

FACIAL AGE CLASSIFICATION

Astrid Alins, Miquel Bisbe, Emma Climent





DEFINITION OF THE PROBLEM

Age estimation from facial images → 0-99 years old

Challenging: aging patterns → vary greatly between individuals

Even for humans

How? By transfer learning

DATA AND MODEL USED

UTKFaceDataset:

23,000 + facial images [**age**, gender, ethnicity]

Base CNN model:

Kaggle → by @mohamedchahed

7 classes for **facial emotion recognition**

Big challenge:

7 → 100 (more than tenfold increase in outputs)



STATE OF THE ART



CNNs achieve over **75%** accuracy on **Facial Emotion Recognition (FER)**

Main Age Estimation Approaches:

- **Classification:** Treats each age as a separate class
- Regression: Predicts continuous age values
- Ordinal Regression: Models age order explicitly (CORAL method)

METHODOLOGY: DATA ANALYSIS

Each image (200,200,3) of utkdataset has:

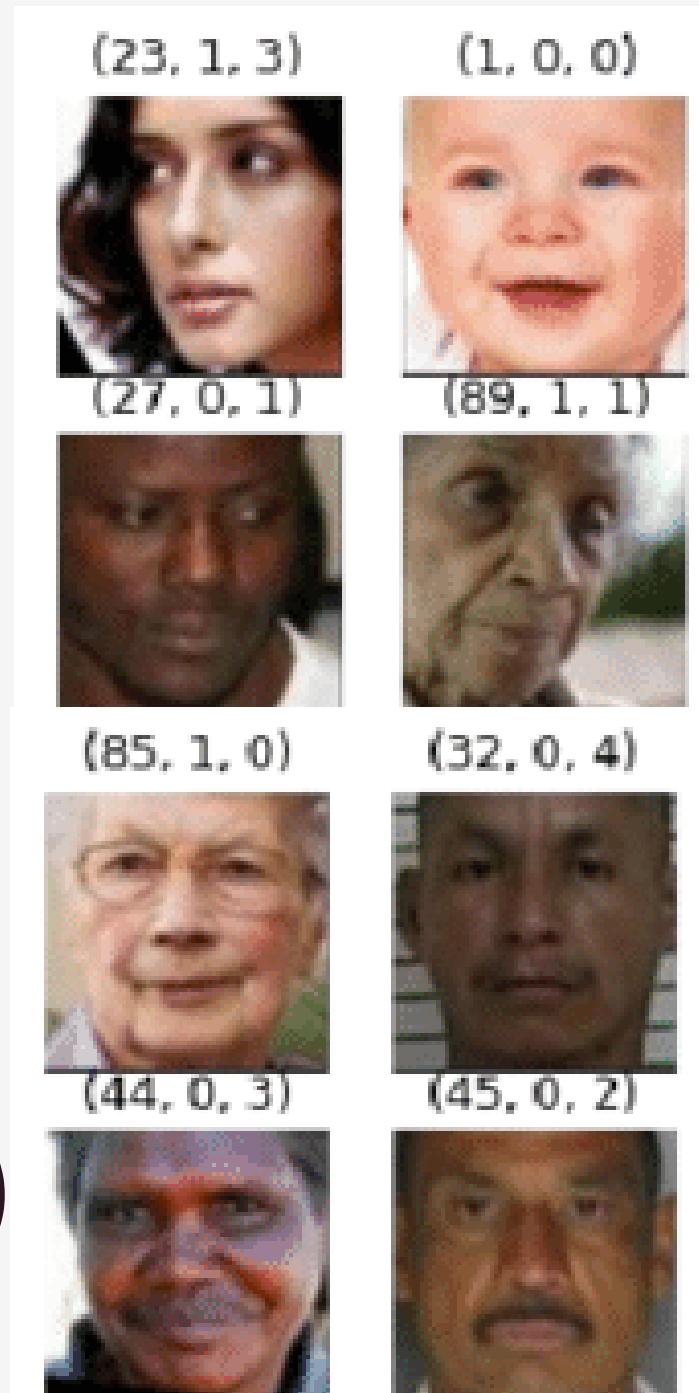
- 3 color channels (RGB)
- 200×200 pixels.

No null values and no duplicates.

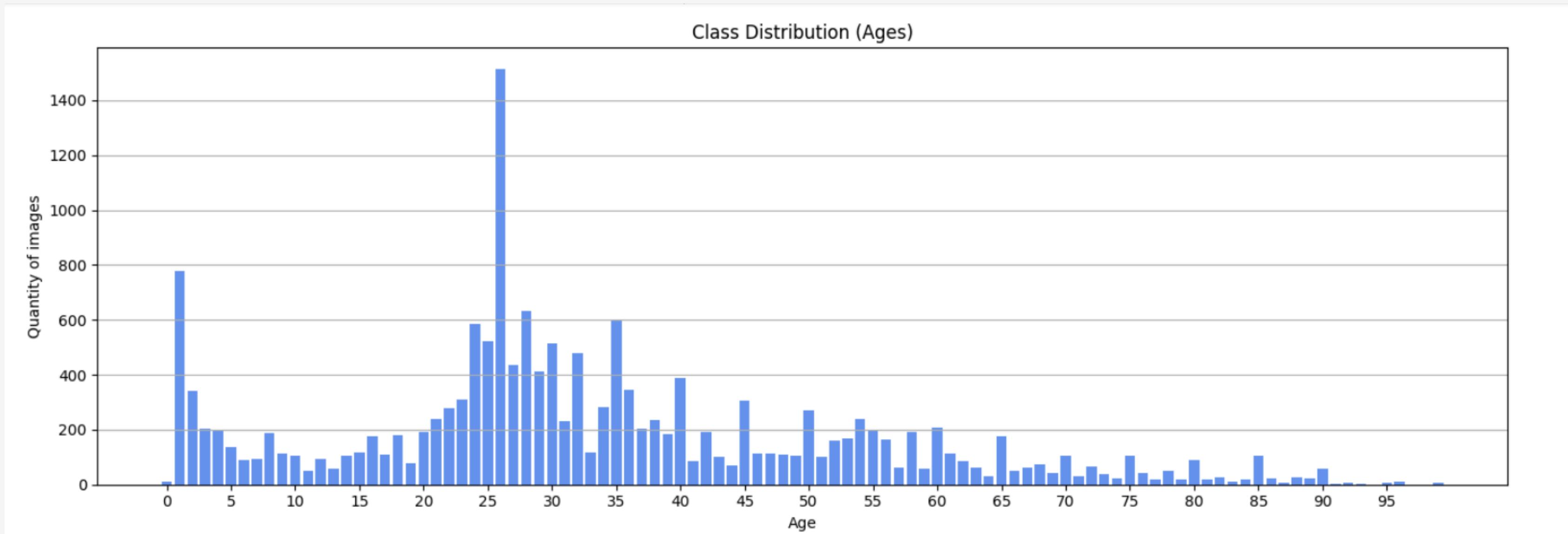
Transformations applied:

- Resized to 48×48 pixels.
- Converted to grayscale.
- Applied one-hot encoding to age labels (to match FER2013 format)

Data split: 70% for training, 15% for validation, 15% for testing

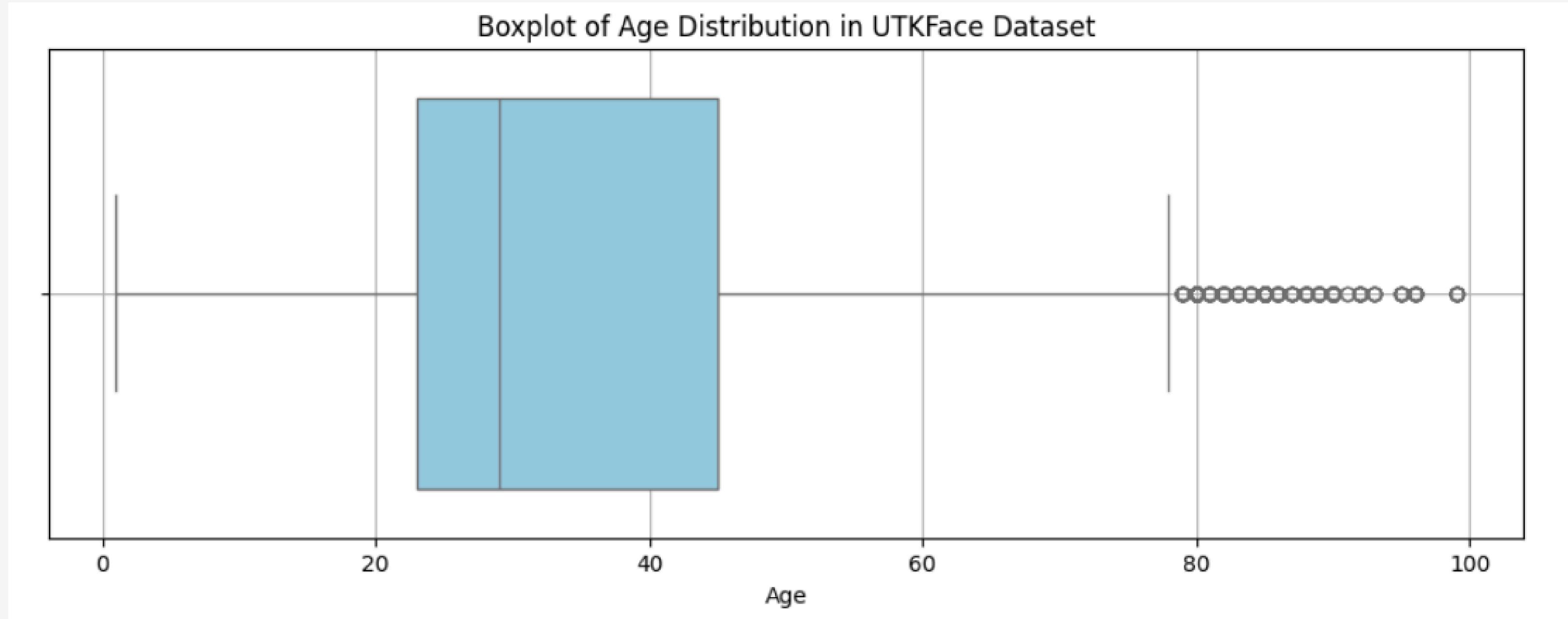


PLOT DISTRIBUTION



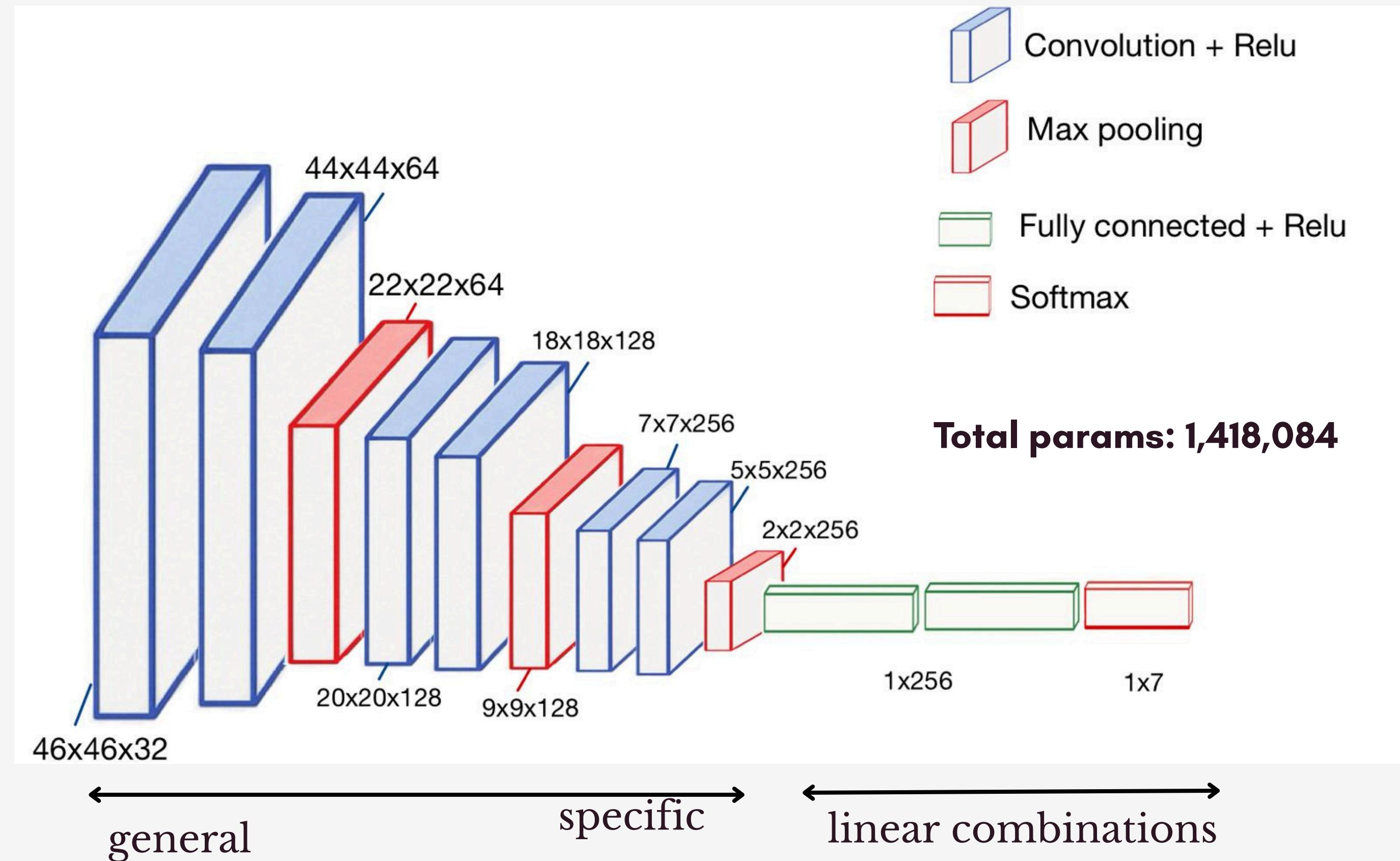
General age distribution, generated in project notebook

BOX PLOT



Boxplot about age distribution, generated in project notebook

METHODOLOGY: ORIGINAL CNN



CROSS ENTROPY

$$\text{LOSS} = - \sum_{i=1}^C y_i \cdot \log(\hat{y}_i)$$

Standard and theoretically choice

- true one-hot encoded label
- SOFTMAX

TRAINING SCHEDULE

Two-phase

- Initial short cycles lr = 0.001
- Longer cycles lr = 1e-5

EXPERIMENTS

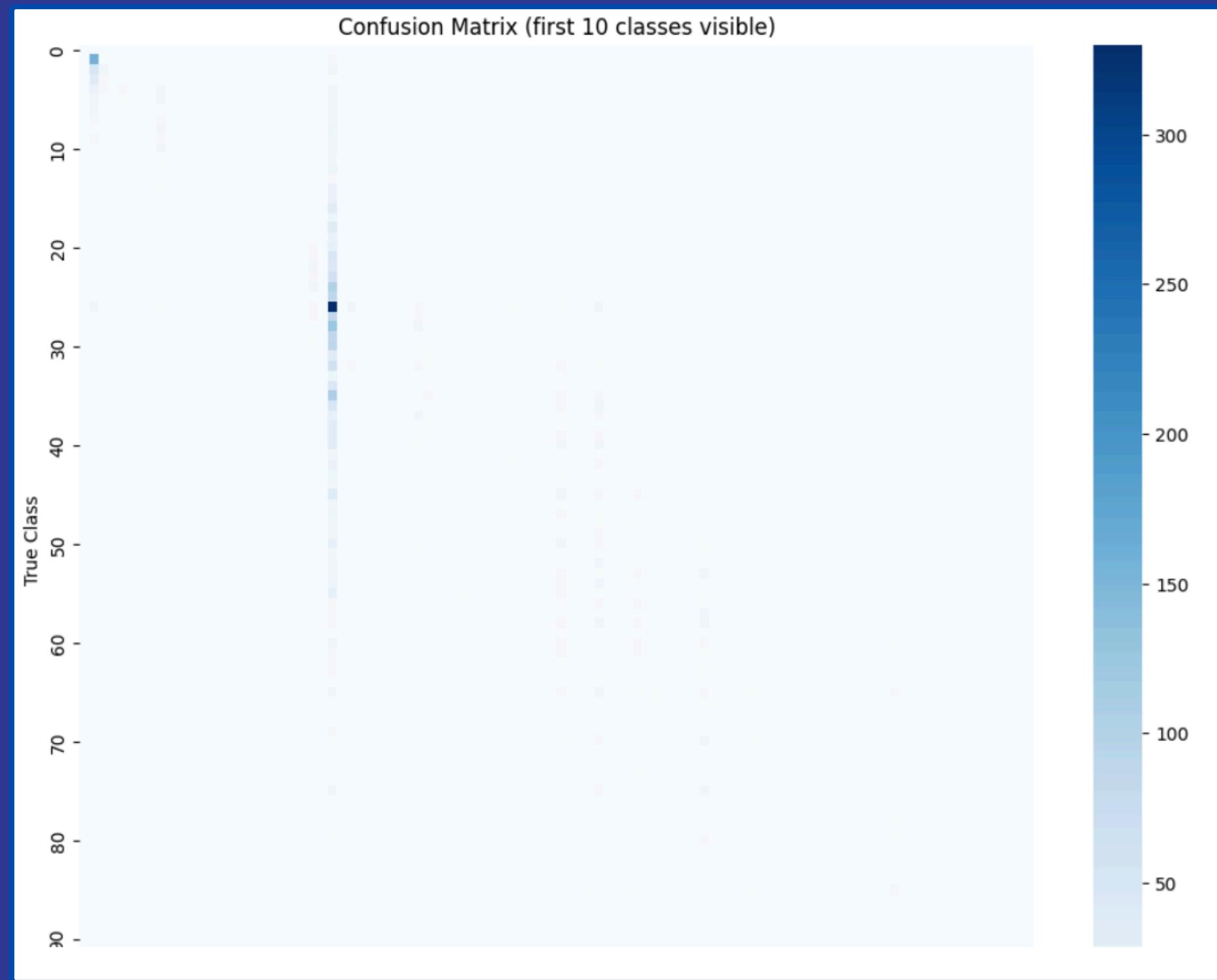
SP: Replicating Mohammed's **FER** approach

- 1- UNDERSAMPLING
- 2- FREEZING
- 3- MODIFYING ARCHITECTURE (ADDING 1/2 FC)
- 4- METRICS AND TOLERANCE ACCURACY
- 5- RANGS
- 6- DATA AUGMENTATION



UNDERSAMPLING

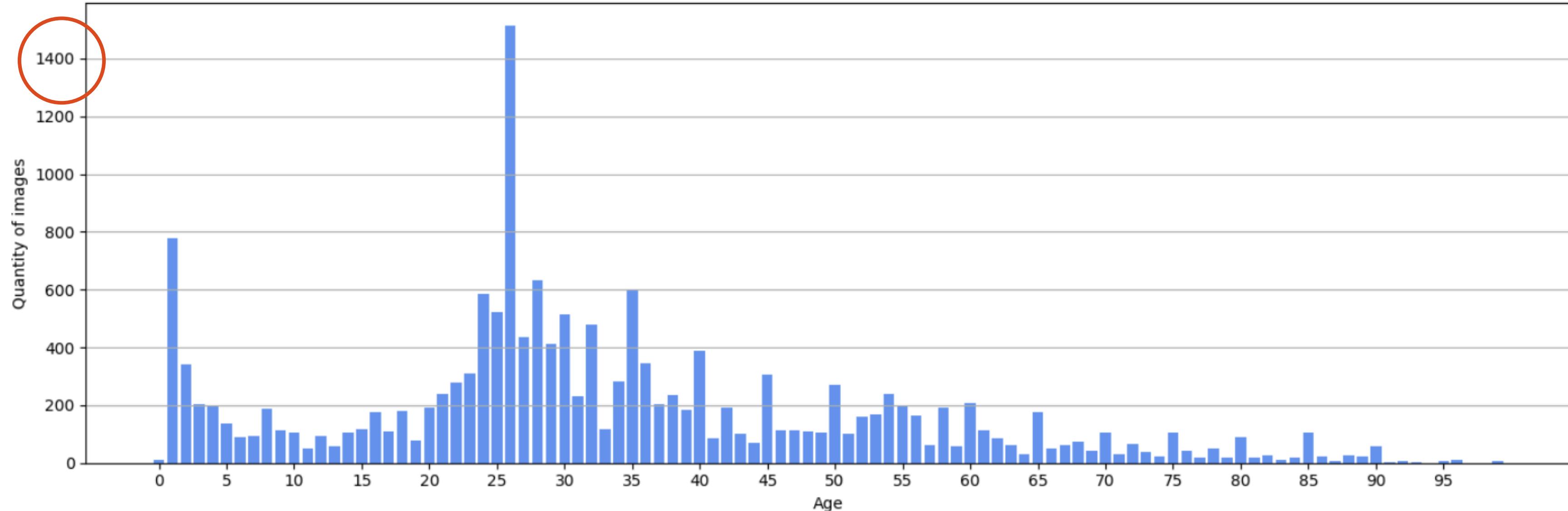
disaster →
predict all as 26 ←



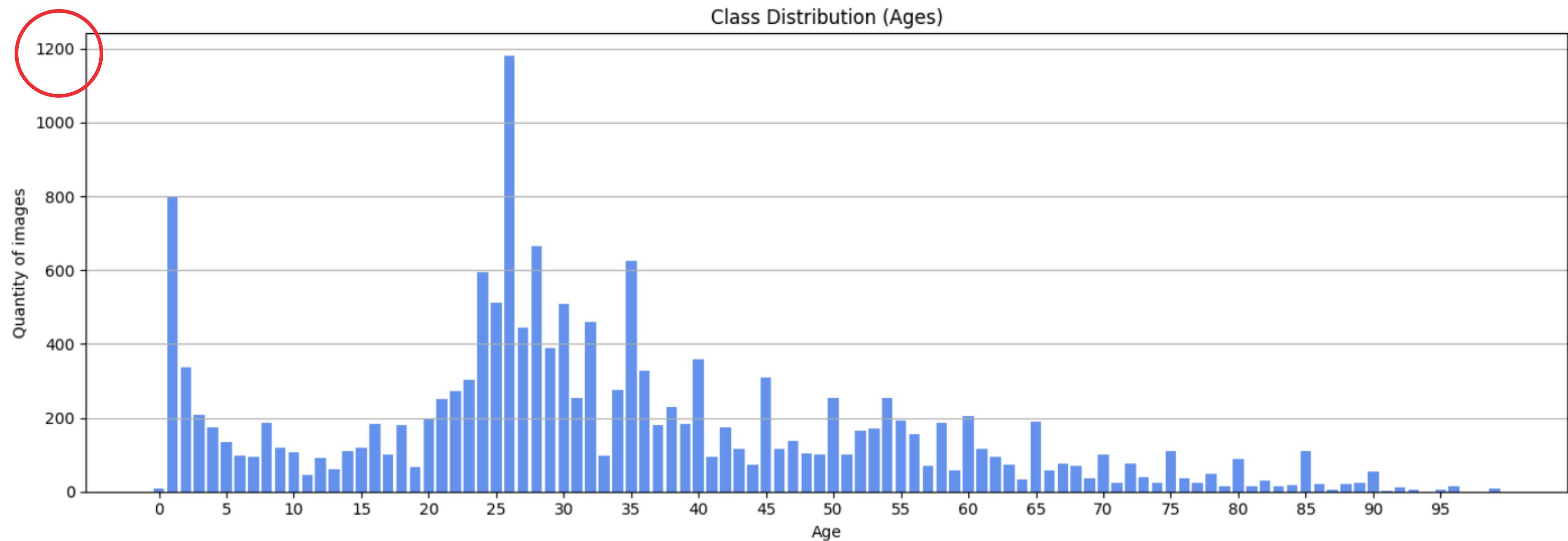
Confusion matrix of orginal model performance, generated in project notebook

class imbalance

Class Distribution (Ages)



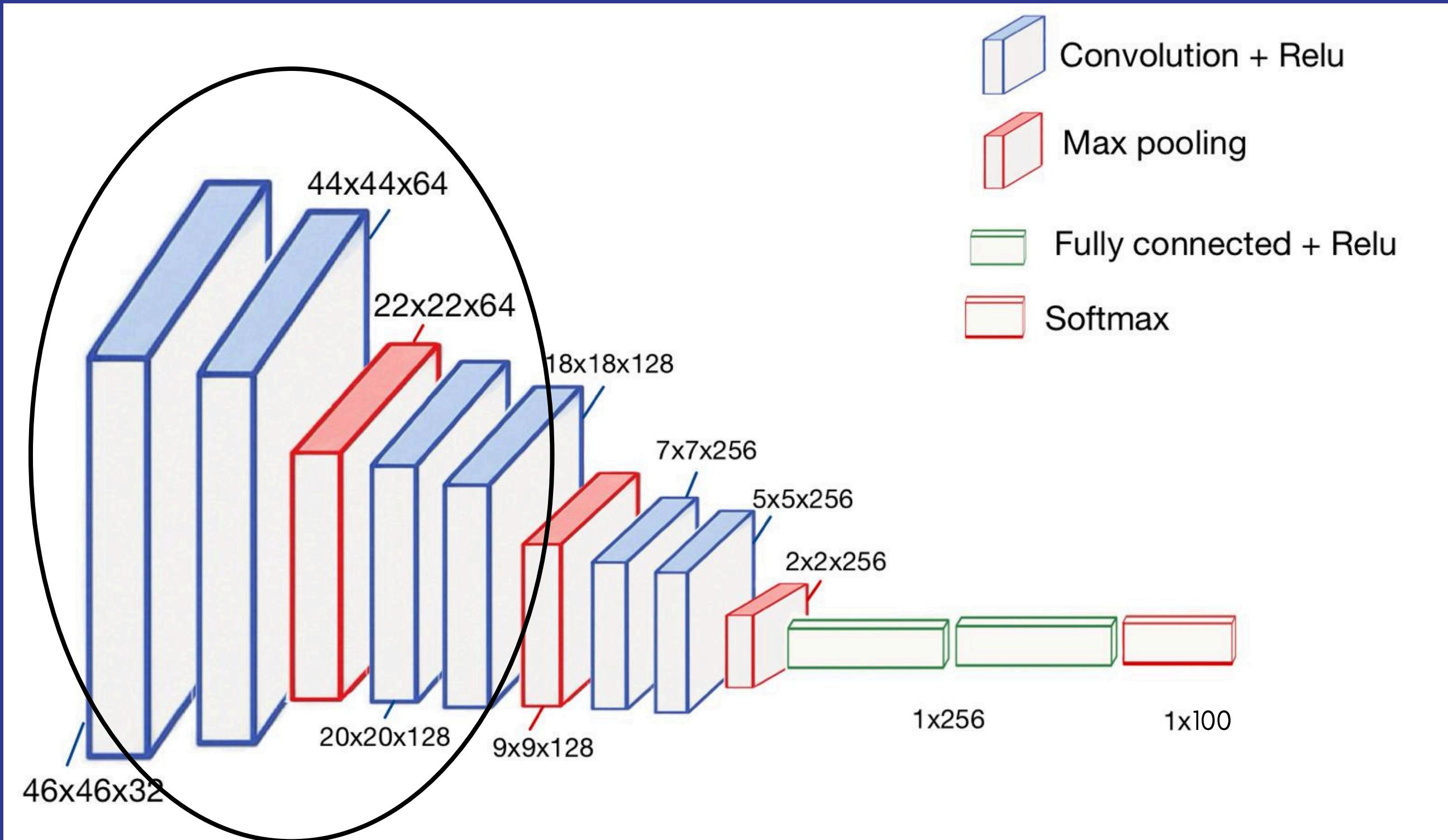
→ undersampling



FREEZING

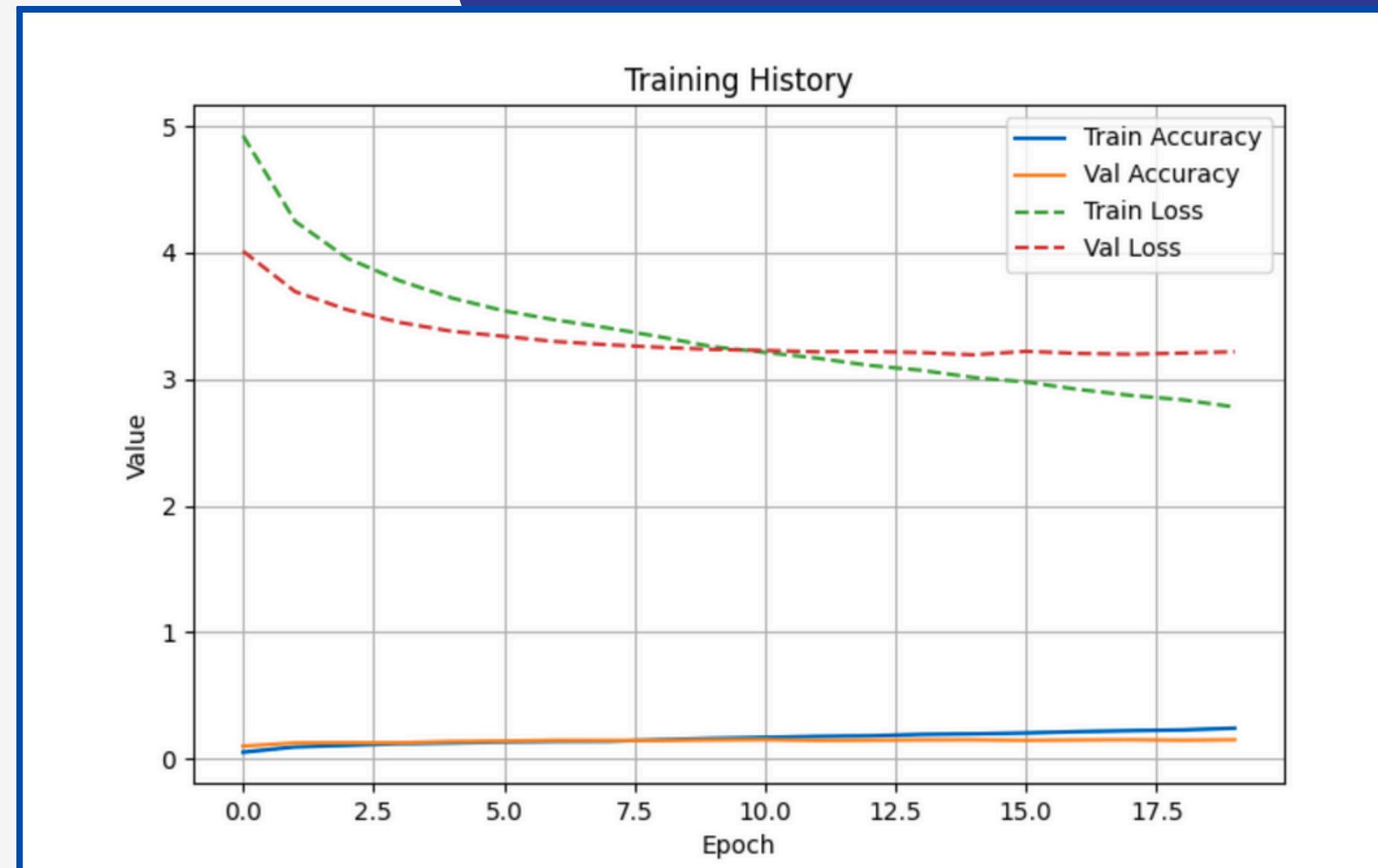
complex TL, how many layers freeze

freezing the first two convolutional block.



CNN model, illustration created by Astrid Alins

FREEZING



Performance analysis 1, generated in project notebook

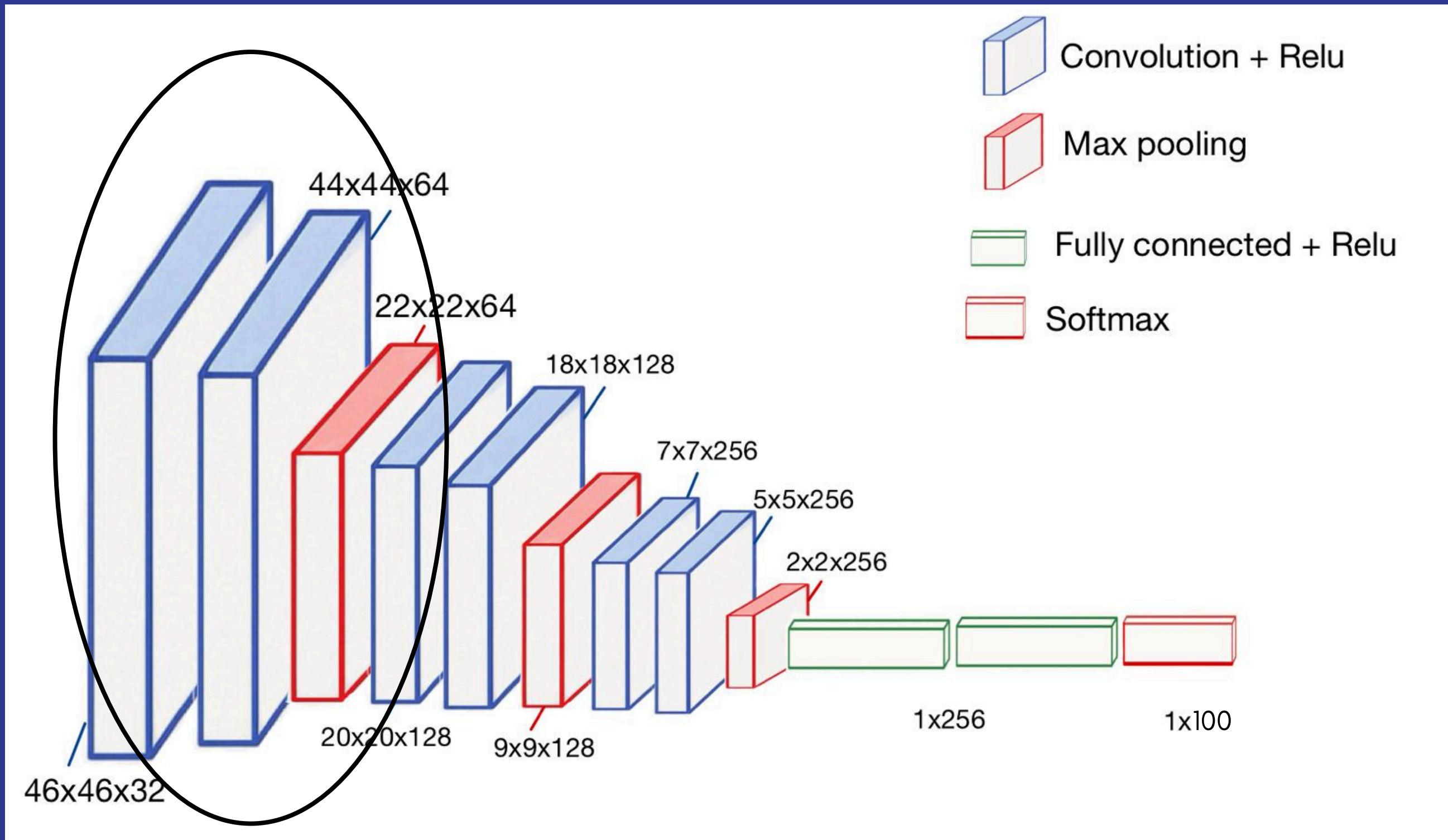
INITIAL accuracy:

- Training : 0.03
- Validation: 0.09

AFTER **TRAINING** accuracy:

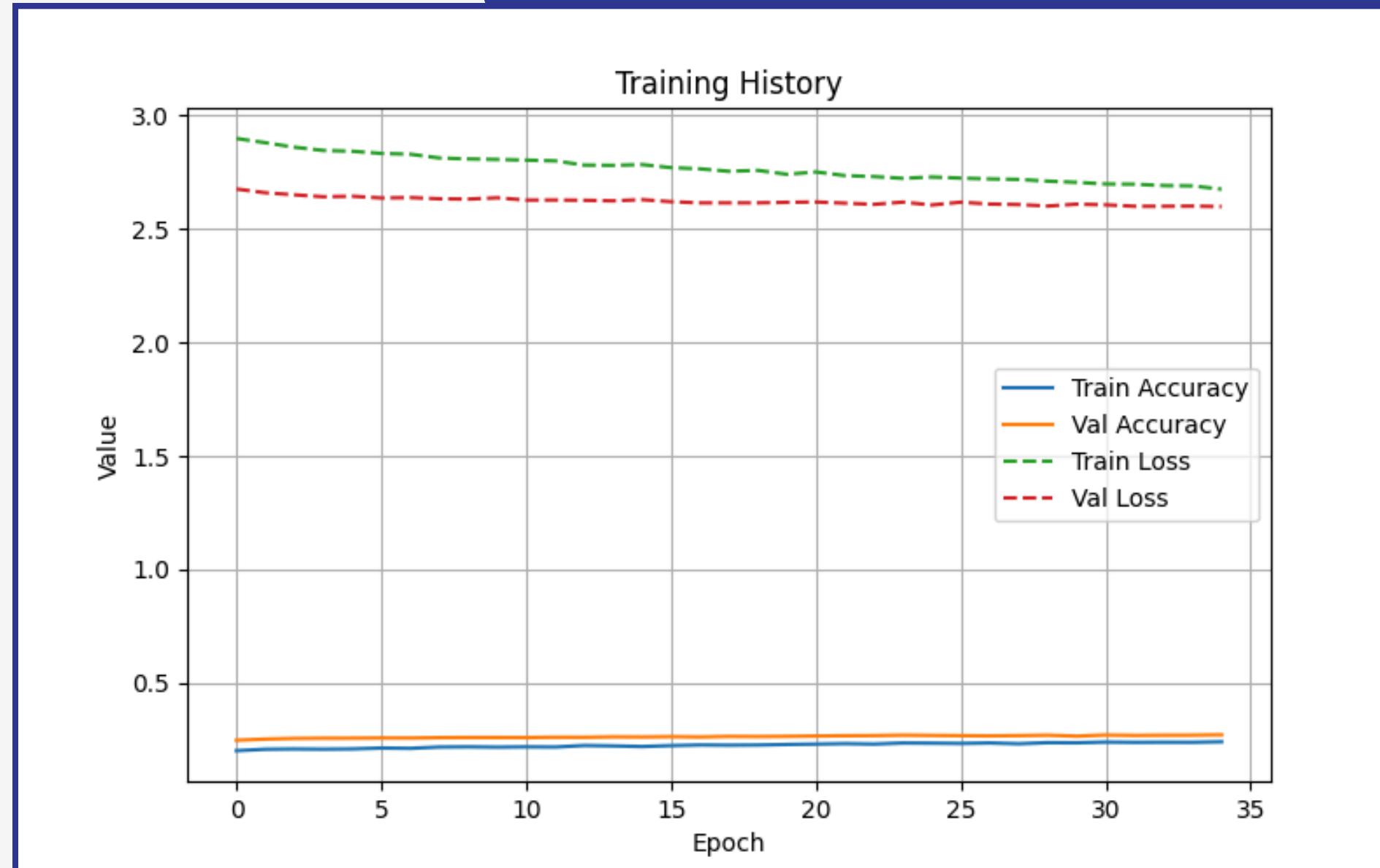
- Training : 0.24
- Validation: 0.15

unfroze another layer



CNN model, illustration created by Astrid Alins

FREEZING



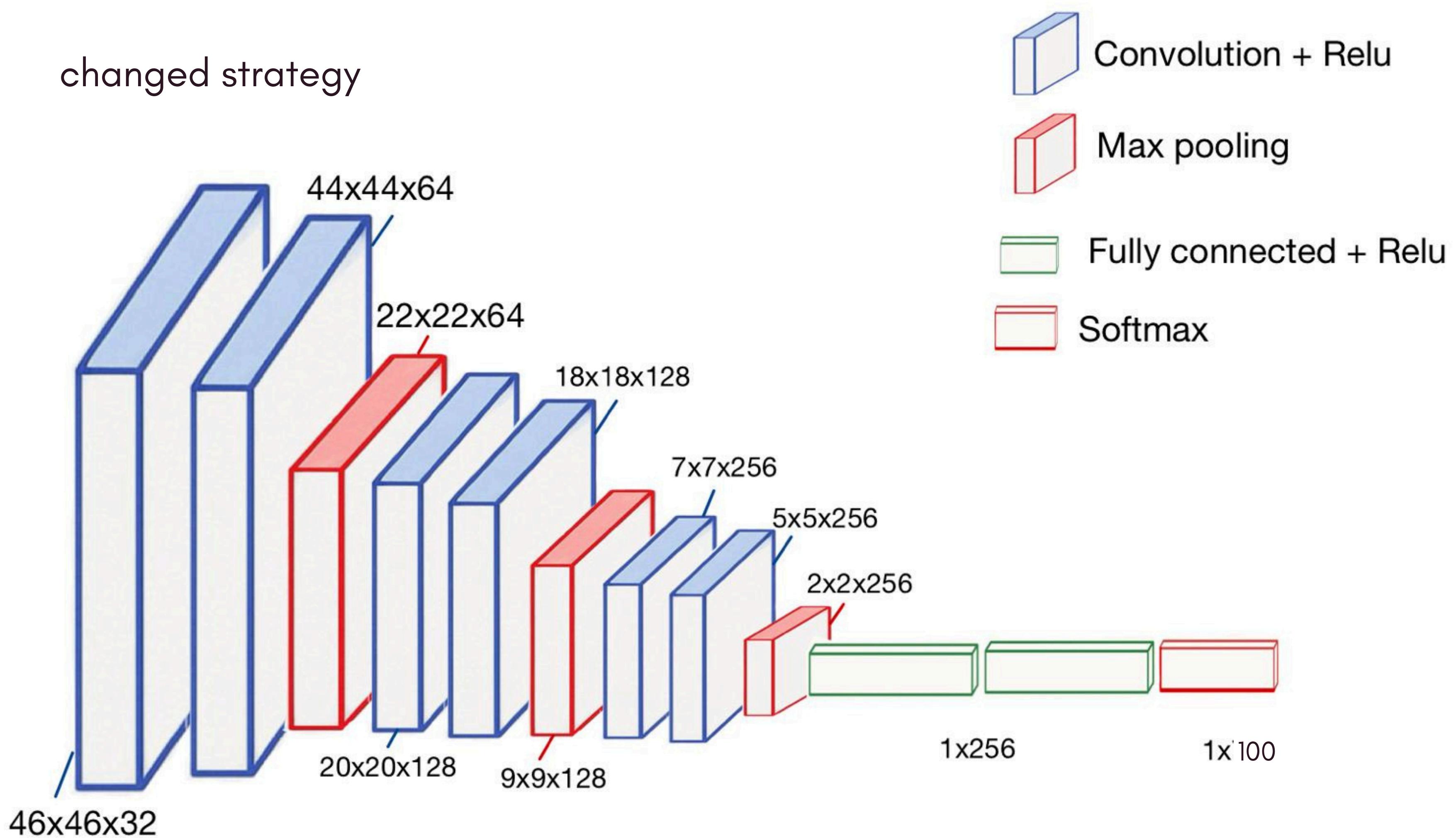
INITIAL accuracy:

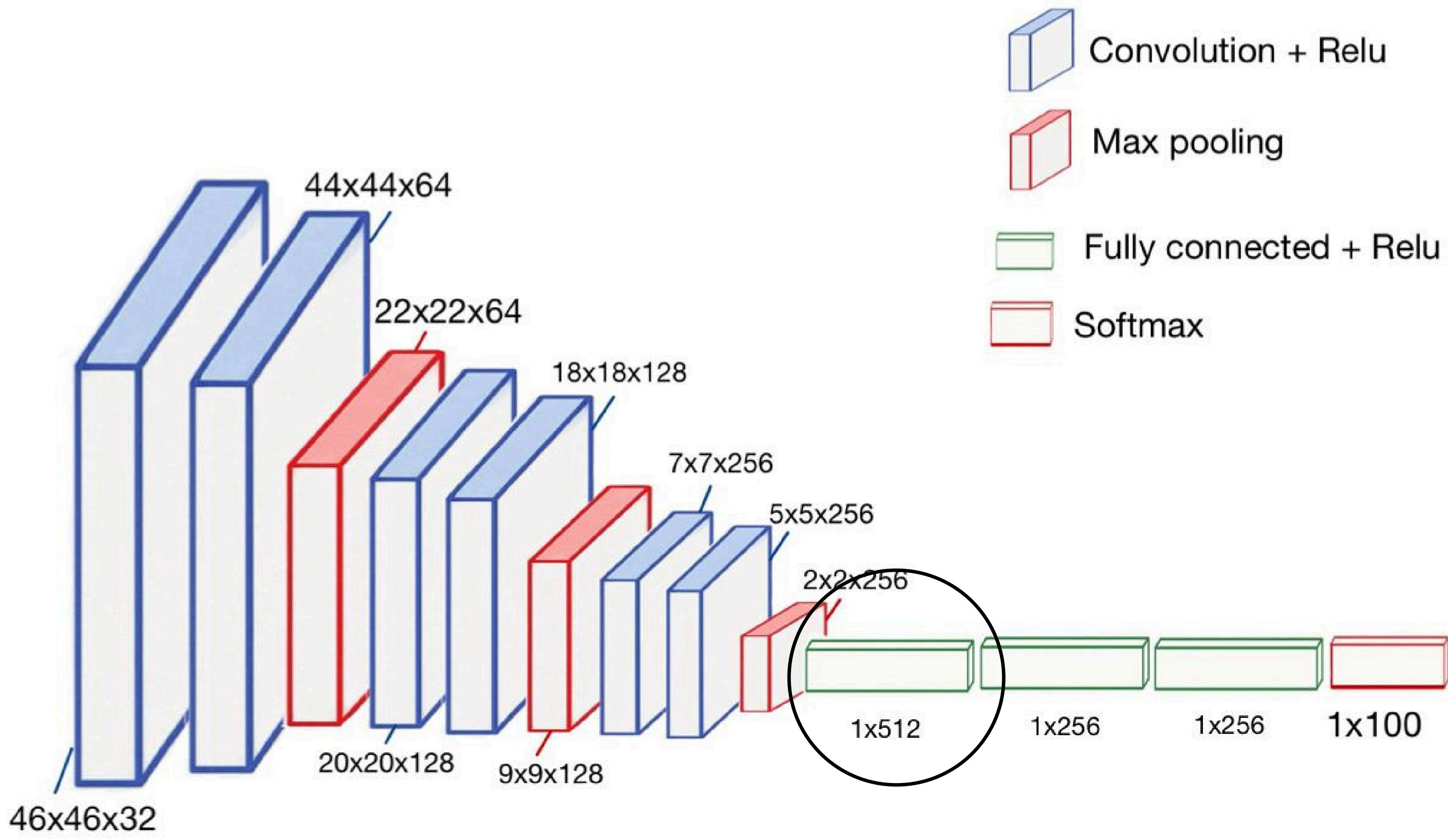
- Training : 0.24
- Validation: 0.15

AFTER **TRAINING** Accuracy

- Training: 0.241
- Validation: 0.27

changed strategy





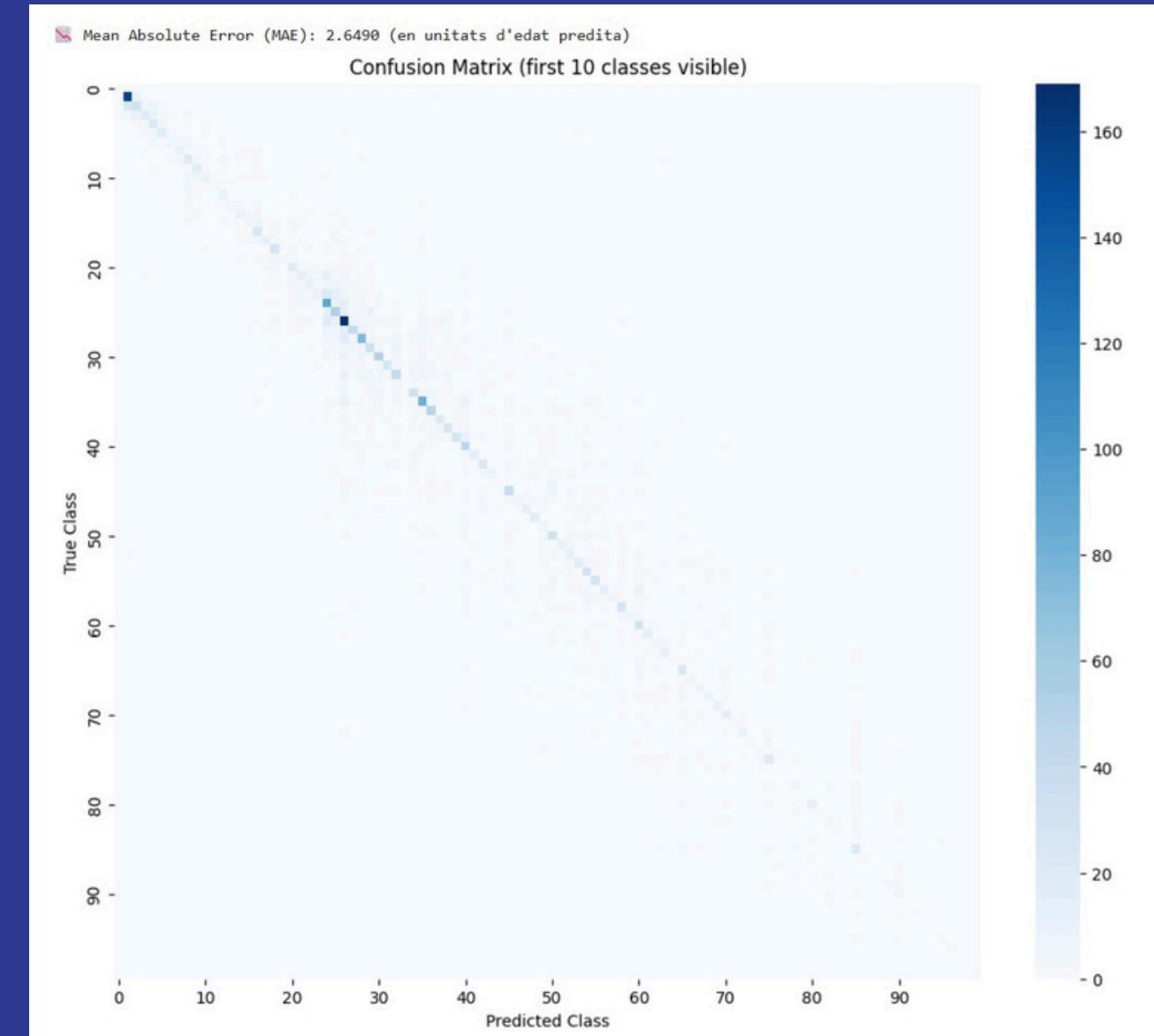
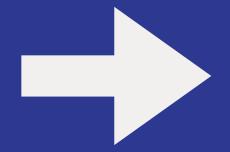
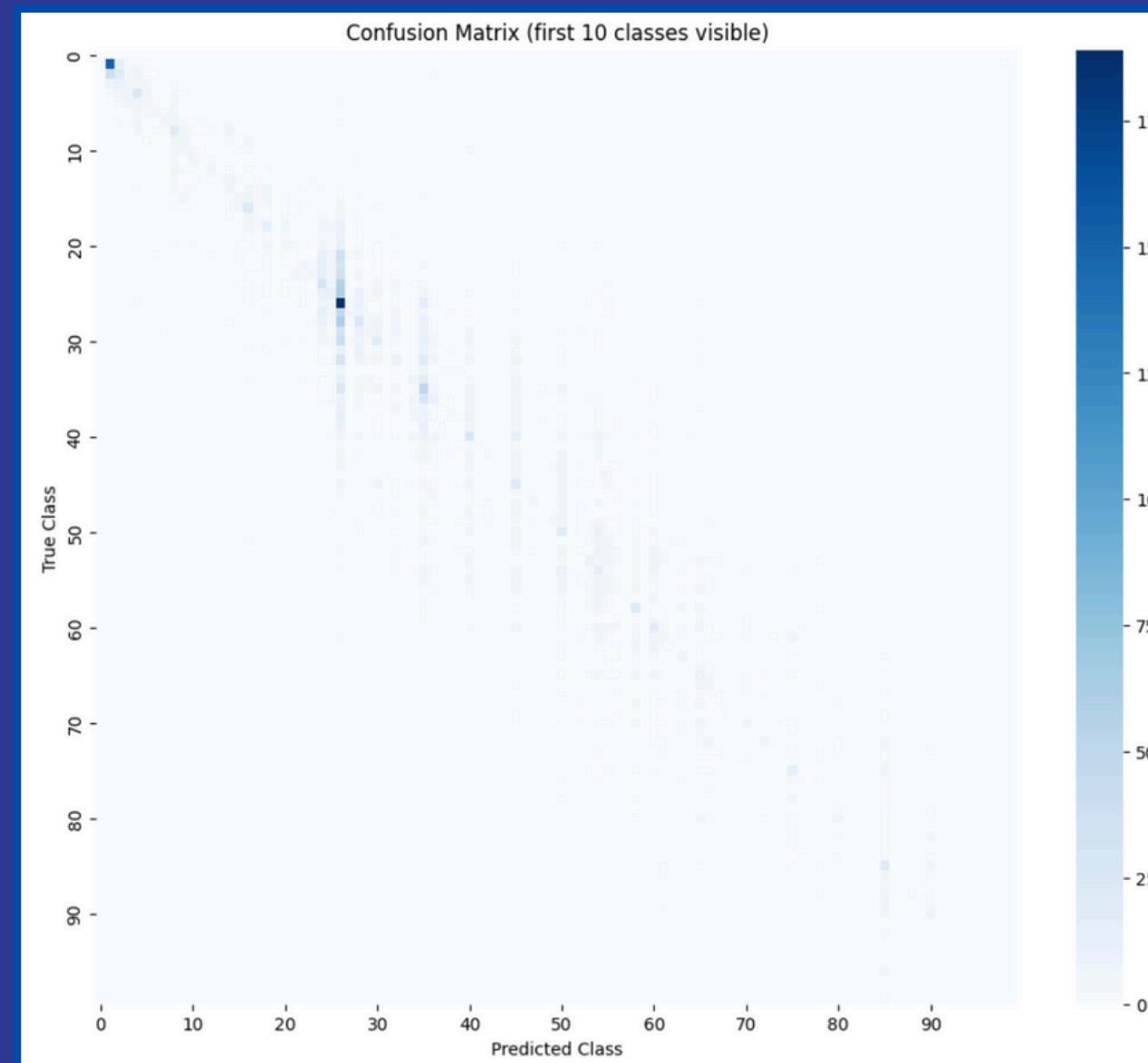
RESULTS OF ADDING 1 FC

INITIAL accuracy:

Training:0.24 Validation:0.27

AFTER TRAINING accuracy:

Training:0.36 Validation:0.54

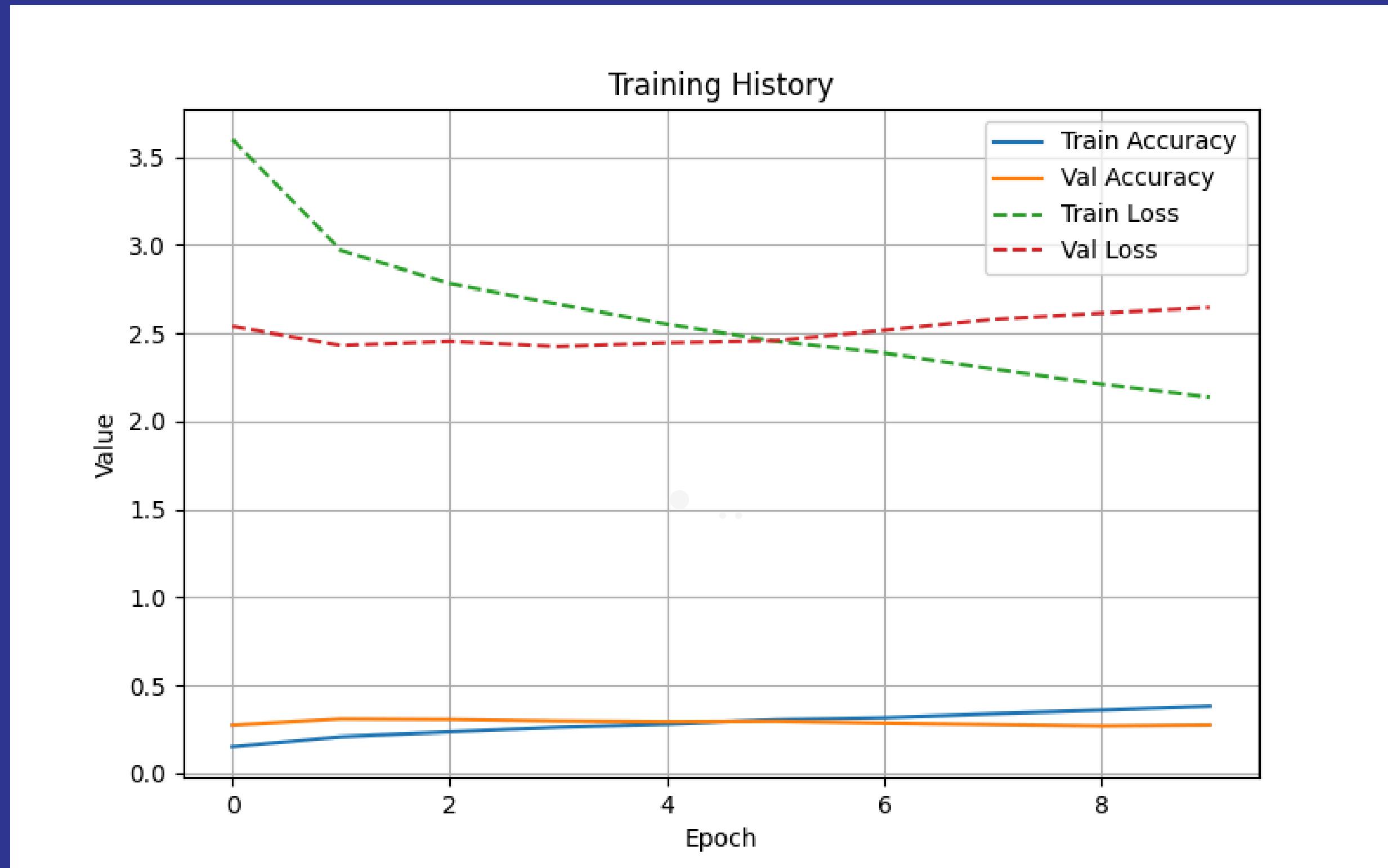


Confusion matrix (1 and 2 FC+), generated in project notebook

tryed add + FC

ADDING F2

gap betwwen train and val
(overfitting)



Performance analysis 3, generated in project notebook

METRICS

Beyond Traditional Accuracy and Confusion Matrix

- Weighted Precision
- Weighted Recall
- Weighted F1-score
- Support
- **Mean Absolute Error: 2.649 years**

CUSTOMED TOLERANCE ACCURACY

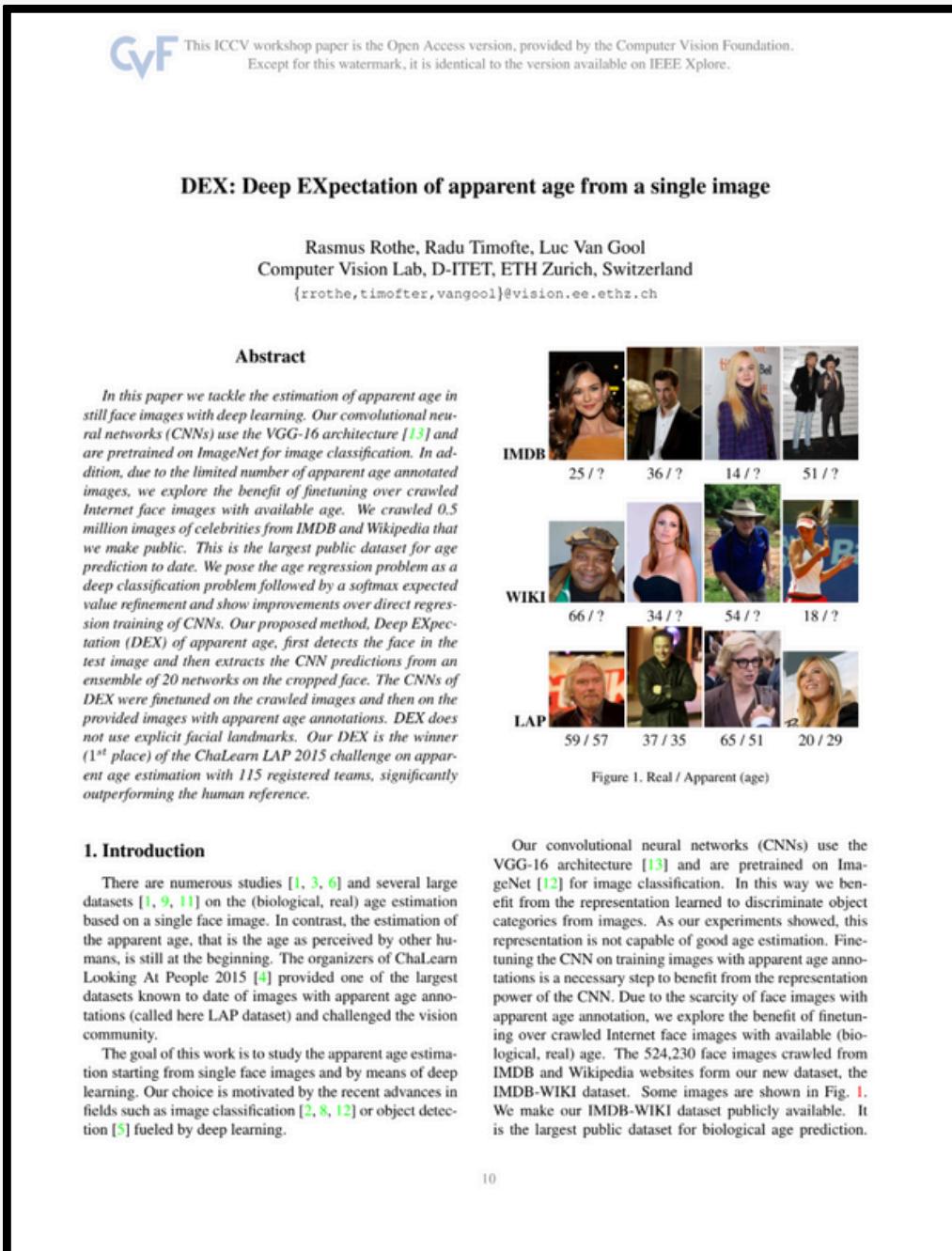
Counts predictions within $\pm k$ years as correct

- ± 0 anys → Accuracy: 54.54%
- ± 1 year → Accuracy: 64.38%
- ± 2 years → Accuracy: 71.06%
- **± 3 years → Accuracy: 76.97%**

BASED ON THESE PROMISING RESULTS...

- ± 0 anys → Accuracy: 54.54%
- ± 1 year → Accuracy: 64.38%
- ± 2 years → Accuracy: 71.06%
- **± 3 years → Accuracy: 76.97%**

HOW TO GROUP PEOPLE?



PROPOSED RANGES:

0-2

3-6

6-9

10-12

13-17

18-24

25-34

35-44

45-54

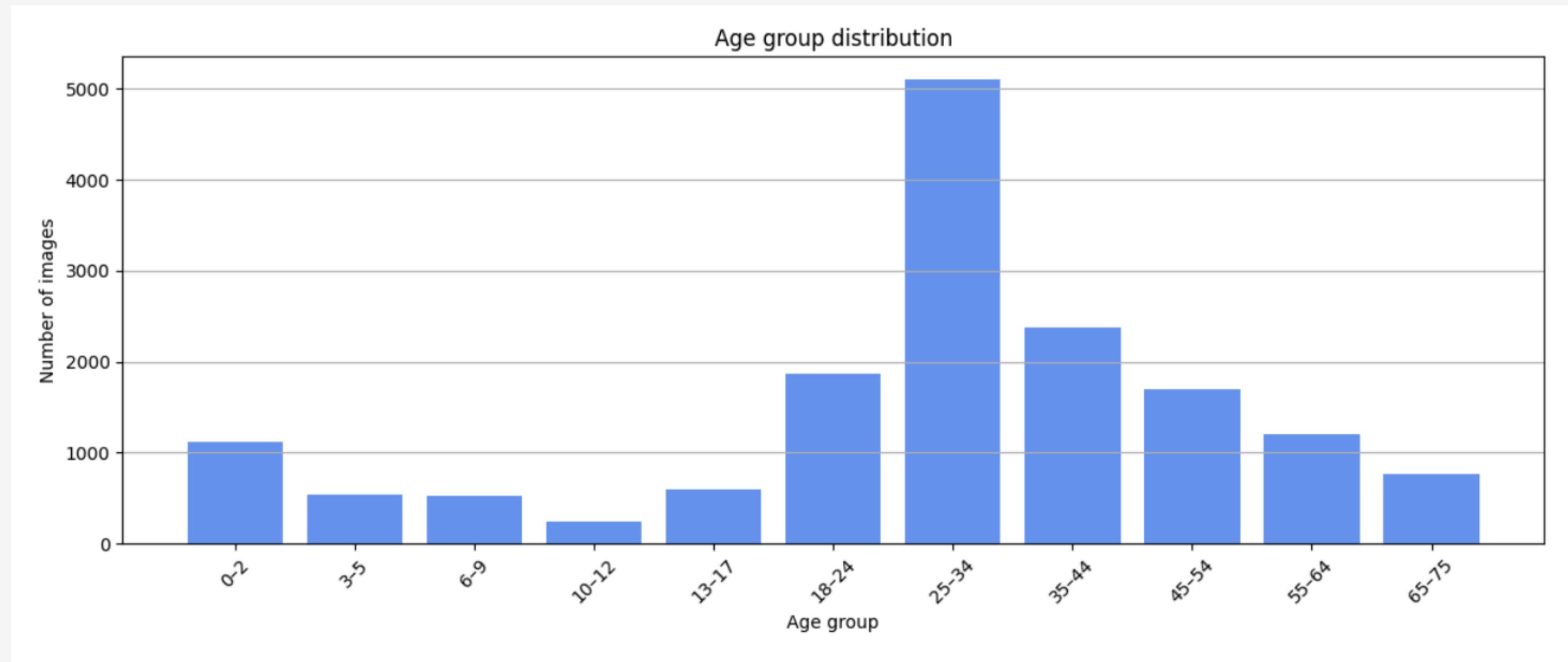
55-64

65-75



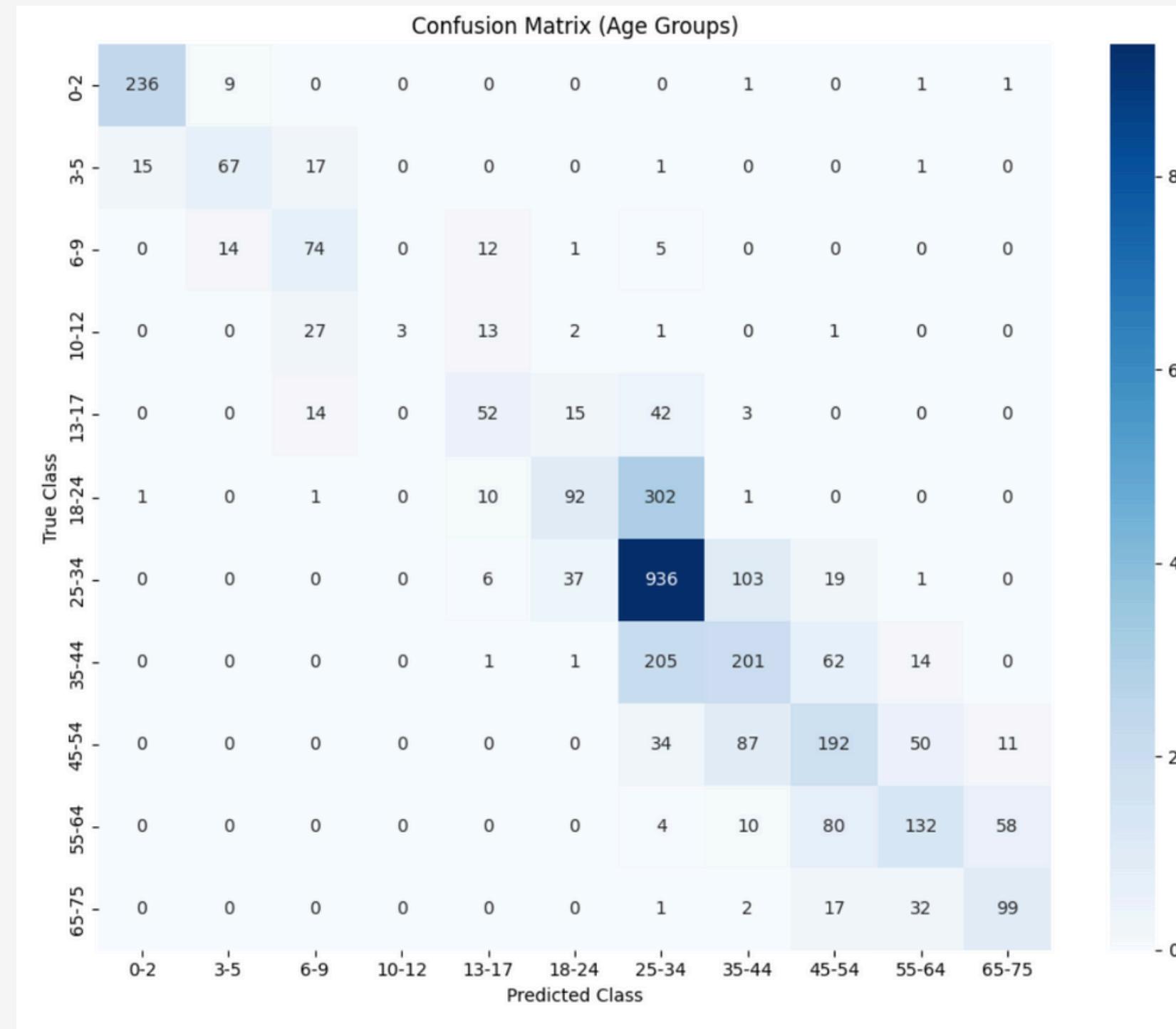
11 CLASSES

GROUPED DATA



Age groups representation generated in project notebook

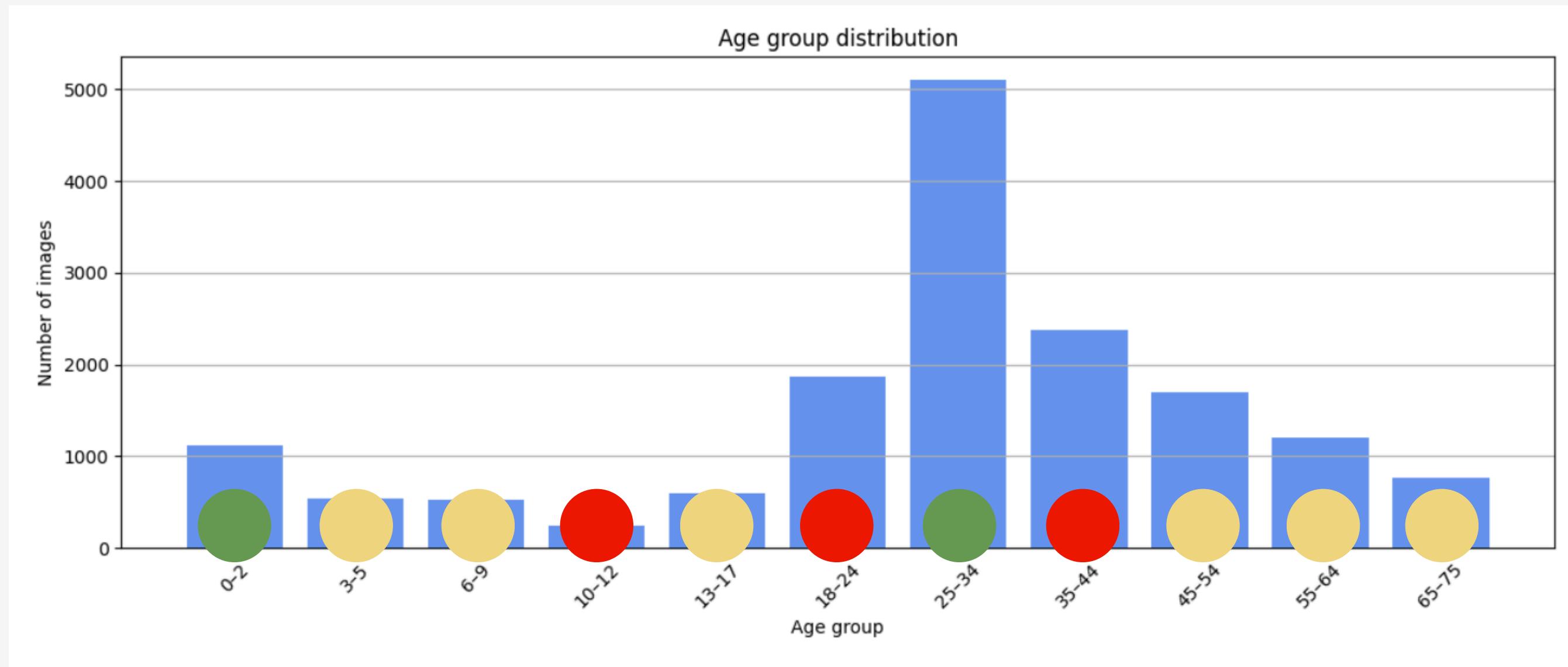
RESULTS



Confusion matrix (ranges), generated in project notebook

Weigthed precision 0.6101
Weighted recall 0.6076
Weighted f1-score 0.5843
Support 3430
Accuracy 0.6076

AGE BY AGE DISCUSSION



Age group representation, generated in project notebook

REALY GOOD

MODERATE

BAD

LET'S
ANALIZE
THE
FINAL
RESULTS



OUR BEST SETUP

One FC added

Parameters re-trained : 1.569.448

CUSTOMED TOLERANCE ACCURACY

Counts predictions within $\pm k$ years as correct

- ± 0 anys → Accuracy: 54.54%
- ± 1 year → Accuracy: 64.38%
- ± 2 years → Accuracy: 71.06%
- **± 3 years → Accuracy: 76.97%**

IF YOU THINK
IT'S NOT FAIR

LET'S EVALUATE
HUMAN CAPACITY

SPECIFICALLY,
YOUR CAPACITY

SOMETIMES IS NOT
THAT HARD BUT
OTHERS...

TRY TO GUESS



Image extracted from UTKFace dataset

10



Image extracted from UTKFace dataset

25



Image extracted from UTKFace dataset

17

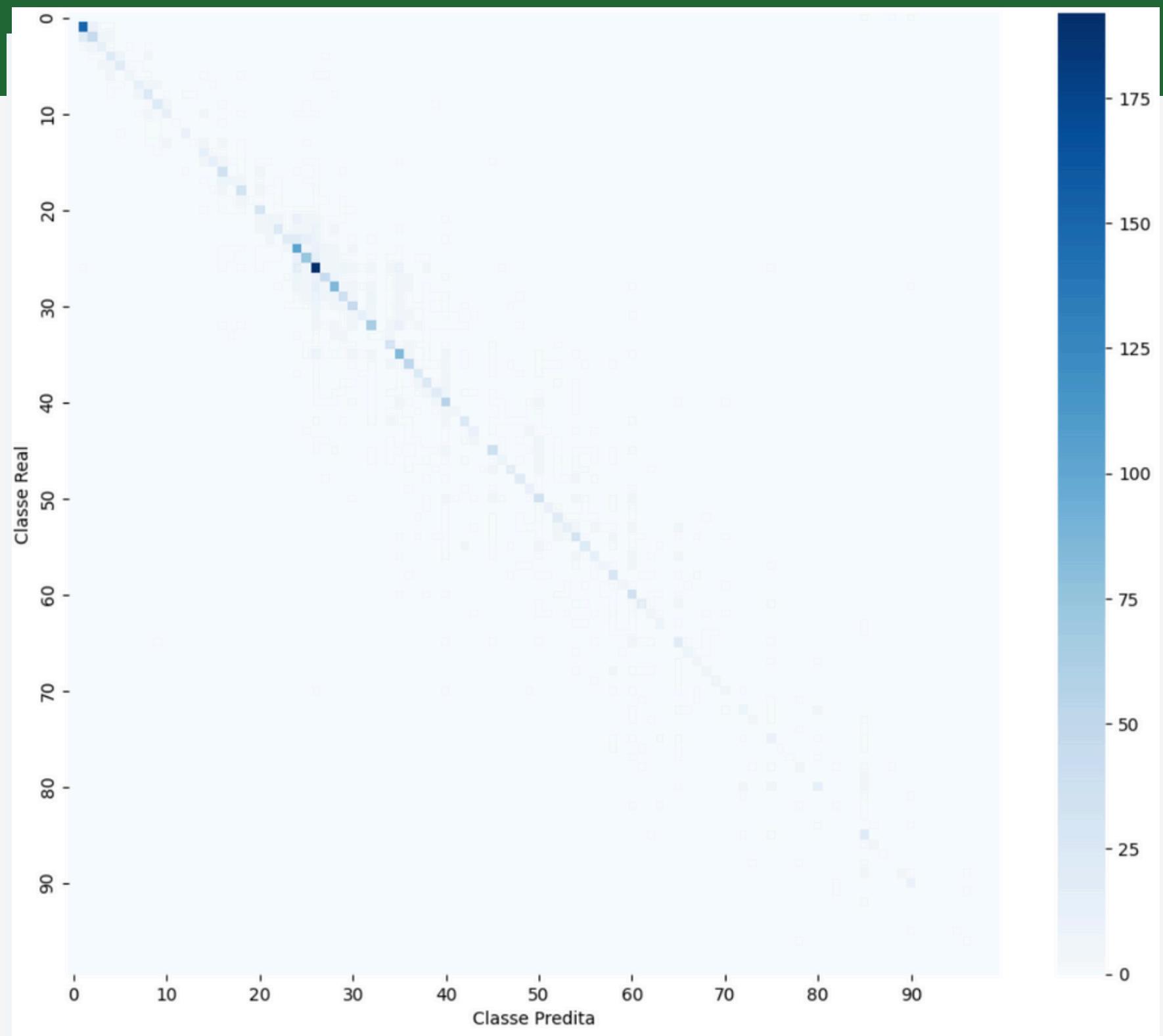
RESULTS



NO TOLERANCE → **WITH TOLERANCE ±3**

Weighted Accuracy:	57%	→	76.97%
Weighted Precision:	0.59	→	0.78
Weighted Recall:	0.55	→	0.77
Weighted F1-score:	0.56	→	0.77

RESULTS

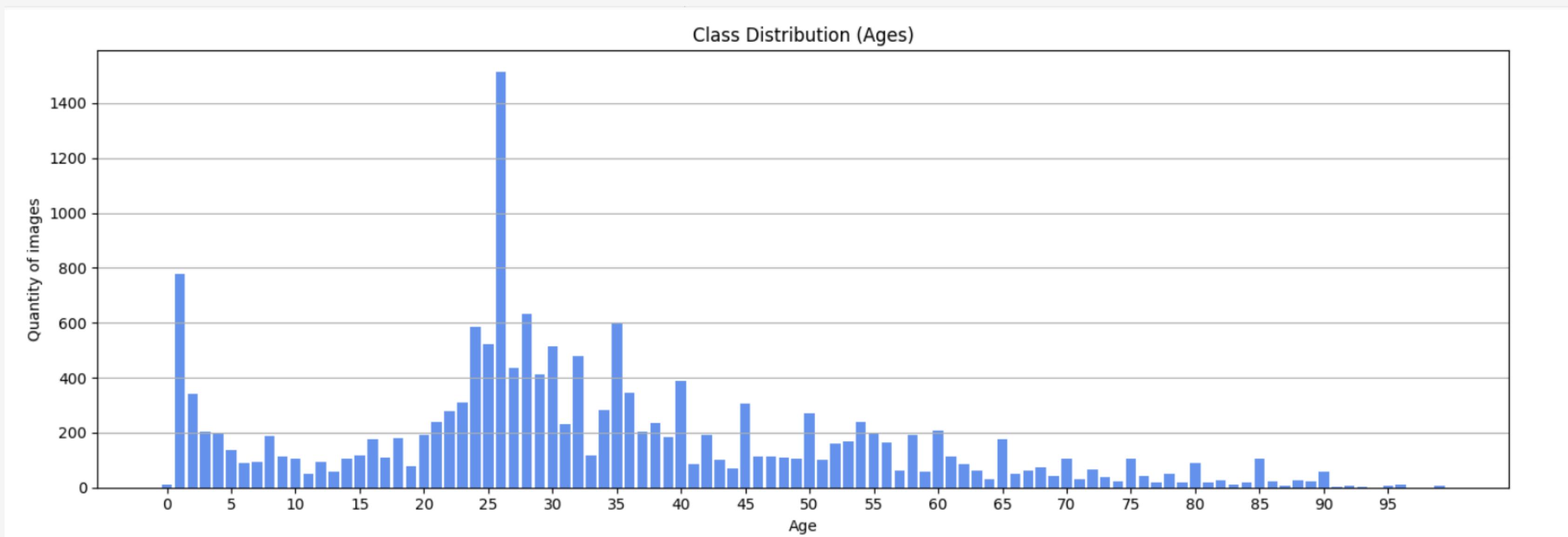


Confusion matrix (best performance), generated in project notebook

DISCUSSION

F1-score range	Performance	Ages	# Ages
0.85 – 1.00	Excellent	1, 2, 3, 5, 23, 26, 86, 87, 83 ...	13
0.75 – 0.84	Good	25, 27, 28, 29, 32, 36, 38, 39 ...	15
0.60 – 0.74	Moderate	8, 10, 20, 22, 30, 34, 35, 37 ...	23
0.00 – 0.59	Poor	0, 4, 6, 7, 9, 19, 21, 31, 33 ...	36

DISCUSSION



General age distribution, generated in project notebook

- Train on a less biased dataset for better generalization
- Modify CNN architecture: adding convolutional layers using padding
- Analyze whether the model predicts ages equally well across different ethnicities.

FUTURE IMPROVEMENTS



CONCLUSIONS

Class imbalance is a critical issue.

The **amount of freezed** layers play a key role in transfer learning.

Right amount of **FC is essential**

CONCLUSIONS



Multiple and problem-specific metrics are essential.

Data augmentation is only effective when sufficient base data is available.

Thanks!

