



Dipartimento di Ingegneria dell'Informazione
Laurea Magistrale in Ingegneria Informatica

Reti Neurali Convoluzionali per la Classificazione di Geroglifici Egizi

Candidato
Marco Loschiavo

Relatori
Prof. Fabrizio Argenti
Prof. Alessandro Piva
Prof. Andrea Barucci
Prof. Massimiliano Franci

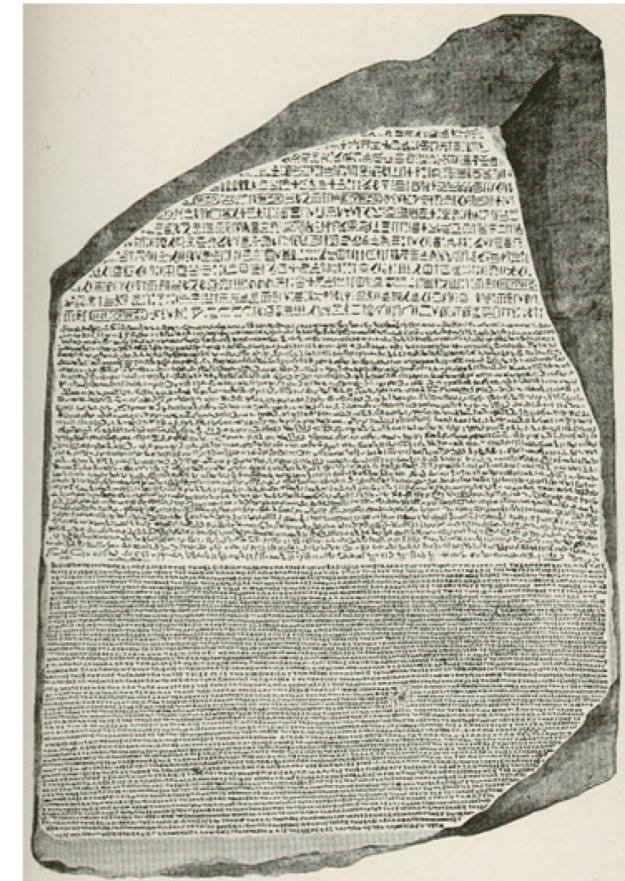
- ① Introduzione
- ② Architetture per la classificazione di immagini
- ③ Architettura di rete convoluzionale proposta
- ④ Datasets di immagini di geroglifici egizi
- ⑤ Transfer learning e training from scratch
- ⑥ Risultati sperimentali
- ⑦ Conclusioni



Introduzione: contesto ed obiettivo

Antichi Geroglifici Egizi:

- Tipo di scrittura antica risalente al 3000 a.c
- Indecifrati fino al 1799 d.c, anno della scoperta della stele di Rosetta
- La decifrazione ha consentito la comprensione della storia dell'antica civiltà egizia



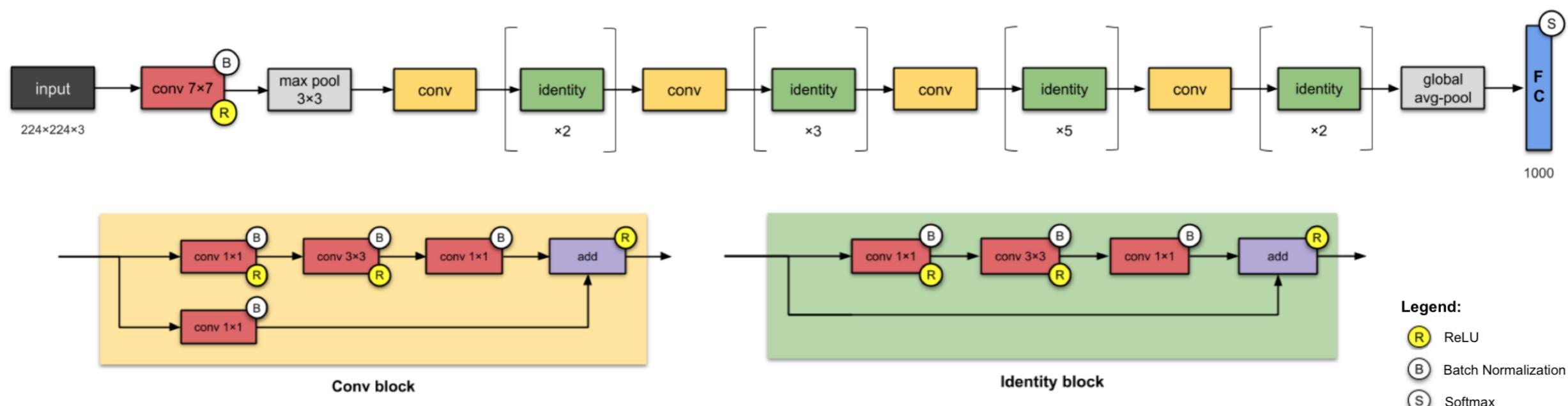
Stele di Rosetta

Obiettivo della tesi:

- Utilizzo di nuovi approcci basati su deep learning in un contesto storico di riconoscimento di linguaggi di testo antichi

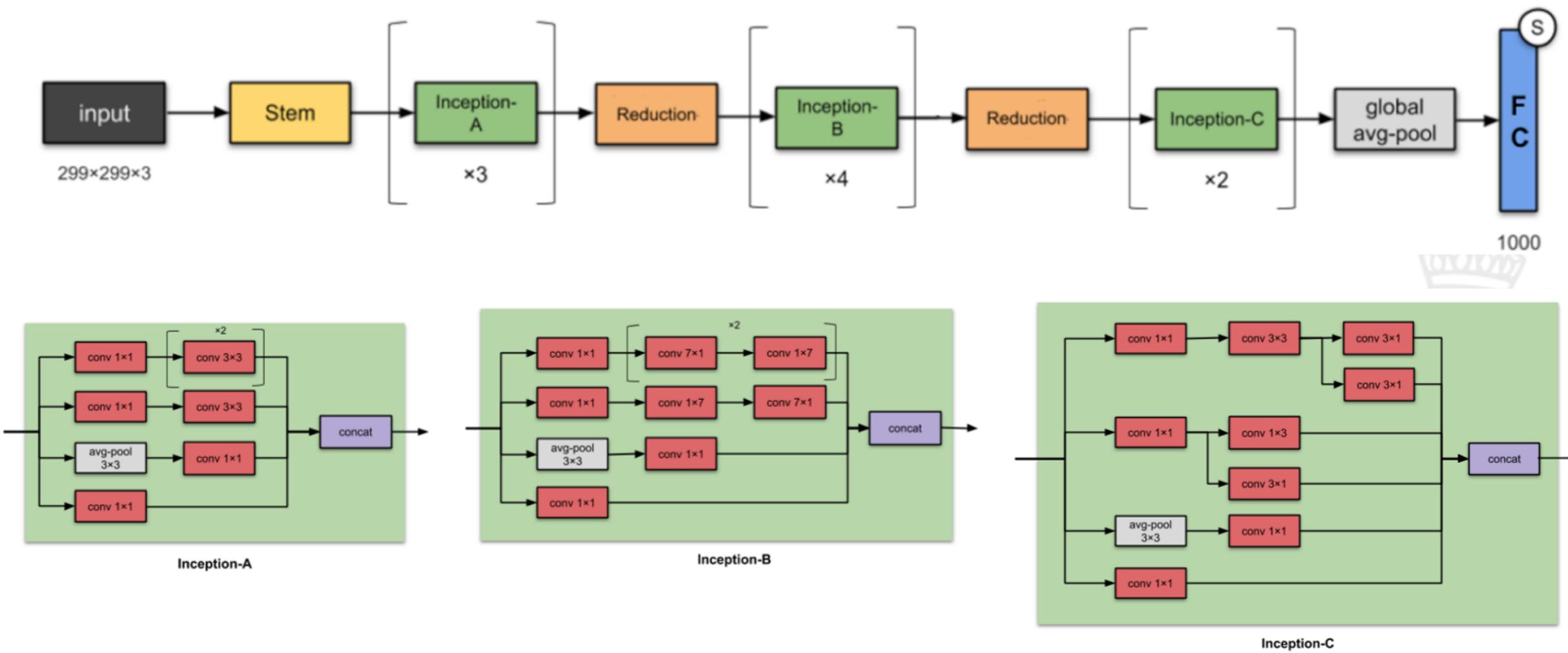
Architettura ResNet50

- Ingresso immagini 224x224
- Caratterizzata da due tipologie di blocchi:
 - Blocco convoluzionale
 - Identity block
- Profondità di 50 layer
- 26 Milioni di parametri addestrabili



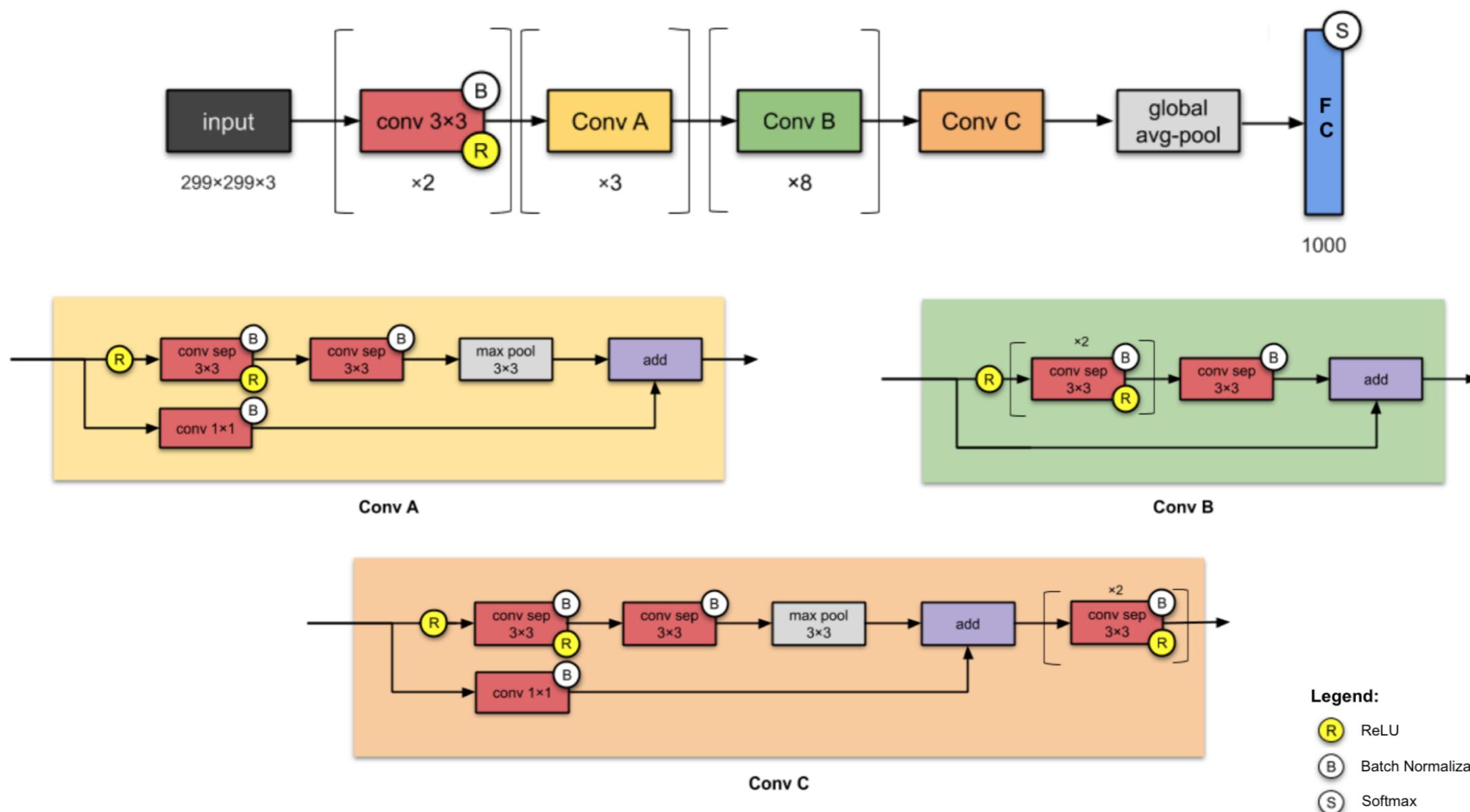
Architettura InceptionV3

- Terza versione del modello GoogleNet
- Ingresso immagini 299x299
- Caratterizzata dai moduli Inception
- Profondità di 48 layer
- 24 Milioni di parametri addestrabili



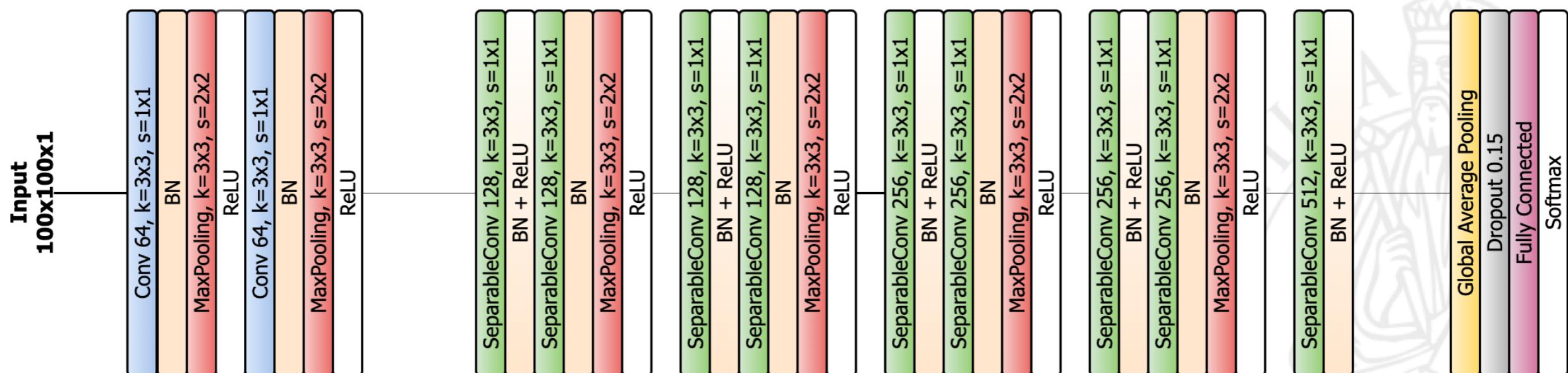
Architettura Xception

- Xception - Extreme Inception
- Ingresso immagini 299x299
- Profondità di 71 layer
- 23 Milioni di parametri addestrabili



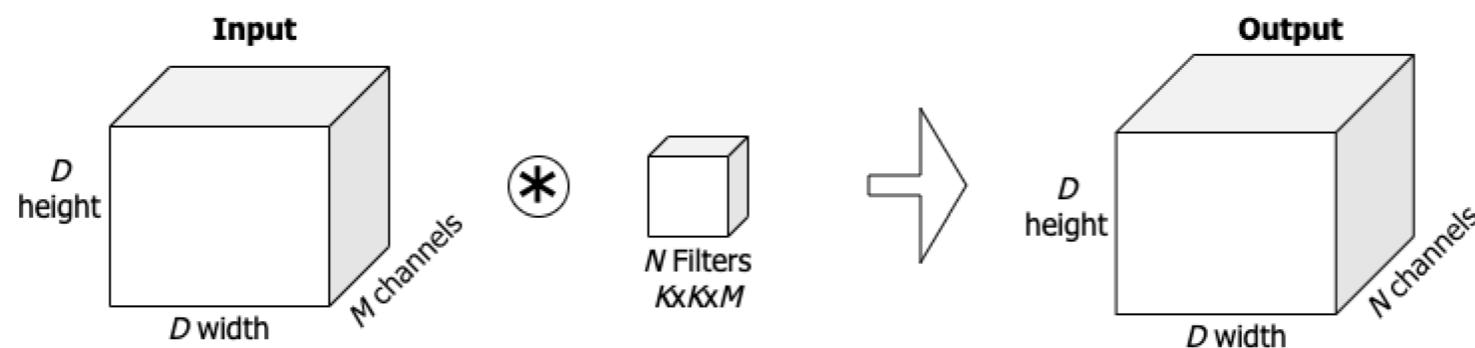
Architettura proposta

- ATCNet: Ancient Texts Classification Network
- Ingresso immagini di dimensione 100x100
- 11 layer convoluzionali (2 Conv + 9 Sep) più 1 Dense layer
 - blocco di ingresso: Conv, MaxPooling
 - blocchi intermedi: SepConv, MaxPooling (eccetto l'ultimo)
 - blocco di uscita: GlobalAveragePooling, Dropout e FC con Softmax
 - Batch Normalization, ReLU
- 500 mila parametri addestrabili

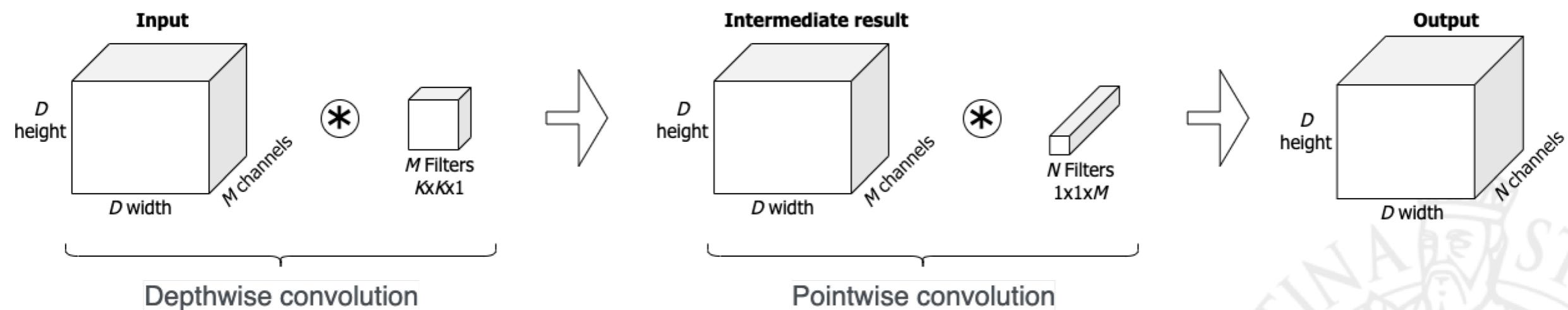


Separable convolution

- Standard Convolution



- Separable Convolution

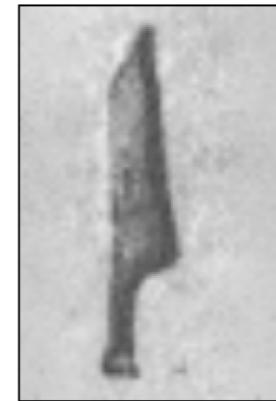


	Standard convolution	Separable convolution	rate (sep/stan)
# Param	$K^2 MN$	$K^2 M + MN$	$\frac{1}{M} + \frac{1}{K^2}$
Cost	$K^2 MND^2$	$K^2 MD^2 + MND^2$	$\frac{1}{M} + \frac{1}{K^2}$

Datasets

- Dataset-1 (D_1)

- immagini 50x75 in scala di grigi
- prese dal libro *Pyramid of Unas* di Alexandre Piankoff e scattate dentro la piramide di Unas



- Dataset-2 (D_2)

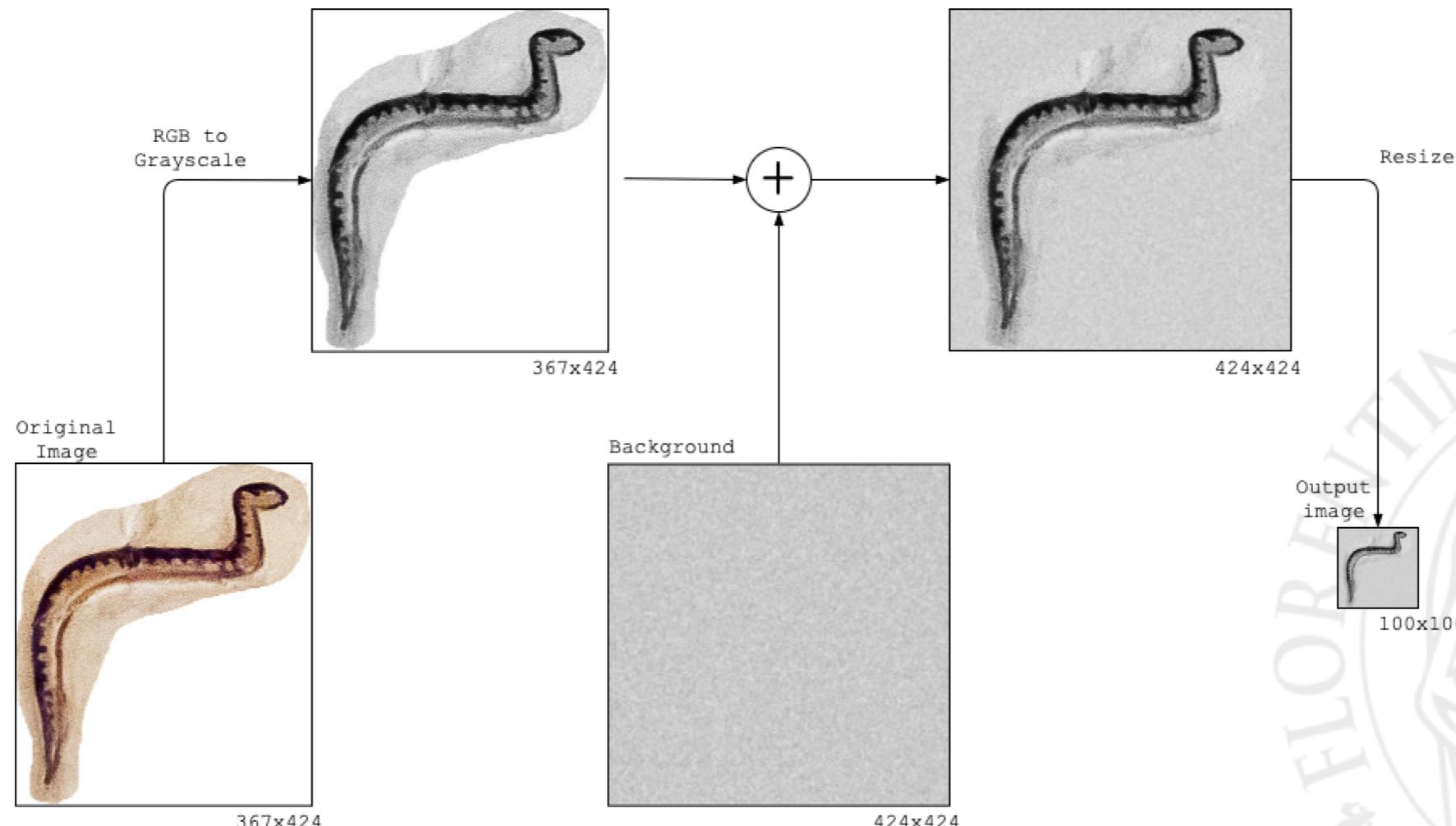
- immagini di varie dimensioni in formato RGB
- incisioni e dipinti



- Codifica mediante Gardiner Sign List

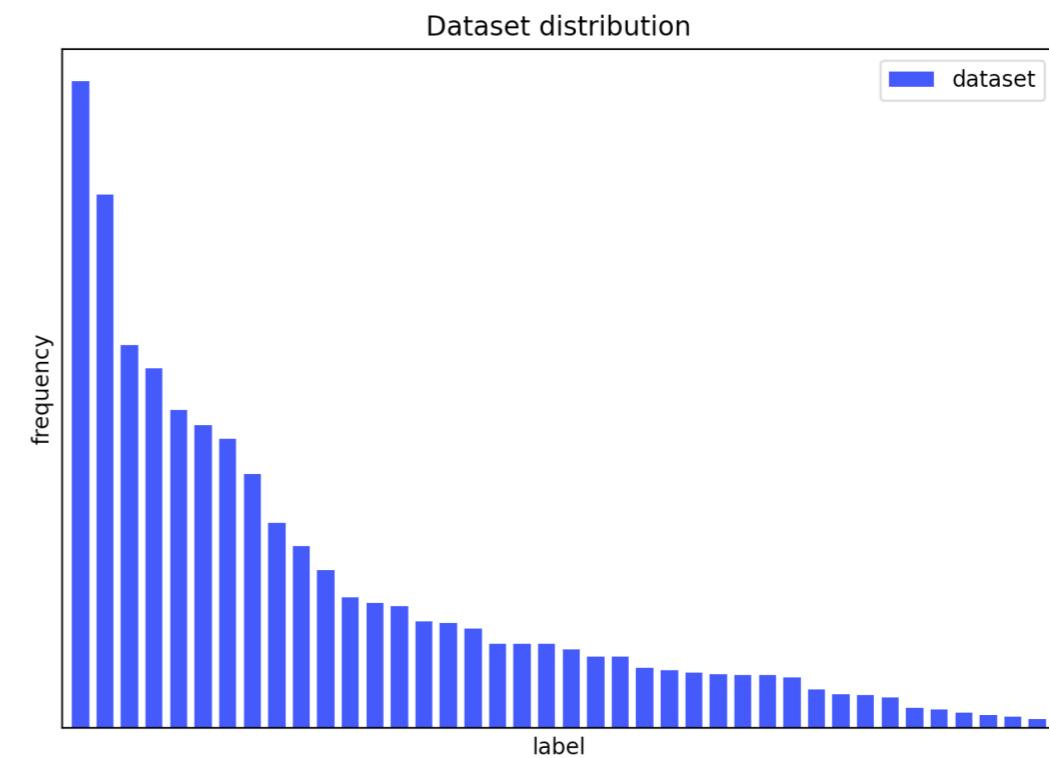
Preprocessing

- Trasformazione Spazio di Colori
 - D_2 da RGB a grayscale
- Riempimento immagine originale
 - generazione immagine di background
 - dimensione quadrata da contenere l'immagine originale
- Resizing
 - senza alterazione delle proporzioni delle immagini originali

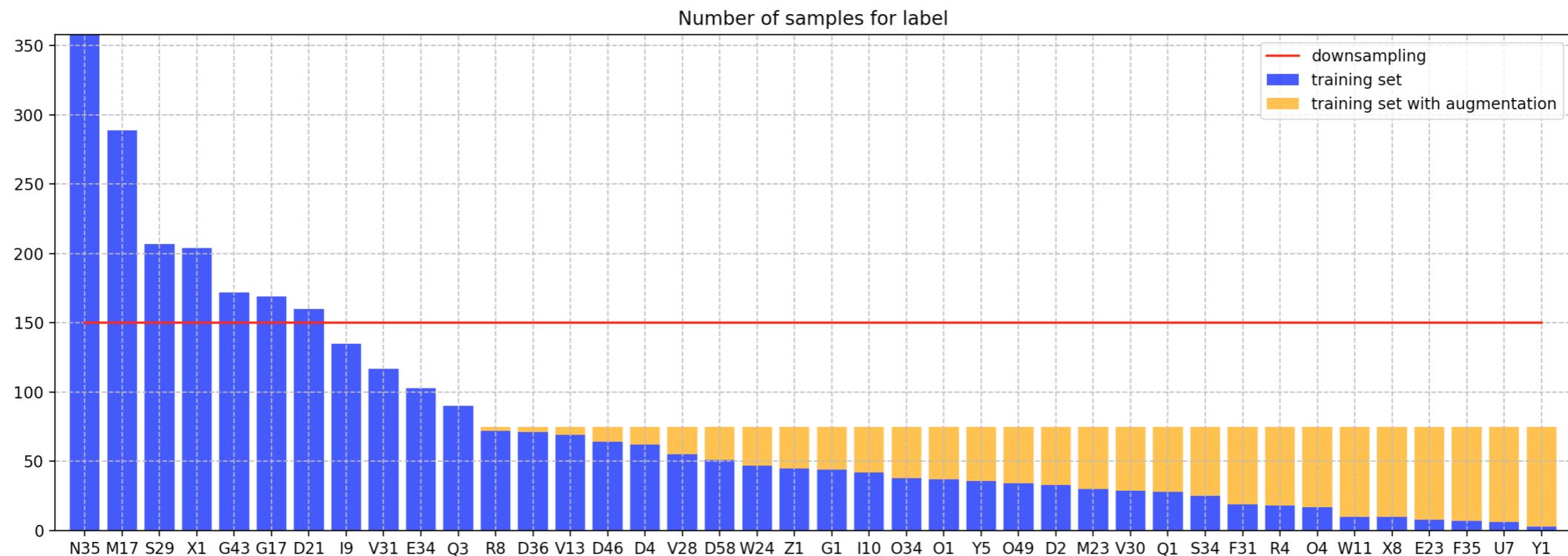


Dataset

- $Dataset = D_1 \cup D_2$
- Dataset sbilanciato
- Suddivisione del dataset in modo stratificato
 - 70% Training set
 - 15% Validation set
 - 15% Test set

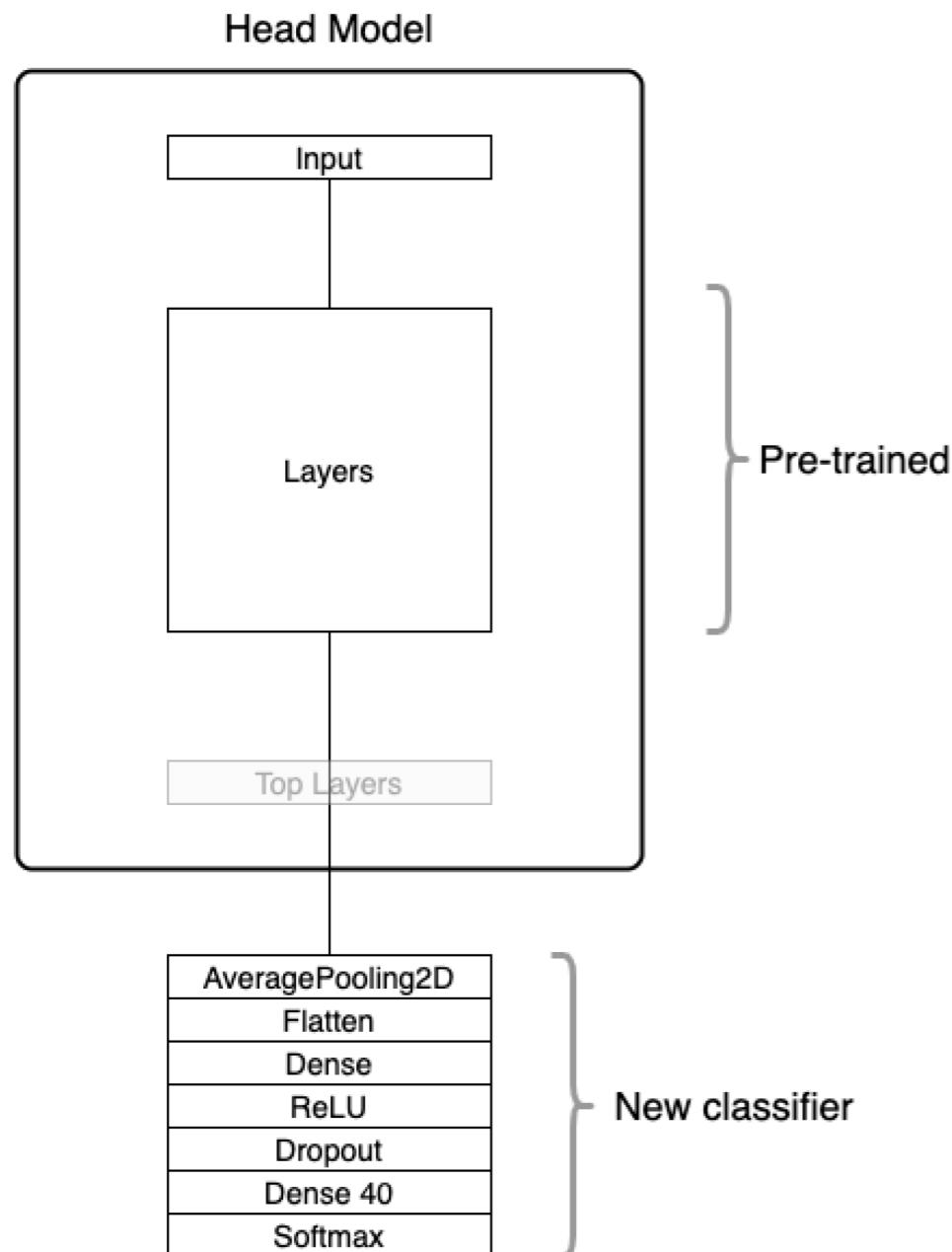


Data Augmentation: training-set



- Ribilanciamento training-set
- Data Augmentation
 - Traslazione, Rotazione, Zoom in/out, Ribaltamento orizzontale
 - $n_{aug}^{(C_i)} = \begin{cases} \frac{\sum_{i=0}^N |C_i|}{N} - |C_i| = \mu - |C_i| & \text{if } |C_i| < \mu \\ 0 & \text{otherwise} \end{cases}$
- Downsampling
 - $\max|C_i| \leq 2 \cdot \mu$

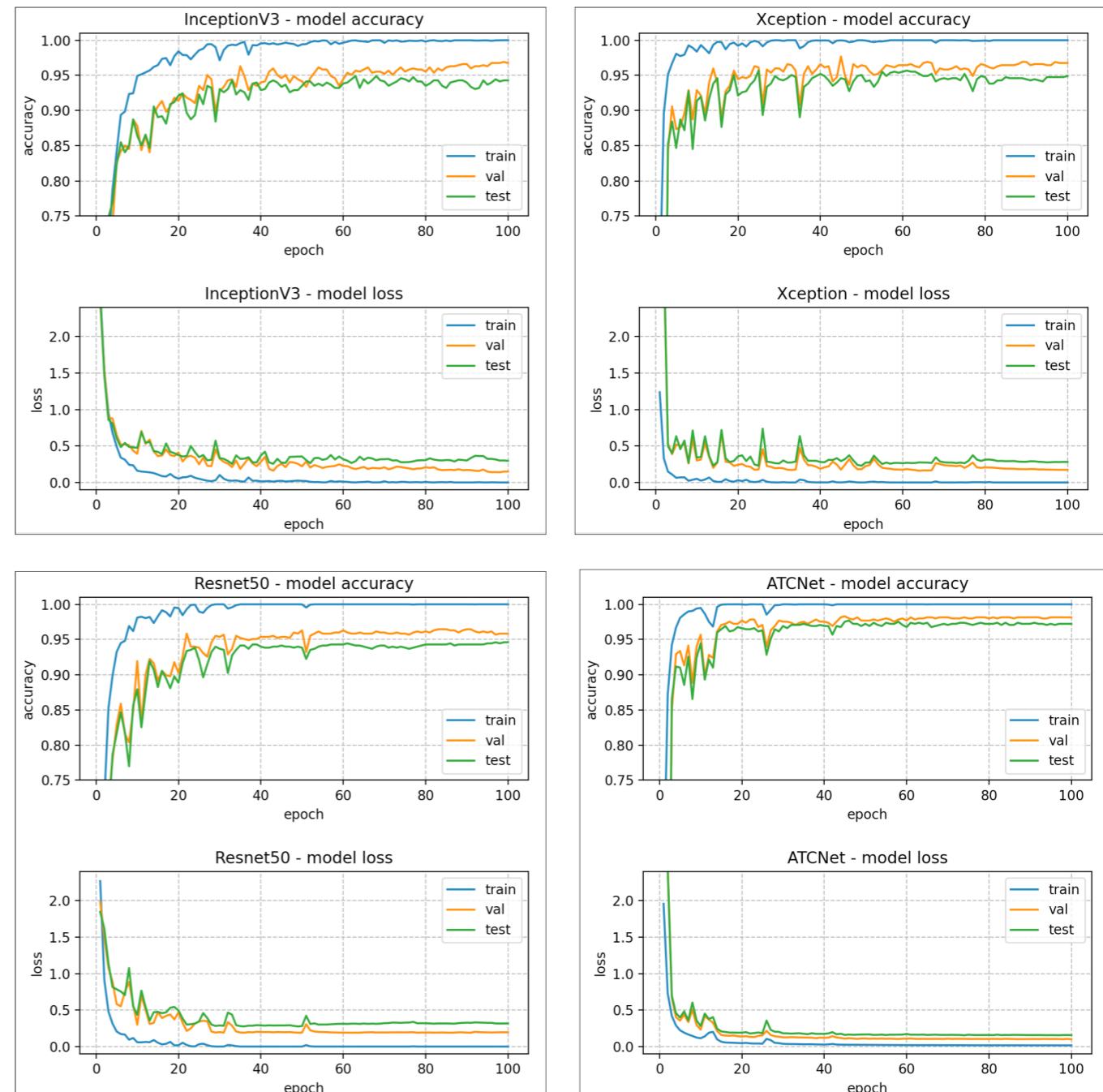
Transfer Learning



- Head Model
 - ResNet50, InceptionV3, Xception
 - Pre-addestrate su ImageNet
 - Rimossi top layers
- Nuovo classificatore
 - AveragePooling2D variabile (7x7, 8x8, 10x10) + Flatten
 - Dense, ReLU, Dropout, Dense con Softmax
- Training
 - Head model layers non addestrabili
 - Classificatore addestrabile
 - Input: 224, 299, 299
 - learning rate iniziale 0.0001
 - decremento esponenziale

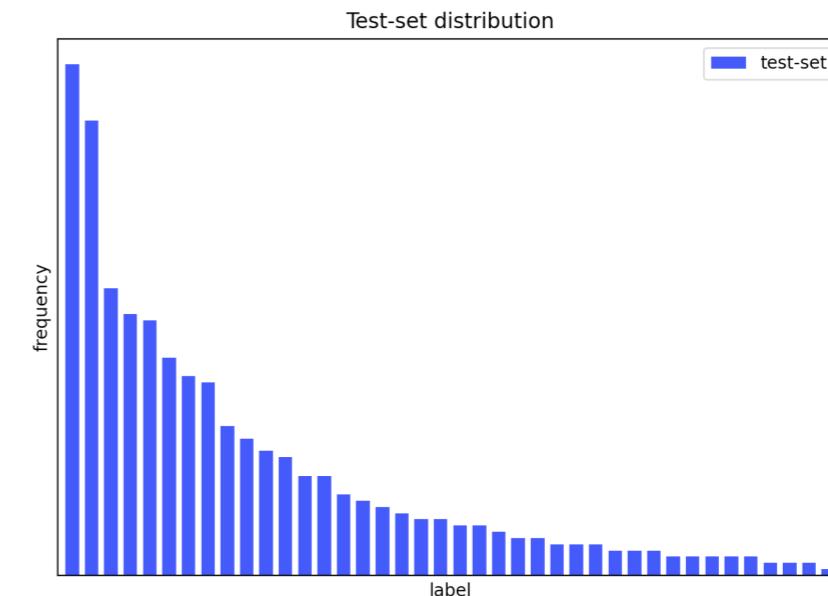
Training

- In ingresso immagine di dimensione:
 - InceptionV3, Xception 299x299
 - Resnet50 224x224
 - ATCNet 100x100
- Cross entropy
- ADAM
 - initial learning rate (*init*) 0.001
 - $lr = init \cdot (1/2)^{(epoch/15)}$
 - batch size 32
- 100 epochhe



Risultati: metriche di valutazione

- Test-set sbilanciato
- Utilizzo di diverse metriche per valutare i modelli



- x_{ij} numero di elementi di classe i classificati come classe j
- $Accuracy = \frac{1}{N} \cdot \sum_i x_{ii}$
- $Recall = \frac{\sum_i R_i}{n}$ dove $R_i = \frac{x_{ii}}{\sum_j x_{ij}}$
- $Precision = \frac{\sum_i P_i}{n}$ dove $P_i = \frac{x_{ii}}{\sum_j x_{ij}}$
- $F1score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$

		Predicted Number			
		Class 1	Class 2	...	Class n
Actual Number	Class 1	x_{11}	x_{12}	...	x_{1n}
	Class 2	x_{21}	x_{22}	...	x_{2n}

	Class n	x_{n1}	x_{n2}	...	x_{nn}

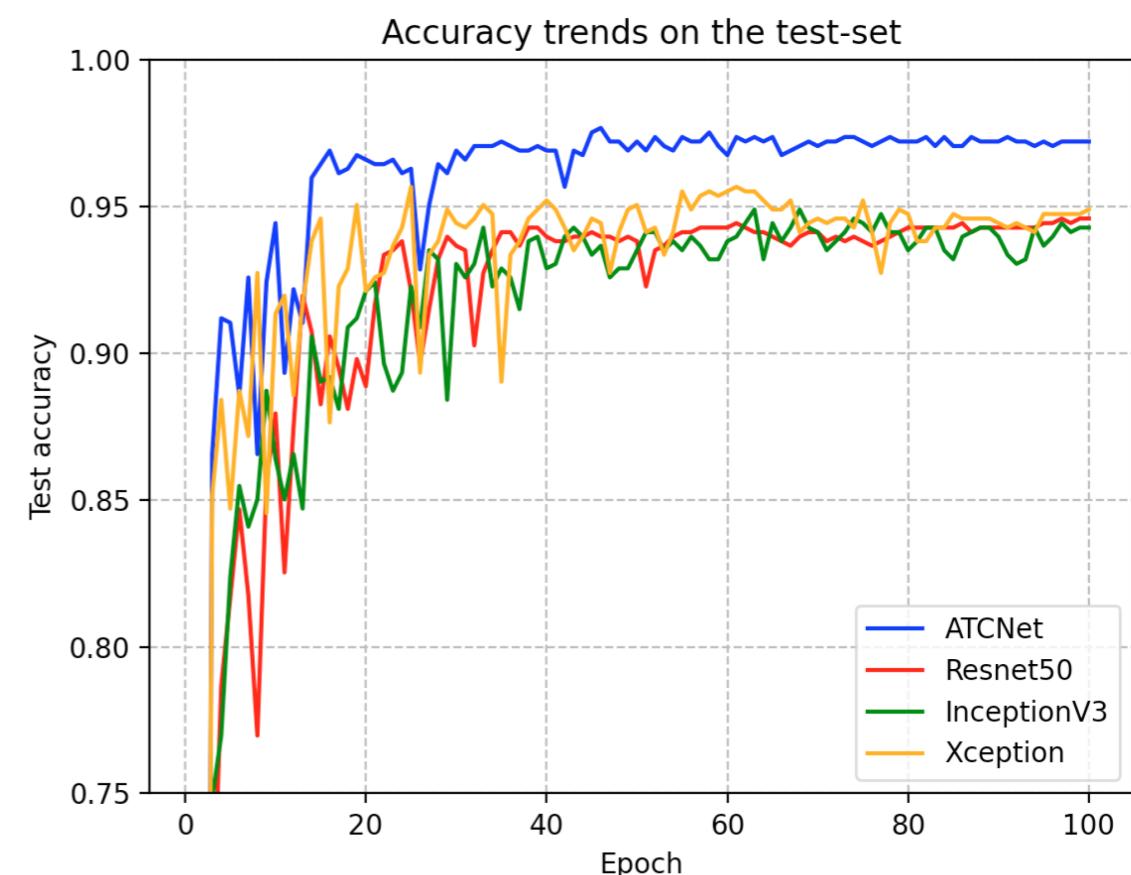
Confusion Matrix

Risultati: comparazioni delle prestazioni

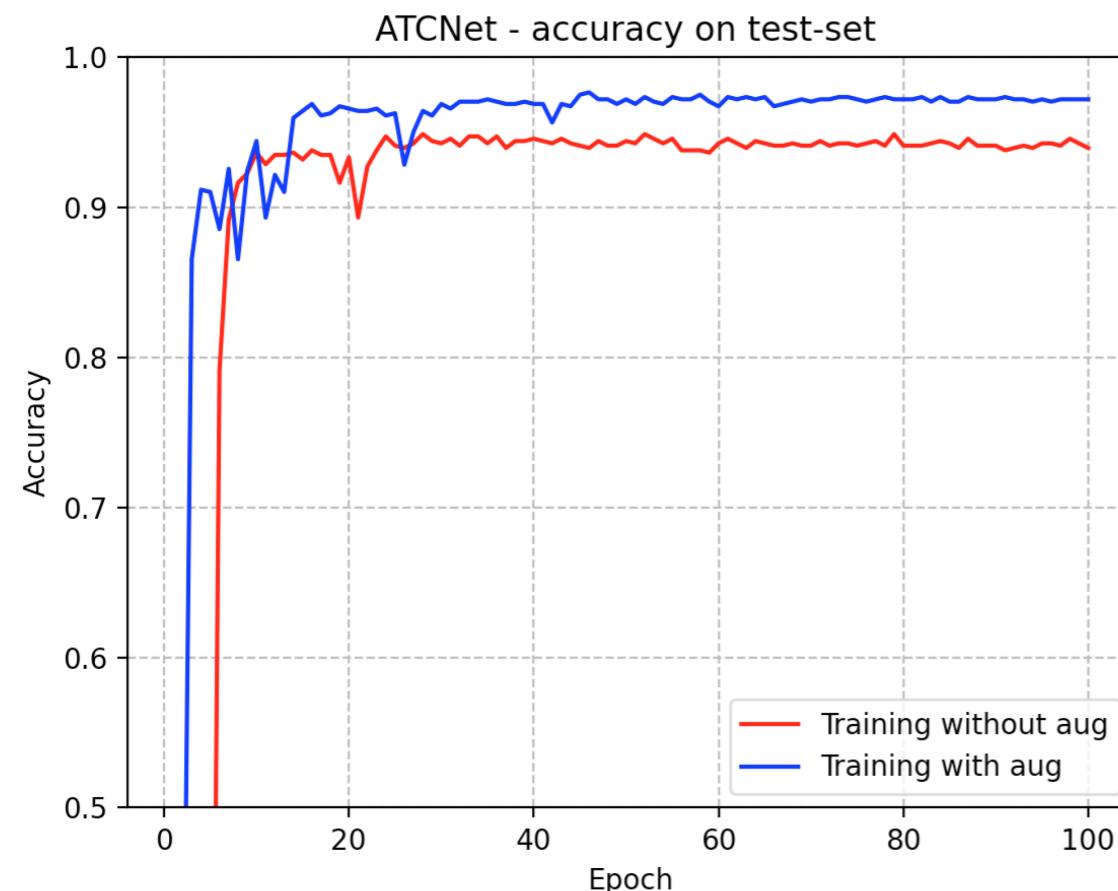
Transfer learning evaluation				
Architecture	Metrics			
	Accuracy	Precision	Recall	F1-Score
Resnet50	0.906	0.882	0.825	0.84
Xception	0.834	0.715	0.72	0.703
InceptionV3	0.864	0.73	0.737	0.717

Training from scratch evaluation				
Architecture	Metrics			
	Accuracy	Precision	Recall	F1-Score
ATCNet	0.976	0.975	0.965	0.968
Resnet50	0.945	0.919	0.905	0.903
Xception	0.956	0.919	0.93	0.919
InceptionV3	0.948	0.917	0.904	0.9

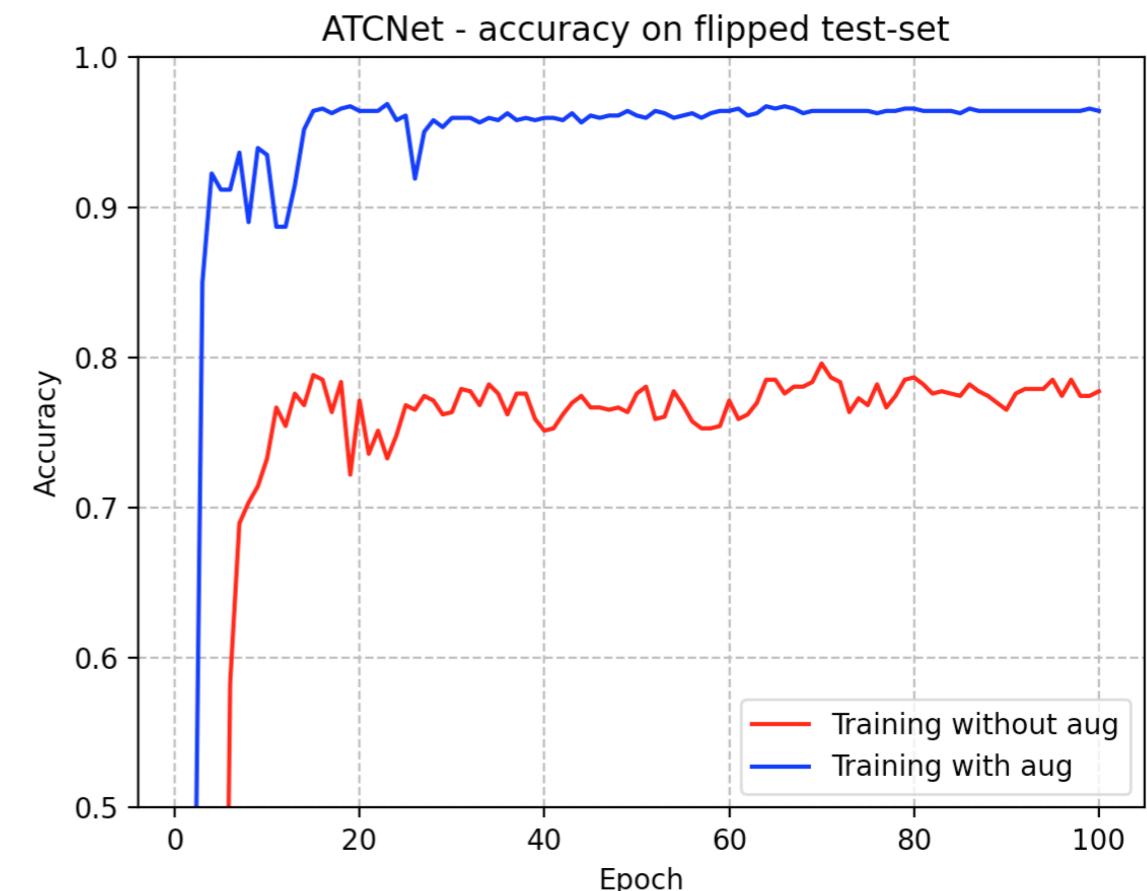
Modelli addestrati from scratch



Risultati: effetti del Data Augmentation



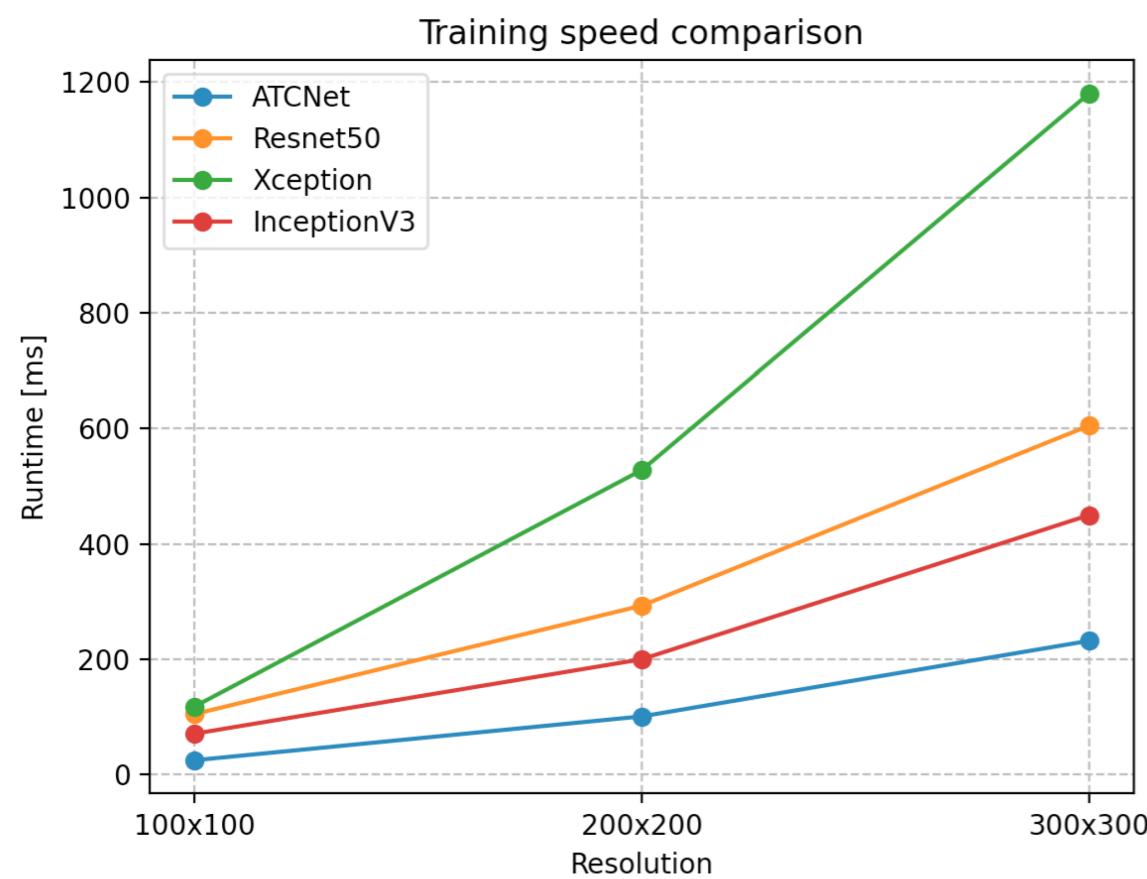
Test-set senza alterazione



Test-set interamente ribaltato
orizzontalmente

Risultati: comparazione costi computazionali

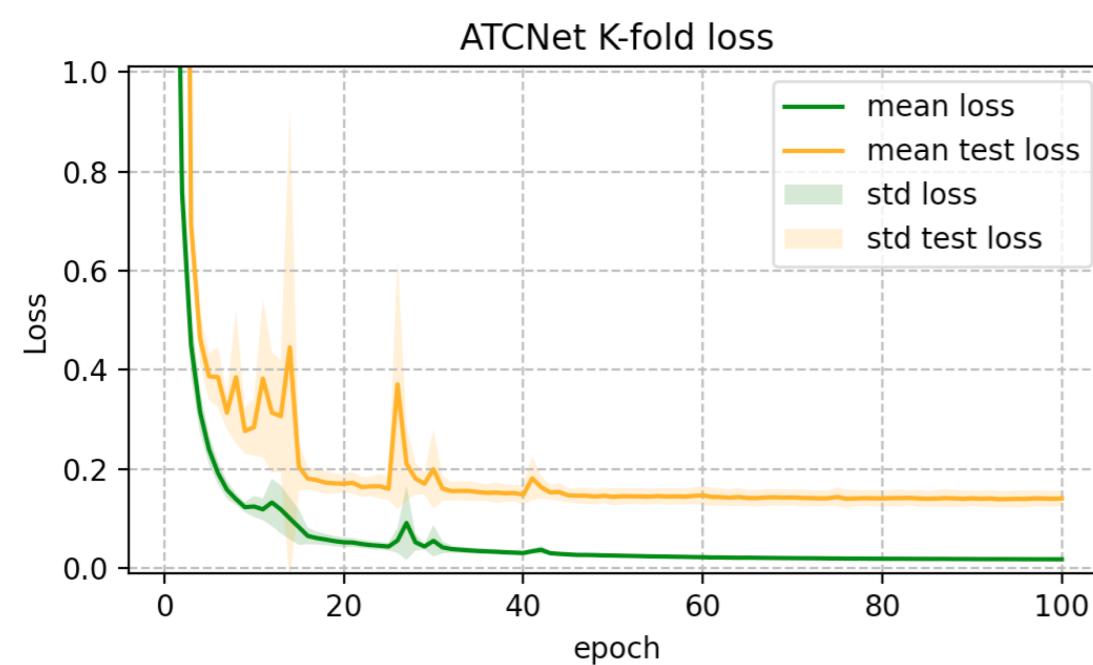
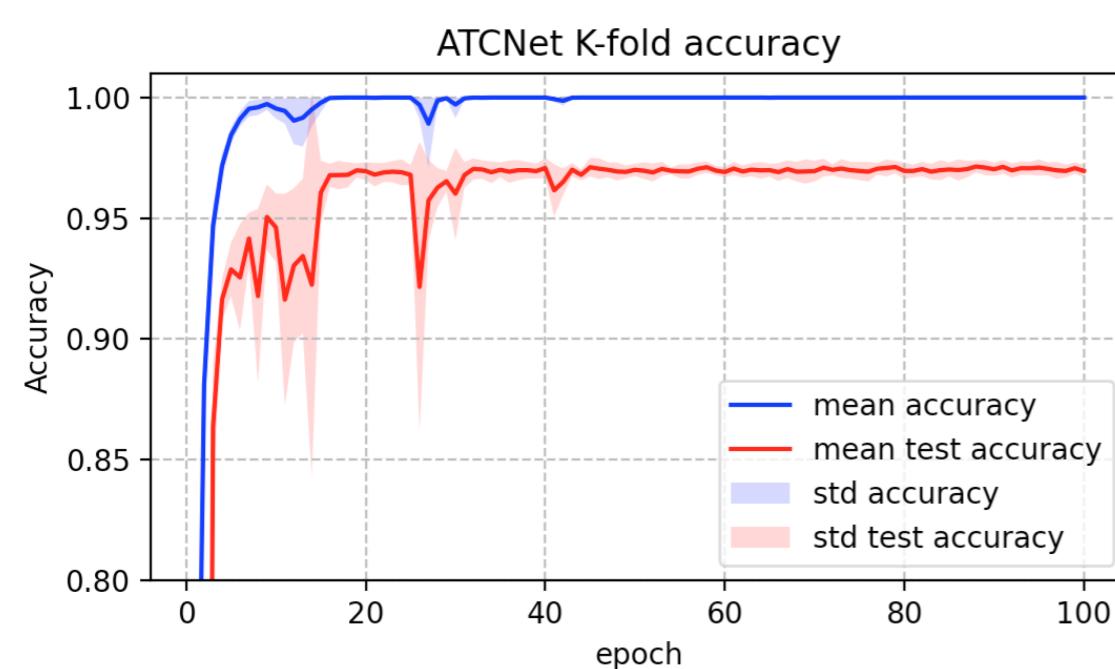
- GPU - Nvidia Tesla T4
- Training: ms per training steps
- Predizioni: ms per singola predizione



Prediction runtime [ms]

Architecture	Resolution		
	100x100	200x200	300x300
ATCNet	3	3	4
Resnet50	8	10	13
Xception	7	10	11
InceptionV3	12	13	14

Risultati: Stratified K-fold Cross-Validation



- Suddivisione dataset in $K = 5$ parti
 - ogni parte mantiene la stessa distribuzione del dataset completo
- Per ogni parte i :
 - Training su $K - 1$ parti esclusa i -esima
 - Data Augmentation
 - Testing sulla i -esima parte

Conclusioni

- Riconoscimento di simboli in testi antichi mediante tecniche di deep learning
- In generale le reti danno buoni risultati nel riconoscimento dei geroglifici
- Modello proposto:
 - Più efficace
 - Minor costo computazionale
- Sviluppi futuri:
 - Incremento del numero di classi nel dataset
 - Integrazione con SW per l'estrazione di geroglifici da scatto fotografico



Dipartimento di Ingegneria dell'Informazione
Laurea Magistrale in Ingegneria Informatica

Reti Neurali Convoluzionali per la Classificazione di Geroglifici Egizi

Candidato
Marco Loschiavo

Relatori
Prof. Fabrizio Argenti
Prof. Alessandro Piva
Prof. Andrea Barucci
Prof. Massimiliano Franci