Matching emotions: portraits and songs

Stefanos Vlachos | dit22002dsc

Efstathios Zaragkas | dit22003dsc

Introduction

What is the goal of this project ?

What approach was adopted ?

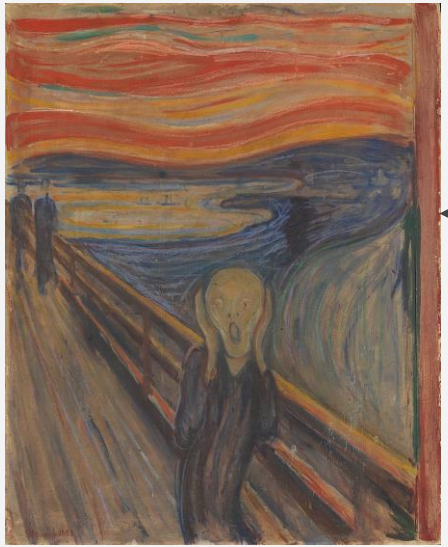
Intrroduction | **Goal**

- Implement an application that utilizes the basic steps of Deep Learning

Intrroduction | **Goal**

- Implement an application that utilizes the basic steps of Deep Learning

Initial Idea



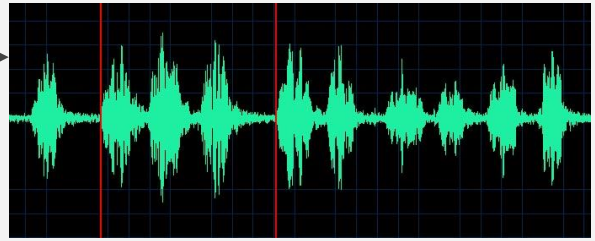
Match based on:

Artistic Movement

Era

Techniques

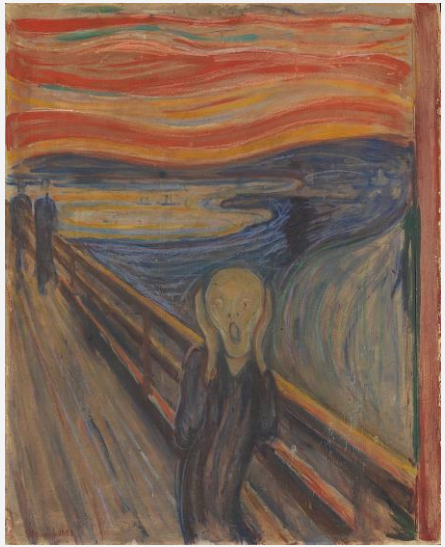
Emotion



Intrroduction | **Goal**

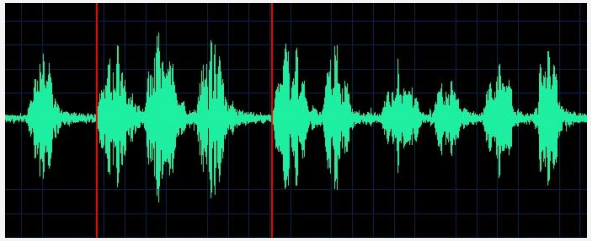
- Implement an application that utilizes the basic steps of Deep Learning

Initial Idea



However:

- Hard to find large volumes of **annotated data**
- **Demanding** task



Intrroduction | **Goal**

Viable Idea

Detect face

Intrroduction | **Goal**

Viable Idea

Detect face

Sentiment Analysis on face

Intrroduction | **Goal**

Viable Idea

Detect face

Sentiment Analysis on face

Title	Time	Artist	Album	Genre	Plays
All of Me (Bumper's Audition)	1:28	Adam DeVine	Pitch Perfect 2 (Or...	Soundtrack	
Girl On Fire	3:45	Alicia Keys	Girl On Fire - Single	R&B/Soul	
The Man	4:15	Aloe Blacc	Lift Your Spirit	Pop	
Cups (Movie Version)	1:17	Anna Kendrick	Pitch Perfect (Orig...	Soundtrack	
Problem (feat. Iggy Azalea)	3:14	Ariana Grande	Problem (feat. Igg...	Pop	
The Way (feat. Mac Miller)	3:47	Ariana Grande	Yours Truly	Pop	
Wake Me Up	4:10	Avicii	Wake Me Up - Sin...	Dance	
Airplanes (feat. Hayley Williams...	2:59	B.o.B	Airplanes (feat. Ha...	Hip-Hop/R...	
John Doe (feat. Priscilla)	3:32	B.o.B	John Doe (feat. Pri...	Hip-Hop/R...	
Get'cha Head In the Game (Pop...	2:44	B5	High School Musical	Soundtrack	
Bellas Regionals: The Sign / Ete...	2:40	The Barden Bellas	Pitch Perfect (Orig...	Soundtrack	
Pool Mashup: Just the Way You...	1:39	The Barden Bellas	Pitch Perfect (Orig...	Soundtrack	
Party In the U.S.A.	1:03	The Barden Bellas	Pitch Perfect (Orig...	Soundtrack	
Bellas Finals: Price Tag / Do...	3:37	The Barden Bellas	Pitch Perfect (Orig...	Soundtrack	
Kennedy Center Performance	2:27	The Barden Bellas	Pitch Perfect 2 (Or...	Soundtrack	
Convention Performance	1:41	The Barden Bellas	Pitch Perfect 2 (Or...	Soundtrack	
Back to Basics	1:31	The Barden Bellas	Pitch Perfect 2 (Or...	Soundtrack	
Cups (When I'm Gone) [Campfir...	0:46	The Barden Bellas	Pitch Perfect 2 (Or...	Soundtrack	
World Championship Finale 2	4:16	The Barden Bellas	Pitch Perfect 2 (Or...	Soundtrack	
Riff Off: Mickey / Like a Virgin /...	3:44	The Barden Bellas,...	Pitch Perfect (Orig...	Soundtrack	
Pompeii	3:34	Bastille	Bad Blood (Bonus...	Alternative	
God Only Knows	2:53	The Beach Boys	Sounds of Summe...	Rock	
Michelle	2:42	The Beatles	The Beatles 1962-...	Rock	
Yellow Submarine	2:39	The Beatles	The Beatles Box Set	Rock	
Shower	3:26	Becky G.	Shower - Single	Pop	
For Forever	5:01	Ben Platt	Dear Evan Hansen...	Soundtrack	
Words Fail	5:52	Ben Platt	Dear Evan Hansen...	Soundtrack	
If I Could Tell Her	4:09	Ben Platt & Laura...	Dear Evan Hansen...	Soundtrack	
Waving Through A Window	3:56	Ben Platt & Origin...	Dear Evan Hansen...	Soundtrack	

Match song with same emotion

Intrroduction | **Approach**

01

Collect annotated **face images**

Collect **portrait images**

Collect annotated **songs**

Intrroduction | **Approach**

01

Collect annotated **face images**

Collect **portrait images**

Collect annotated **songs**

02

Apply **Transfer Learning** on faces

Intrroduction | **Approach**

01

Collect annotated **face images**

Collect **portrait images**

Collect annotated **songs**

02

Apply **Transfer Learning** on faces

03

Crop faces inside the portraits

Extraction of sentiment using the model

Intrroduction | **Approach**

01 — Collect annotated **face images** — Collect **portrait images** — Collect annotated **songs**

02 — Apply **Transfer Learning** on faces

03 — **Crop faces** inside the portraits — **Extraction** of sentiment using the model

04 — **Match** predicted face emotion with song emotion from database

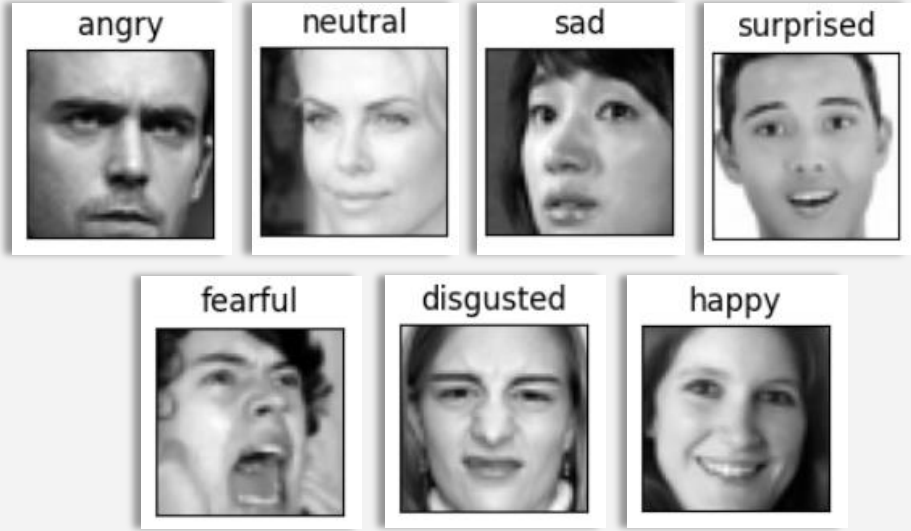
Datasets

Which datasets were used for each modality ?

Datasets | **Face Images**

Emotion Detection | Kaggle Dataset

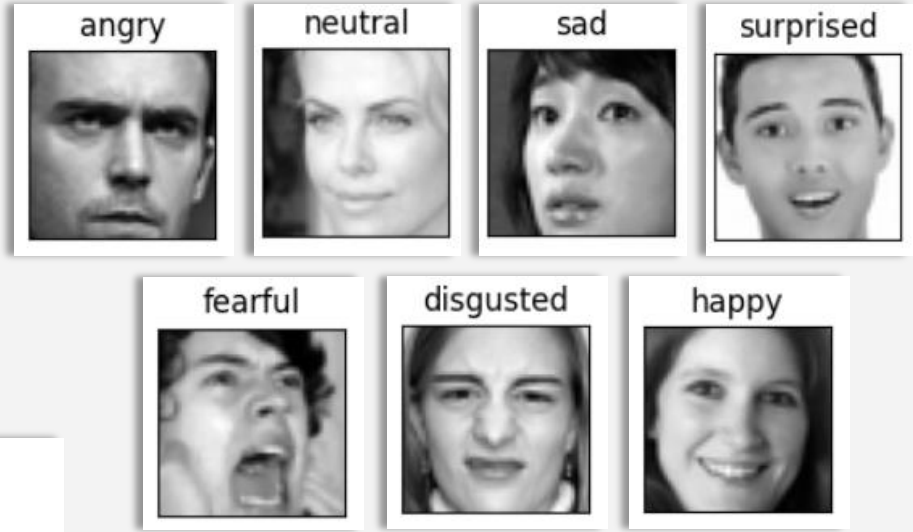
- **35.902** Total images
- **48 X 48** pixels
- Greyscale
- **Annotated:**
  - 4002 **Surprised**
  - 6198 **Neutral**
  - 6082 **Sad**
  - 5121 **Fearful**
  - 4980 **Angry**
  - 574 **Disgusted**
  - 8999 **Happy**



Datasets | **Face Images**

Emotion Detection | Kaggle Dataset

- **35.902** Total images
- **48 X 48** pixels
- Greyscale
- **Annotated:**
  - 4002 **Surprised**
  - 6198 **Neutral**
  - 6082 **Sad**
  - 5121 **Fearful**
  - 4980 **Angry**
  - 574 **Disgusted**
  - 8999 **Happy**



Used for:

Train, Validation and Test

Datasets | **Face Images**

Art Portraits | Kaggle Dataset

- **4.117** Total images
- **Random** resolutions
- **RGB**
- **Not Annotated**

Datasets | **Face Images**

Art Portraits | Kaggle Dataset

- **4.117** Total images
- **Random** resolutions
- **RGB**
- **Not Annotated**

Used for:

Demonstration purposes

Datasets | **Songs**

4Q Audio Emotion Dataset

- **900** Audio Clips
- **~30** seconds
- **Annotated:**

  Q1, Q2, Q3, Q4

Datasets | **Songs**

4Q Audio Emotion Dataset

- **900** Audio Clips
- **~30** seconds
- **Annotated:**

  Q1, Q2, Q3, Q4

PMEmo

- **794** Chorus Clips
- **Random** durations
- **Annotated:**

  Arousal, Valence

Datasets | **Songs**

4Q Audio Emotion Dataset

- **900** Audio Clips
- **~30** seconds
- **Annotated:**

Q1, Q2, Q3, Q4

PMEmo

- **794** Chorus Clips
- **Random** durations
- **Annotated:**

Arousal, Valence

WCMED

- **200** Classical Songs
- **~8-20** seconds
- **Annotated:**

Arousal, Valence

Datasets | **Songs**

4Q Audio Emotion Dataset

- **900** Audio Clips
- **~30** seconds
- **Annotated:**

Q1, Q2, Q3, Q4

PMEmo

- **794** Chorus Clips
- **Random** durations
- **Annotated:**

Arousal, Valence

WCMED

- **200** Classical Songs
- **~8-20** seconds
- **Annotated:**

Arousal, Valence

Emotify

- **400** Audio Clips
- **1** minute
- **Annotated:**

Amazement, Solemnity, Power etc

Annotations

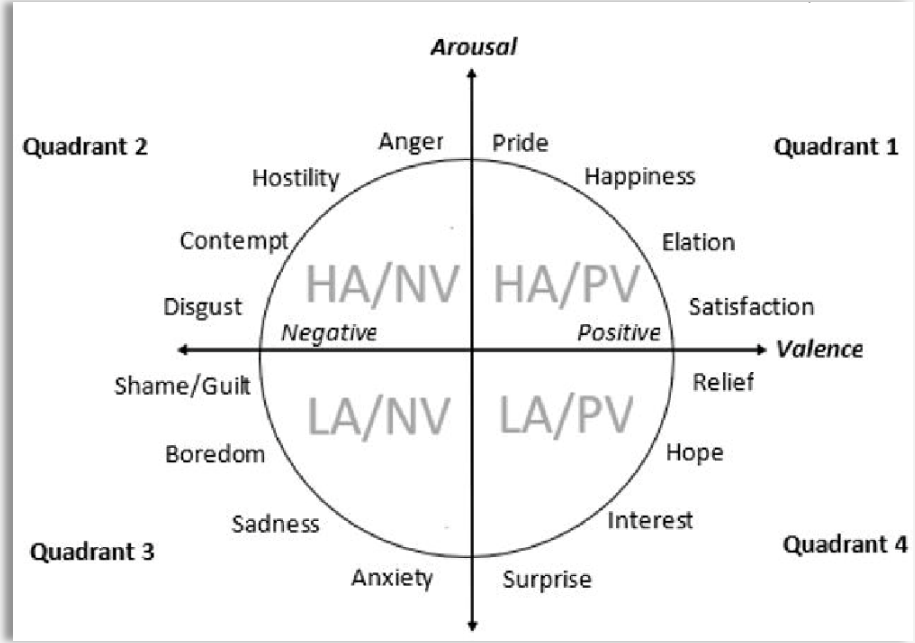
How can we find a common ground for different annotations ?

Russell's Four-Quadrant Emotion Prediction

Class-Based Analysis of Russell's Four-Quadrant Emotion Prediction in Virtual Reality using Multi-Layer Feedforward ANNs

Authors:
J. Mountstephens, N. S. Suhaimi, J. Teo

- Experiment using VR and ML
- **Four-class** emotion classification



Russell's Four-Quadrant Emotion Prediction

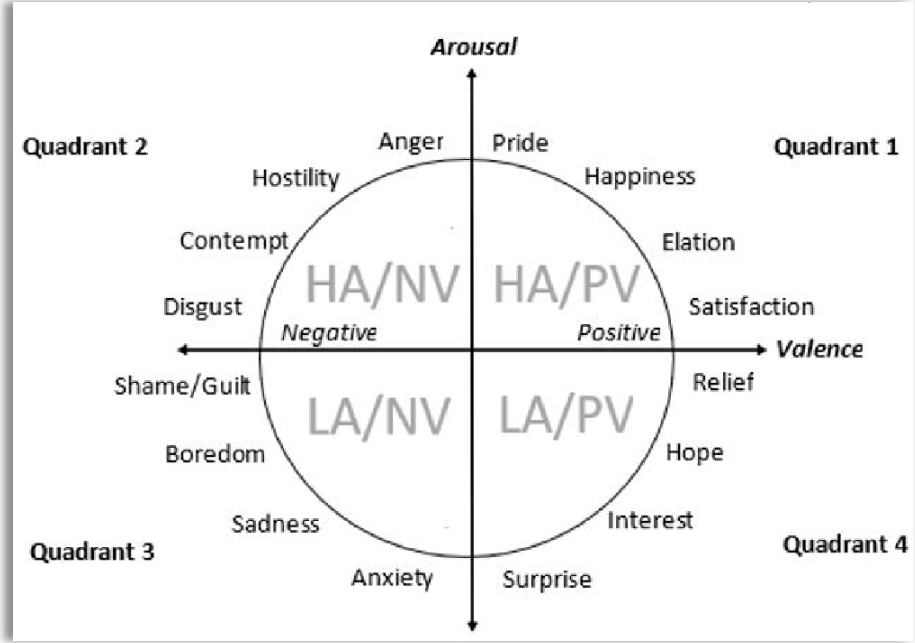
Class-Based Analysis of Russell's Four-Quadrant Emotion Prediction in Virtual Reality using Multi-Layer Feedforward ANNs

Authors:
J. Mountstephens, N. S. Suhaimi, J. Teo

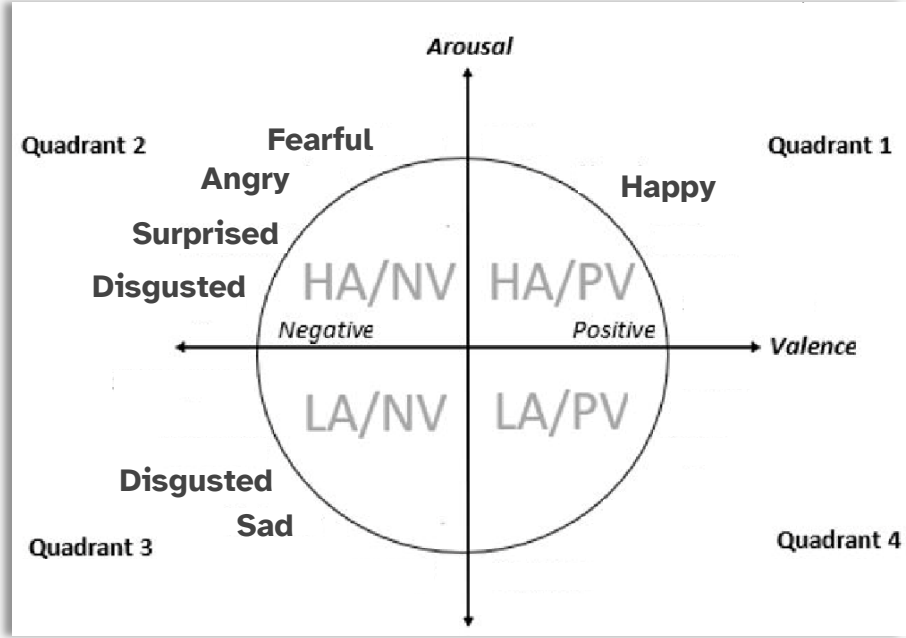
- Experiment using VR and ML
- **Four-class** emotion classification

Used for:

Forming a common ground



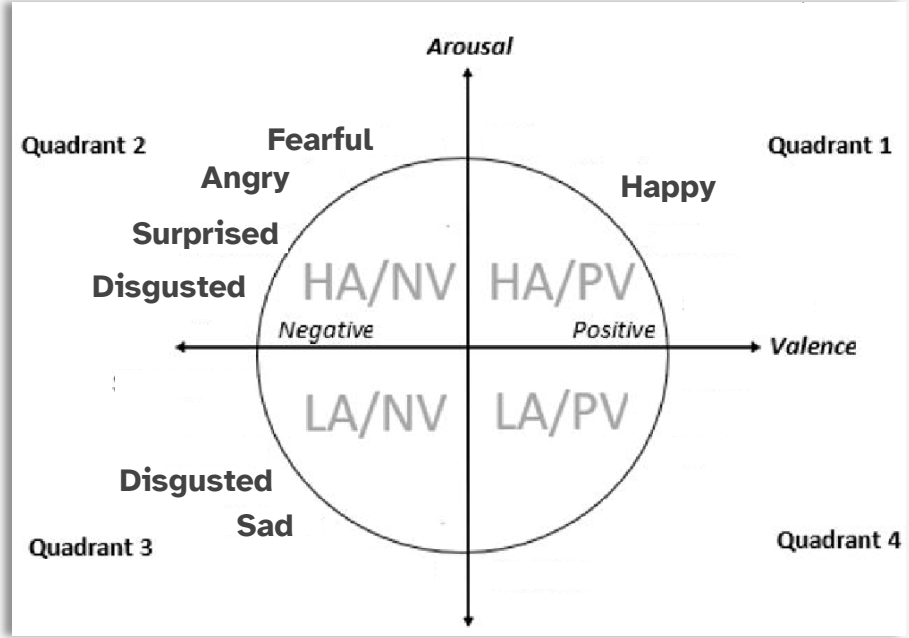
Quadrants based categorization



Quadrants based categorization

Additional Class:

Neutral (N)



Images Preparation

How were images prepared for training?

Images preparation

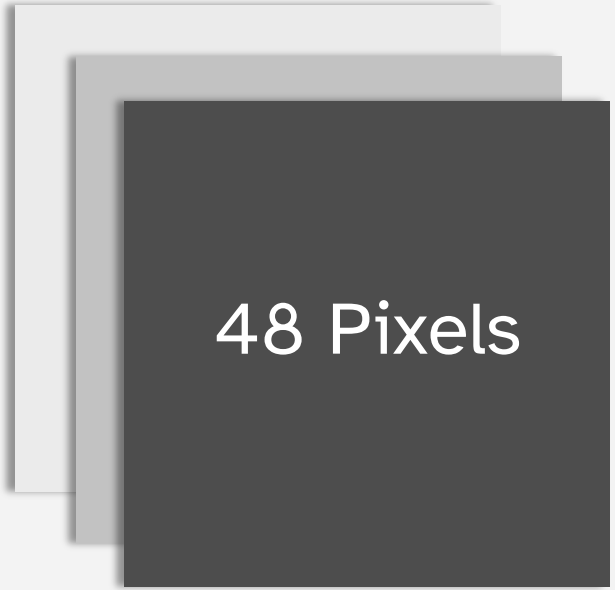
Loaded images to 48x48x3 NumPy arrays

One-Hot encoded labels

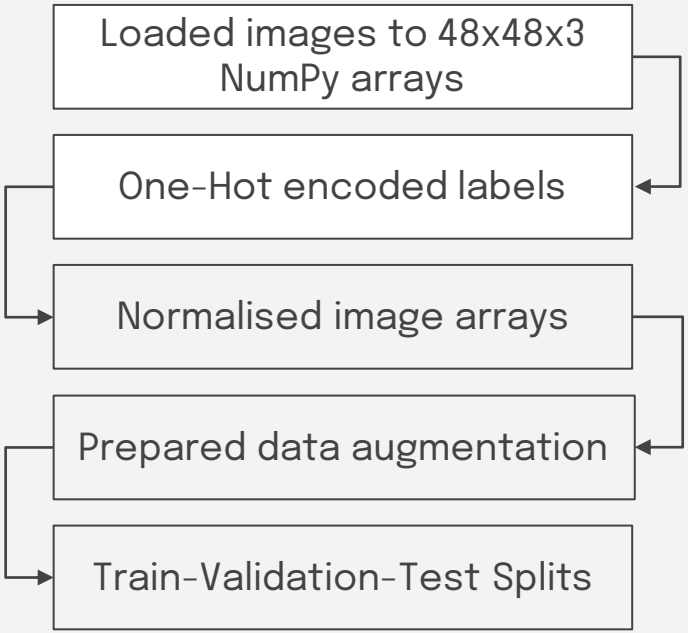
Normalised image arrays

Prepared data augmentation

Train-Validation-Test Splits

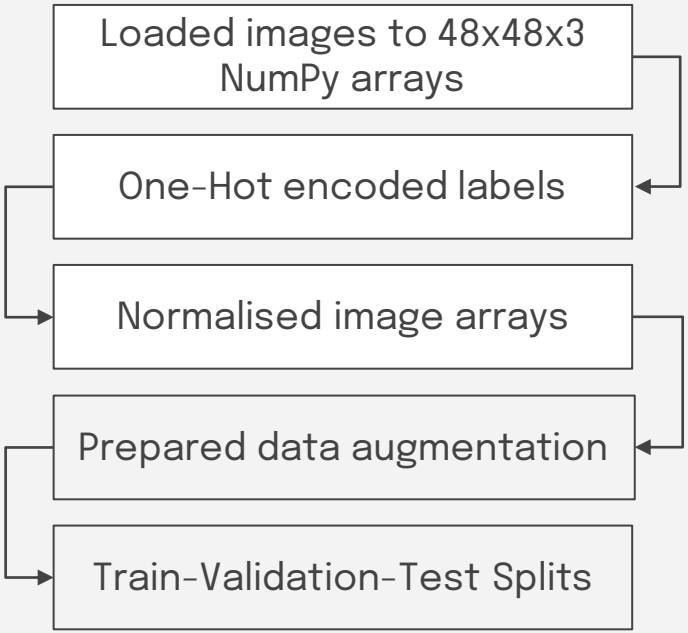


Images preparation



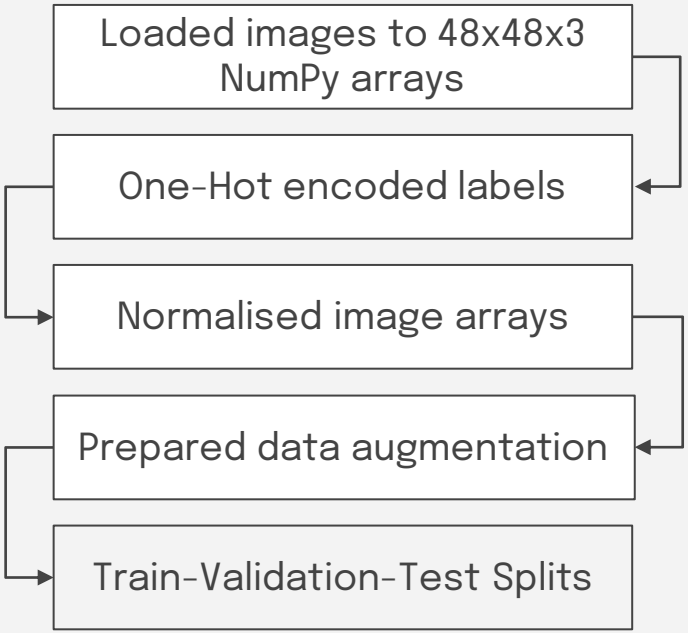
- **Q1:** (1, 0, 0, 0)
- **Q2:** (0, 1, 0, 0)
- **Q3:** (0, 0, 1, 0)
- **N:** (0, 0, 0, 1)

Images preparation



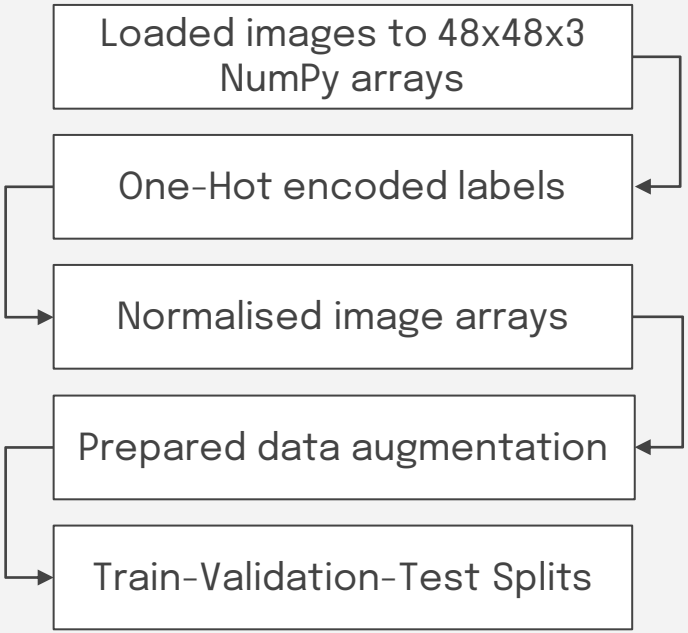
- Devided all pixels **by 255**
- Preserved **similar distributions**
- Enhanced **generalization**

Images preparation



- Used **ImageDataGenerator**
- Transformations:
  - **Rotation**
  - **Flip**
  - **Zoom**

Images preparation



Train Set: 70%

Validation Set: 20%

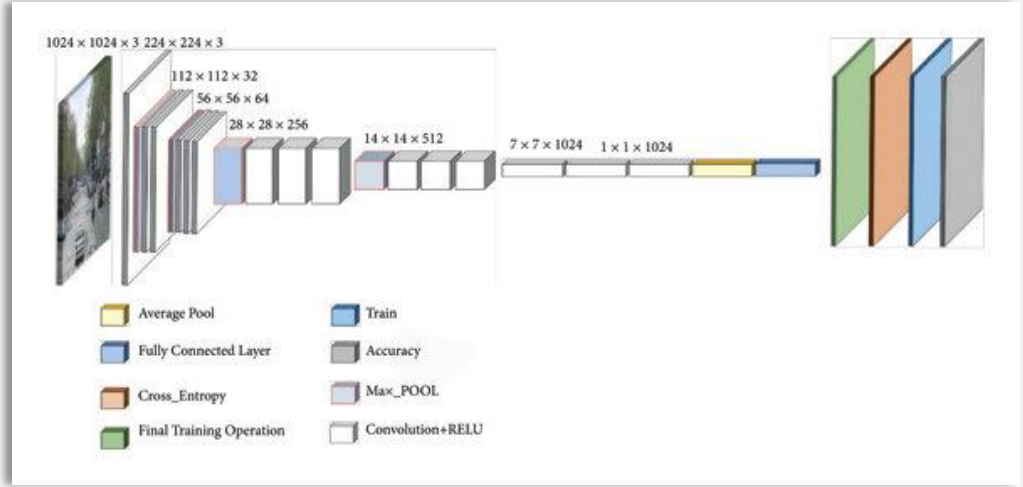
Test Set: 10%

Model Architecture

What is the architecture of the model used for Transfer Learning?

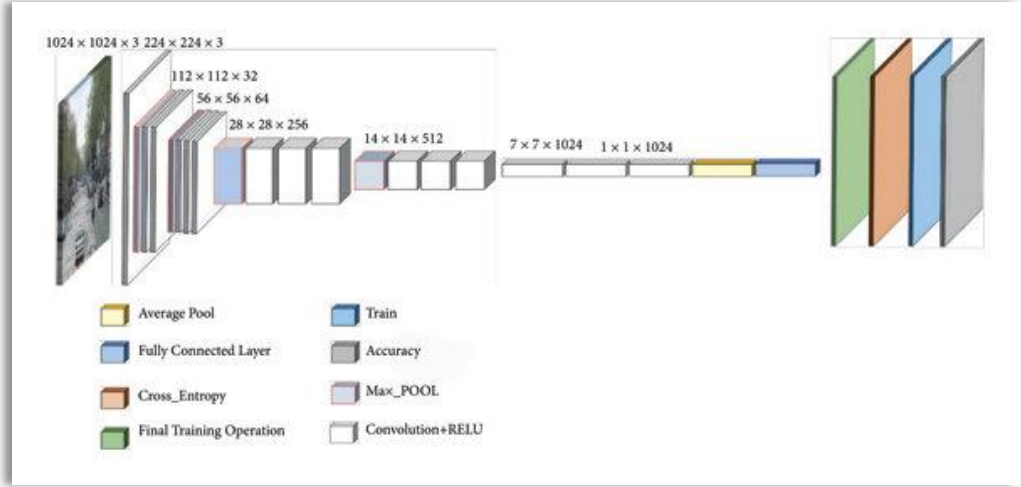
Base Model | **MobileNet**

- **Lightweight** deep **convolutional** neural network
- Designed for **mobile devices**
- **224 × 224 × 3** input
- **Fully-connected** at the top



Base Model | **MobileNet**

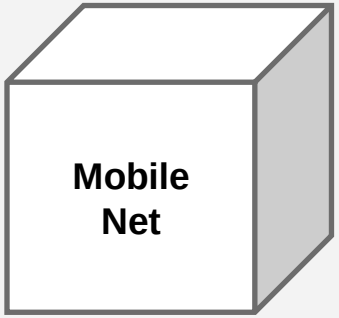
- **Lightweight** deep **convolutional** neural network
- Designed for **mobile devices**
- **224 × 224 × 3** input
- **Fully-connected** at the top



Also tried:

VGG16, DenseNet169, DenseNet201, ResNet50

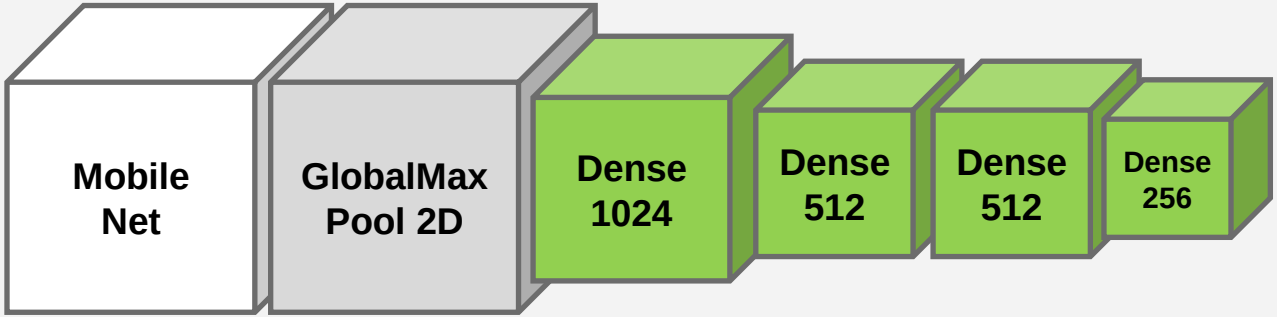
Final Model



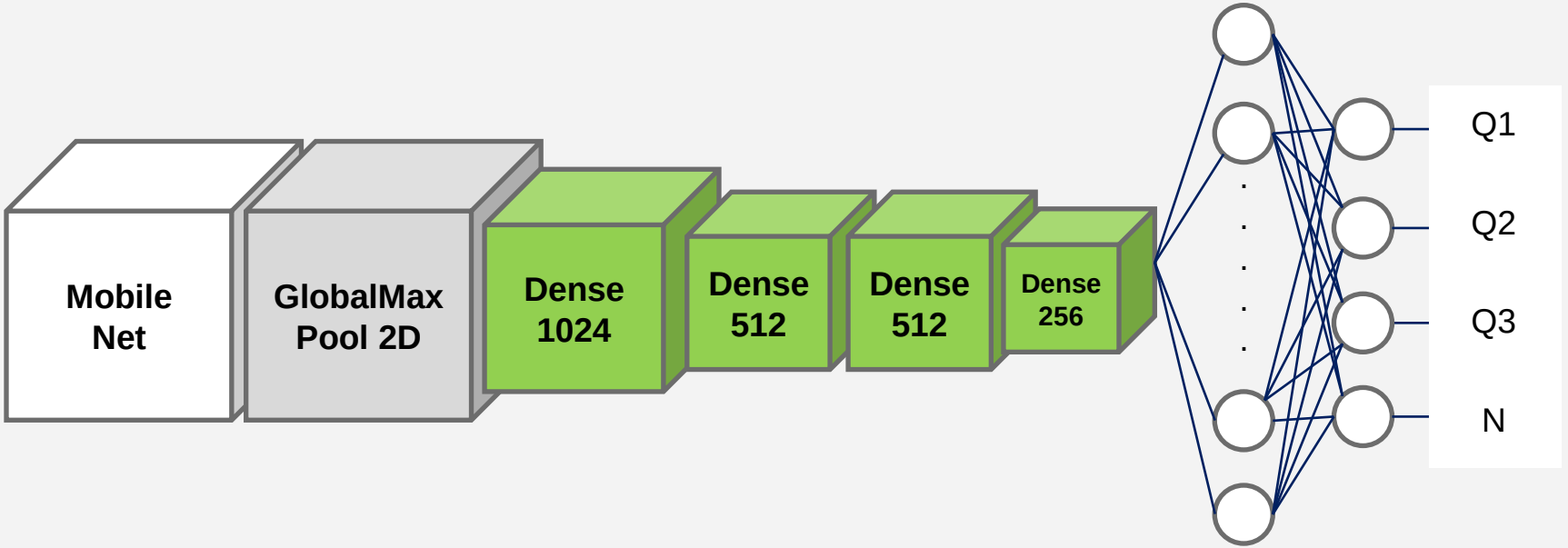
Mention:

Full-connected output was cut out

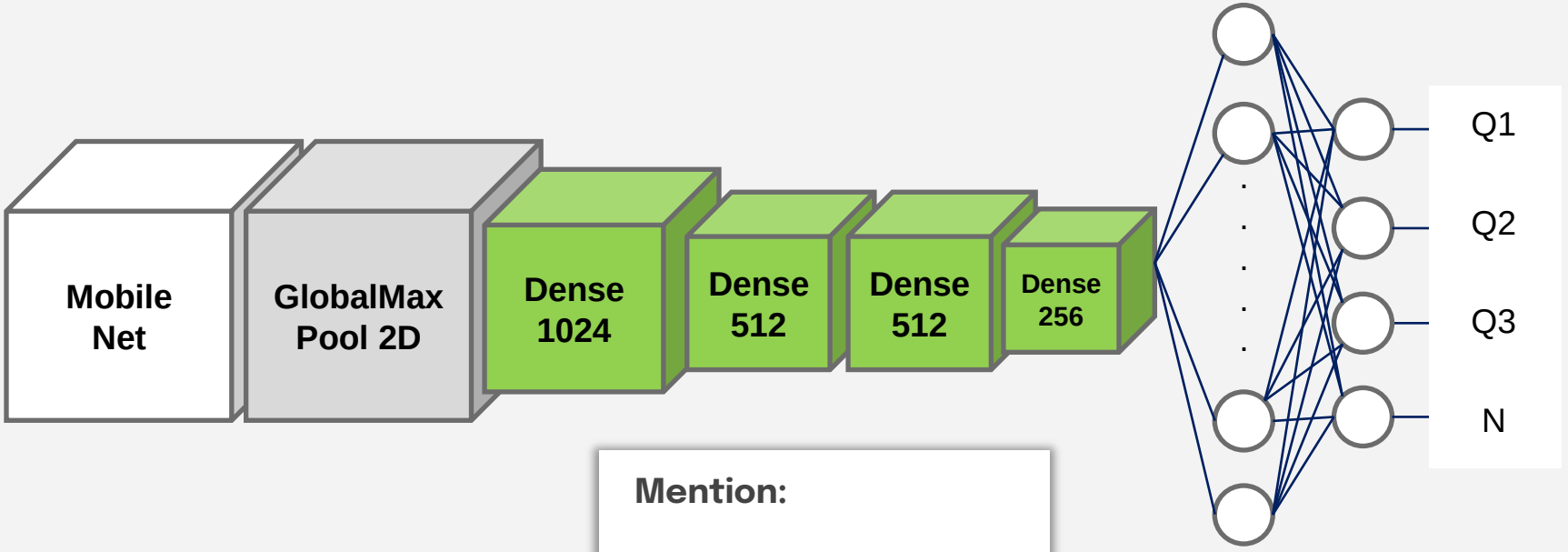
Final Model


Mobile Net
GlobalMax Pool 2D
Dense 1024
Dense 512
Dense 512
Dense 256

Final Model


Mobile Net
GlobalMax Pool 2D
Dense 1024
Dense 512
Dense 512
Dense 256
Q1
Q2
Q3
N

Final Model



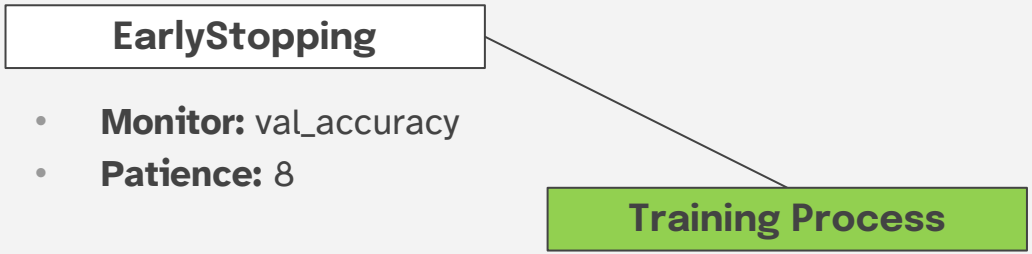
Mention:

First 13 layers were frozen

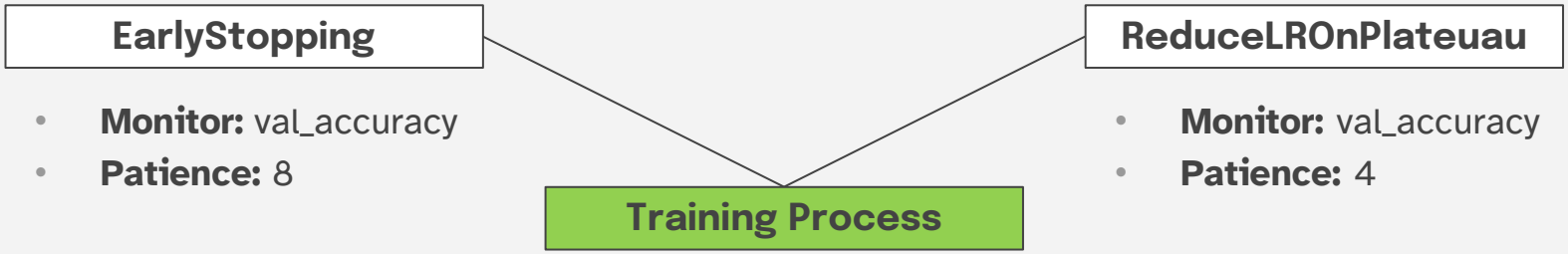
Model Training

How was the model trained ?

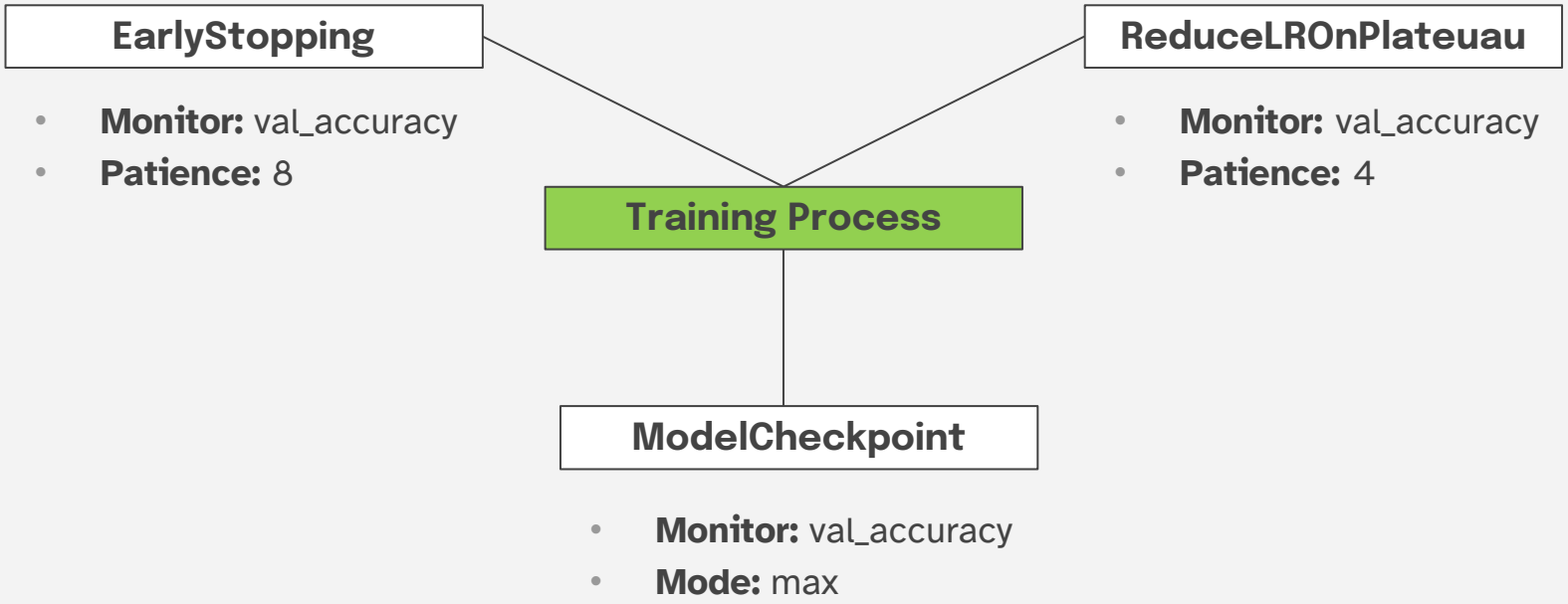
Callbacks & Hyperparameters



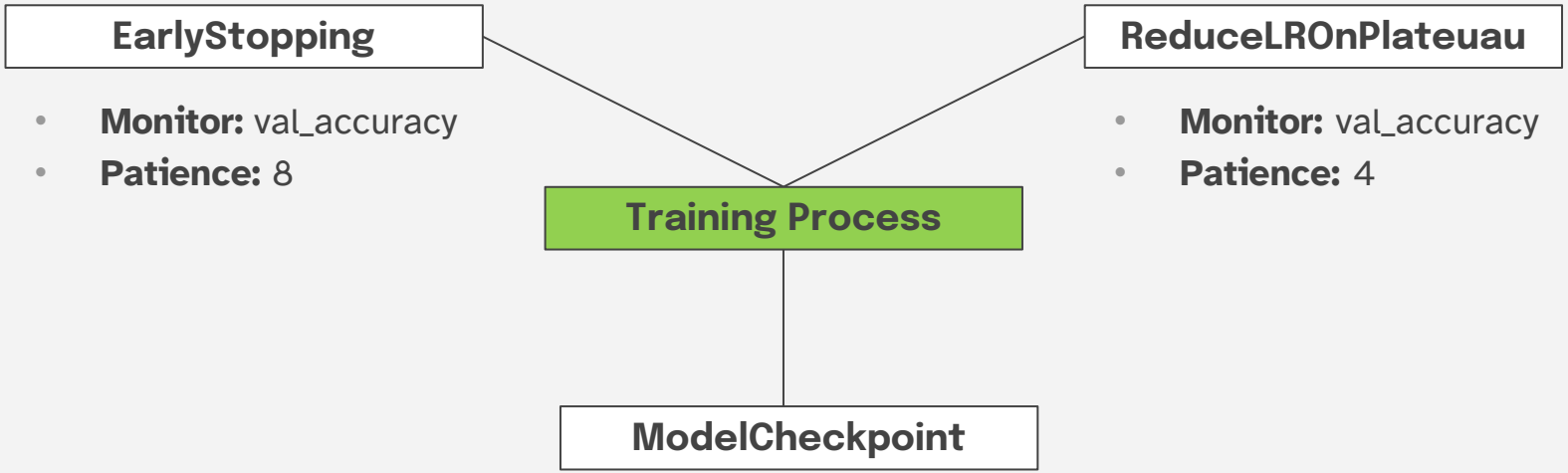
Callbacks & Hyperparameters



Callbacks & Hyperparameters



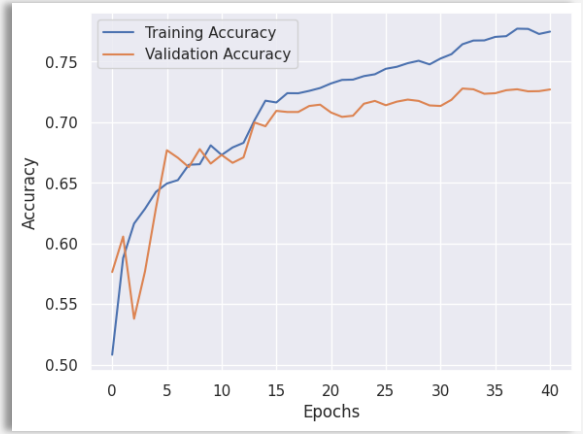
Callbacks & Hyperparameters

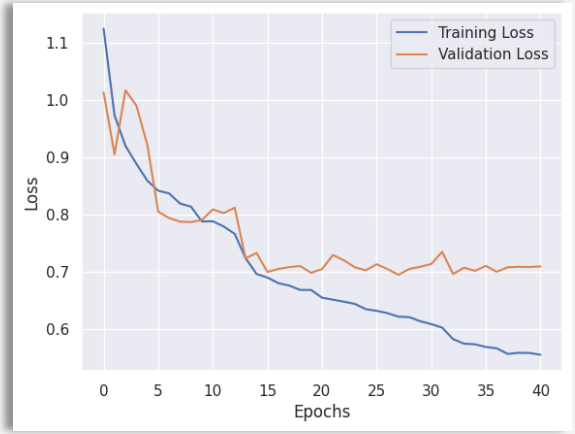


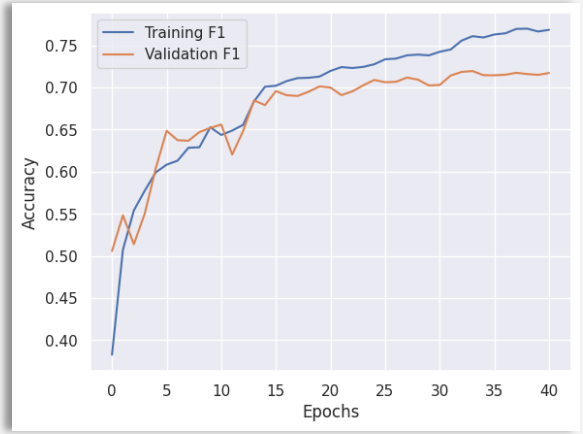
Hyperparameters:

Batch Size: 64, **Epochs**: 80, **LR**: 0,001

Training Results





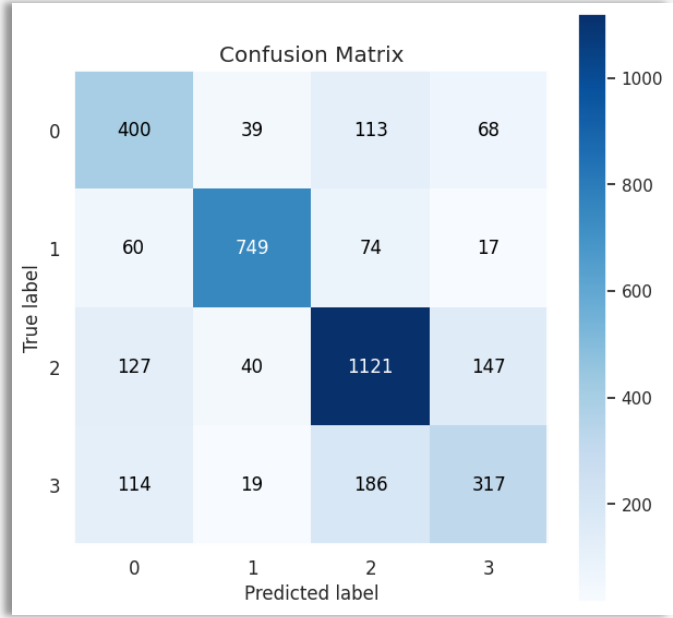


Model Evaluation

What is the performance of the model on test data ?

Testing Results

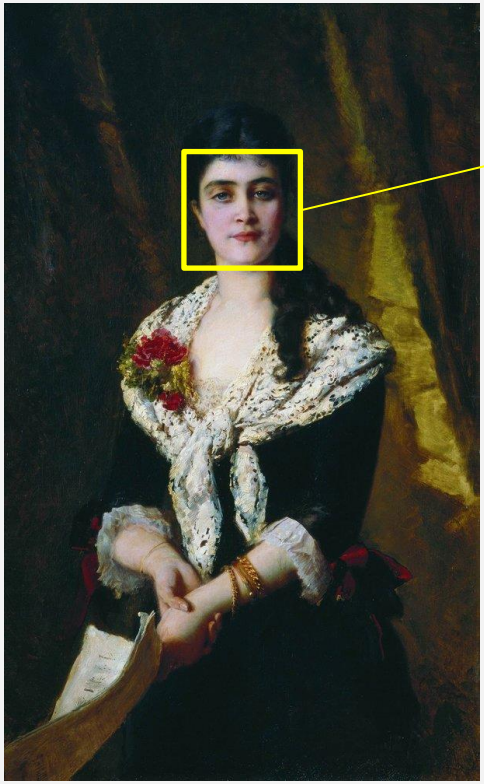
	precision	recall	f1-score	support
0	0.57	0.65	0.61	620
1	0.88	0.83	0.86	900
2	0.75	0.78	0.77	1435
3	0.58	0.50	0.54	636
accuracy			0.72	3591
macro avg	0.70	0.69	0.69	3591
weighted avg	0.72	0.72	0.72	3591



Sentiment Analysis on portraits

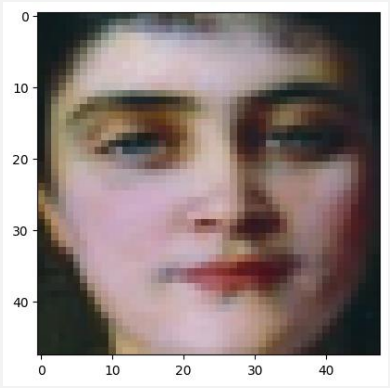
How is sentiment extracted from portraits ?

Sentiment Analysis on Portraits



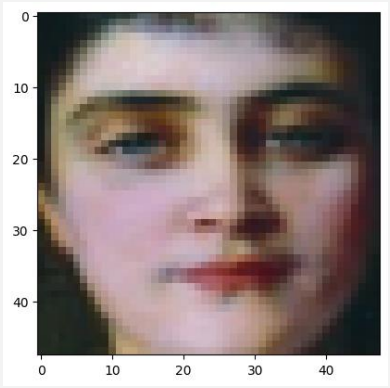
- Use of **MTCNN** to detect face
- Crops **48 x 48** pixels face

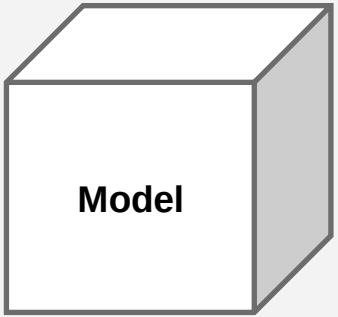
Sentiment Analysis on Portraits



Crop is **normalized**

Sentiment Analysis on Portraits





Prediction is made

['Q1']

Thank you.

Stefanos Vlachos
dit2202dsc

Efstathios Zaragkas
dit2203dsc