

# UTILIZZO DI UN AUTOENCODER UNET PER IL DENOISING DI IMMAGINI

Tesi di Laurea in Modelli e Metodi per la Sicurezza nelle Applicazioni

Relatore:

Prof. Donato Impedovo

*Co-relatore:*

*Dott. Davide Veneto*

Laureando:

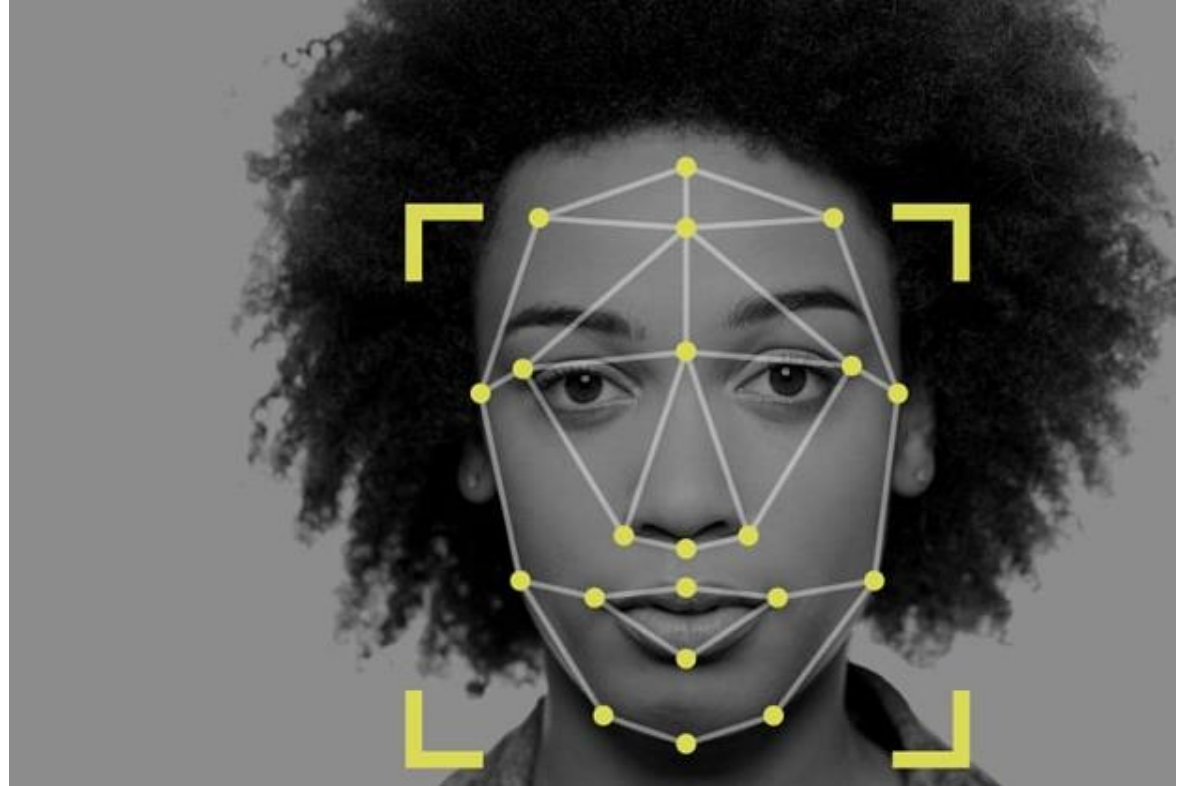
Matteo Esposito

A.A. 2021/2022



# STUDIO DELLO STATO DELL'ARTE

1. FACE RECOGNITION
2. DATASET FACCIALI
3. RETI NEURALI
4. AUTOENCODER
5. UNET

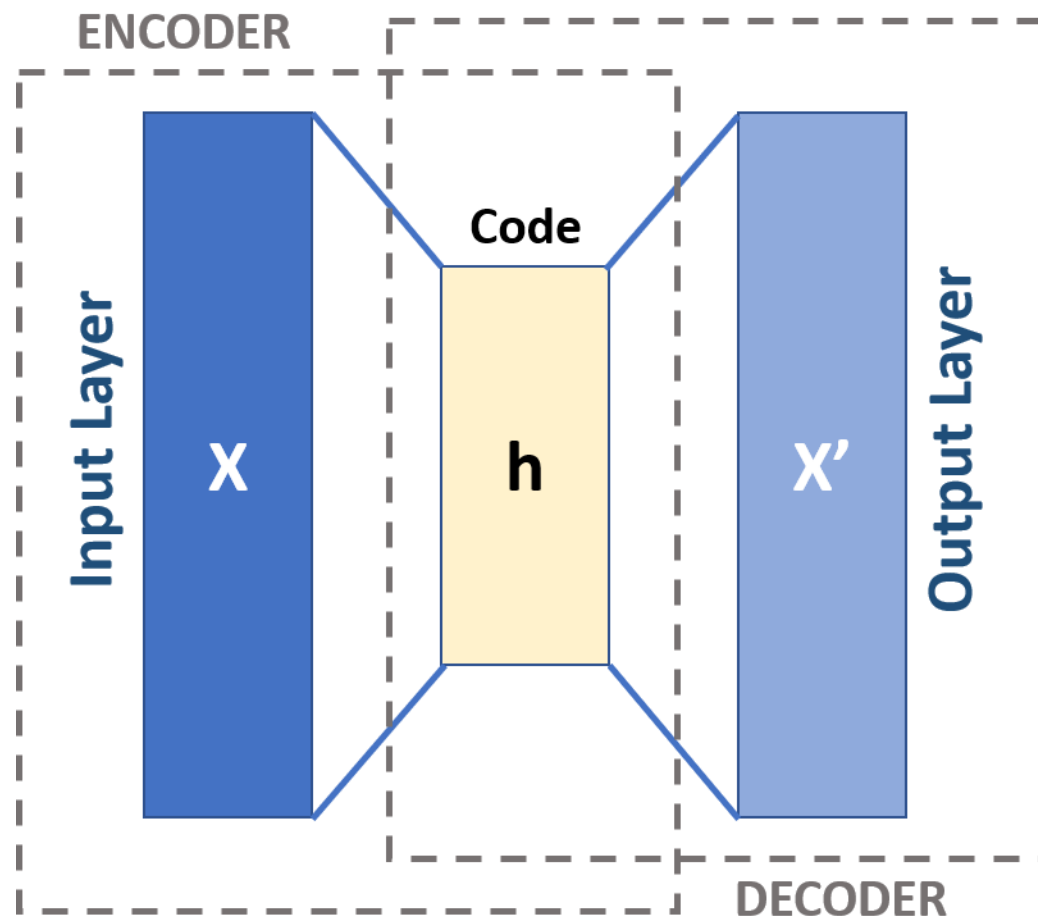


# COSA È UN AUTOENCODER

Gli autoencoder sono reti neurali artificiali addestrate in modo non supervisionato.

Il compito principale di un autoencoder è quello di essere capace di ricostruire i dati giacenti in una varietà.

Un autoencoder è composto da due parti, l'encoder che codifica i dati in input e il decoder che genera l'output in base alla codifica precedentemente generata.

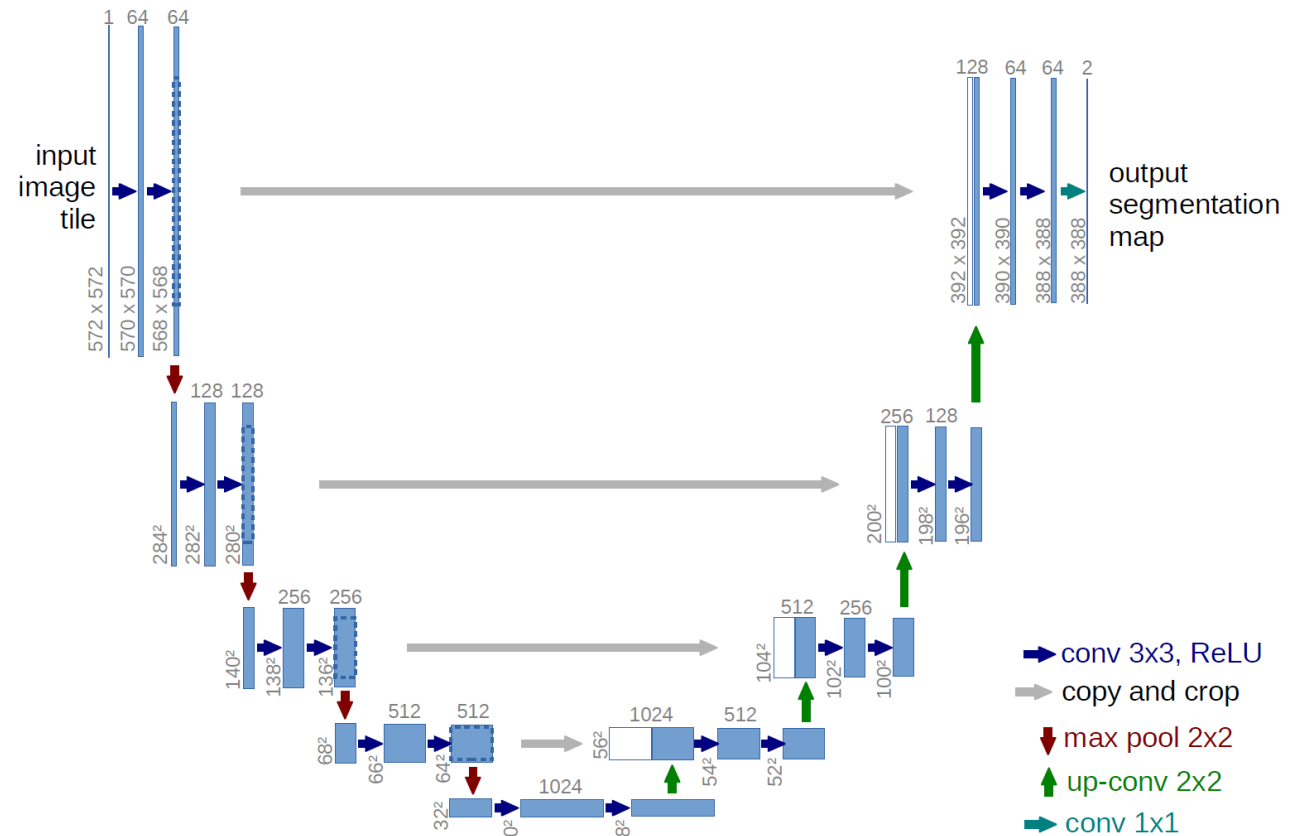


# UNET

La UNet è stata progettata presso il dipartimento di informatica dell'Università di Friburgo nel 2015 per elaborare immagini biomediche.

Nei casi biomedici viene richiesto non solo di distinguere se c'è una malattia, ma anche di localizzare l'area dell'anomalia.

La UNet si dedica a risolvere questo problema andando ad eseguire la classificazione su ogni pixel dell'immagine data in input, a differenza delle classiche reti neurali che effettuano solo la classificazione delle immagini.



# TASK DI UTILIZZO UNET

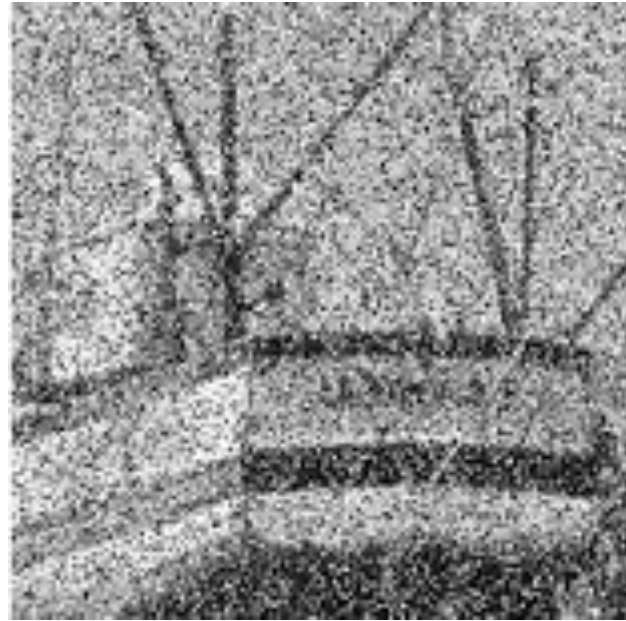
SEGMENTAZIONE DI  
IMMAGINI

GENERAZIONE DI  
IMMAGINI

RICOSTRUZIONI DI  
IMMAGINI  
(DENOISING,  
DEBLURRING)

# IMAGE DENOISING

L'immagine denoising è una tecnica di elaborazione delle immagini che mira a rimuovere il rumore presente in un'immagine digitale, migliorando così la qualità dell'immagine.



# LABELED FACES IN THE WILD

Il dataset LFW contiene circa 13.000 immagini di volti di circa 5.000 individui diversi, catturate in contesti naturali, come ad esempio feste, manifestazioni pubbliche, luoghi di lavoro e istituzioni, è stato creato e mantenuto dai ricercatori dell'Università del Massachusetts.

Le immagini sono state raccolte da diverse fonti, tra cui fotografie di notizie, album fotografici online e video di YouTube.



# CELEBRITIES ATTRIBUTES

Il dataset CelebA é stato introdotto nel 2015 e contiene circa 202.000 immagini di volti di celebrità di età, genere ed etnie diversi. Le immagini sono state catturate in contesti diversi, come eventi pubblici, spettacoli televisivi e set di film.

Le immagini sono state annotate manualmente con diverse etichette, tra cui la posizione degli occhi, del naso e della bocca, la forma del viso, l'età e il genere.





# AGGIUNTA DEL RUMORE

Ogni dataset È stato suddiviso in:

- Training set (70%)
- Validation set (10%)
- Test set (20%)



# SPLIT DEI DATASET

Ogni dataset È stato suddiviso in:

- Training set (70%)
- Validation set (10%)
- Test set (20%)

	LFW	CELEBA
TRAINING SET	18.000 *	140.000
VALIDATION SET	2.000 *	20.000
TEST SET	5.000 *	40.000
TOTALE	26.000 *	200.000

\* VALORI APPROSSIMATIVI

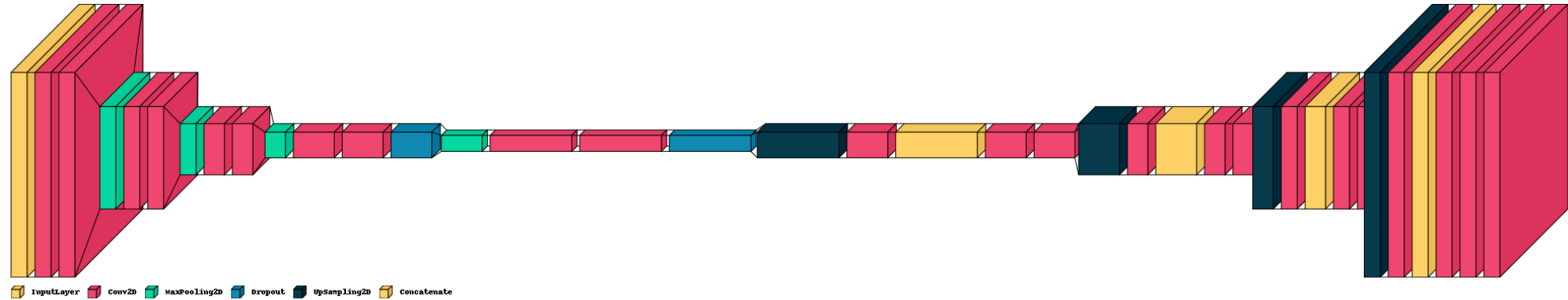
# SPERIMENTAZIONE

—

Gli esperimenti sono stati effettuati utilizzando la libreria Tensorflow su Python.  
Il modello è stato addestrato su entrambi i dataset per 100 epoche con un batch size di 16 elementi e l'ottimizzatore adam.



# SPERIMENTAZIONE: MODELLO



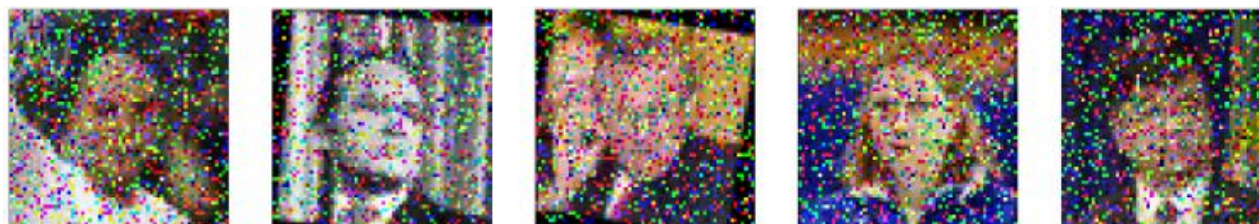
- Il primo livello in giallo è l'*Input Layer*;
- In rosso sono rappresentati i *Conv2D Layer*;
- In verde i *MaxPooling2D Layer*;
- Con il blu chiaro le fasi di *Dropout*;
- In blu scuro gli *UpSampling2D Layer*;
- In giallo, ad eccezione del primo livello, le fasi di *Concatenate*.

# RISULTATI LFW

Immagini originali



Immagini con rumore



Immagini ricostruite



# RISULTATI CELEBA

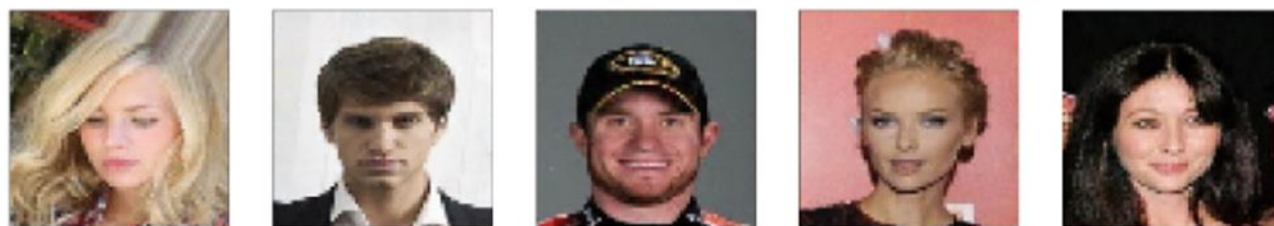
Immagini originali



Immagini con rumore



Immagini ricostruite



# METRICHE

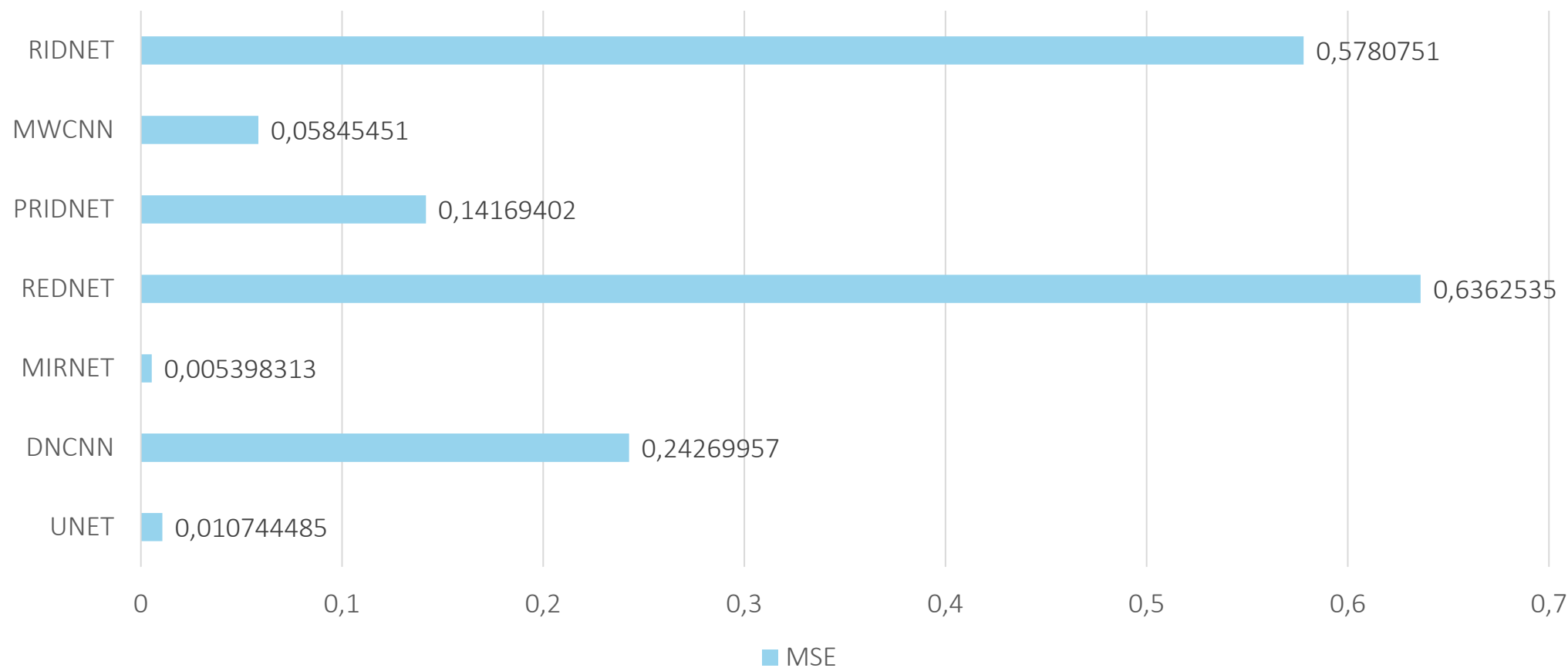
Per quantificare la discrepanza tra l'immagine originale e l'output prodotto dalla rete abbiamo utilizzato le seguenti metriche di errore.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}$$

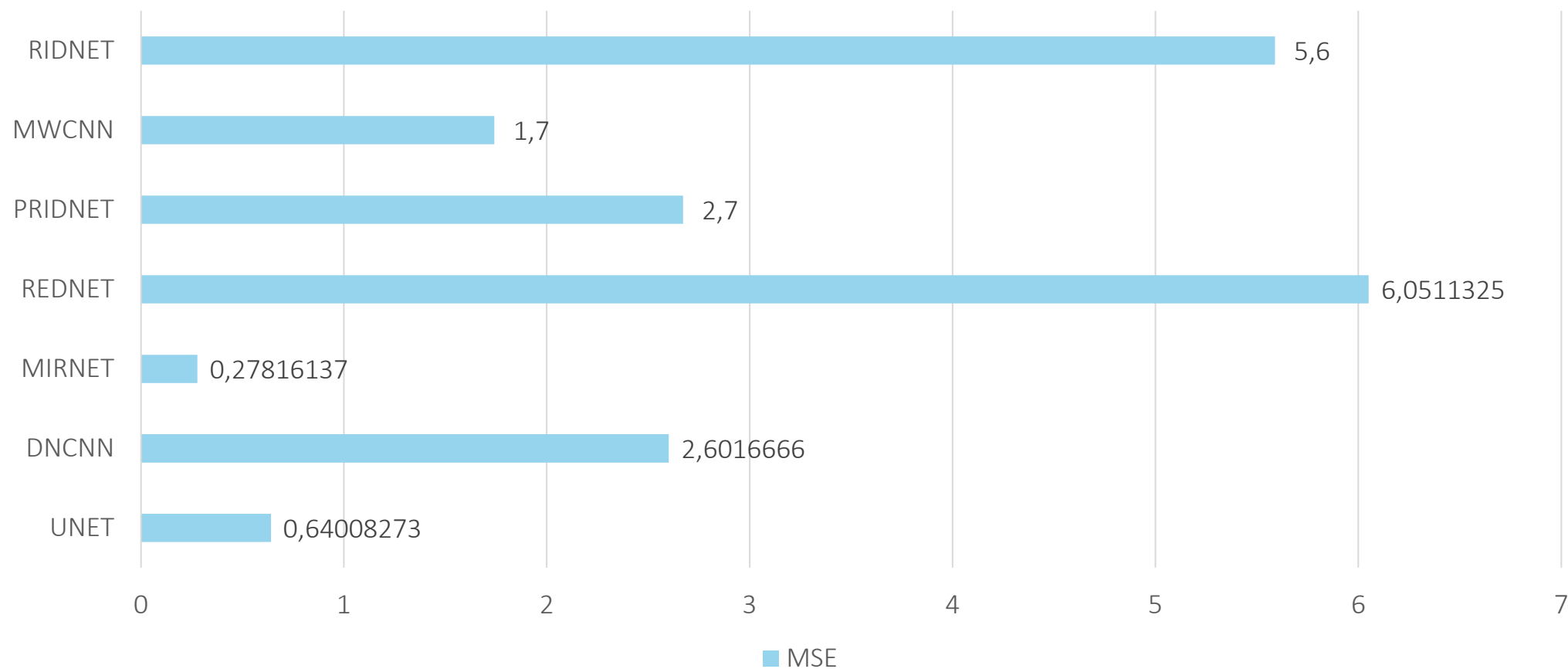
# RISULTATI MAE LFW



\* I VALORI SONO STATI MOLTIPLICATI PER CENTO

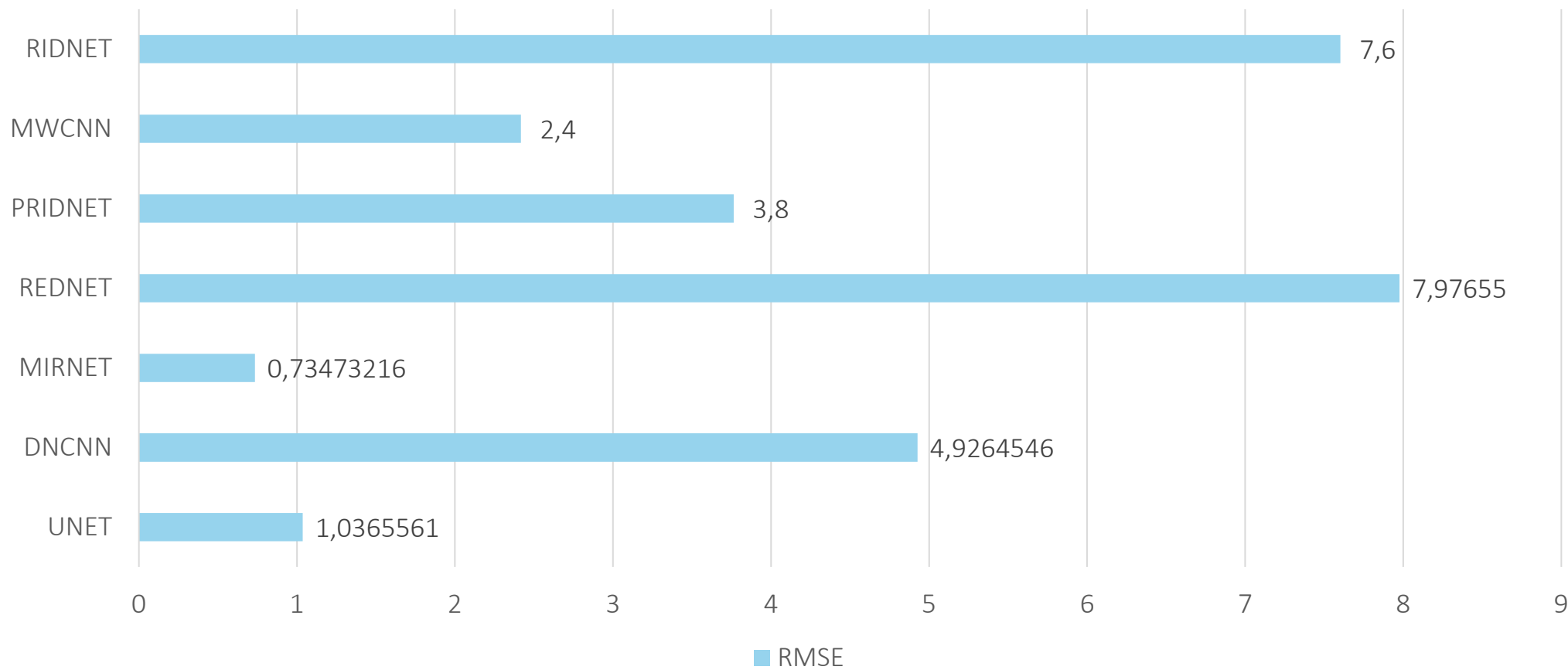


# RISULTATI MSE LFW



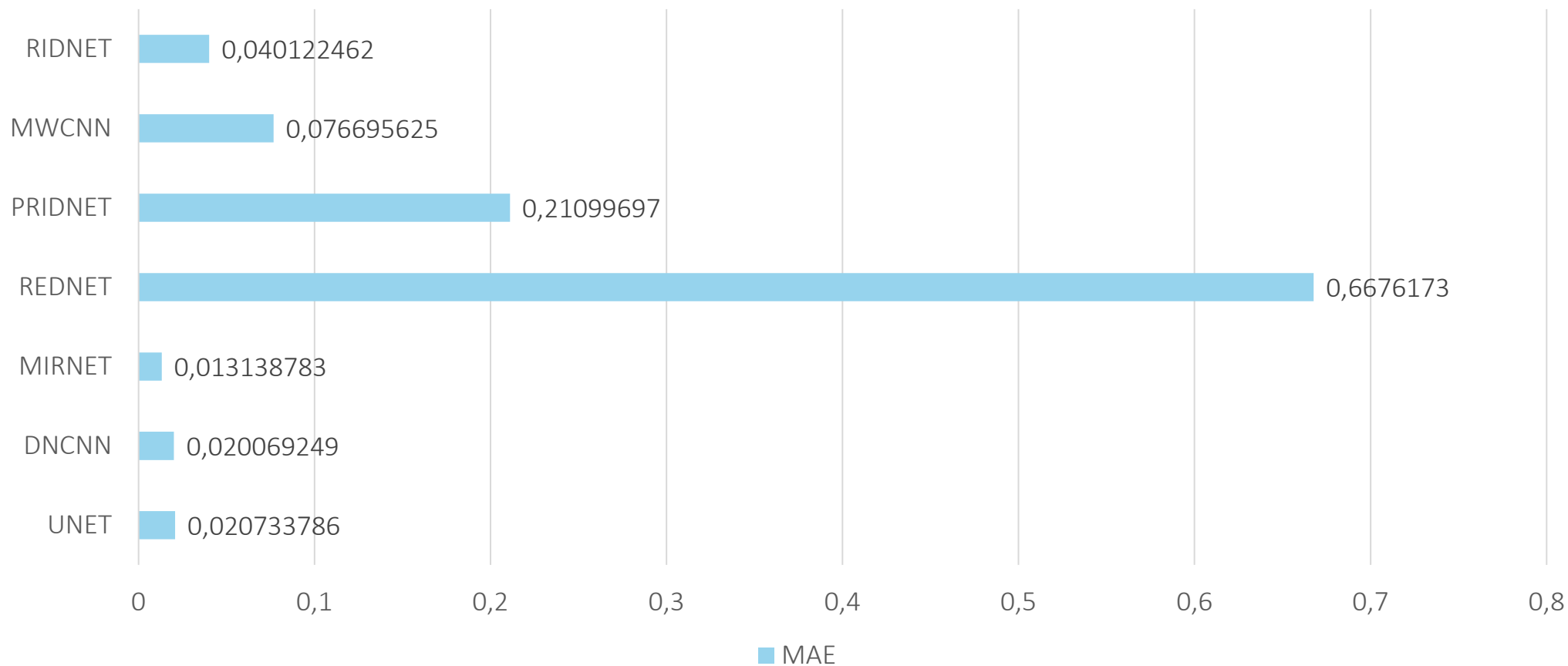
\* I VALORI SONO STATI MOLTIPLICATI PER CENTO

# RISULTATI RMSE LFW



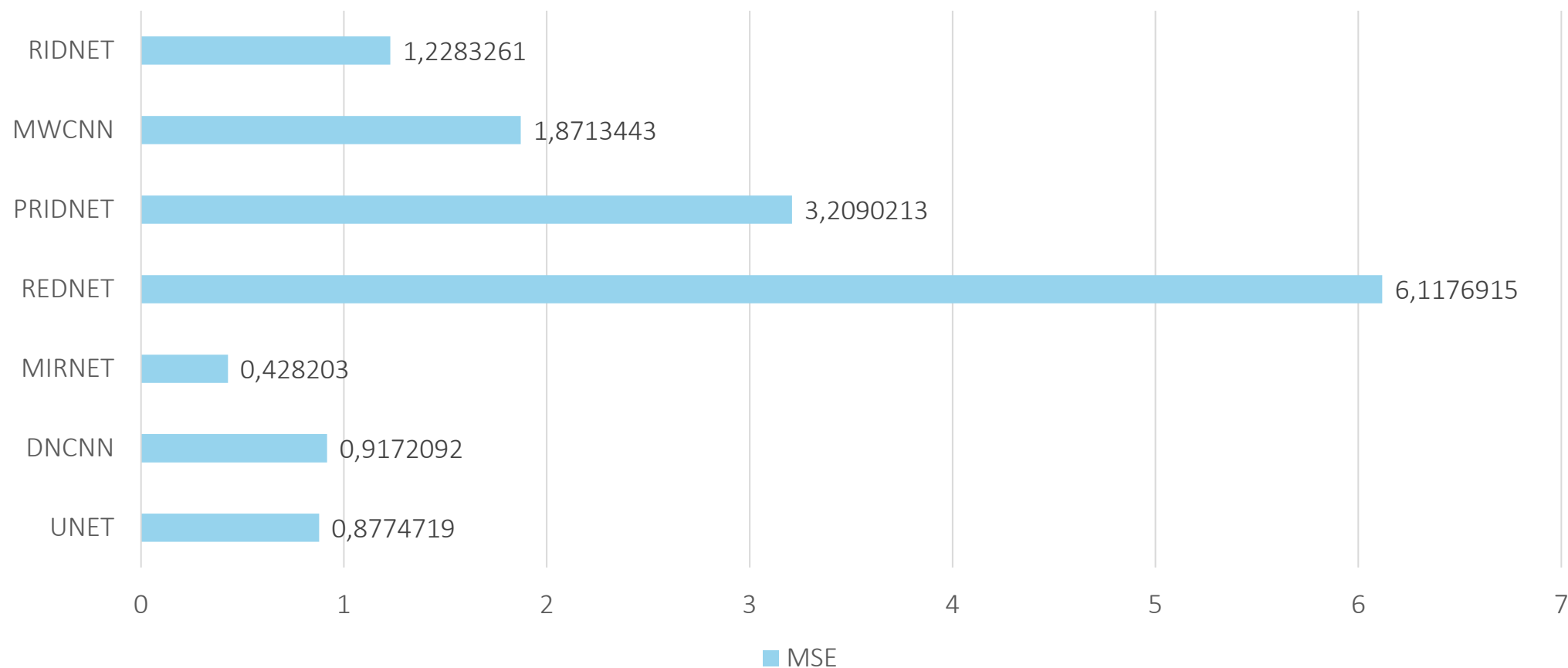
\* I VALORI SONO STATI MOLTIPLICATI PER CENTO

# RISULTATI MAE CELEBA

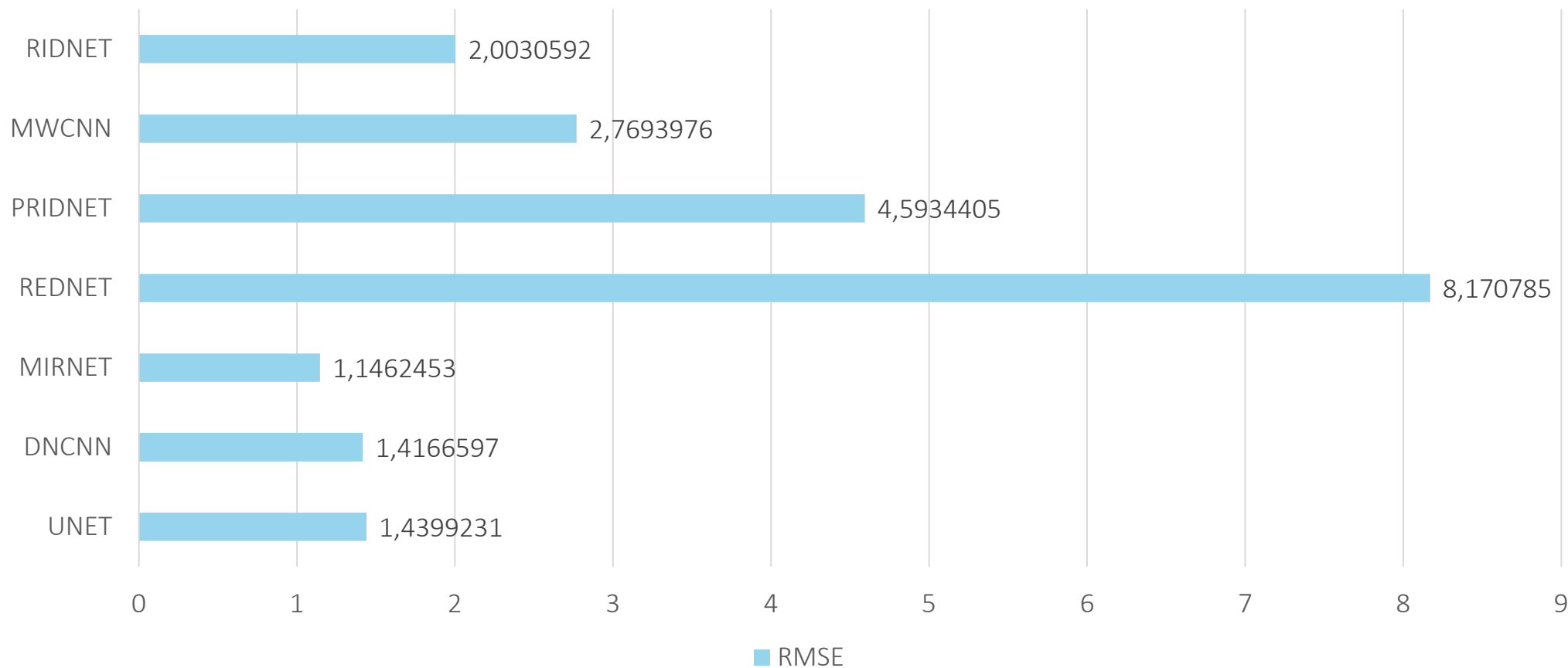


\* I VALORI SONO STATI MOLTIPLICATI PER CENTO

# RISULTATI MSE CELEBA



# RISULTATI RMSE CELEBA



\* I VALORI SONO STATI MOLTIPLICATI PER CENTO

# CONCLUSIONI



I risultati mostrano come l'autoencoder UNet riesca ad avere degli ottimi risultati in confronto alle reti neurali allo stato dell'arte per il denoising di immagini, la UNet può quindi essere una valida alternativa alle reti precedentemente citate per il denoising e la ricostruzione di immagini.

# GRAZIE DELL'ATTENZIONE

MATTEO ESPOSITO

