte

VIDTIMIT

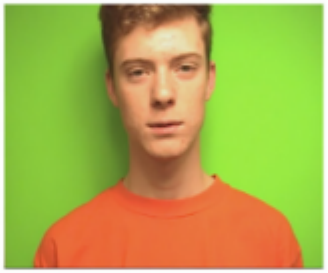
AVLetters

OuluVS2

AVICAR

LILiR

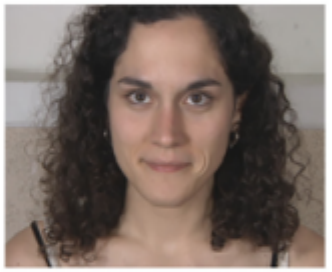
AV@CAR



CUAVE

GRID

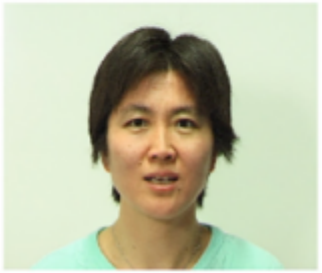
TCD-TIMIT



VLRF

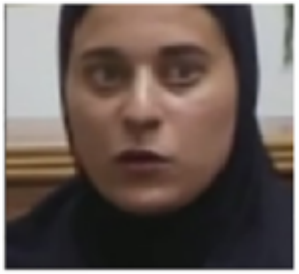
LRW

MOBIO

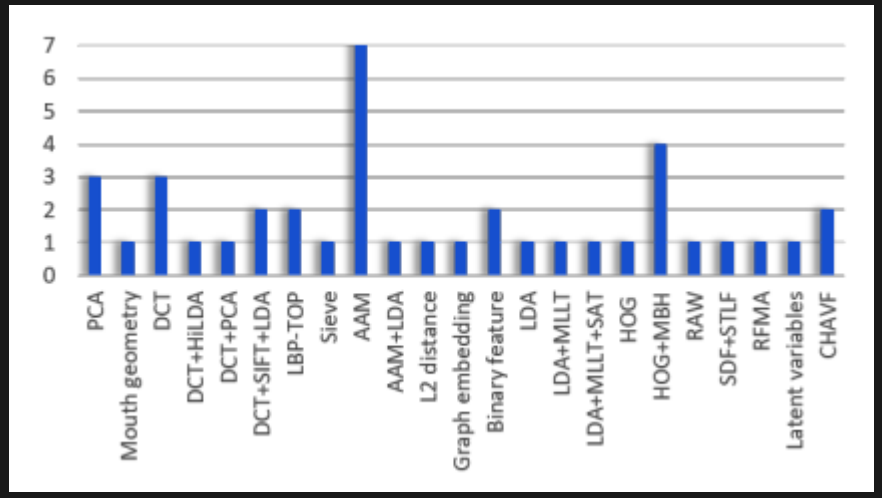


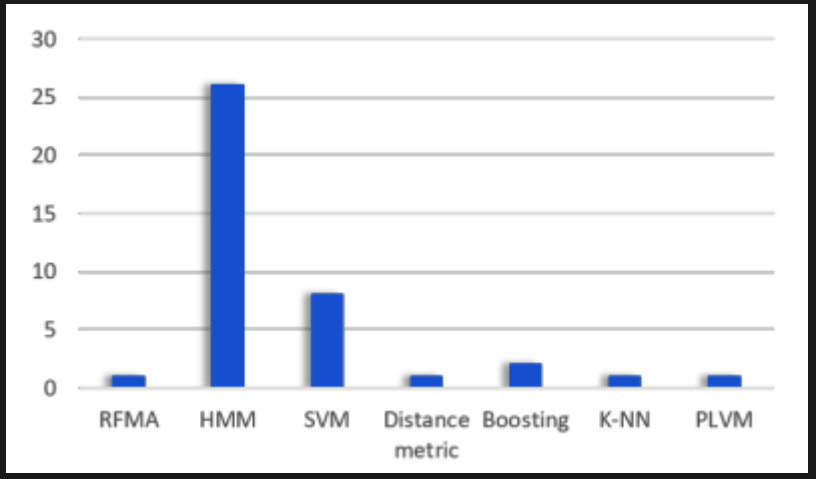
OuluVS

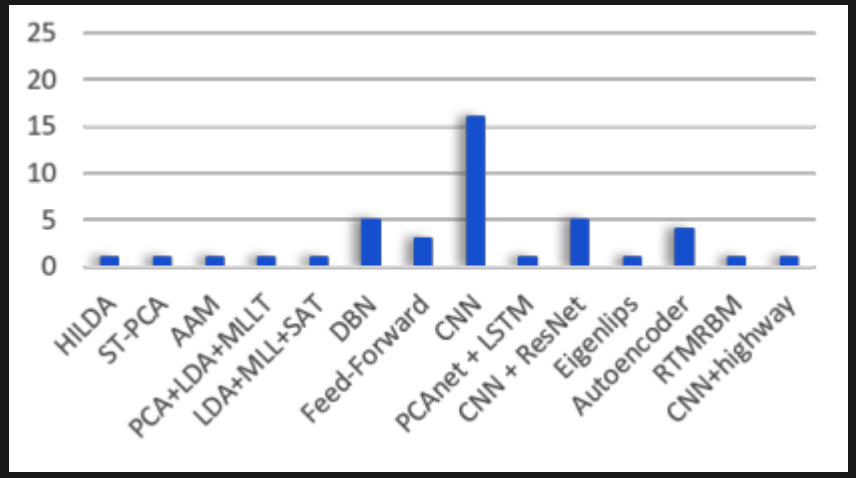
XM2VTSDB

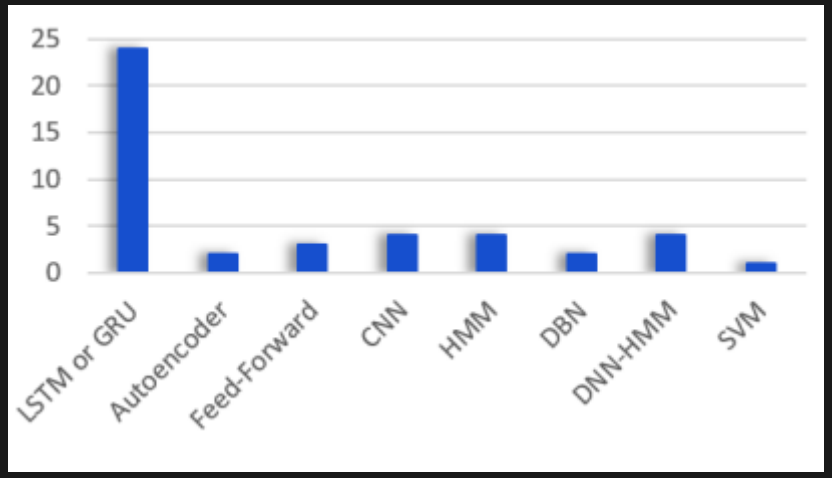


MV-LRS


7
6
5
4
3
2
1
0
PCA
Mouth geometry
DCT
DCT+HiLDA
DCT+PCA
DCT+SIFT+LDA
LBP-TOP
Sieve
AAM
AAM+LDA
L2 distance
Graph embedding
Binary feature
LDA
LDA+MLLT
LDA+MLLT+SAT
HOG
HOG+MBH
RAW
SDF+STLF
RFMA
Latent variables
CHAVF

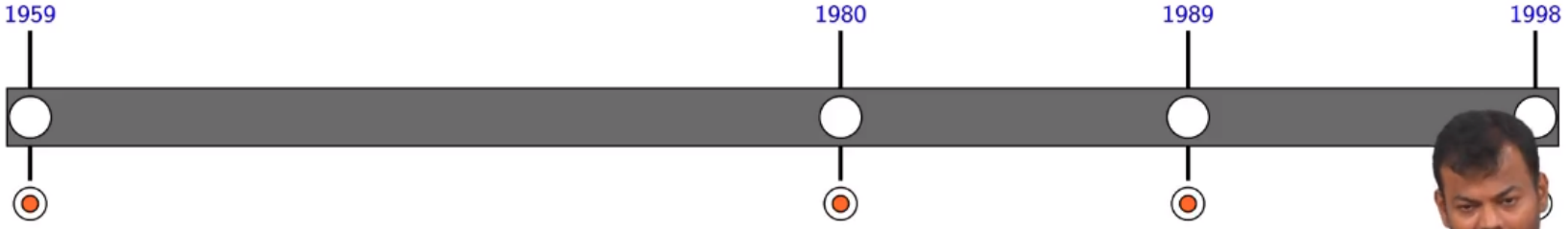

30
25
20
15
10
5
0
RFMA
HMM
SVM
Distance metric
Boosting
K-NN
PLVM

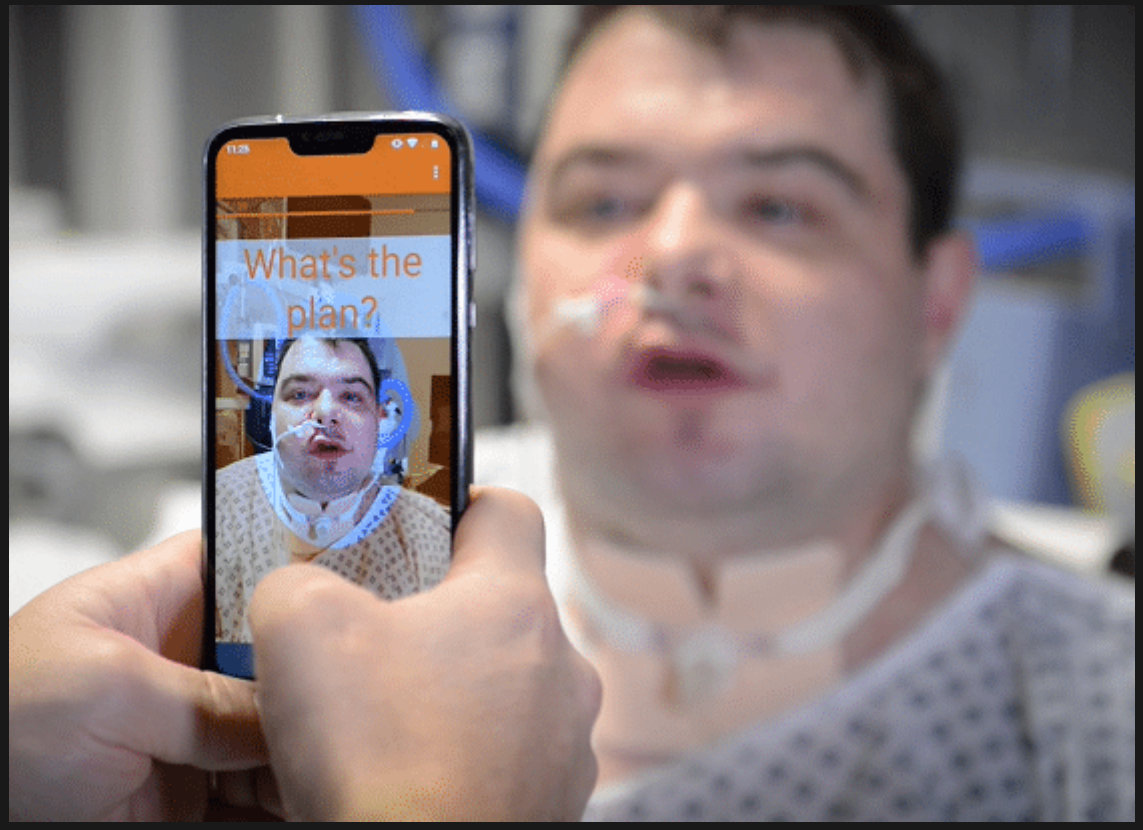

25
20
15
10
5
0
HILDA
ST-PCA
AAM
PCA+LDA+MLLT
LDA+MLL+SAT
DBN
Feed-Forward
CNN
PCAnet + LSTM
CNN + ResNet
Eigenlips
Autoencoder
RTMRBM
CNN+highway


25
20
15
10
5
0
LSTM or GRU
Autoencoder
Feed-Forward
CNN
HMM
DBN
DNN-HMM
SVM

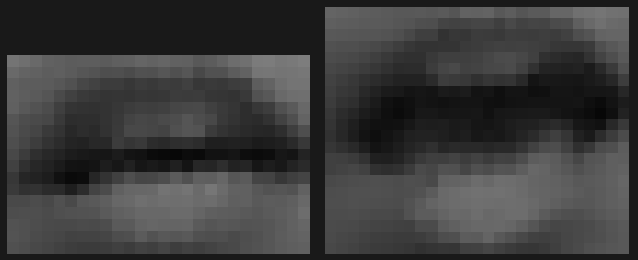
LeNet-5

Introduced the (now famous) MNIST dataset (LeCun et. al.)[21]

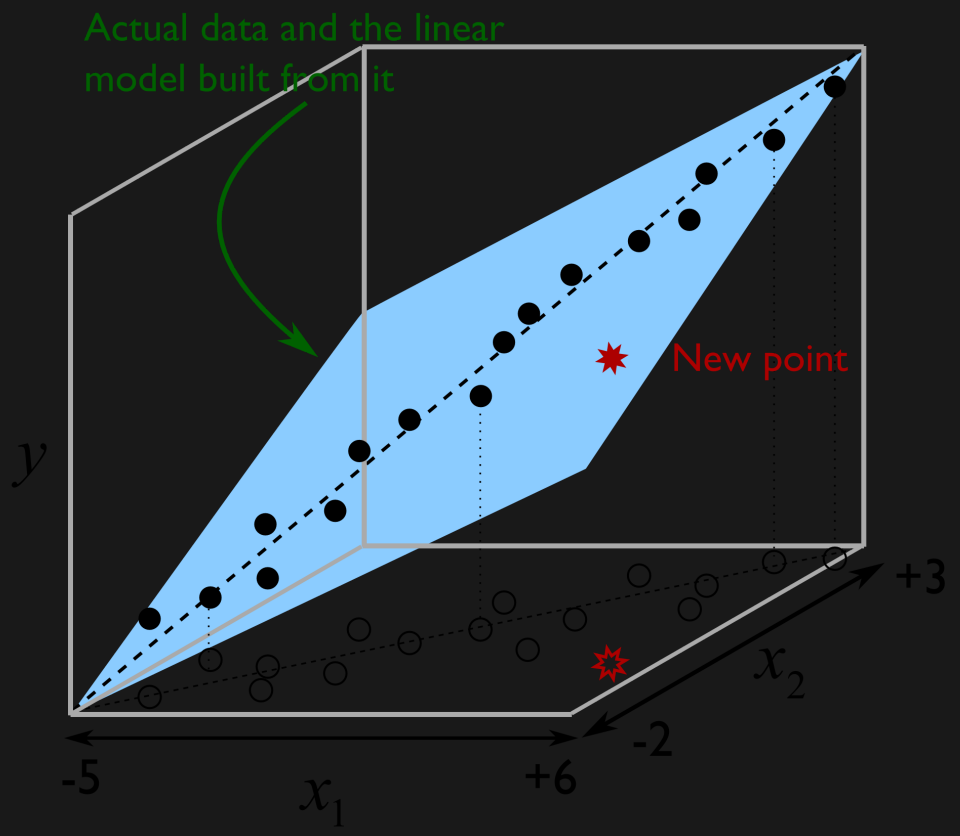


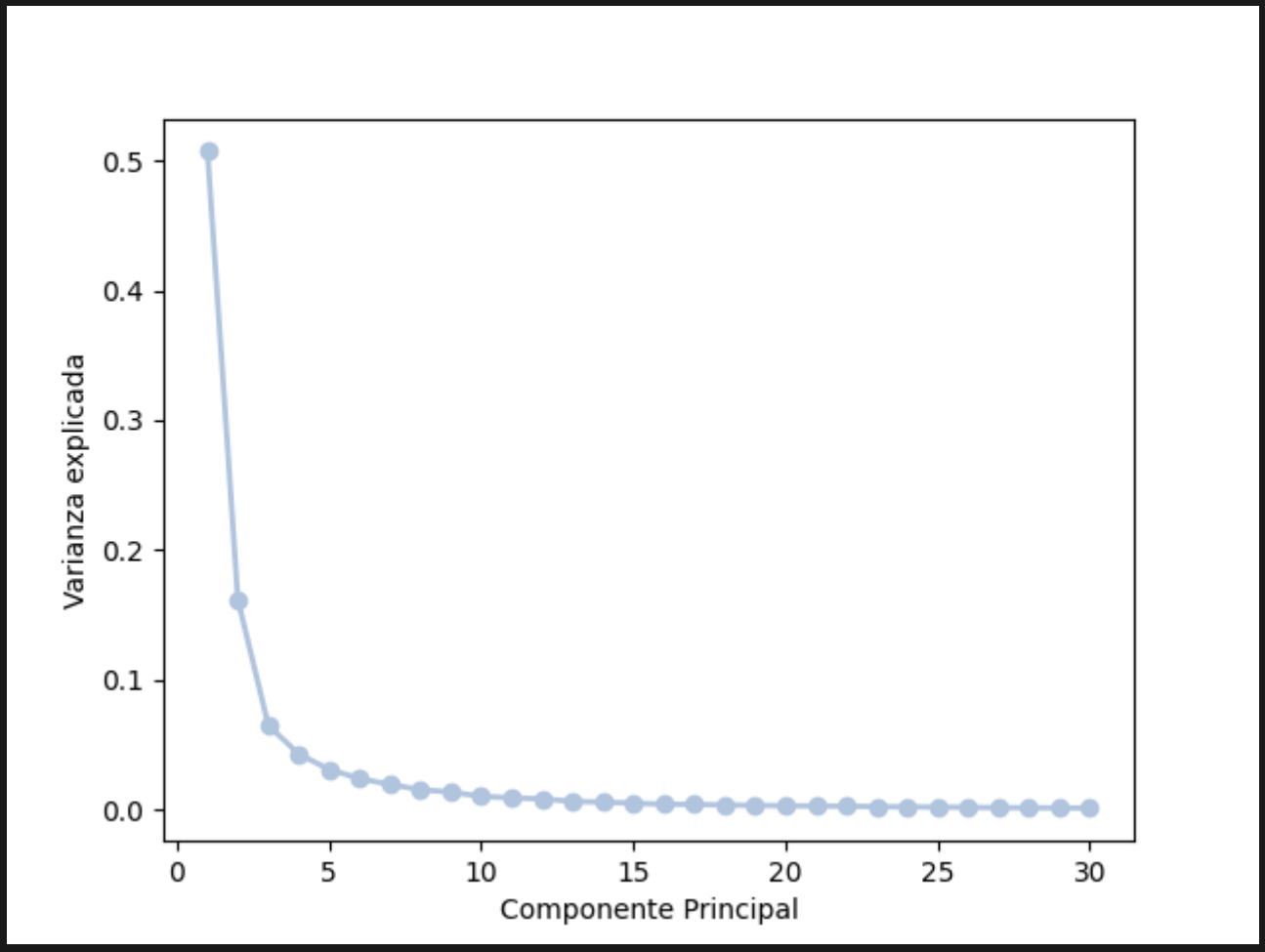

What's the plan?

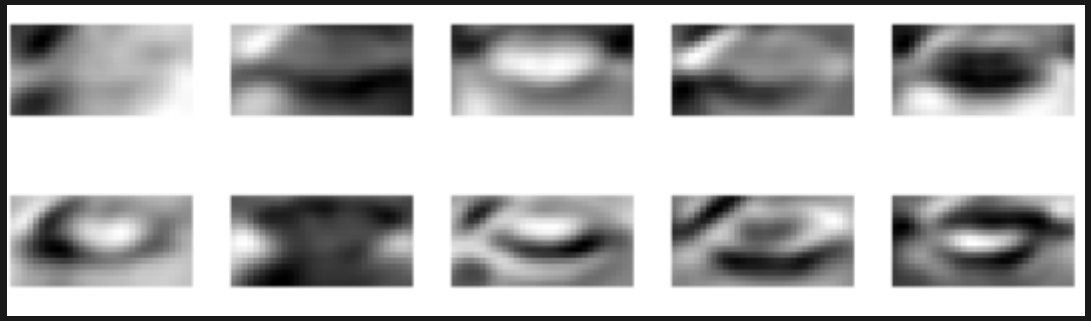
Sistemas Tradicionales



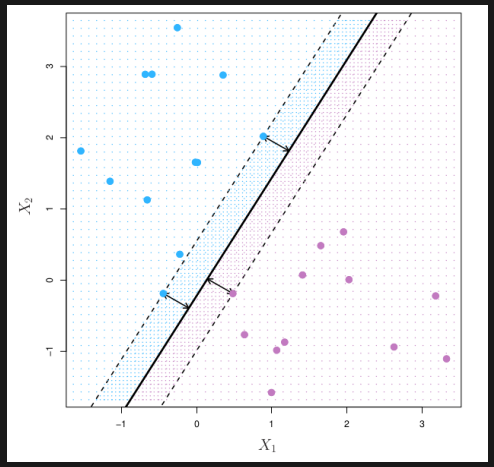
Análisis de Componentes Principales

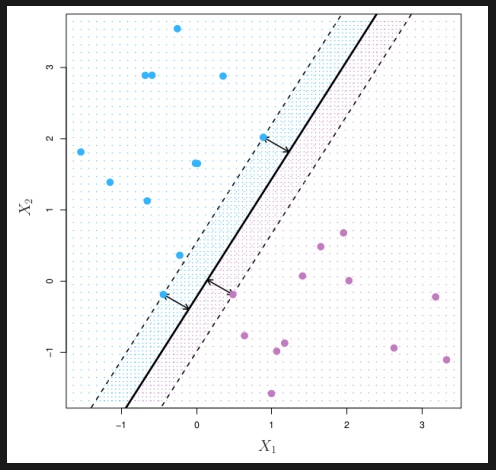


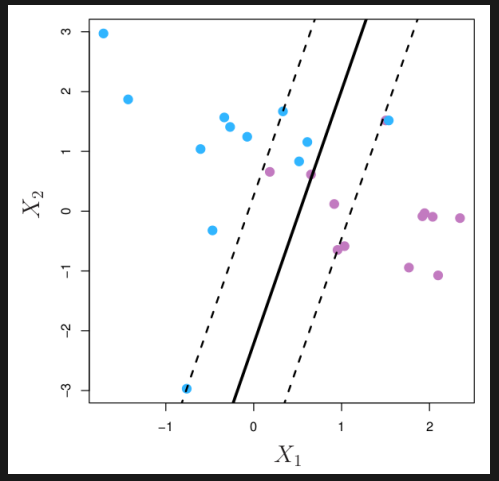

0.5
0.4
0.3
0.2
0.1
0.0
Varianza explicada
0
5
10
15
20
25
30
Componente Principal



Máquina de Soporte Vectorial

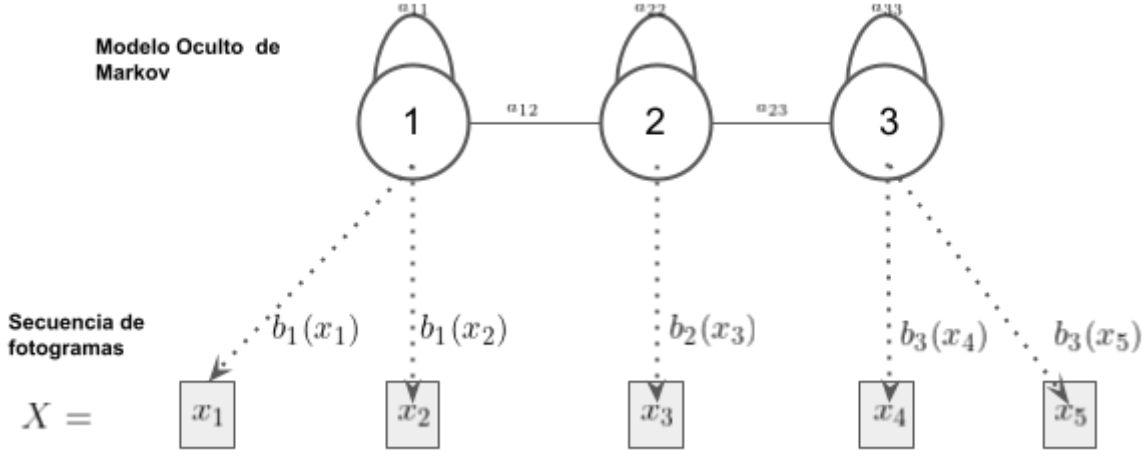

X_1
X_2
−1
0
1
2
3
−1
0
1
2
3


X_1
X_2


X_1
X_2

	Palabras	Frases
Entrenamiento	99.58%	100%
Validación	61%	58%
Test	34%	41%

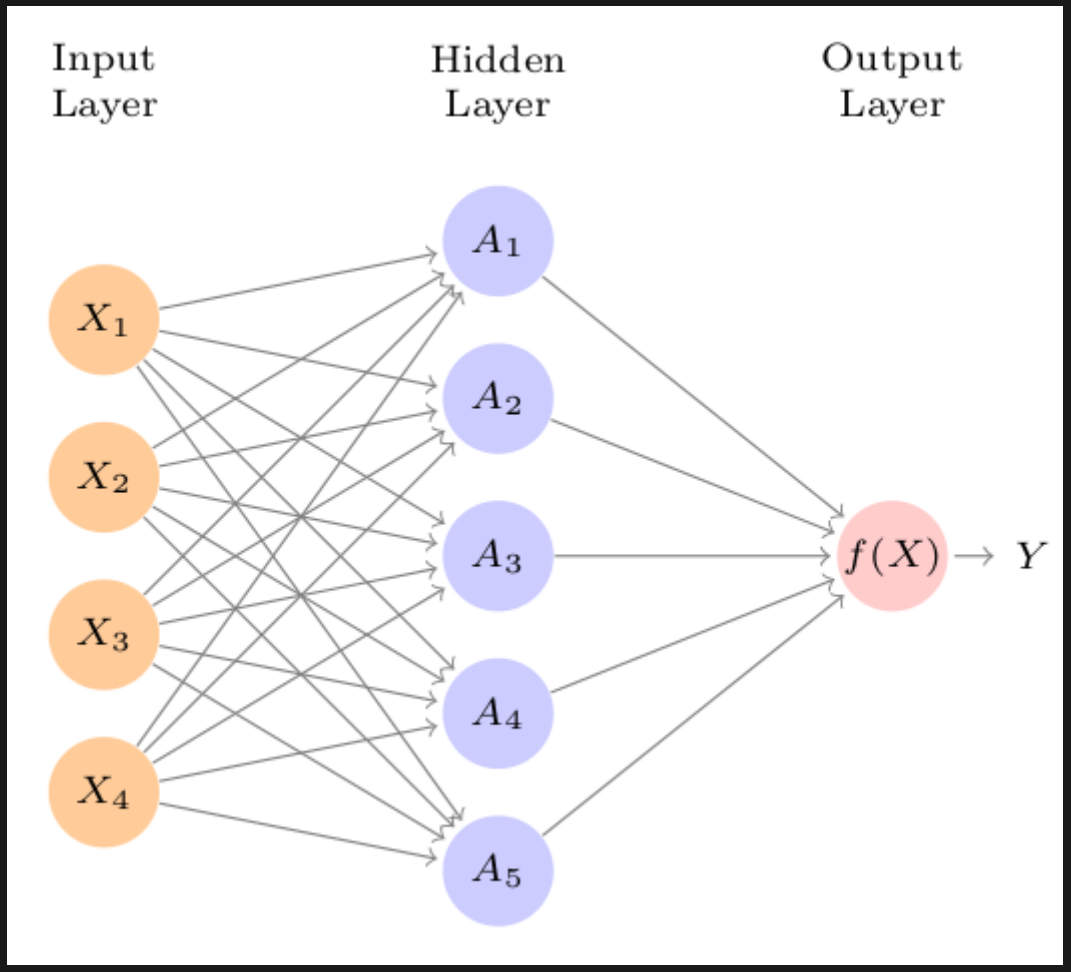
Modelo Oculto de Markov

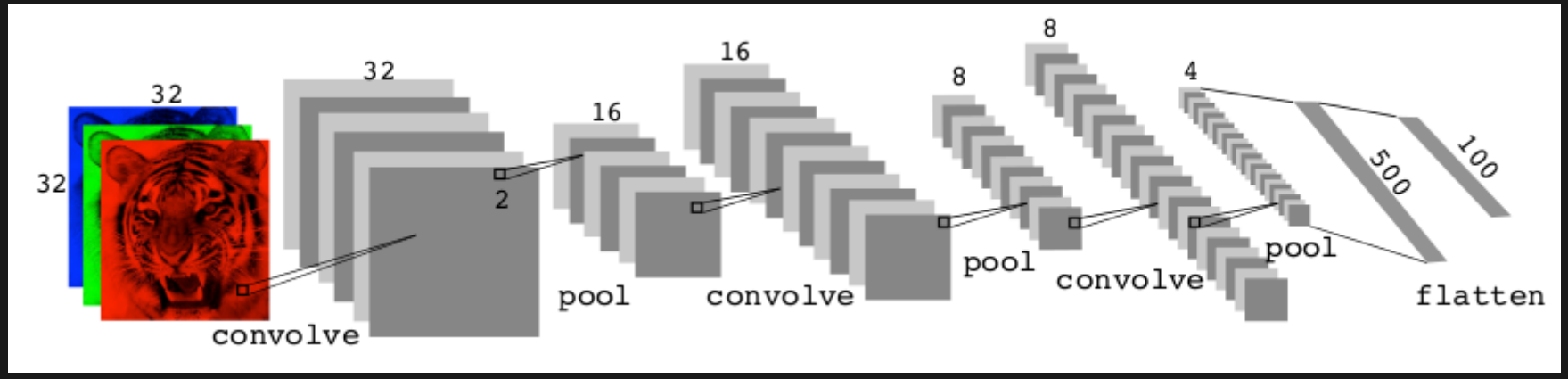

Modelo Oculto de Markov
a_{11}
a_{22}
a_{33}
1
a_{12}
2
a_{23}
3
Secuencia de fotogramas
$b_1(x_1)$
$b_1(x_2)$
$b_2(x_3)$
$b_3(x_4)$
$b_3(x_5)$
$X =$
x_1
x_2
x_3
x_4
x_5

$$P(X|\lambda)$$

	Palabras	Frases
Entrenamiento	41%	34%
Validación	20%	10%
Test	16%	22%

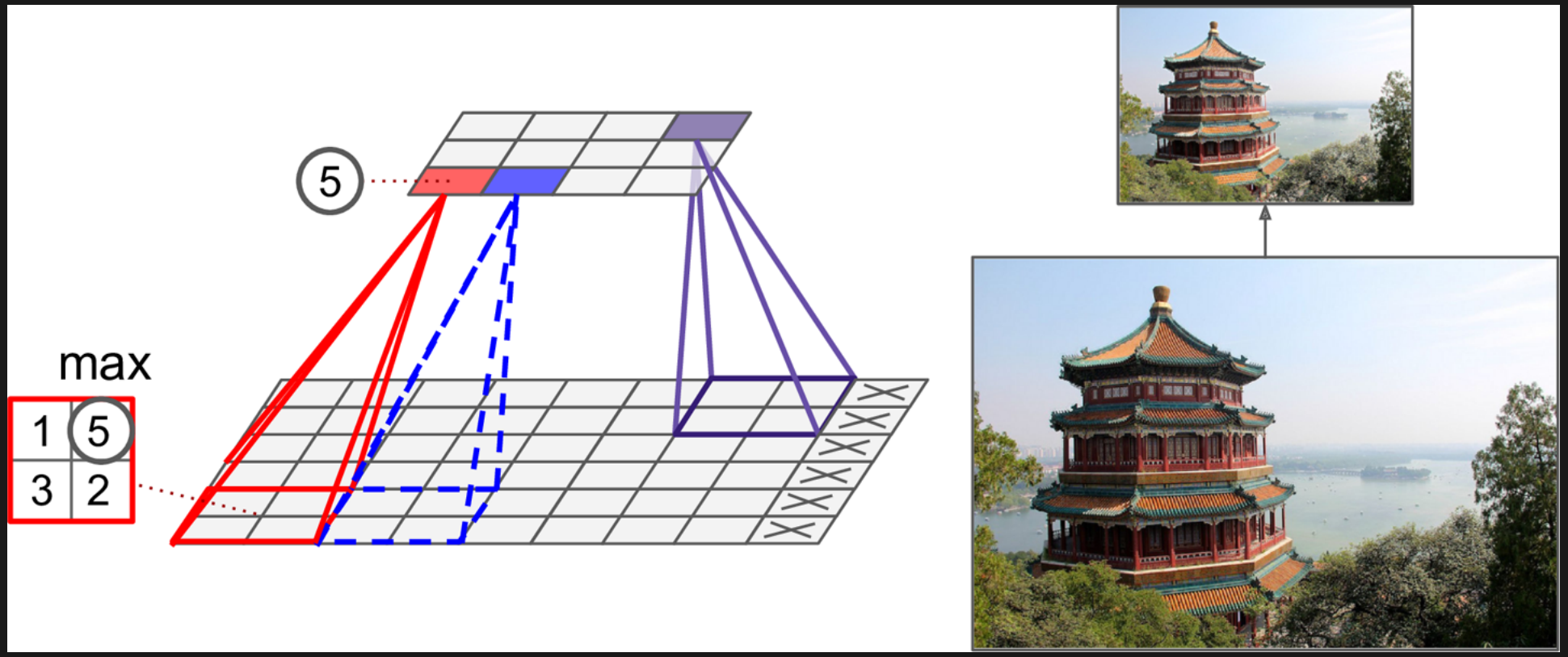
Deep Learning

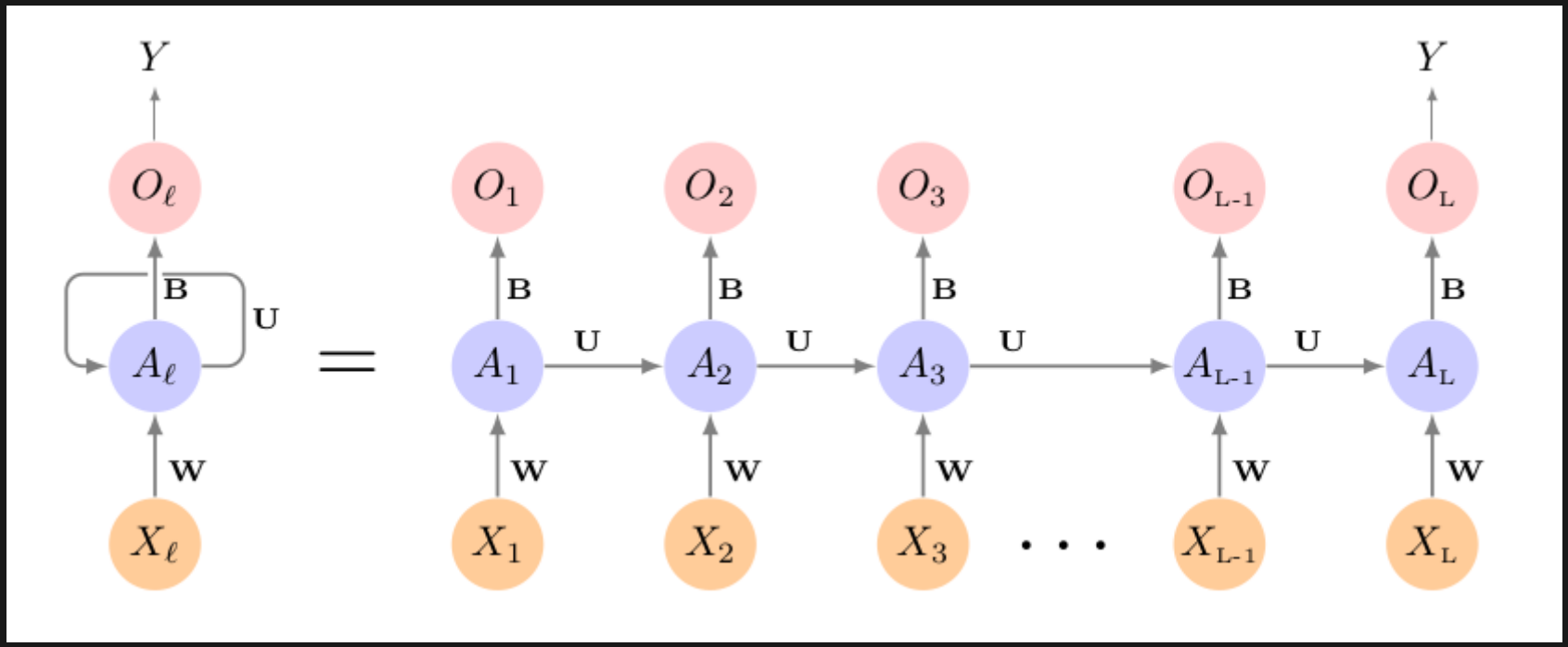

Input Layer
Hidden Layer
Output Layer
X_1
X_2
X_3
X_4
A_1
A_2
A_3
A_4
A_5
$f(X)$
Y

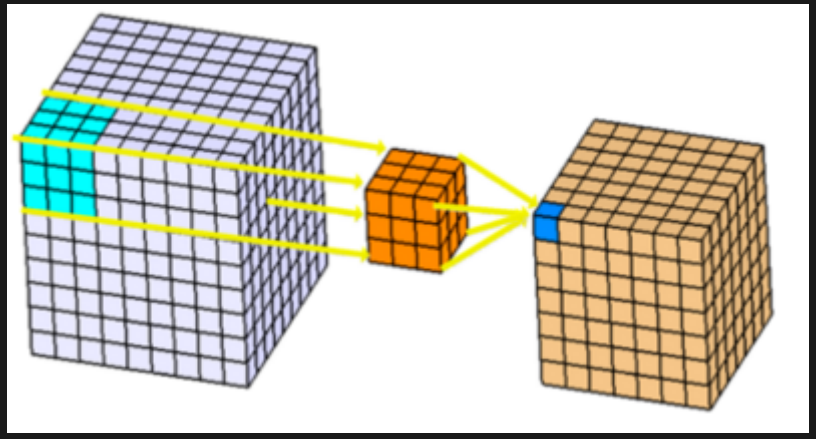

32
32
32
convolve
16
2
16
pool
8
16
convolve
8
pool
4
8
convolve
pool
500
100
flatten

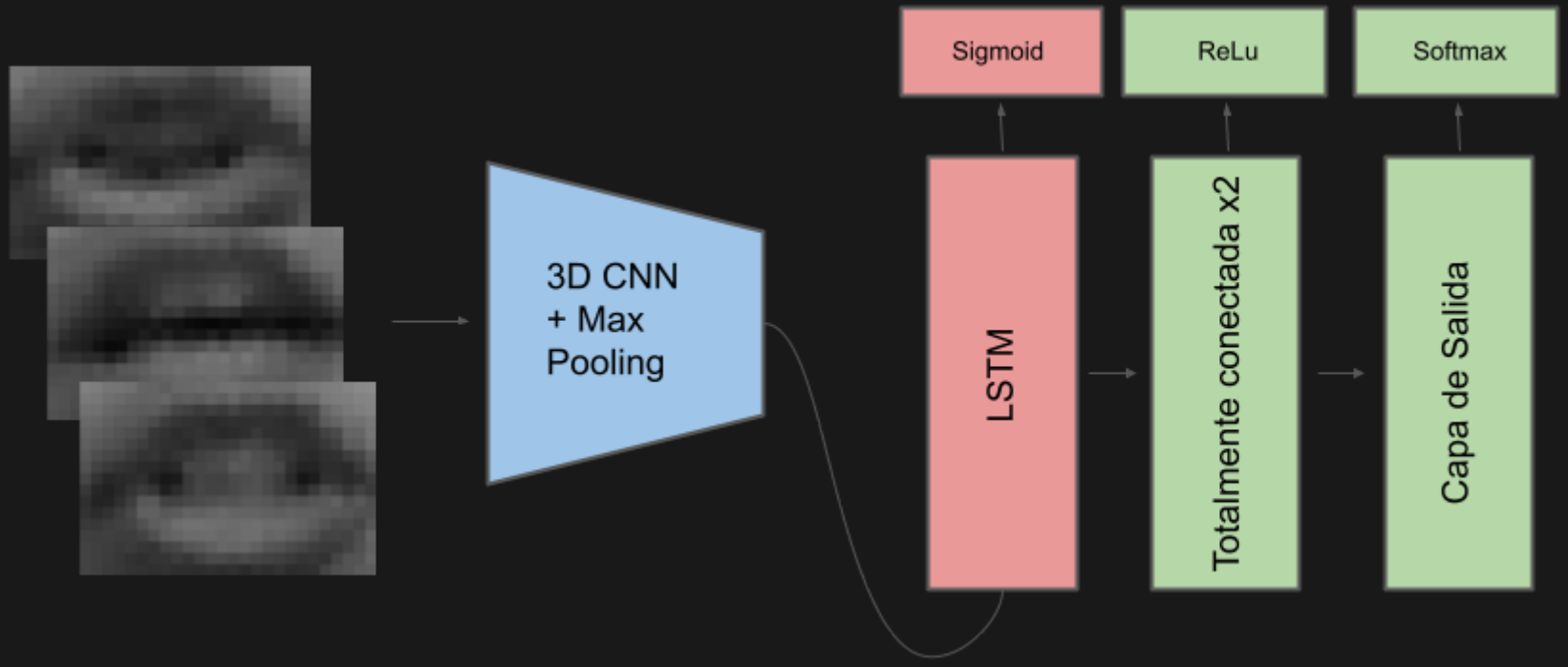
-1	0	1
-2	0	2
-1	0	1

-1	-2	-1
0	0	0
1	2	1


5
max
1
5
3
2


Y
O_ℓ
B
U
A_ℓ
W
X_ℓ
=
O_1
O_2
O_3
O_{L-1}
O_L
A_1
A_2
A_3
A_{L-1}
A_L
X_1
X_2
X_3
X_{L-1}
X_L




Sigmoid
ReLu
Softmax
3D CNN
+ Max
Pooling
LSTM
Totalmente conectada x2
Capa de Salida

	Palabras	Frases
Entrenamiento	86%	71%
Validación	40%	42%
Test	45%	30%

Conclusiones