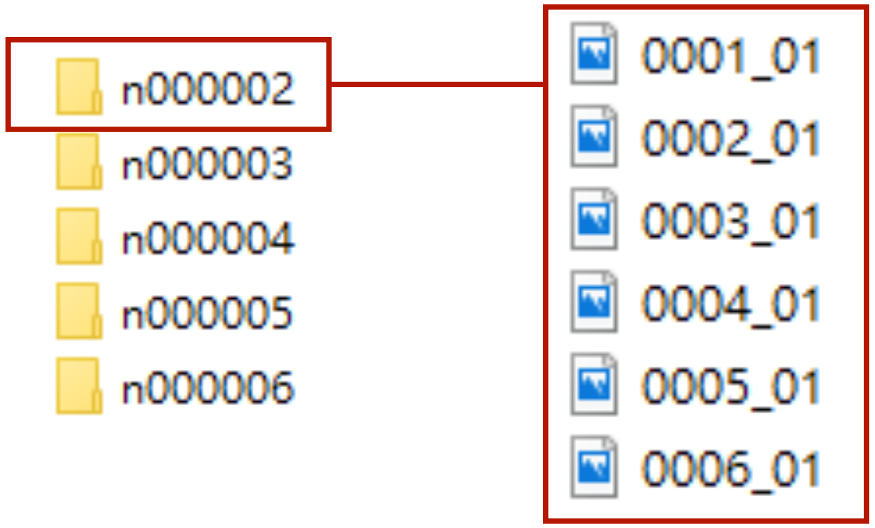
Monograph - Facenet

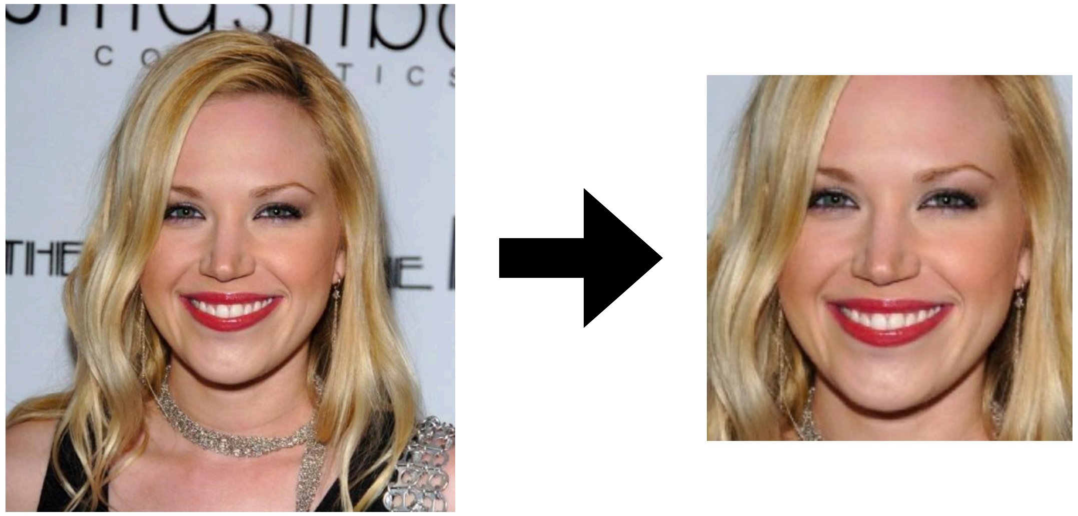
林怡瑄

Data Preprocess

- ~3.3 M face and ~9000 class
- a person has ~367 photos



- Resize to 224x224
- Made the photo close to face



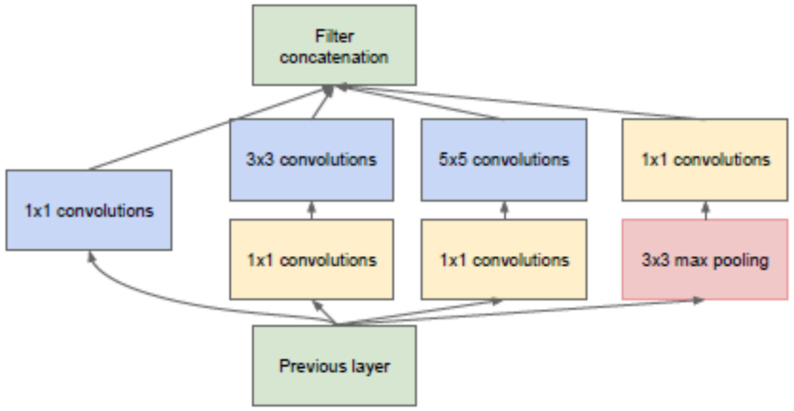
Problem

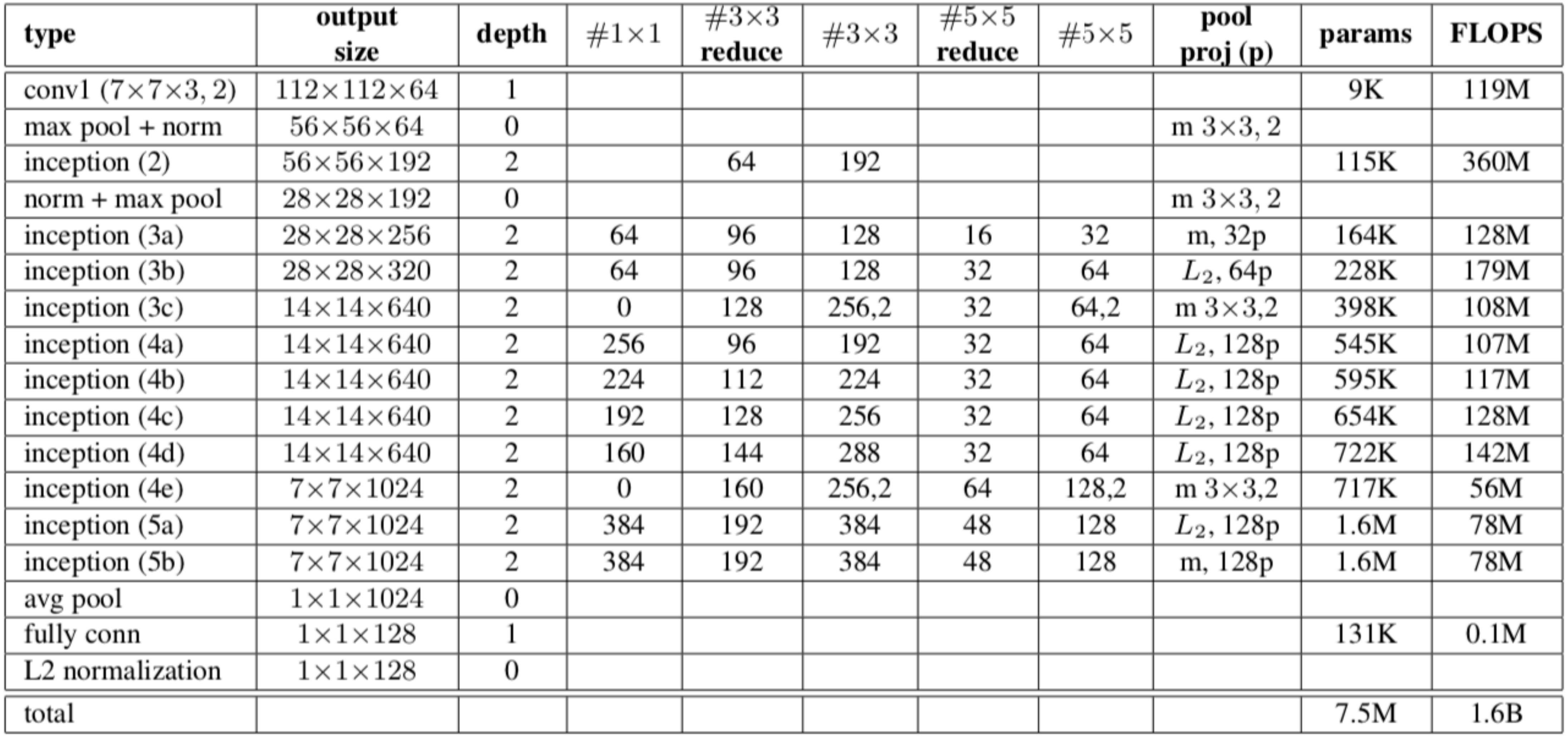
1. **The number of photo is to much**
   - **Data put in 186 mount to 180**
   - **Multiprocess**
2. **Face detect**

Build Model

I build a model as same as paper first.

- **lambda**
- **Inception**
- **Layer name can’t contain 1x1**





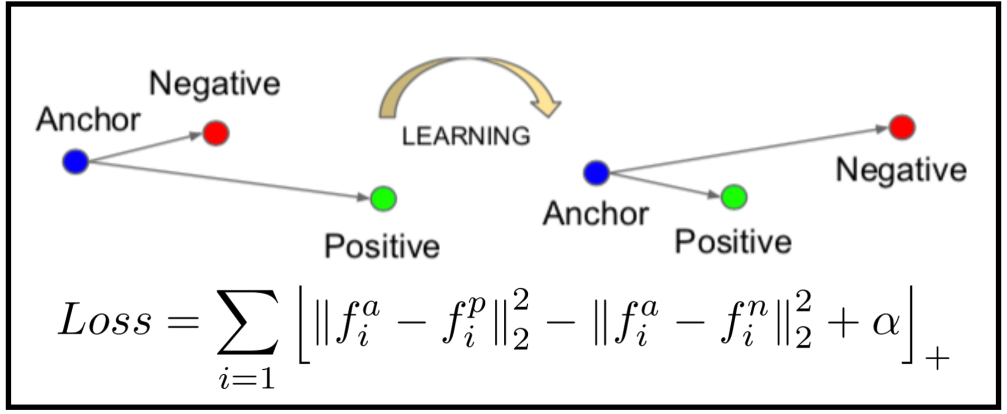
type	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj (p)	params	FLOPS
conv1 (7×7×3, 2)	112×112×64	1							9K	119M
max pool + norm	56×56×64	0						m 3×3, 2		
inception (2)	56×56×192	2		64	192				115K	360M
norm + max pool	28×28×192	0						m 3×3, 2		
inception (3a)	28×28×256	2	64	96	128	16	32	m, 32p	164K	128M
inception (3b)	28×28×320	2	64	96	128	32	64	L_2, 64p	228K	179M
inception (3c)	14×14×640	2	0	128	256,2	32	64,2	m 3×3,2	398K	108M
inception (4a)	14×14×640	2	256	96	192	32	64	L_2, 128p	545K	107M
inception (4b)	14×14×640	2	224	112	224	32	64	L_2, 128p	595K	117M
inception (4c)	14×14×640	2	192	128	256	32	64	L_2, 128p	654K	128M
inception (4d)	14×14×640	2	160	144	288	32	64	L_2, 128p	722K	142M
inception (4e)	7×7×1024	2	0	160	256,2	64	128,2	m 3×3,2	717K	56M
inception (5a)	7×7×1024	2	384	192	384	48	128	L_2, 128p	1.6M	78M
inception (5b)	7×7×1024	2	384	192	384	48	128	m, 128p	1.6M	78M
avg pool	1×1×1024	0								
fully conn	1×1×128	1							131K	0.1M
L2 normalization	1×1×128	0								
total									7.5M	1.6B

Loss function

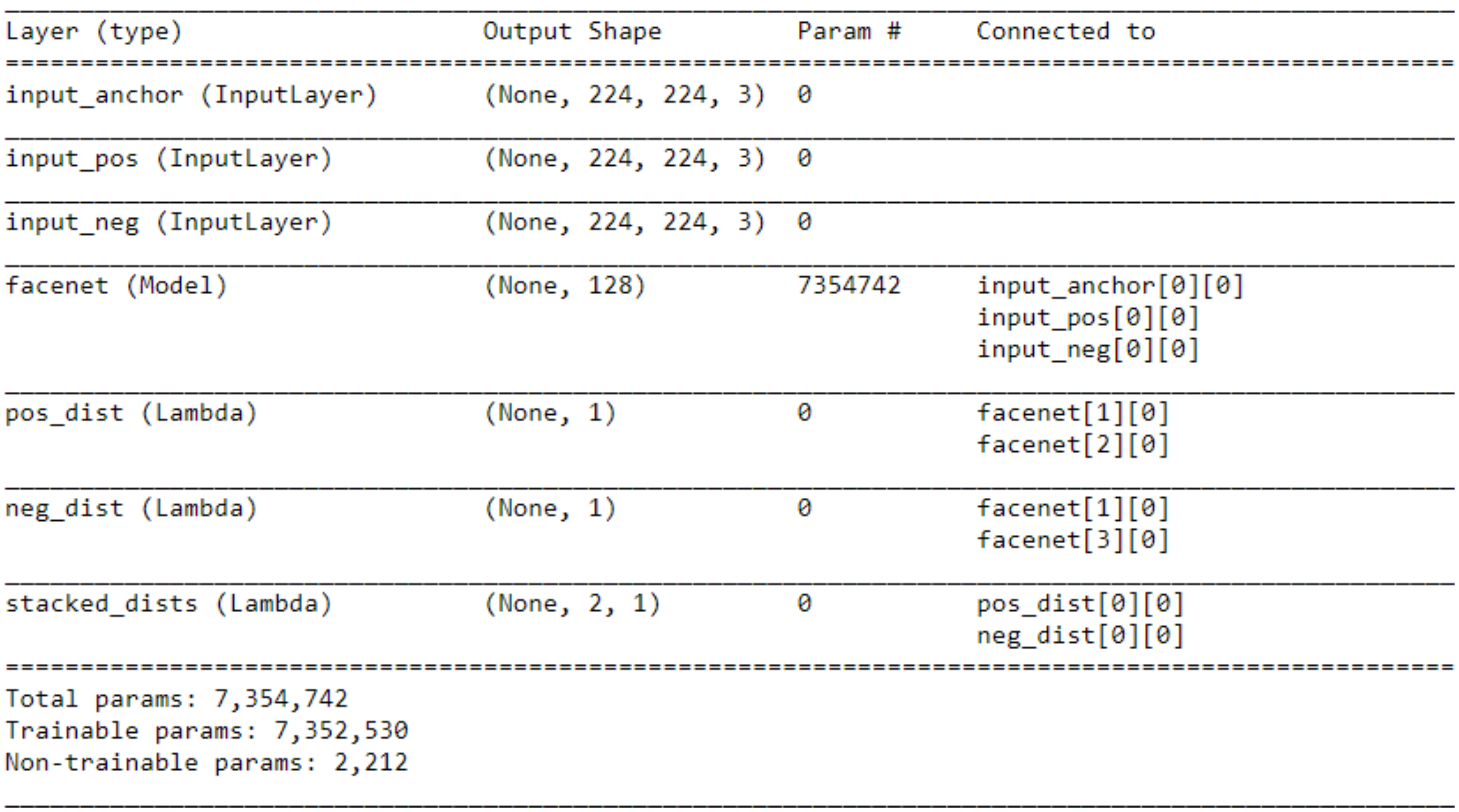
Because facet paper use Triplet loss .

- **Code loss myself**
- **Reuse model**
- **Size error**

Triplet loss :



$$Loss = \sum_{i=1} \left\lfloor \|f_i^a - f_i^p\|_2^2 - \|f_i^a - f_i^n\|_2^2 + \alpha \right\rfloor_+$$



```
____________________________________________________________________________________________________
Layer (type)                     Output Shape          Param #     Connected to
====================================================================================================
input_anchor (InputLayer)        (None, 224, 224, 3)   0
____________________________________________________________________________________________________
input_pos (InputLayer)           (None, 224, 224, 3)   0
____________________________________________________________________________________________________
input_neg (InputLayer)           (None, 224, 224, 3)   0
____________________________________________________________________________________________________
facenet (Model)                  (None, 128)           7354742     input_anchor[0][0]
                                                                   input_pos[0][0]
                                                                   input_neg[0][0]
____________________________________________________________________________________________________
pos_dist (Lambda)                (None, 1)             0           facenet[1][0]
                                                                   facenet[2][0]
____________________________________________________________________________________________________
neg_dist (Lambda)                (None, 1)             0           facenet[1][0]
                                                                   facenet[3][0]
____________________________________________________________________________________________________
stacked_dists (Lambda)           (None, 2, 1)          0           pos_dist[0][0]
                                                                   neg_dist[0][0]
====================================================================================================
Total params: 7,354,742
Trainable params: 7,352,530
Non-trainable params: 2,212
____________________________________________________________________________________________________
```

Final model

Fine tune

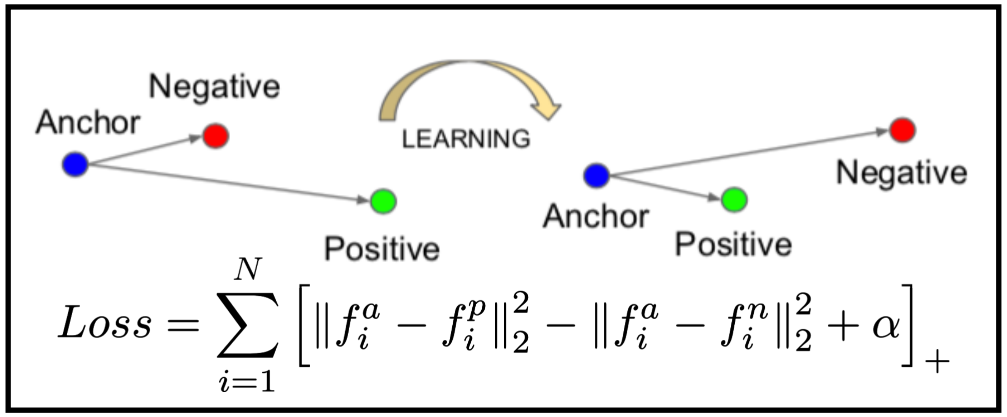
Small dataset
100 person ~ 36600 photos

- **Problem 1 :**

Loss par set a = 1

↓

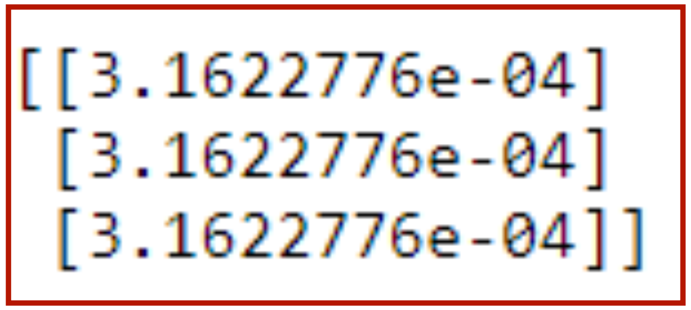
Loss = 1, acc = 0



$$Loss = \sum_{i=1}^{N} \left[\|f_i^a - f_i^p\|_2^2 - \|f_i^a - f_i^n\|_2^2 + \alpha \right]_+$$

$\|f_i^a - f_i^p\|_2^2$ →

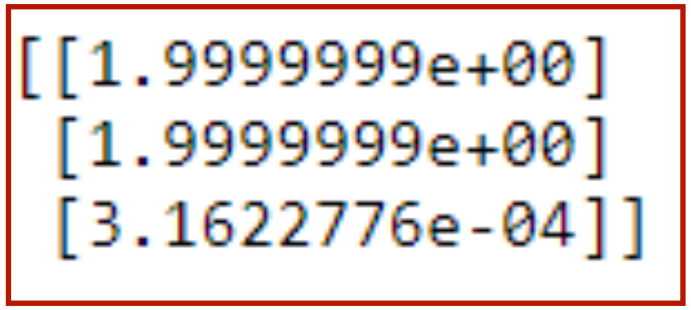
$\|f_i^a - f_i^n\|_2^2$ →



```
[[3.1622776e-04]
 [3.1622776e-04]
 [3.1622776e-04]]
```

$\|f_i^a - f_i^p\|_2^2$ →

$\|f_i^a - f_i^n\|_2^2$ →



```
[[1.9999999e+00]
 [1.9999999e+00]
 [3.1622776e-04]]
```

→

1. Smaller batch size
2. Smaller learning rate
3. Bigger margin ✓

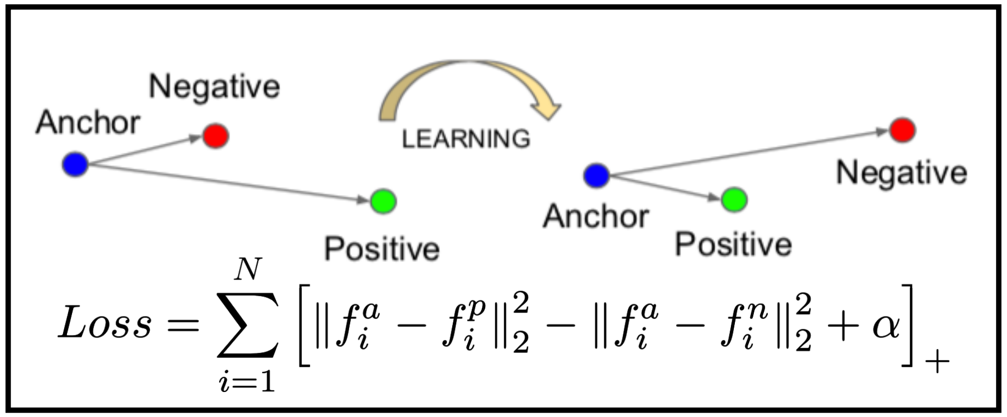
Fine tune

Small dataset
100 person ~ 36600 photos

- **Problem 2 :**

Loss par set a = 4

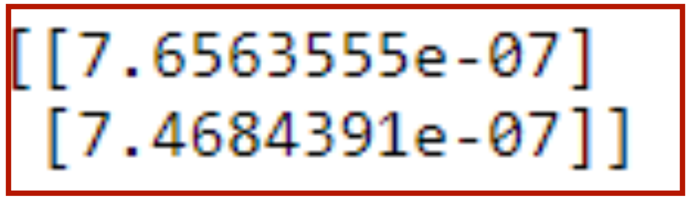
Loss = 4, acc ~ 0.5



$$Loss = \sum_{i=1}^{N} \left[\|f_i^a - f_i^p\|_2^2 - \|f_i^a - f_i^n\|_2^2 + \alpha \right]_+$$

$\|f_i^a - f_i^p\|_2^2 \longrightarrow$

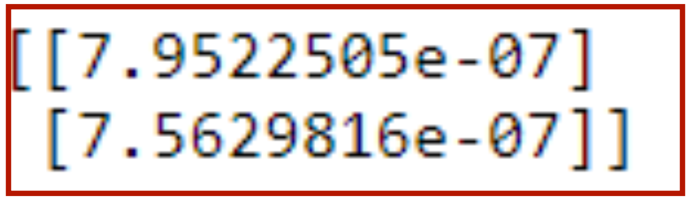
$\|f_i^a - f_i^n\|_2^2 \longrightarrow$



```
[[7.6563555e-07]
 [7.4684391e-07]]
```

$\|f_i^a - f_i^p\|_2^2 \longrightarrow$

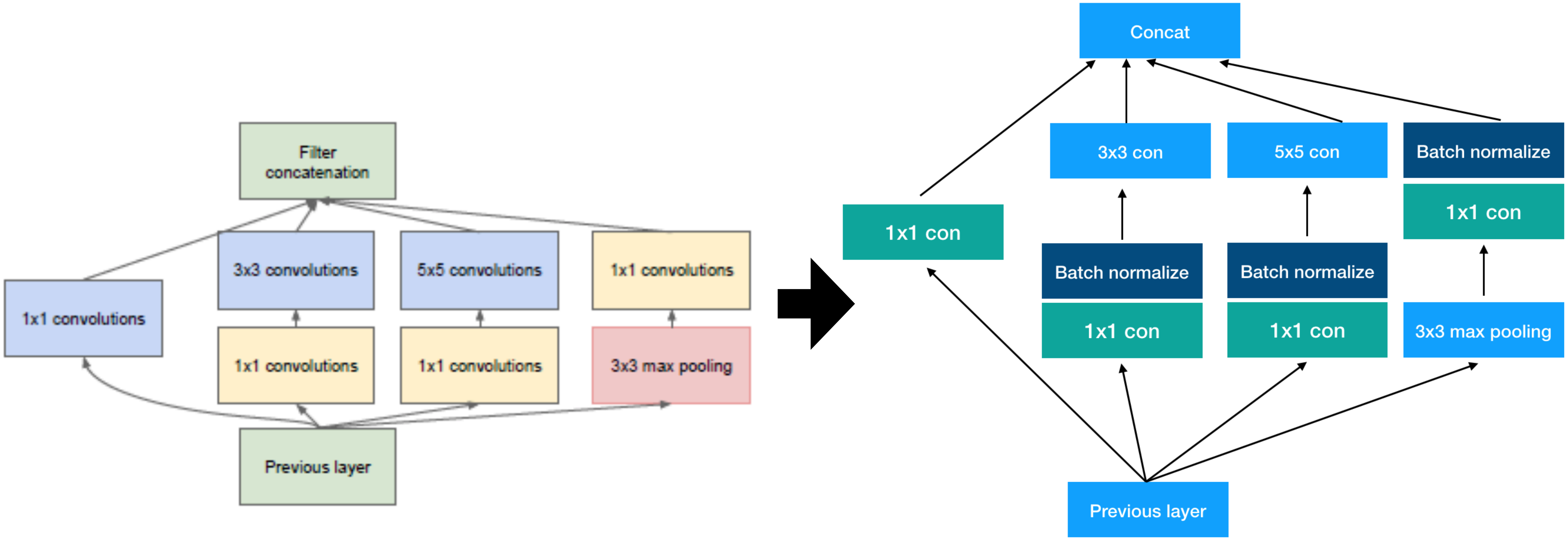
$\|f_i^a - f_i^n\|_2^2 \longrightarrow$



```
[[7.9522505e-07]
 [7.5629816e-07]]
```

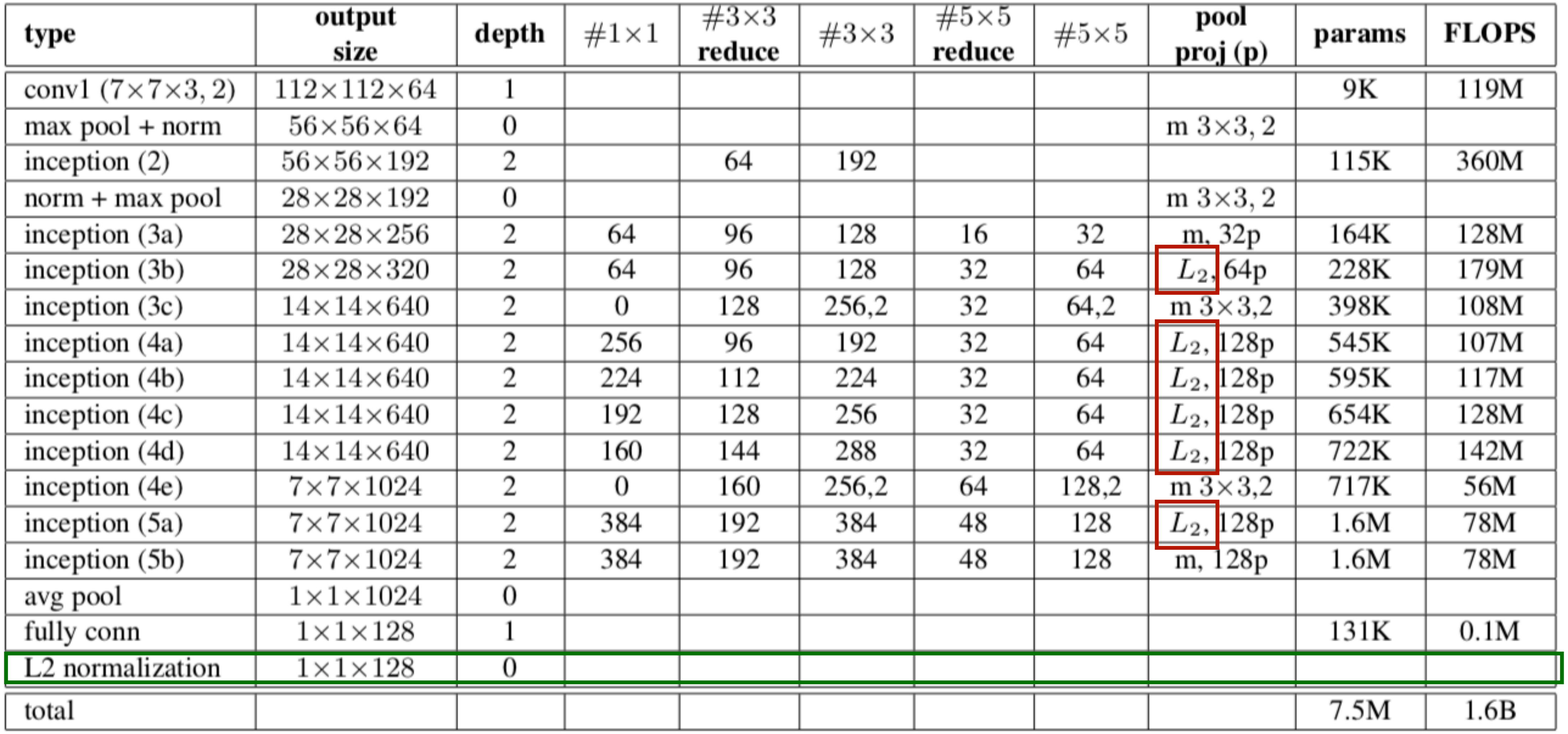
1. Batch normalize
2. Replace L2 nor as max pooling
3. Remove last L2 layer

Inception



Inception modify

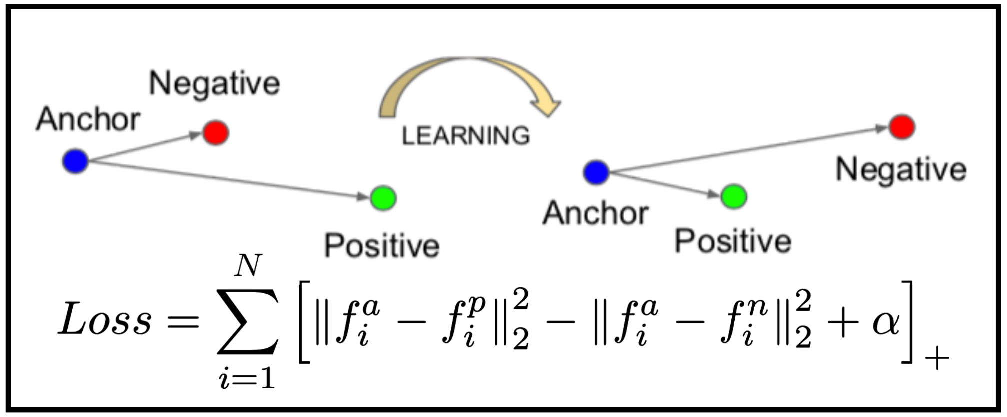
Facenet model



type	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj (p)	params	FLOPS
conv1 (7×7×3, 2)	112×112×64	1							9K	119M
max pool + norm	56×56×64	0						m 3×3, 2		
inception (2)	56×56×192	2		64	192				115K	360M
norm + max pool	28×28×192	0						m 3×3, 2		
inception (3a)	28×28×256	2	64	96	128	16	32	m, 32p	164K	128M
inception (3b)	28×28×320	2	64	96	128	32	64	L_2, 64p	228K	179M
inception (3c)	14×14×640	2	0	128	256,2	32	64,2	m 3×3,2	398K	108M
inception (4a)	14×14×640	2	256	96	192	32	64	L_2, 128p	545K	107M
inception (4b)	14×14×640	2	224	112	224	32	64	L_2, 128p	595K	117M
inception (4c)	14×14×640	2	192	128	256	32	64	L_2, 128p	654K	128M
inception (4d)	14×14×640	2	160	144	288	32	64	L_2, 128p	722K	142M
inception (4e)	7×7×1024	2	0	160	256,2	64	128,2	m 3×3,2	717K	56M
inception (5a)	7×7×1024	2	384	192	384	48	128	L_2, 128p	1.6M	78M
inception (5b)	7×7×1024	2	384	192	384	48	128	m, 128p	1.6M	78M
avg pool	1×1×1024	0								
fully conn	1×1×128	1							131K	0.1M
L2 normalization	1×1×128	0								
total									7.5M	1.6B

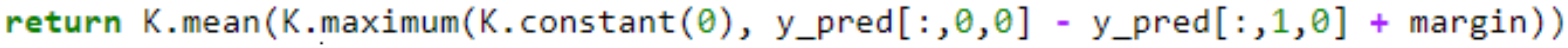
Facenet model

Fine tune



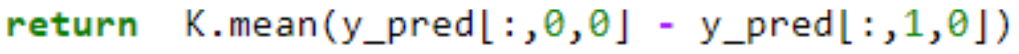
$$Loss = \sum_{i=1}^{N}\left[\|f_i^a - f_i^p\|_2^2 - \|f_i^a - f_i^n\|_2^2 + \alpha\right]_+$$

- **Different loss test**

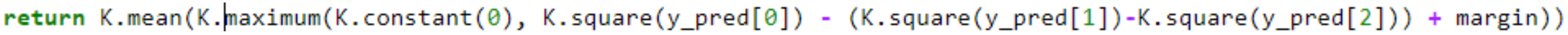


```
return K.mean(K.maximum(K.constant(0), y_pred[:,0,0] - y_pred[:,1,0] + margin))
```

✓

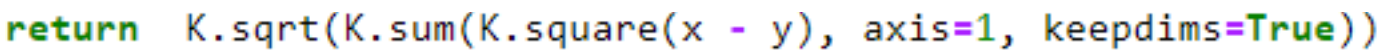


```
return  K.mean(y_pred[:,0,0] - y_pred[:,1,0])
```



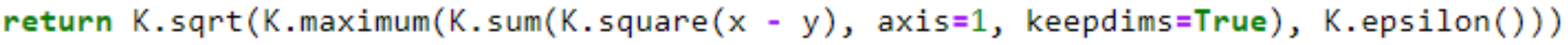
```
return K.mean(K.maximum(K.constant(0), K.square(y_pred[0]) - (K.square(y_pred[1])-K.square(y_pred[2])) + margin))
```

- **Different distance calculate**



```
return  K.sqrt(K.sum(K.square(x - y), axis=1, keepdims=True))
```

✓



```
return K.sqrt(K.maximum(K.sum(K.square(x - y), axis=1, keepdims=True), K.epsilon()))
```

Parameter Setting

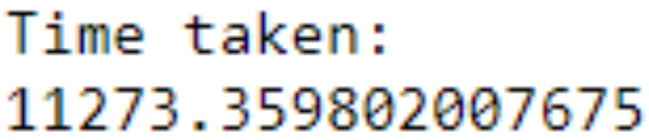
Data :

- **Person : 100**
- **Photo num : 33600**

Train setting :

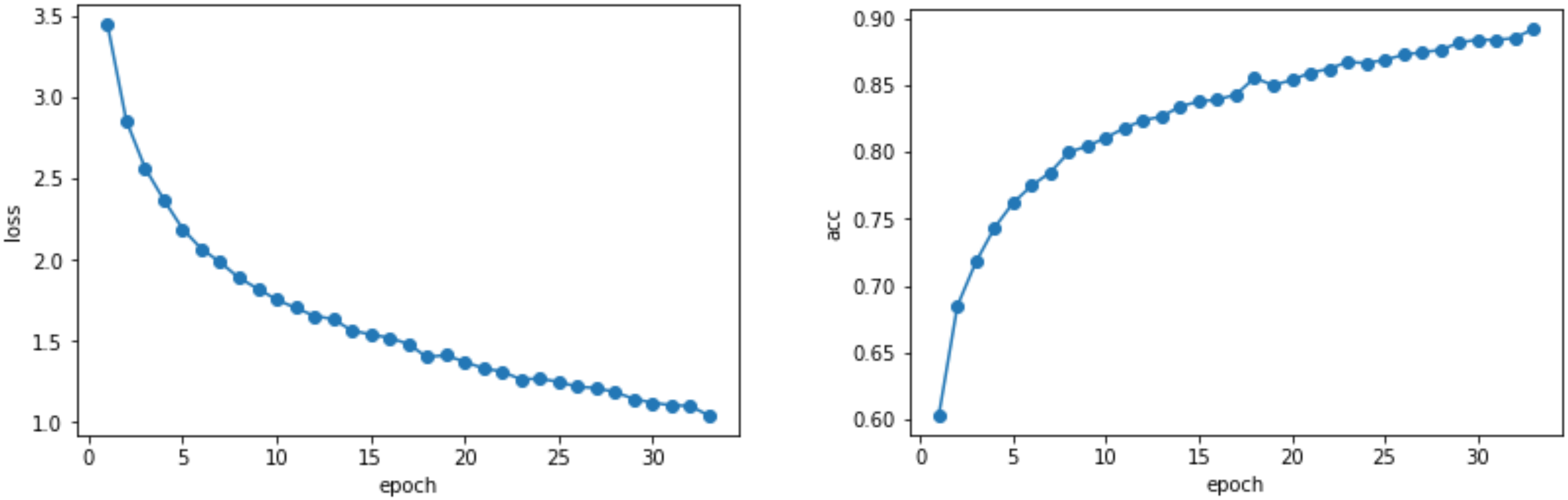
- **Trunk : 1000 photo**
- **Batchsize : 100**
- **Epoch : 100**
- **A epoch need 33 trunk**

Time taken a epoch :



```
Time taken:
11273.359802007675
```

Result



Loss

Acc

Training result per epoch

Parameter Setting

Data :

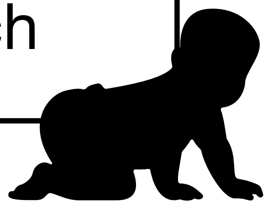
- **Person : 4775**
- **Photo num : 1747650**

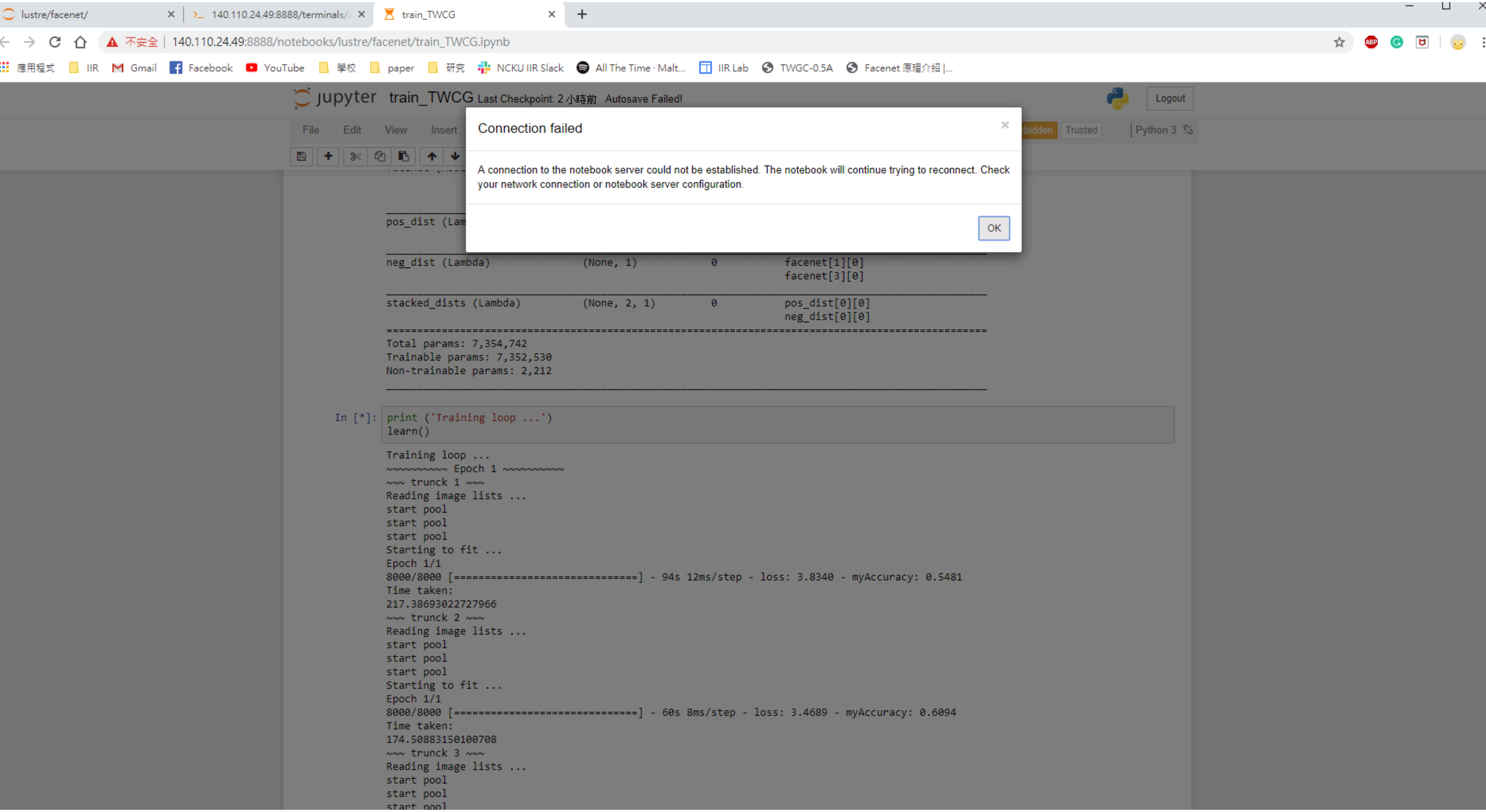
Train setting :

- **Trunk : 8000 photo**
- **Batchsize : 80**
- **Epoch : 100**
- **A epoch need 6 trunk**

Time taken a epoch :

A trunk need 1206 s
~7.37 hr per epoch





Connection failed

A connection to the notebook server could not be established. The notebook will continue trying to reconnect. Check your network connection or notebook server configuration.

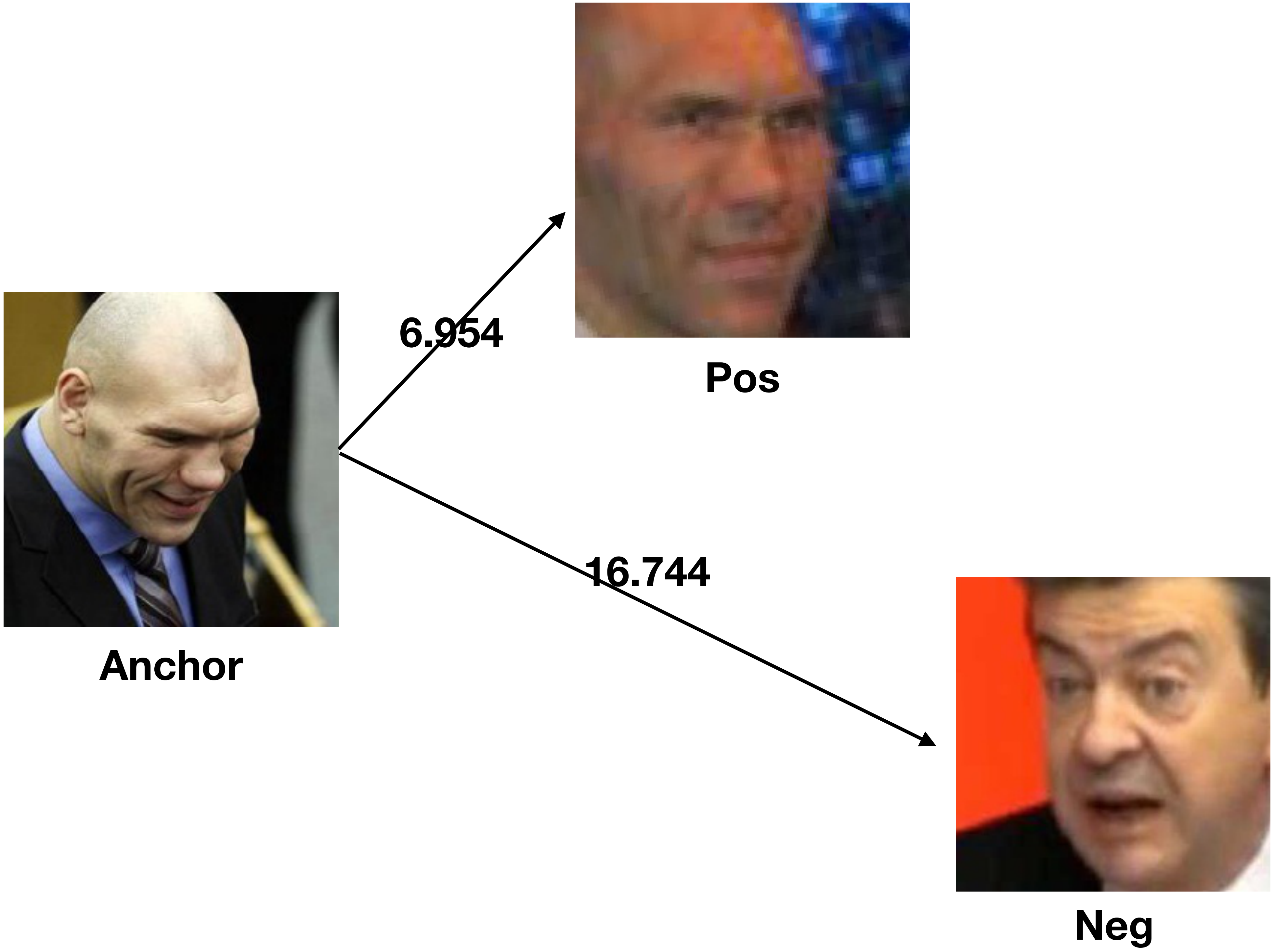
```
____________________________________________________________________________________________________
pos_dist (Lam

____________________________________________________________________________________________________
neg_dist (Lambda)               (None, 1)            0           facenet[1][0]
                                                                 facenet[3][0]
____________________________________________________________________________________________________
stacked_dists (Lambda)          (None, 2, 1)         0           pos_dist[0][0]
                                                                 neg_dist[0][0]
====================================================================================================
Total params: 7,354,742
Trainable params: 7,352,530
Non-trainable params: 2,212
____________________________________________________________________________________________________
```

```
In [*]: print ('Training loop ...')
        learn()
```

```
Training loop ...
~~~~~~~~~~ Epoch 1 ~~~~~~~~~~
~~~ trunck 1 ~~~
Reading image lists ...
start pool
start pool
start pool
Starting to fit ...
Epoch 1/1
8000/8000 [==============================] - 94s 12ms/step - loss: 3.8340 - myAccuracy: 0.5481
Time taken:
217.38693022727966
~~~ trunck 2 ~~~
Reading image lists ...
start pool
start pool
start pool
Starting to fit ...
Epoch 1/1
8000/8000 [==============================] - 60s 8ms/step - loss: 3.4689 - myAccuracy: 0.6094
Time taken:
174.50883150100708
~~~ trunck 3 ~~~
Reading image lists ...
start pool
start pool
start pool
```

Case Study



6.954

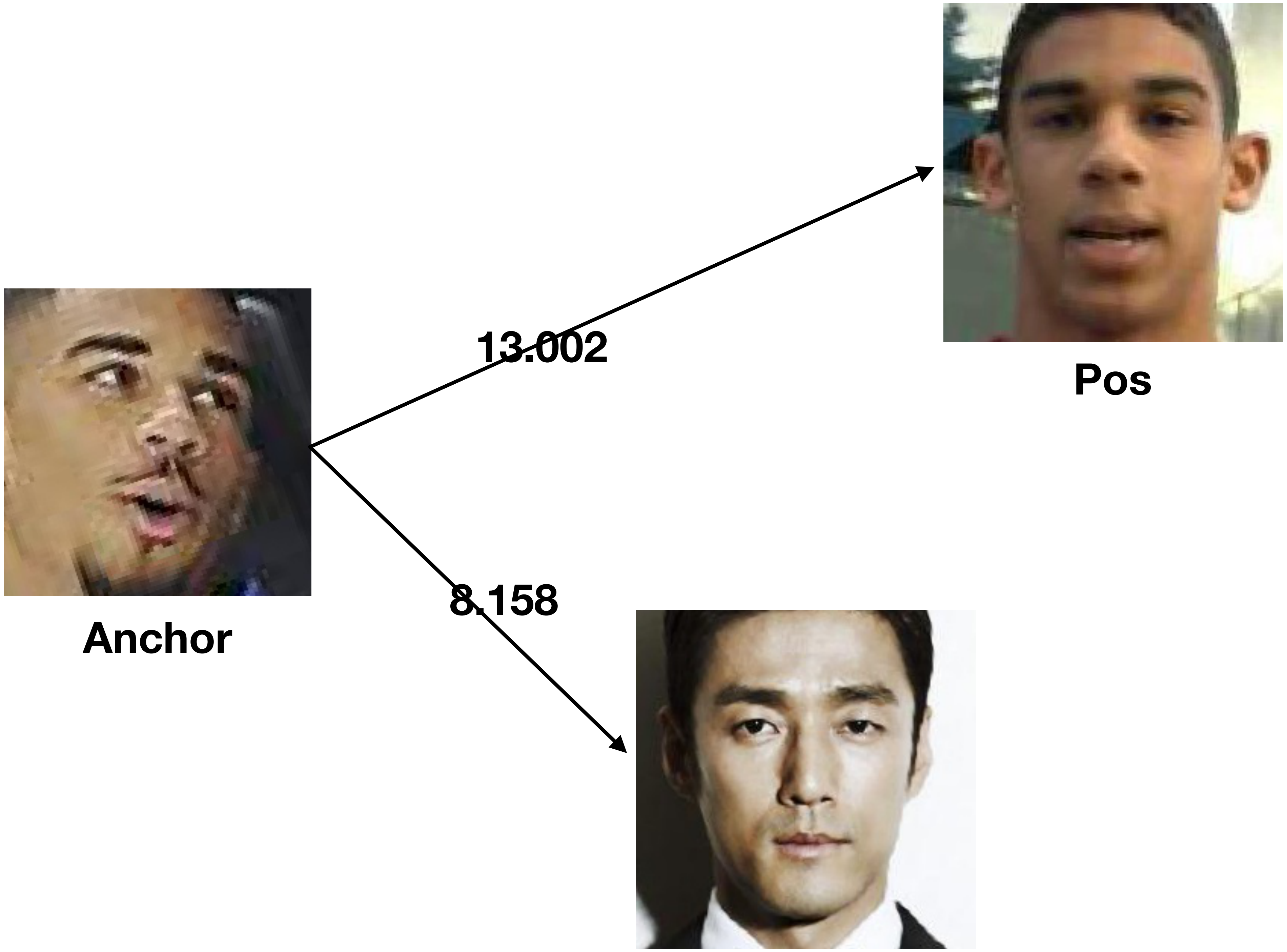
Pos

16.744

Anchor

Neg

Case Study



13.002

Pos

Anchor

8.158

Neg