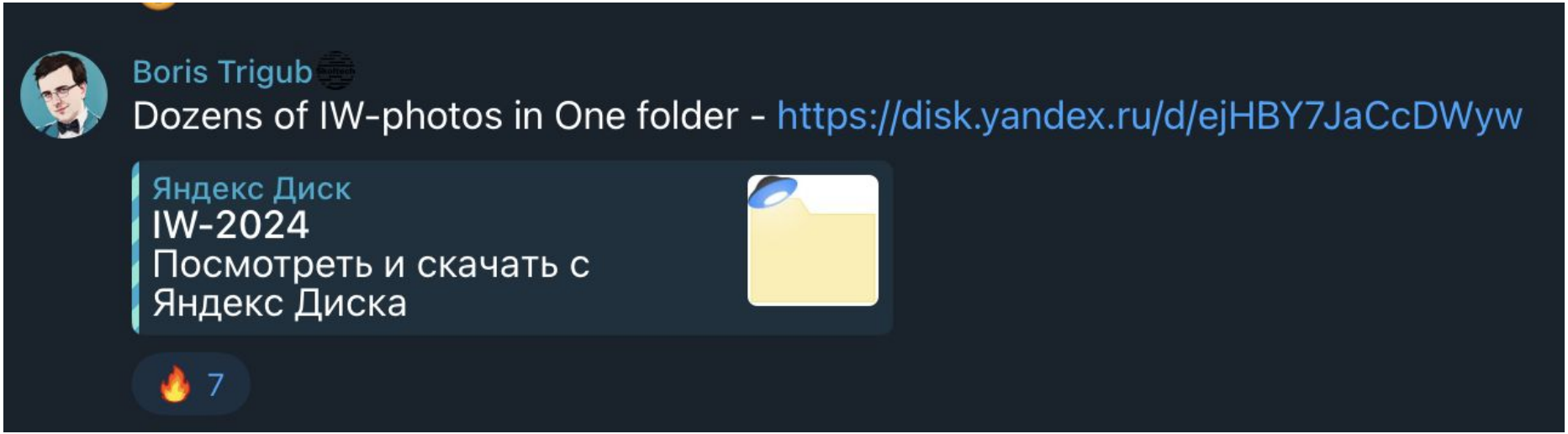
My photo search advice

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

Problem statement



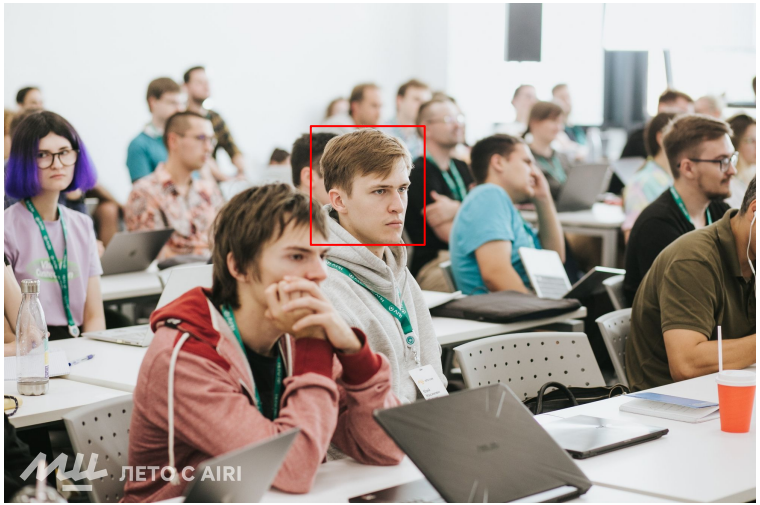
How to **find yourself** in dozens of photos?

What we achieved

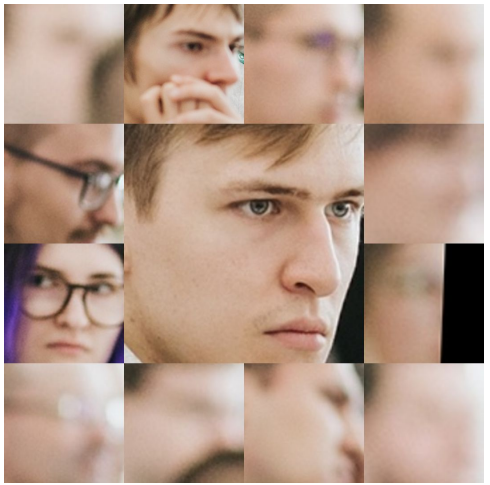
1. **Created dataset** and labelled by hand via detector to find faces
2. Used **2 pre-trained models** from DeepFace as baseline
3. **Trained 2 models**, using transfer learning on our dataset
4. Got **better results** than benchmark models

Data collection

Detection



YOLOv8



Preprocessing: Deblurring

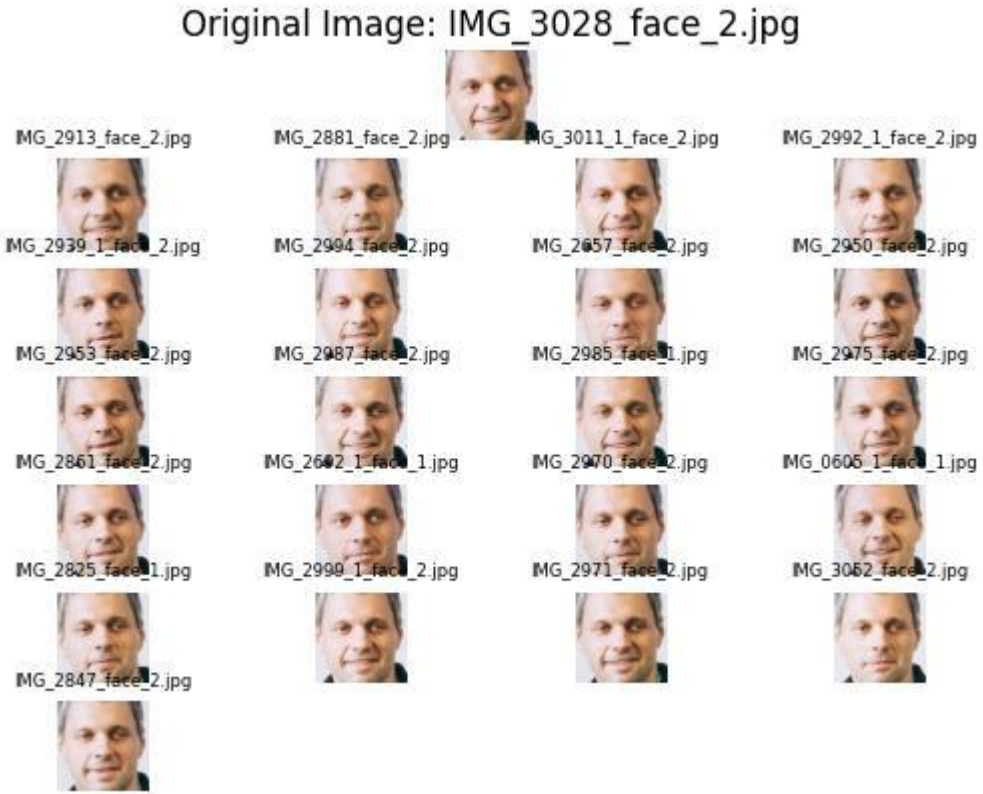
input:

L = Laplacian(image)

If L.var() < threshold

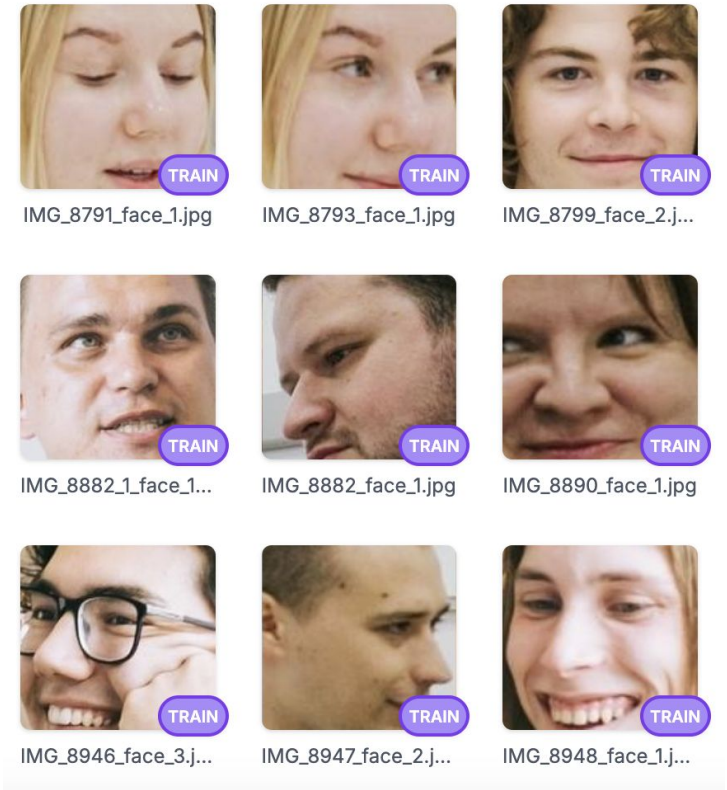
return “blurred image”

Preprocessing: Deduplications



Skoltech

Data labelling



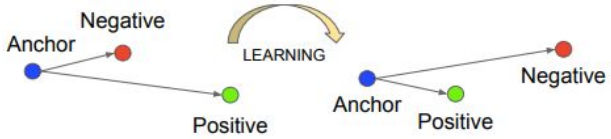
label =
- 0, if not Yurii
- 1, if Yurii
- not valid

Skoltech

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

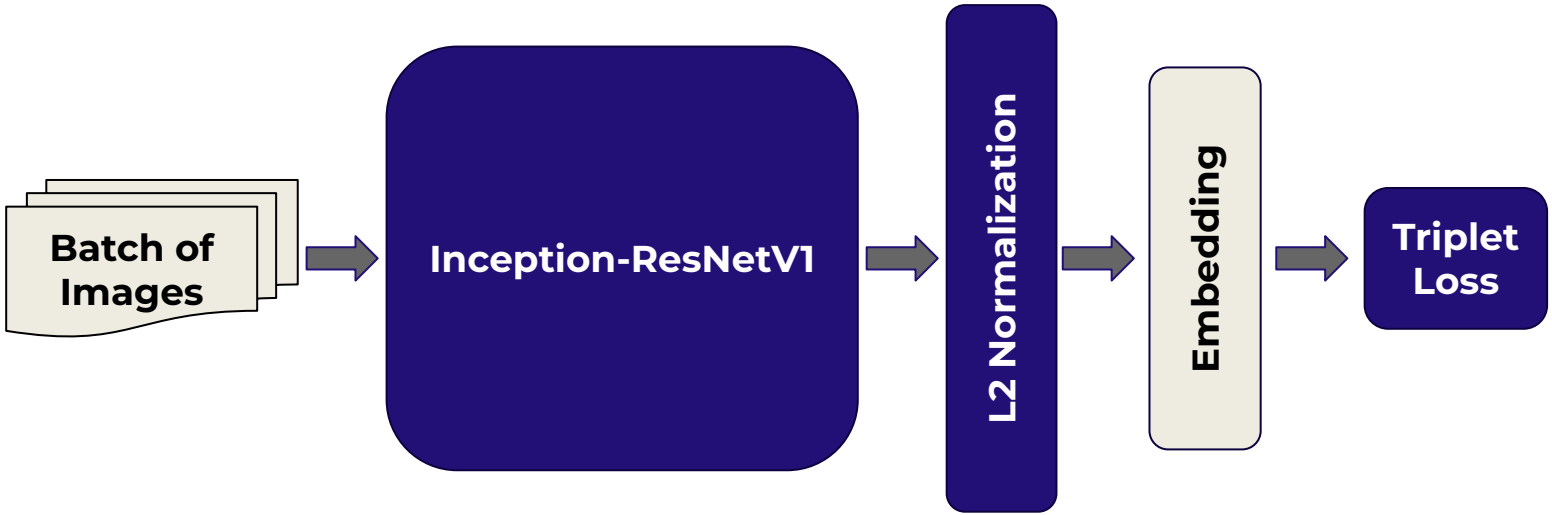
Triplet Loss Visualization



Triplet Loss Definition

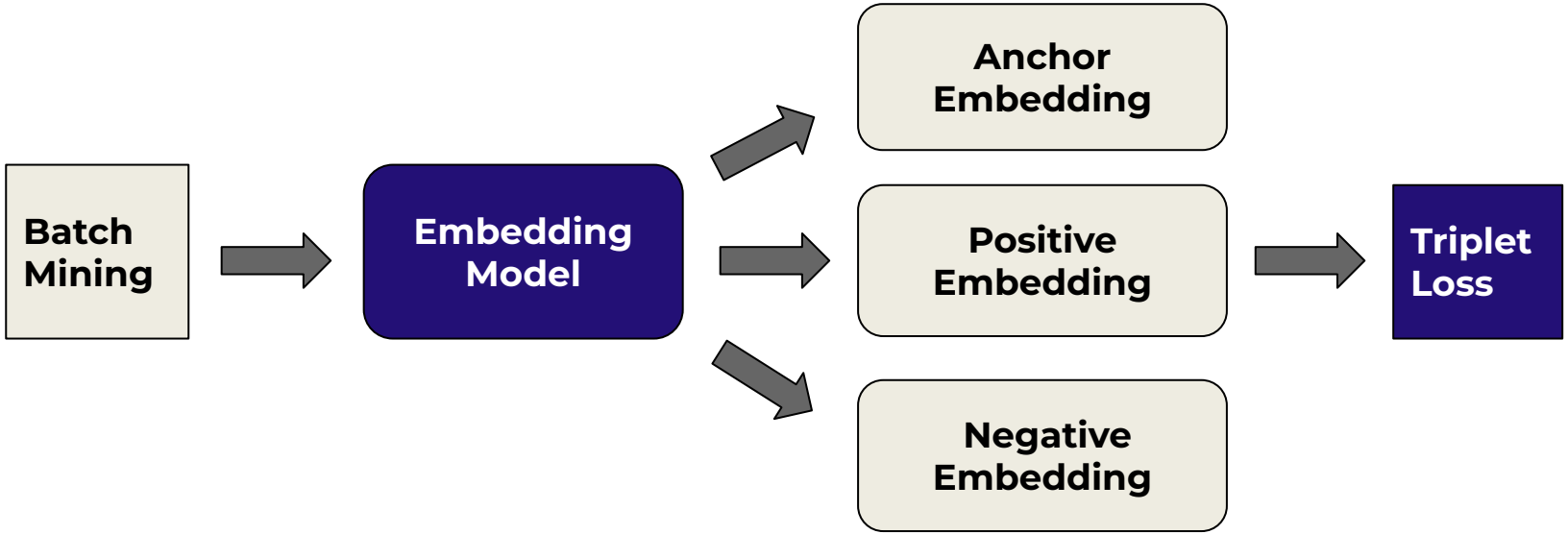
$$L(a,p,n) = \sum_i \max\{\|f(a_i) - f(p_i)\|_2^2 - |f(a_i) - f(n_i)\|_2^2 + \alpha, 0\}$$

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?