

My photo search advice

Ivan Listopadov
Sergey Grozny
Alexander Zaytsev
Yurii Melnik
TA: Petr Sokerin



Problem statement



Boris Trigub

Dozens of IW-photos in One folder - <https://disk.yandex.ru/d/ejHBY7JaCcDWyw>

Яндекс Диск

IW-2024

Посмотреть и скачать с
Яндекс Диска



7

How to **find yourself** in dozens of photos?

What we achieved

1. **Created dataset** and labelled by hand via detector to find faces
2. Used **2 pre-trained models** from DeepFace as baseline
3. **Trained 2 models**, using transfer learning on our dataset
4. Got **better results** than benchmark models

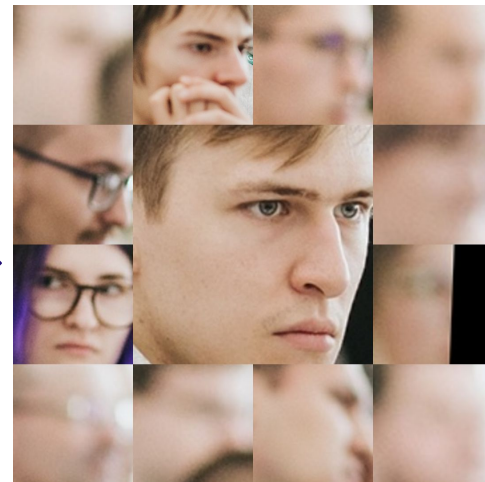
Data collection



Detection

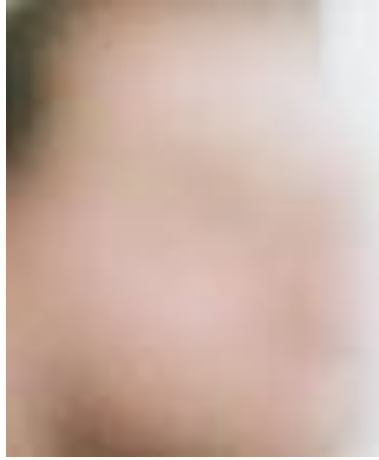


YOLOv8



Preprocessing: Deblurring

input:



$L = \text{Laplacian}(\text{image})$

If $L.\text{var}() < \text{threshold}$

return “blurred image”

Preprocessing: Deduplications

Original Image: IMG_3028_face_2.jpg



IMG_2913_face_2.jpg



IMG_2881_face_2.jpg



IMG_3011_1_face_2.jpg



IMG_2992_1_face_2.jpg



IMG_2939_1_face_2.jpg



IMG_2994_face_2.jpg



IMG_2657_face_2.jpg



IMG_2950_face_2.jpg



IMG_2953_face_2.jpg



IMG_2987_face_2.jpg



IMG_2985_face_1.jpg



IMG_2975_face_2.jpg



IMG_2861_face_2.jpg



IMG_2602_1_face_1.jpg



IMG_2970_face_2.jpg



IMG_0605_1_face_1.jpg



IMG_2825_face_1.jpg



IMG_2999_1_face_2.jpg



IMG_2971_face_2.jpg



IMG_3052_face_2.jpg



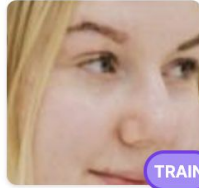
IMG_2847_face_2.jpg



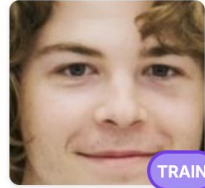
Data labelling



IMG_8791_face_1.jpg



IMG_8793_face_1.jpg



IMG_8799_face_2.j...



IMG_8882_1_face_1...



IMG_8882_face_1.jpg



IMG_8890_face_1.jpg



IMG_8946_face_3.j...



IMG_8947_face_2.j...



IMG_8948_face_1.j...

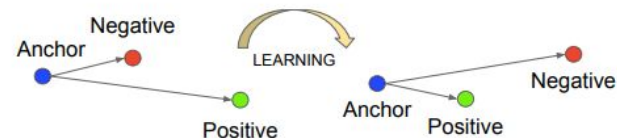


label = $\begin{cases} 0, & \text{if not Yurii} \\ 1, & \text{if Yurii} \\ \text{not valid} \end{cases}$

Triplet Loss

- $f(x)$ – **d-dimensional embedding** with unit euclidean norm of an image x
- a , p and n – **anchor**, **positive** and **negative** examples
- α - **margin** that is enforced between positive and negative pairs

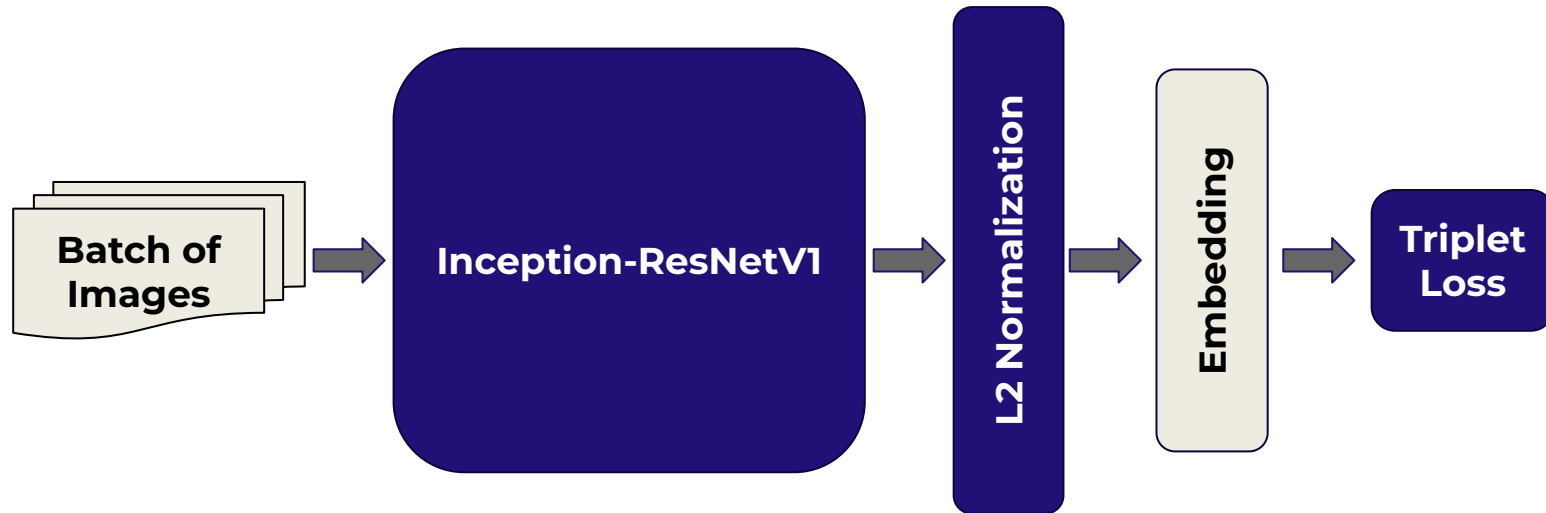
Triplet Loss Visualization



Triplet Loss Definition

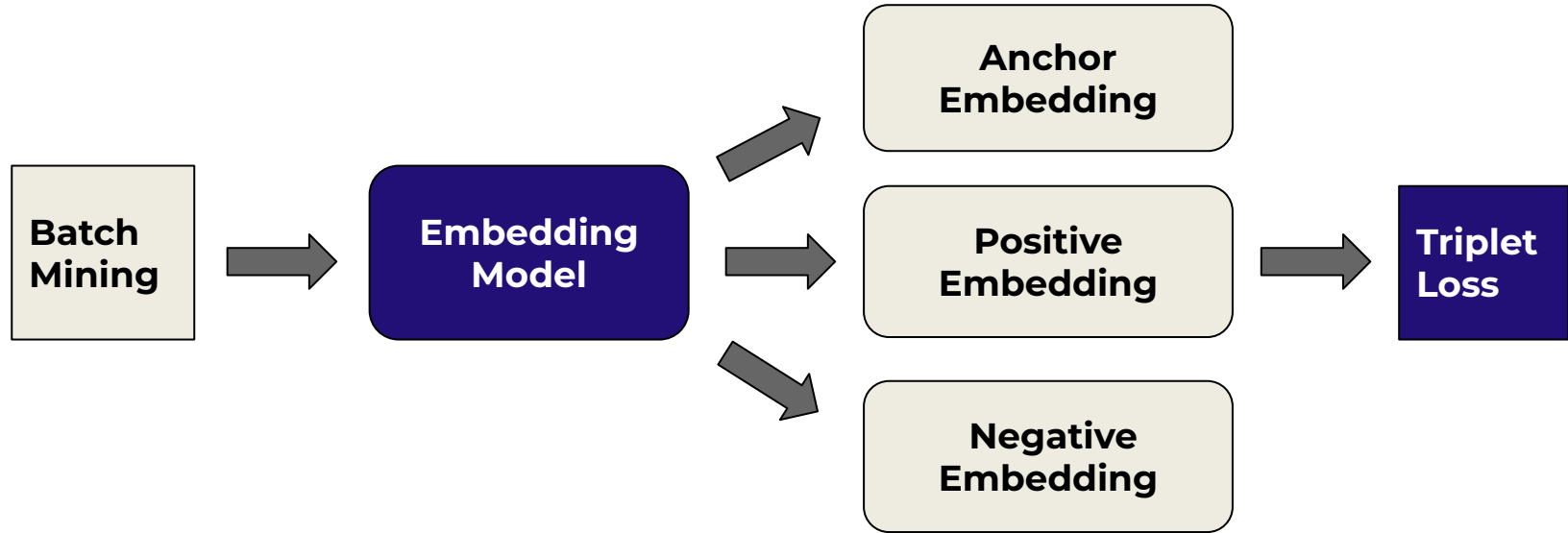
$$L(a, p, n) = \sum_i \max\{\|f(a_i) - f(p_i)\|_2^2 - \|f(a_i) - f(n_i)\|_2^2 + \alpha, 0\}$$

FaceNet



- Inception-ResNetV1 backbone pretrained on VGGFace2 Dataset
- Training method - **Transfer Learning**
- Triplet Selection Strategy - **Offline Selection**

FaceTripleNet



- **All triplets** - all possible anchor-positive-negative combos
- In **semi-hard mining** it selects negatives harder than positives but within margin
- In **hard mining** closest to anchor negative is selected

Evaluation

- Used **baseline** recognition (ArcFace/VGG-Face), and YOLOv8 for detection
- Used **classical metrics**: Precision, Recall & F1
- **Our model outperformed** both VGG-Face and ArcFace
- More data needed for a definitive efficiency comparison

Table 1: Performance of models

Model	F_1	Recall	Precision
YOLOv8 + ArcFace	0.58	0.78	0.46
YOLOv8 + VGG-Face	0.73	0.82	0.66
FaceTripleNet (ours)	0.78	0.85	0.72
FaceNet (ours)	0.74	0.60	0.92

Conclusion

- We constructed **new dataset** and labeled by hand faces.
- Made **4 experiments** (2 baselines, 2 ours)
- Via transfer learning trained our models.
- Achieved **superior performance** compared to benchmarks.

Questions?

