



UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE
CORSO DI LAUREA TRIENNALE

MACHINE LEARNING FOR SOCIAL MEDIA DATA ANALYSIS

Relazione della prova finale di ALESSANDRO MANGONE 793602
Relatore FRANCESCA GASPARINI
Anno Accademico 2016-2017

INTRODUZIONE

Obiettivi principali:

- 1 Creare un modello che data una immagine e la sua descrizione testuale sia in grado di predire l'emozione indotta
- 2 Impiego reti neurali per il riconoscimento delle emozioni da immagini
- 3 Impiego autoencoders per il riconoscimento delle emozioni da descrizioni testuali

DATASET



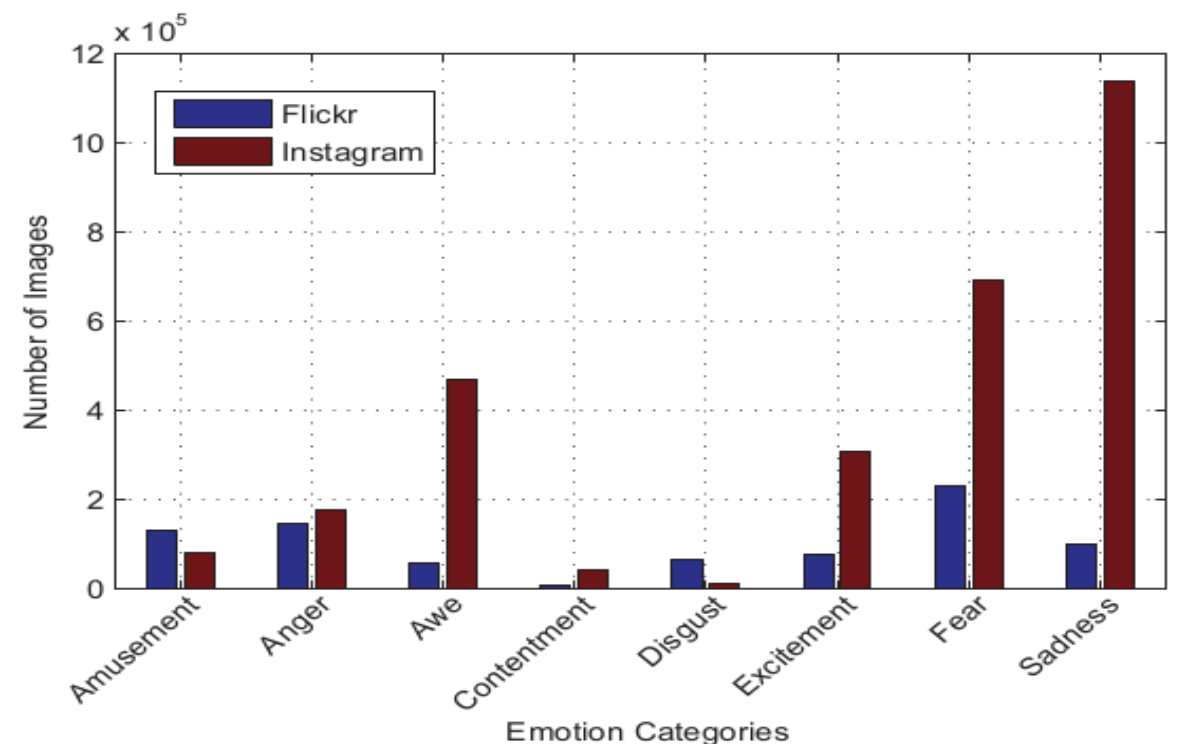
Large Scale Dataset
ottenuto interrogando
l'image search engine di
Flickr e Instagram



Utilizzo delle otto classi
come tags per la ricerca



Dataset weakly labeled
(URLs disponibili)



AMAZON MECHANICAL TURK



“

Do You feel anger
seeing this photo?

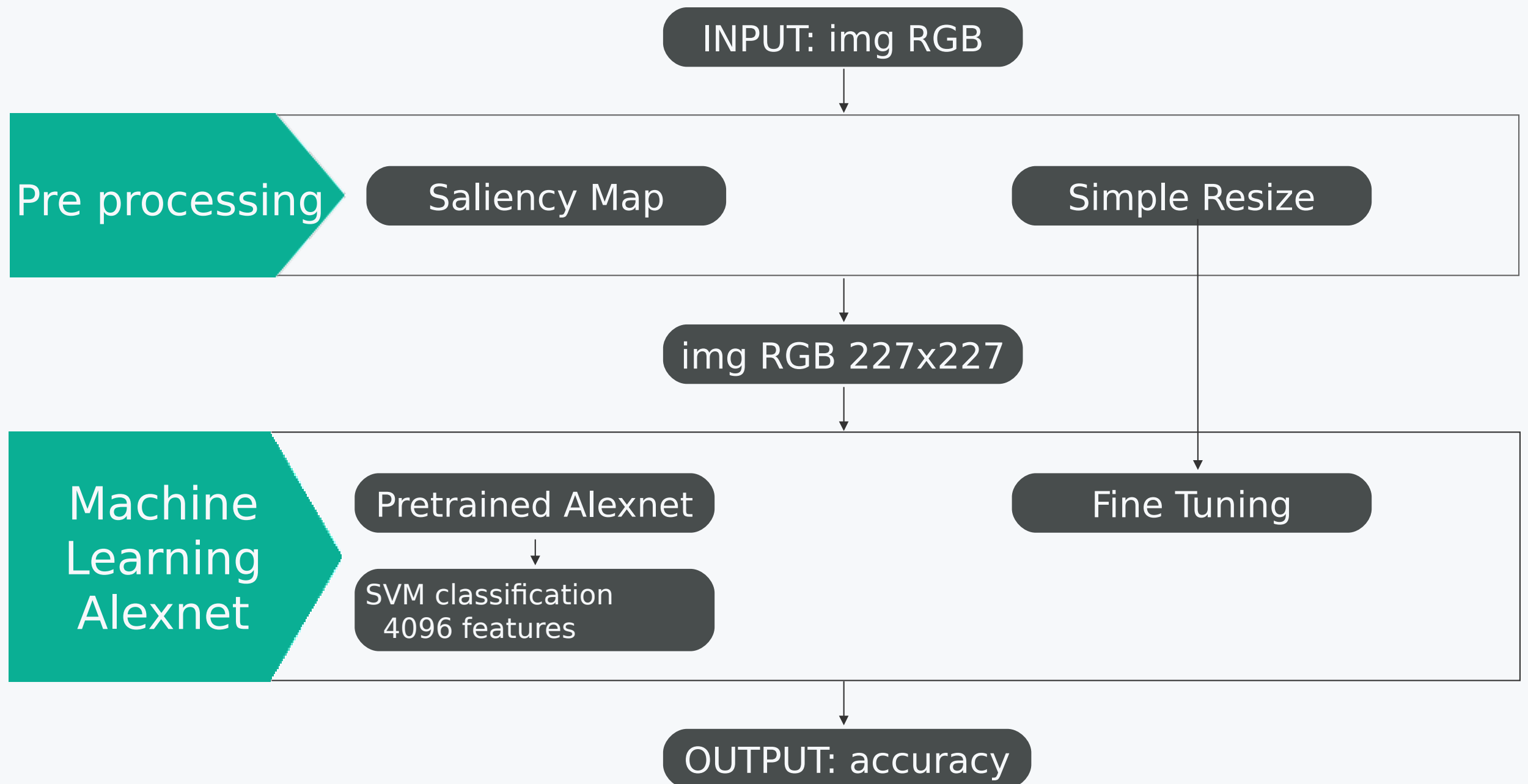
”



Dataset well labeled

Dataset	Submitted	Labeled
Amusement	11.000	4.492
Anger	11.000	1.266
awe	11.000	3.151
Contentment	11.000	5.374
Disgust	11.000	1.568
Excitement	11.000	2.963
Fear	11.000	1.032
Sadness	11.000	2.922
Sum	88.000	23.308

RICONOSCIMENTO EMOZIONI DA IMMAGINI



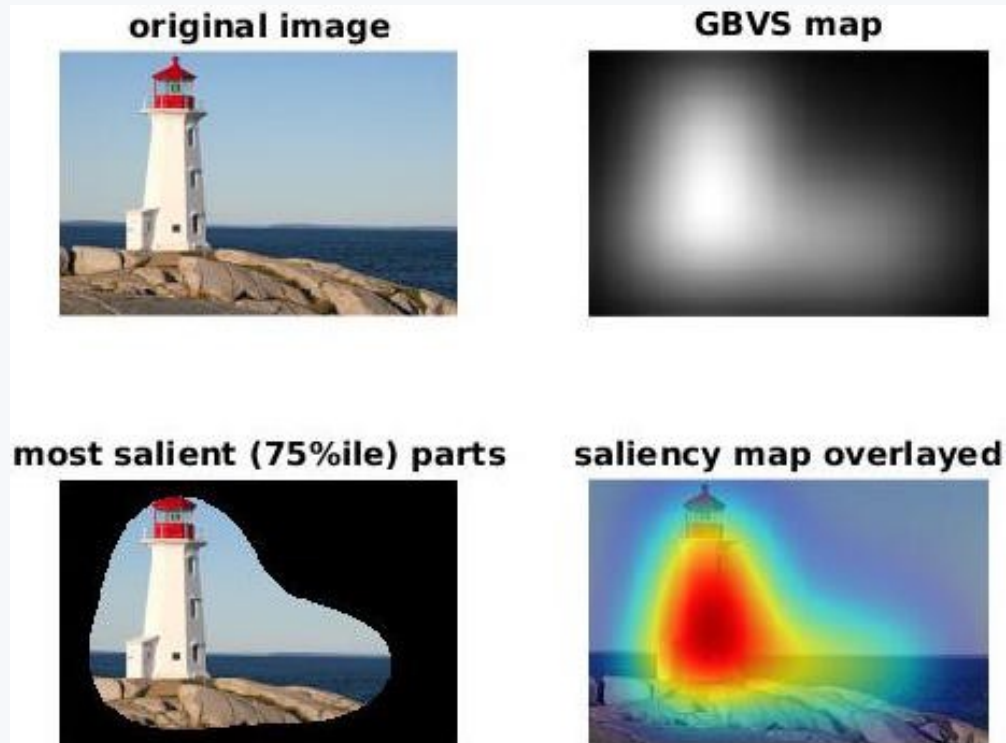
PRE PROCESSING



Mappe di salienza usate per ridurre il numero di distrazioni e ricercare l'elemento di massimo interesse



Algoritmo GBVS che si basa sul modello Itti & Koch



Semplice resize dell'immagine attraverso il comando `matlab imresize()`



PRO: più veloce delle mappe di salienze



CONTRO: immagini distorte



227 x 227

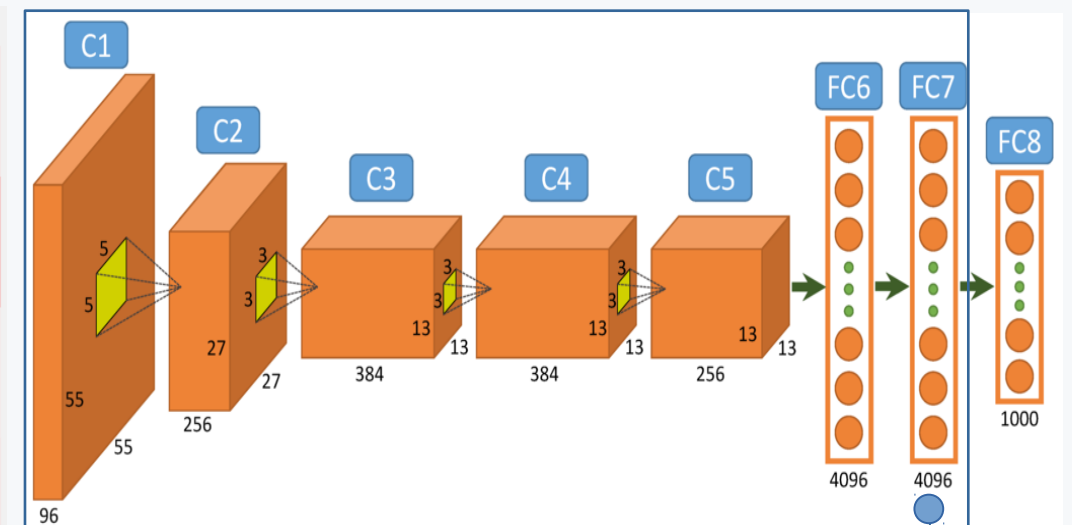


PRETRAINED ALEXNET: simple resize

Model 1

True class \ Predicted class	amusement	anger	awe	contentment	disgust	excitement	fear	sadness	True Positive Rate	False Negative Rate
amusement	84%	1%	2%	3%		10%		1%	84%	16%
anger	12%	15%	1%	21%	1%	9%		41%	15%	85%
awe	3%		91%	3%	1%	2%		1%	91%	9%
contentment	1%	2%	4%	81%	1%	2%		8%	81%	19%
disgust	5%	2%	3%	12%	67%	2%		9%	67%	33%
excitement	29%	1%	3%	8%	1%	55%		2%	55%	45%
fear	7%	7%		14%	14%	7%	3%	48%	3%	97%
sadness	1%	2%	2%	19%	4%	2%		69%	69%	31%

Matrice ottenuta su dataset 5-4-3



SVM

5
(5.191 immagini)

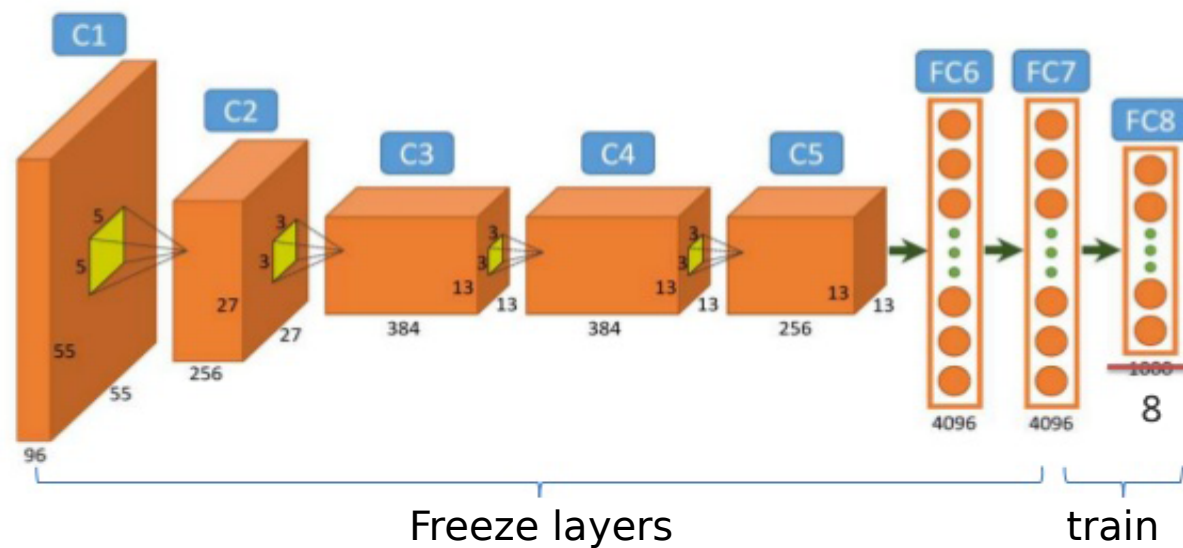
74%

5-4-3
(23.308 immagini)

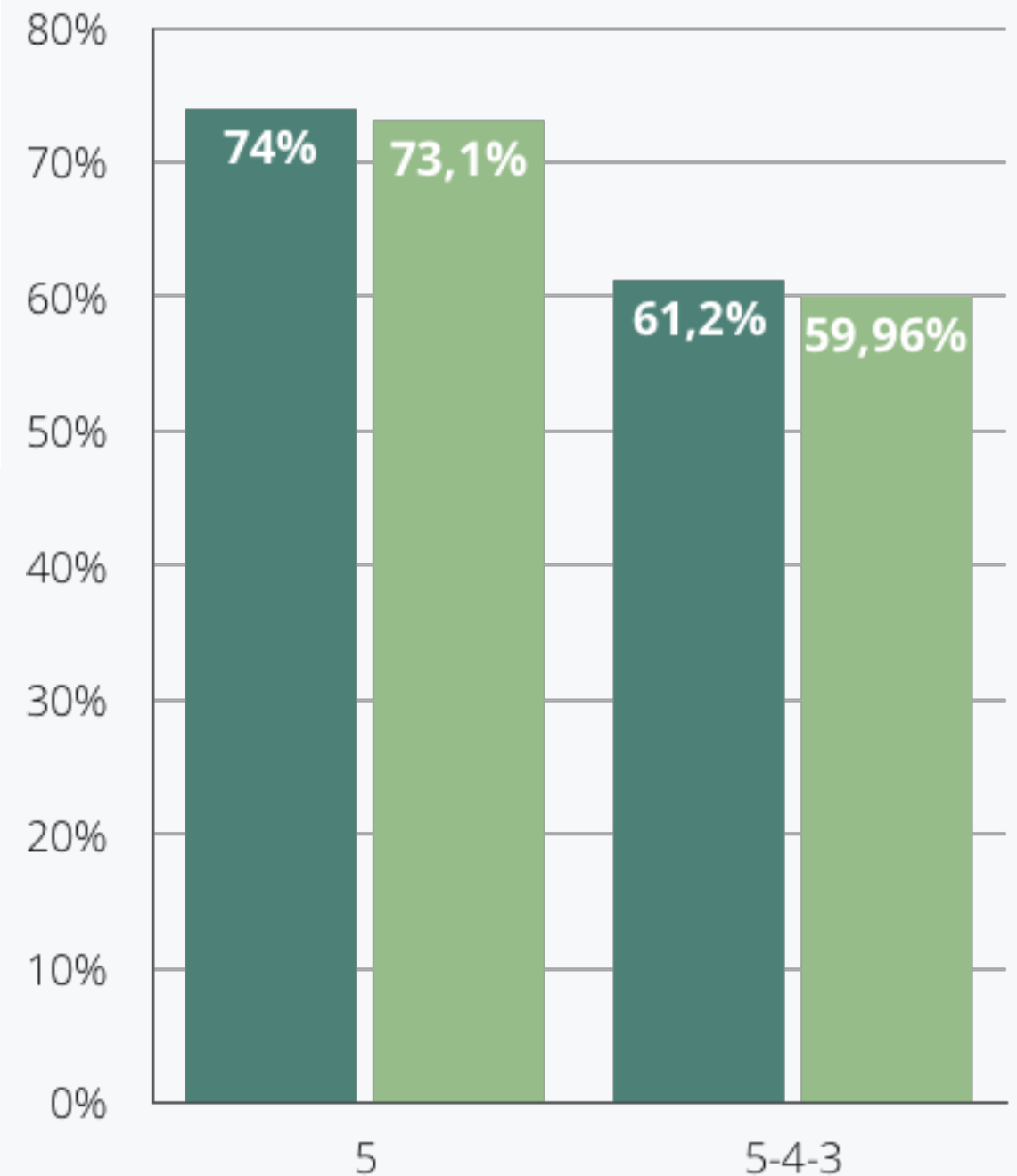
61.20%

FINE TUNING ALEXNET

Fine-tuning Pretrained Network

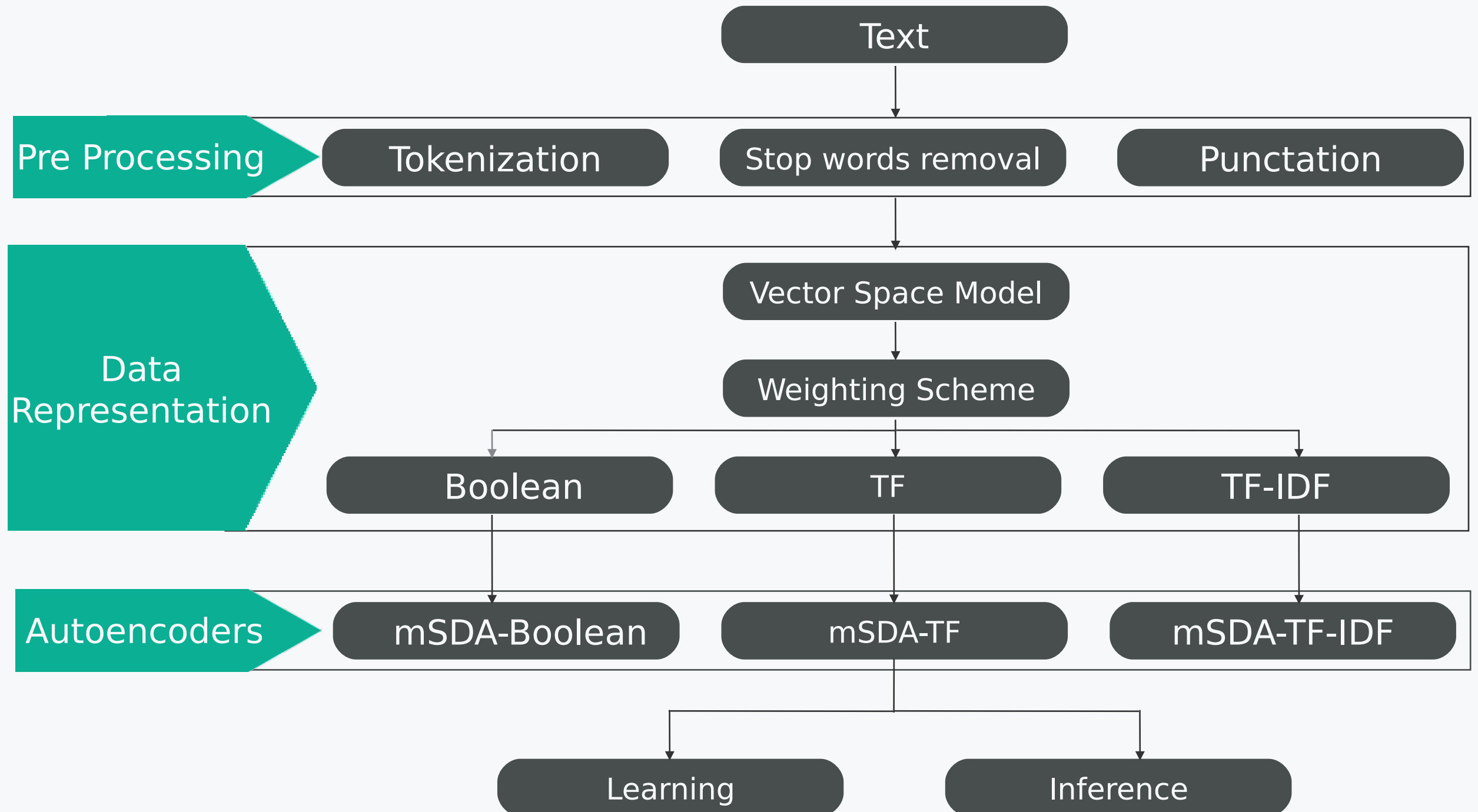


Alexnet Fine Tuning Alexnet



- 1 Sfrutta parte di una rete già addestrata
- 2 Vengono copiati n-layers della rete pretrained
- 3 I layers restanti vengono addestrati da zero

RICONOSCIMENTO EMOZIONI DA DESCRIZIONI TESTUALI



DATA REPRESENTATION

A swimmer likes swimming, thus he swims

a swimmer likes swimming thus he swims

Esempio di tokenization

BINARY REPRESENTATION

Rappresentazione semplice e semplificativa. La matrice dei pesi dei termini che compongono il testo tiene conto della presenza o assenza di una parola nel testo. Il valore 1 indica la presenza del termine, il valore 0 nel caso contrario.

	t_1	t_2	t_3	t_4	t_5	...	...	...	...	...	...	...	...	...	t_n
img_1	0	0	1	0	1	0	1	0	0	1	0	0	0	1	1
img_2	0	1	1	0	0	0	1	1	1	0	1	1	0	0	0
...	1	0	1	1	0	1	0	0	0	0	1	1	1	1	0
img_m	1	0	1	1	0	1	0	0	0	0	1	1	1	1	0

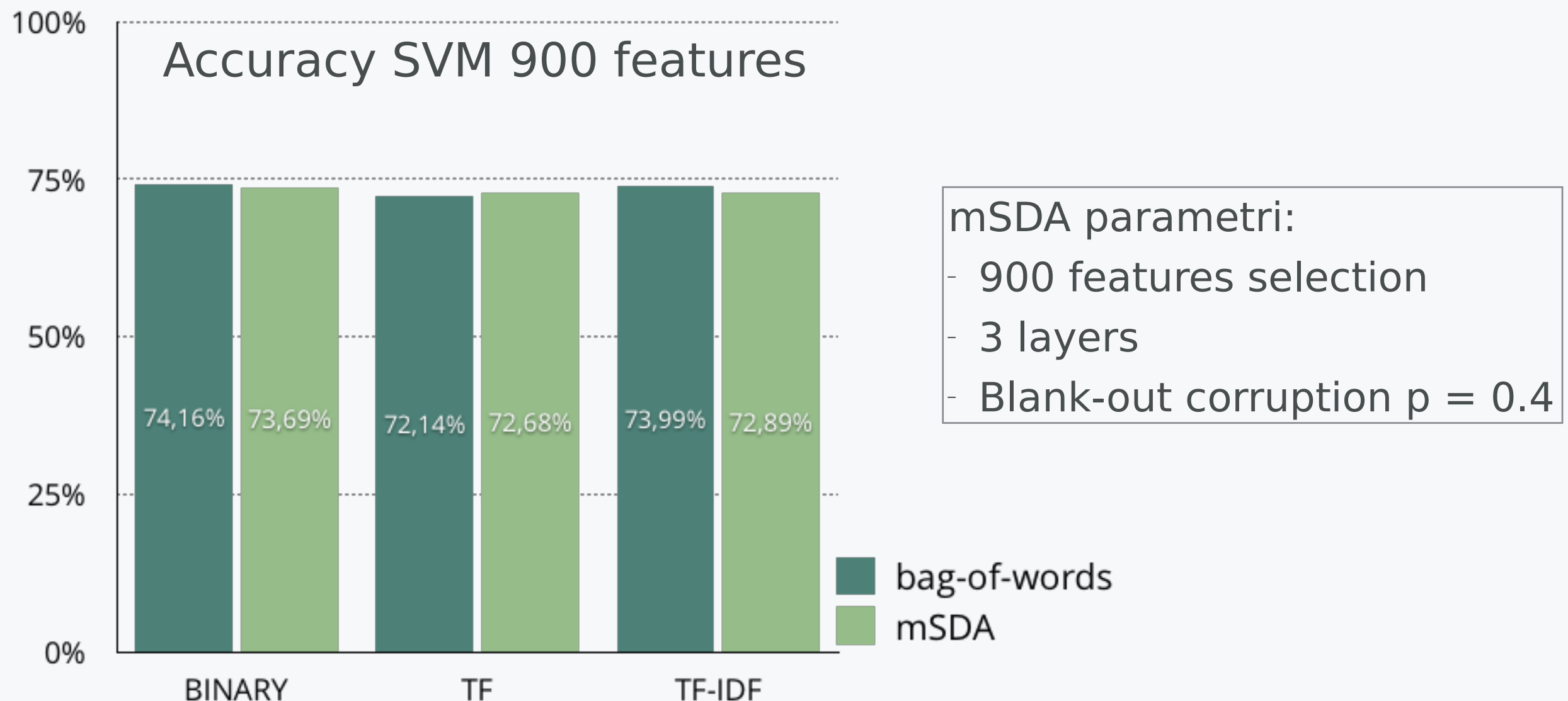
MARGINALIZED STACKED DENOISING AUTOENCODER



Denoising: capacità di ricostruire i dati a partire da input corrotti



Stacked denoising: autoencoder con multipli hidden layers



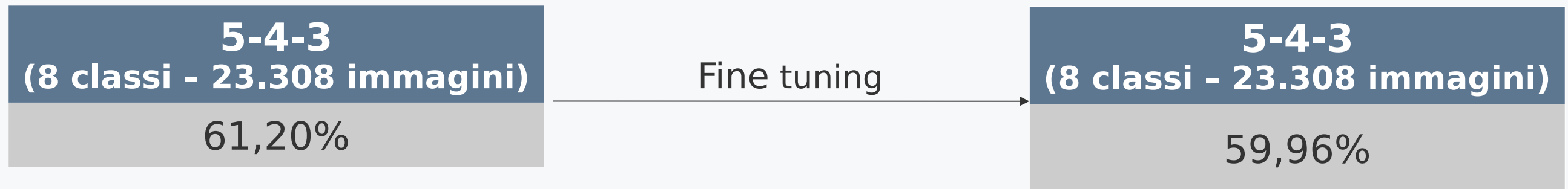
FINE TUNING MSDA

	p=0.1	p=0.2	p=0.3	p=0.4	p=0.5	BOW-BINARY
LAYER 1	74,34%	74,44%	74,48%	74,63%	74,69%	73,69%
LAYER 2	74,46%	74,71%	74,69%	74,45%	74,14%	
LAYER 3	74,63%	74,47%	74,05%	73,69%	73,69%	
LAYER 4	74,62%	74,12%	73,53%	73,39%	73,34%	
LAYER 5	74,49%	73,80%	73,24%	73,21%	73,12%	

CONCLUSIONI

Riconoscimento emozioni da immagini:

Alexnet - simple resize preprocess



Riconoscimento emozioni da descrizioni testuali:

	BINARY	TF	TF-IDF
BOW	74,16%	72%	73,99%
mSDA	73,69%	72,68%	72,89%

	BINARY	TF	TF-IDF
Fine tuning mSDA	74,70%	74,70%	74,45%