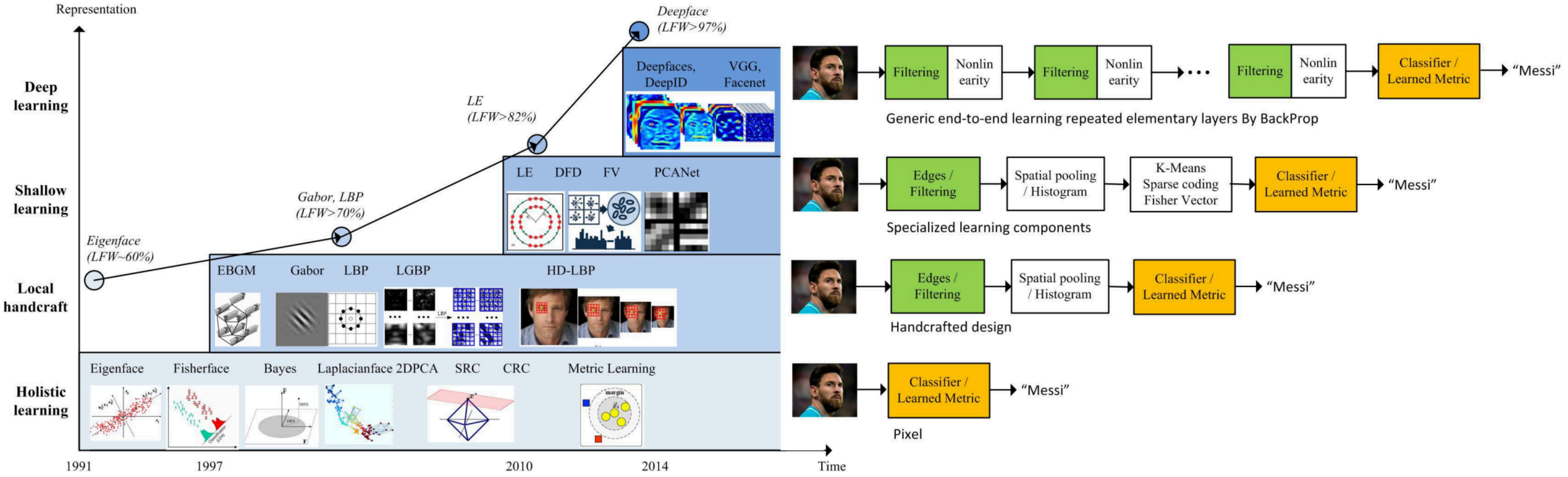
Face Recognition

Roman Vlasov

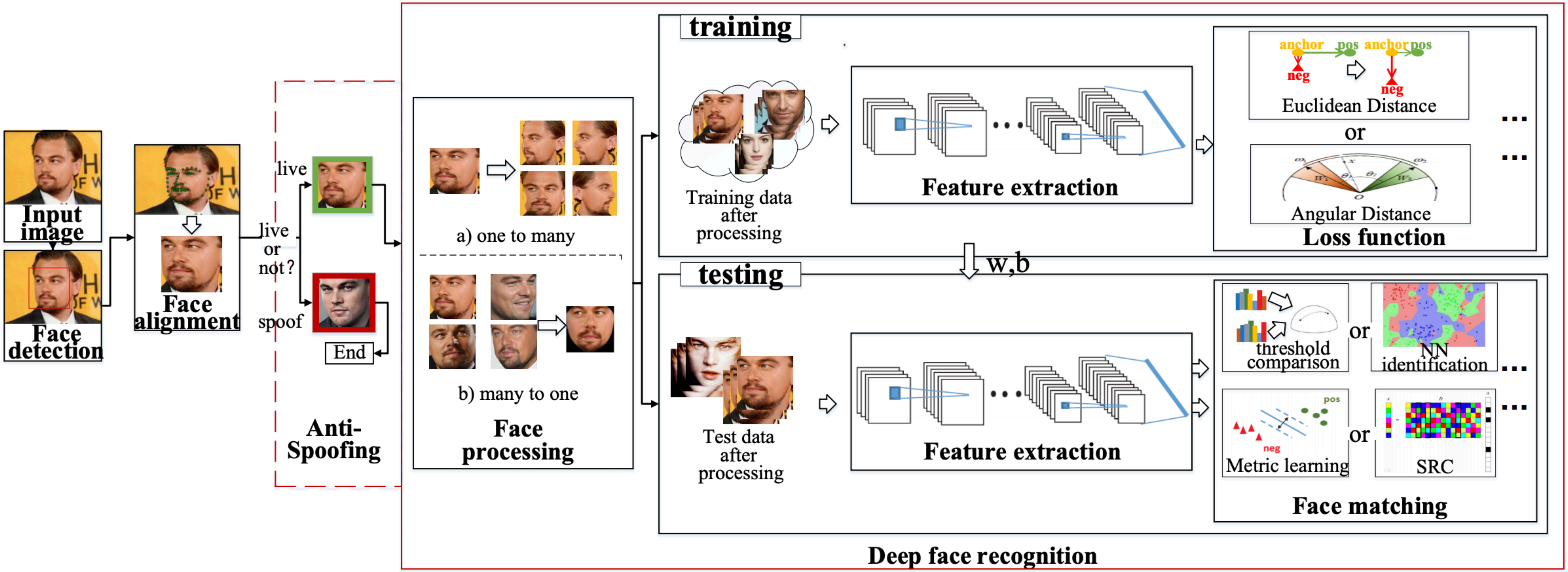
Face Recognition Progress



Milestones of face representation for recognition. The holistic approaches dominated the face recognition community in the 1990s. In the early 2000s, handcrafted local descriptors became popular, and the local feature learning approach were introduced in the late 2000s. In 2014, DeepFace and DeepID achieved a breakthrough on state-of-the-art performance, and research focus has shifted to deep-learning-based approaches. As the representation pipeline becomes deeper and deeper, the LFW (Labeled Face in-the-Wild) performance steadily improves from around 60% to above 90%, while deep learning boosts the performance to 99.80% in just three years.

https://arxiv.org/pdf/1804.06655.pdf

Face Recognition Pipeline

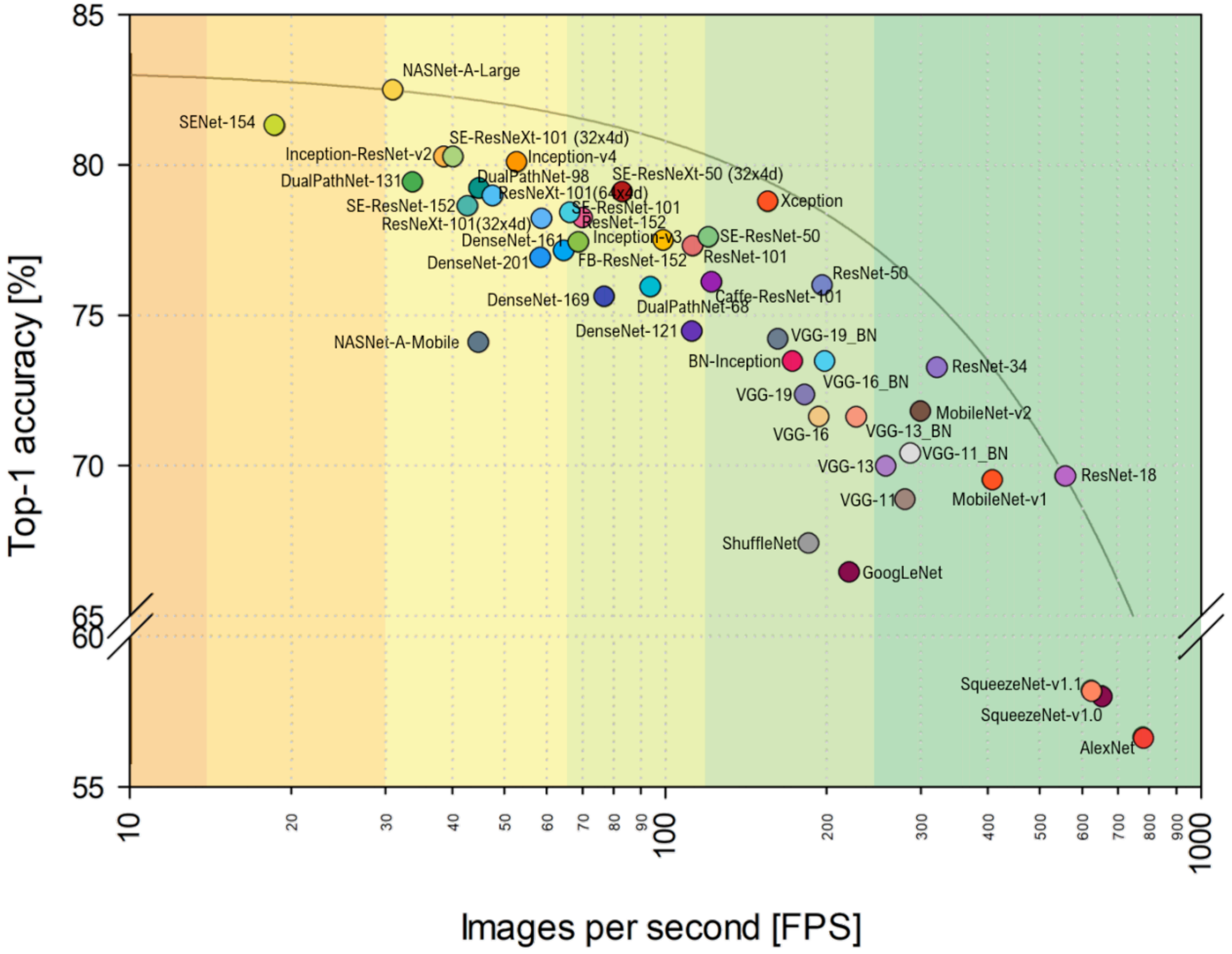


Deep FR system with face detector and alignment. First, a face detector is used to localize faces. Second, the faces are aligned to normalized canonical coordinates. Third, the FR module is implemented. In FR module, face anti-spoofing recognizes whether the face is live or spoofed; face processing is used to handle recognition difficulty before training and testing; different architectures and loss functions are used to extract discriminative deep feature when training; face matching methods are used to do feature classification when the deep feature of testing data are extracted.

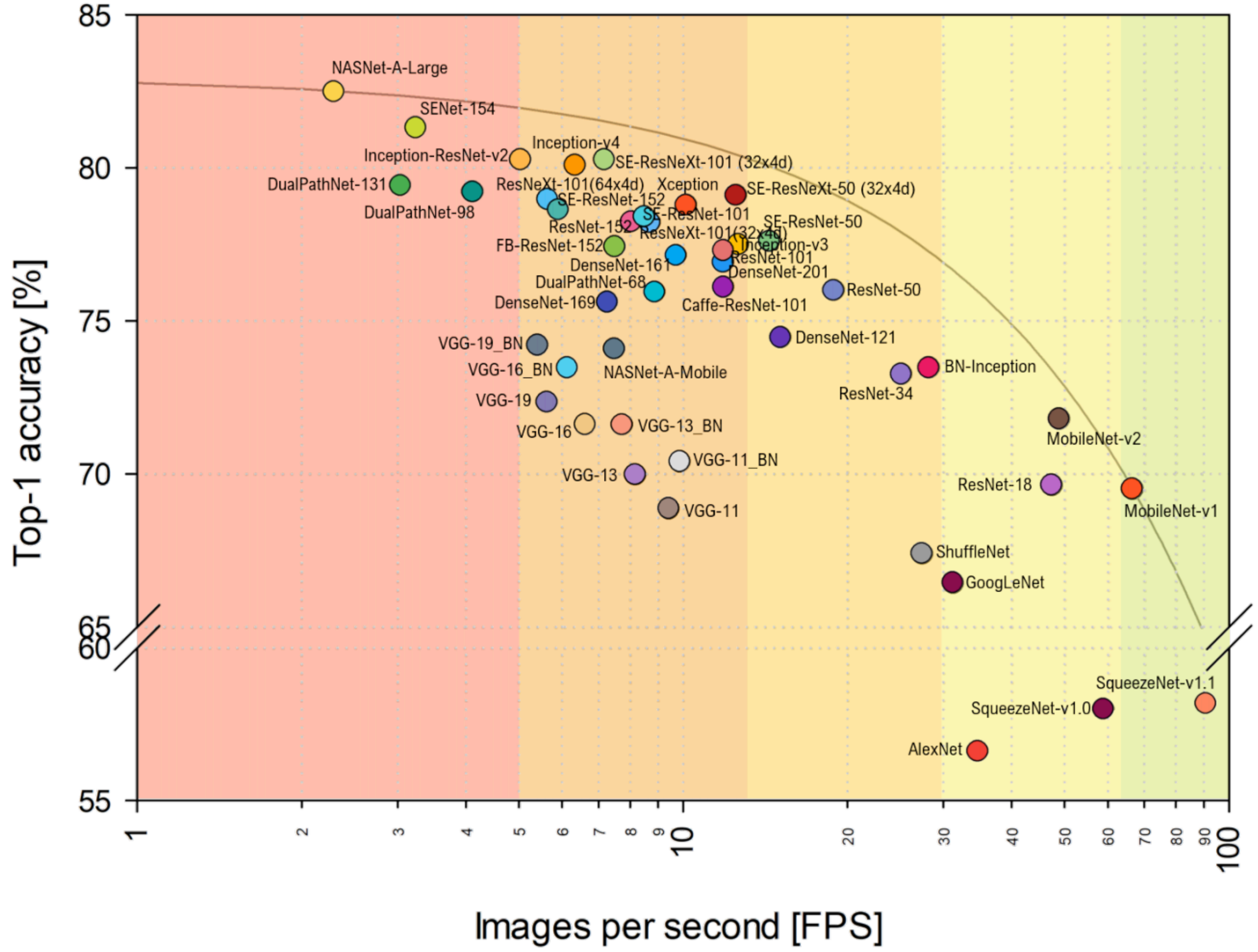
https://arxiv.org/pdf/1804.06655.pdf

Models Overview

Titan Xp

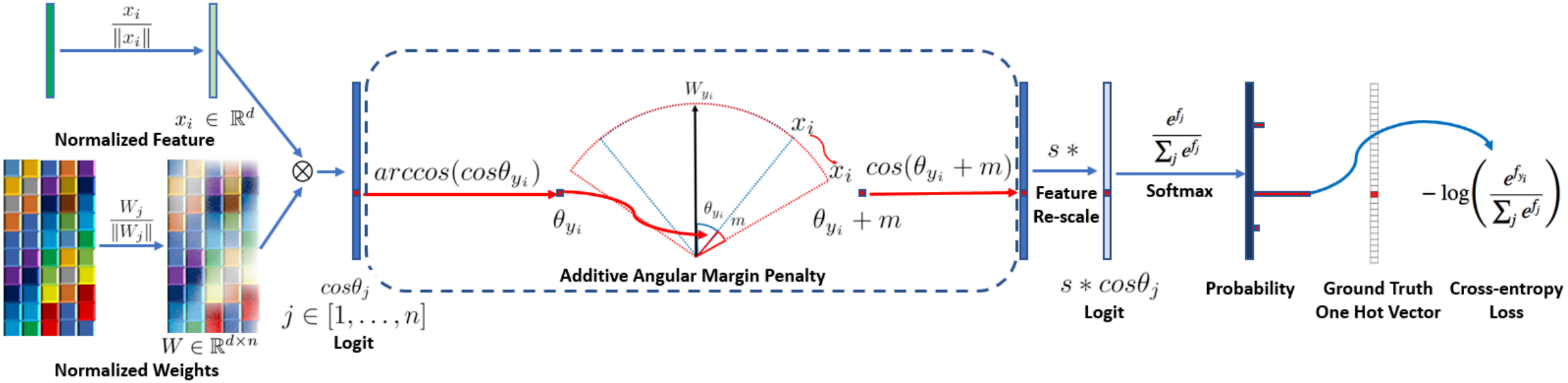


Jetson TX1



https://arxiv.org/pdf/1810.00736.pdf

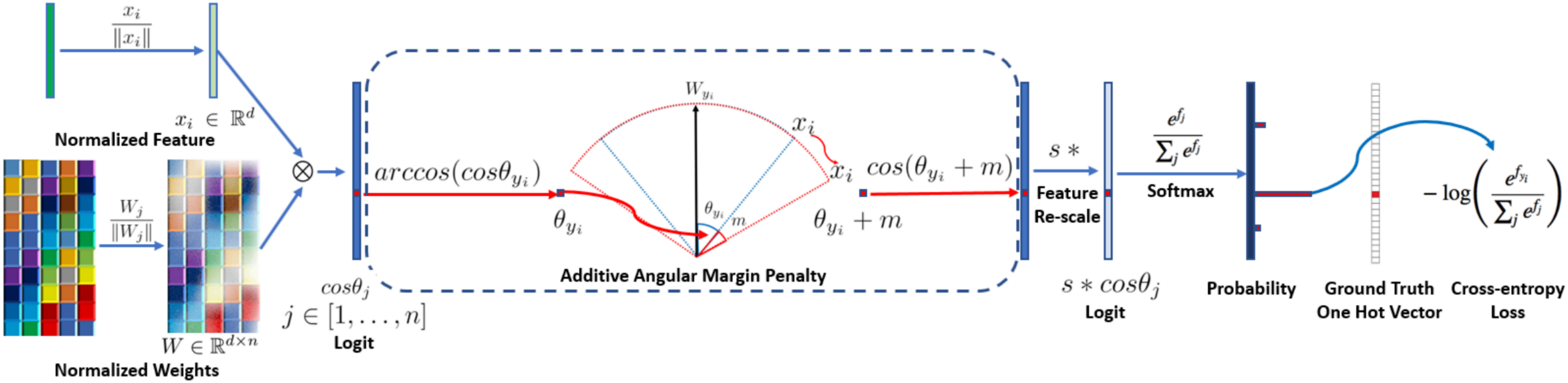
ArcFace



Convolutional Neural Network for face recognition supervised by the ArcFace loss. Based on the feature x_i and weight W normalisation, we get the $cos\theta_i$ (logit) for each class as $W_i^T x_i$. We calculate the $arccos\theta_{yi}$ and get the angle between the feature x_i and the ground truth weight W_{y_i}. In fact, W_j provides a kind of centre for each class. Then, we add an angular margin penalty m on the target (ground truth) angle θ_{y_i}. After that, we calculate $cos(\theta_{y_i} + m)$ and multiply all logits by the feature scale s. The logits then go through the softmax function and contribute to the cross entropy loss.

https://arxiv.org/pdf/1801.07698.pdf

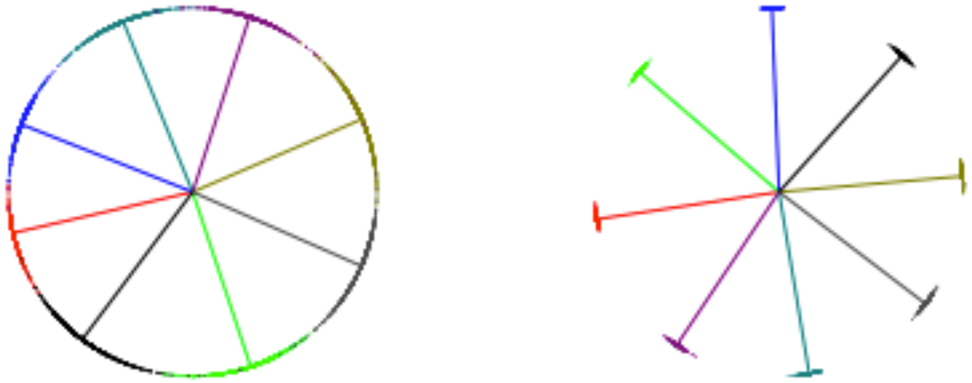
ArcFace



$$L_1 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_j^T x_i + b_j}}$$

$$L_2 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\theta_{y_i}}}{e^{s\cos\theta_{y_i}} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}}$$

$$L_3 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1, j\neq y_i}^{n} e^{s\cos\theta_j}}$$



(a) Softmax (b) ArcFace

https://arxiv.org/pdf/1801.07698.pdf

Face Recognition Results

Encoder	CPU Inference (ms)	LFW (%)	CFP_FP (%)	AGEDB_30 (%)	Mean (%)
MobileNet	61	99.58	95.73	96.50	97.27
ResNet34	113	99.75	97.64	97.60	98.33
SEResNeXt50	232	99.75	98.77	97.92	98.81
InsightFace (Resnet101)	-	99.80	98.27	98.28	98.78

Distillation

Kullback-Leibler divergence

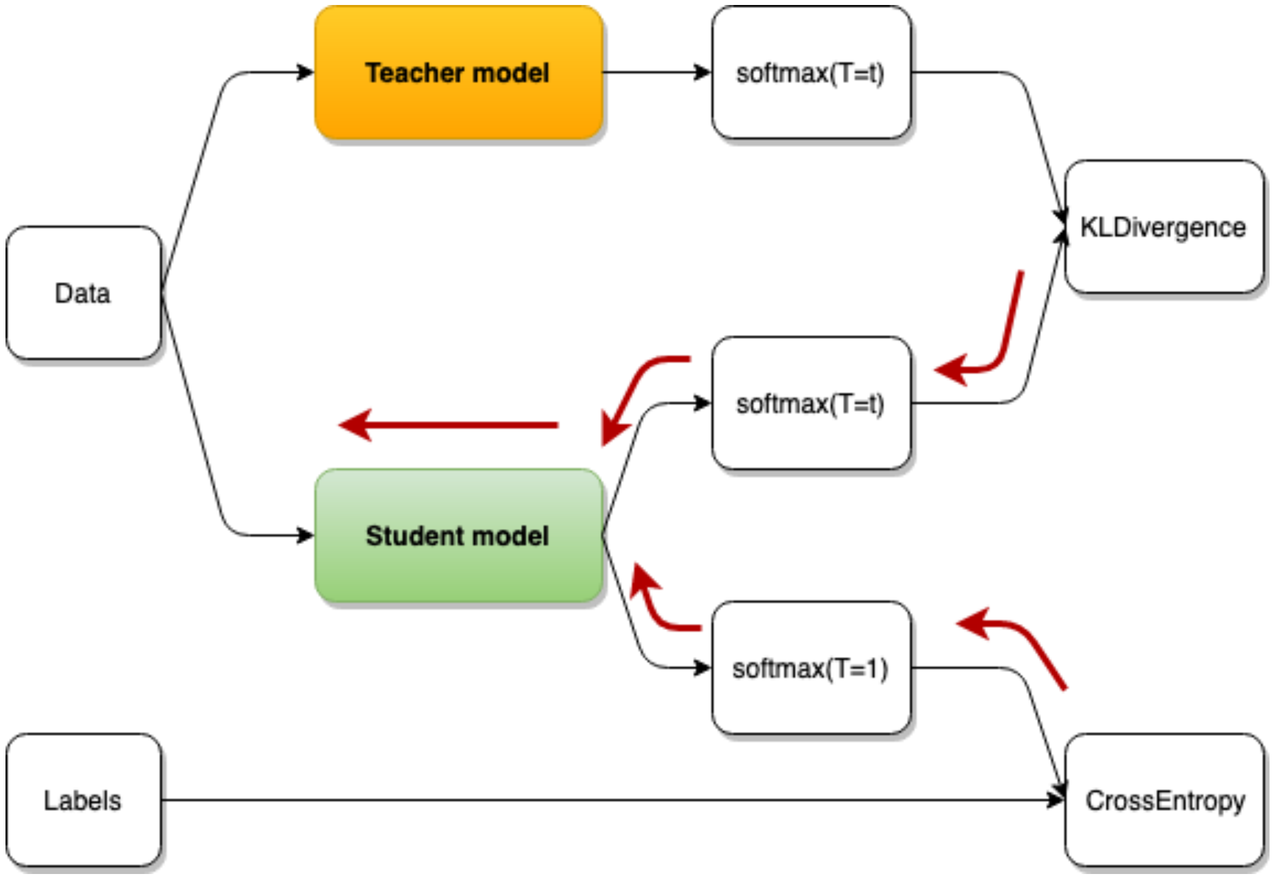
$$D_{KL}(P \parallel Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$$

Cross Entropy

$$\mathrm{H}(p,q) = -\sum_x p(x)\, \log q(x).$$

Softmax

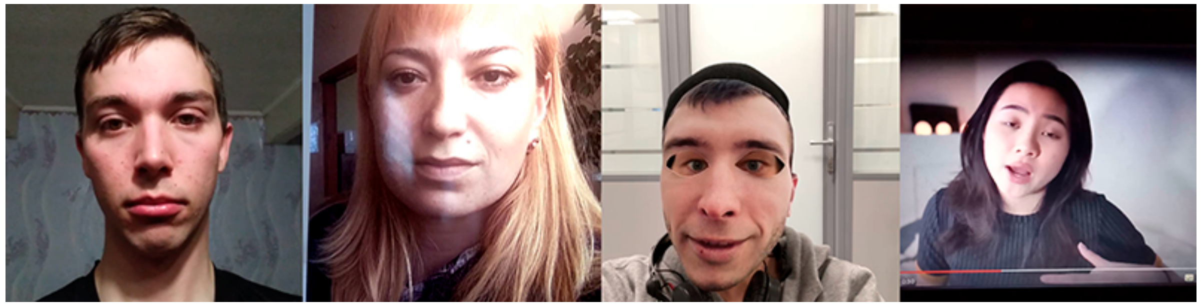
$$q_i = \frac{exp(z_i/T)}{\sum_j exp(z_j/T)}$$



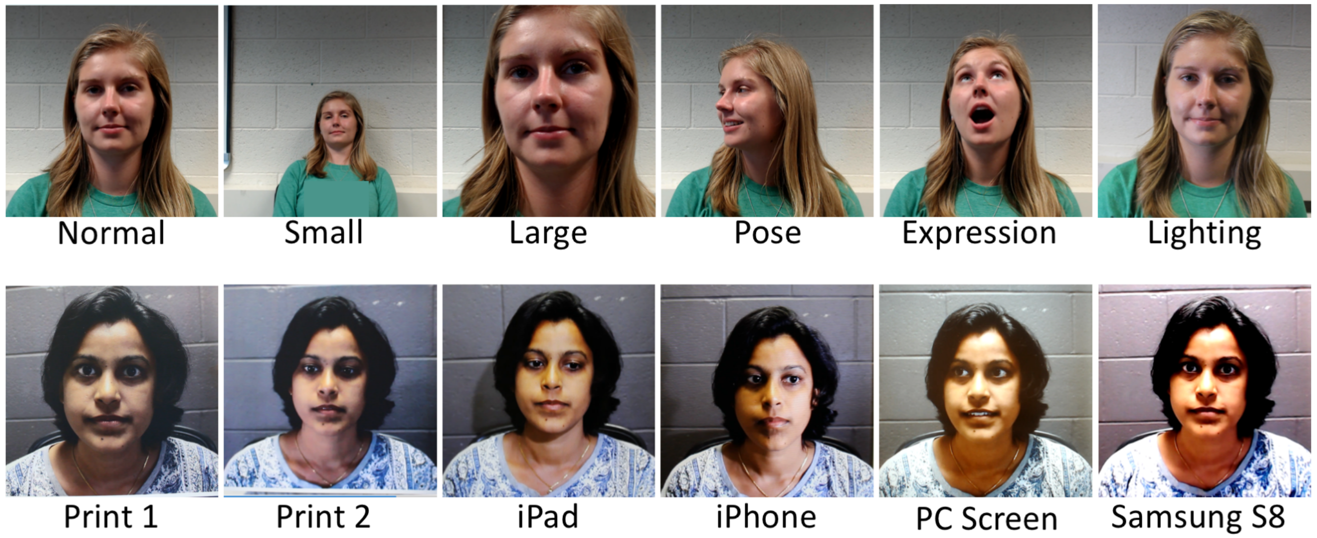
Face Recognition Results

Encoder	**CPU Inference (ms)**	**LFW (%)**	**CFP_FP (%)**	**AGEDB_30 (%)**	**Mean (%)**
MobileNet	61	99.58	95.73	96.50	97.27
MobileNet KD	**61**	**99.70**	**96.29**	**96.68**	**97.57**
ResNet34	113	99.75	97.64	97.60	98.33
SEResNeXt50	232	99.75	98.77	97.92	98.81
InsightFace (Resnet101)	-	99.80	98.27	98.28	98.78

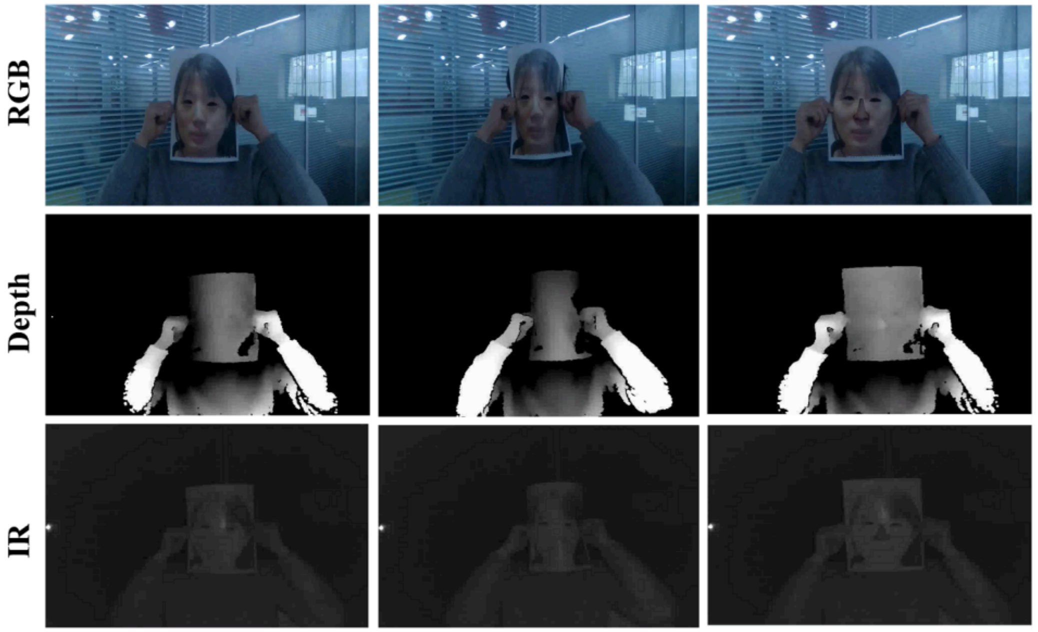
Anti Spoofing



IdR&D dataset provides live and spoof images from 11500 subjects. For each subject, it has 5 frames captured with in interval of 200ms and 1080P HD resolution. The dataset are collected from the wild and has different variation of poses and backgrounds. There are three type of attack in dataset: 2d-mask, replay, printed paper

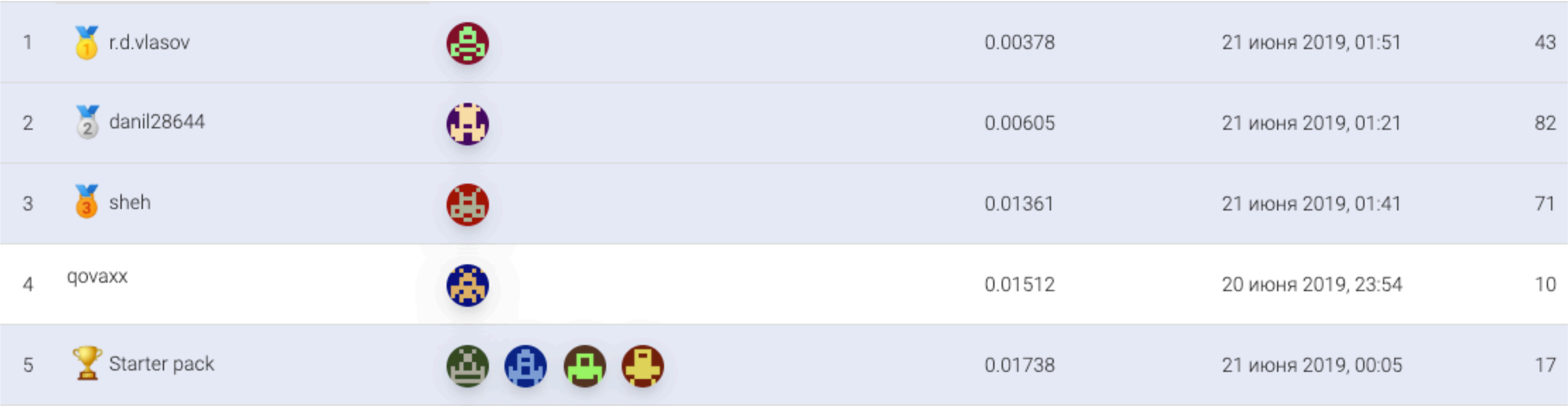


Spoof in the Wild (SiW) provides live and spoof videos from 165 subjects. For each subject, it has 8 live and up to 20 spoof videos, in total 4,478 videos. All videos are in 30 fps, about 15 second length, and 1080P HD resolution. The live videos are collected in four sessions with variations of distance, pose, illumination and expression. The spoof videos are collected with several attacks such as printed paper and replay.



CASIA-SURF dataset containing 1, 000 Chinese people in 21, 000 videos. In the proposed dataset, each sample includes 1 live video clip, and 6 fake video clips under different attack ways. Four video streams including RGB, Depth and IR images were captured at the same time, plus the RGB-DepthIR aligned images using RealSense SDK. The resolution is 1280 × 720 for RGB images, and 640 × 480 for Depth, IR and aligned images.

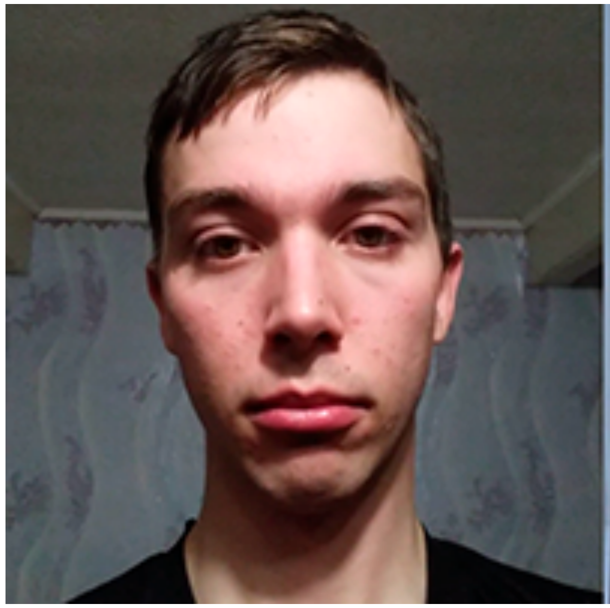
Leader Board (Id R&D anti-spoofing challenge)



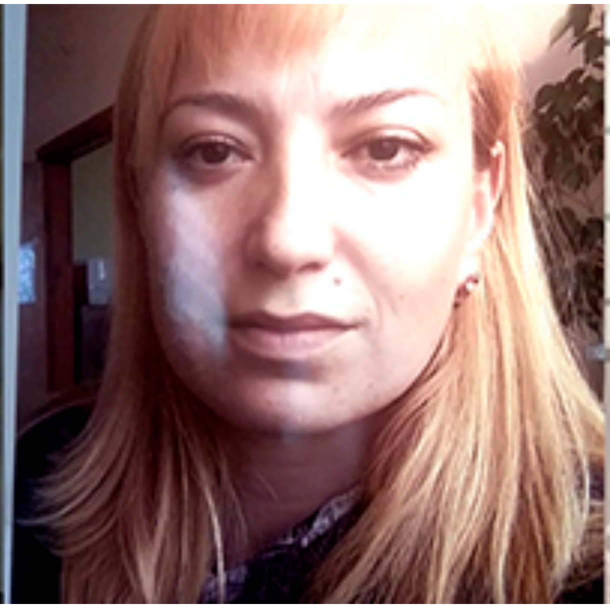
1	🥇 r.d.vlasov		0.00378	21 июня 2019, 01:51	43
2	🥈 danil28644		0.00605	21 июня 2019, 01:21	82
3	🥉 sheh		0.01361	21 июня 2019, 01:41	71
4	qovaxx		0.01512	20 июня 2019, 23:54	10
5	🏆 Starter pack		0.01738	21 июня 2019, 00:05	17

https://datasouls.com/c/idrnd-antispoof/leaderboard

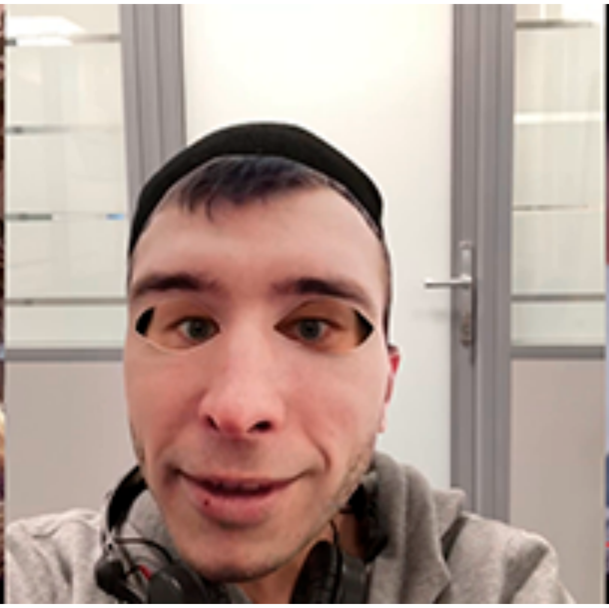
Data



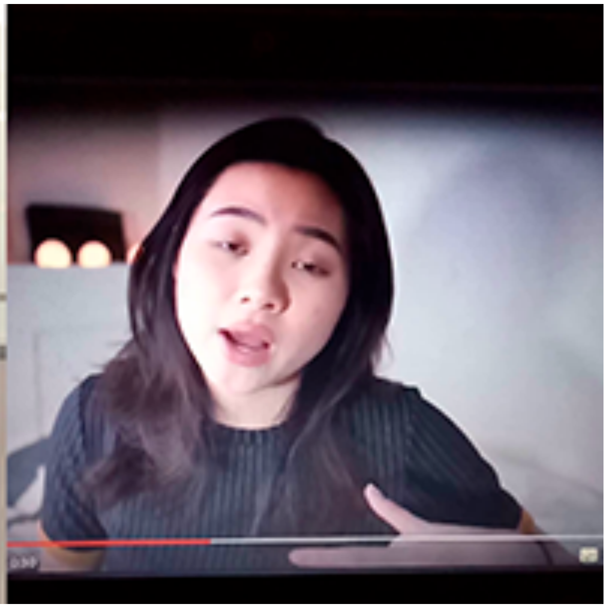
live



printed paper



2d-mask



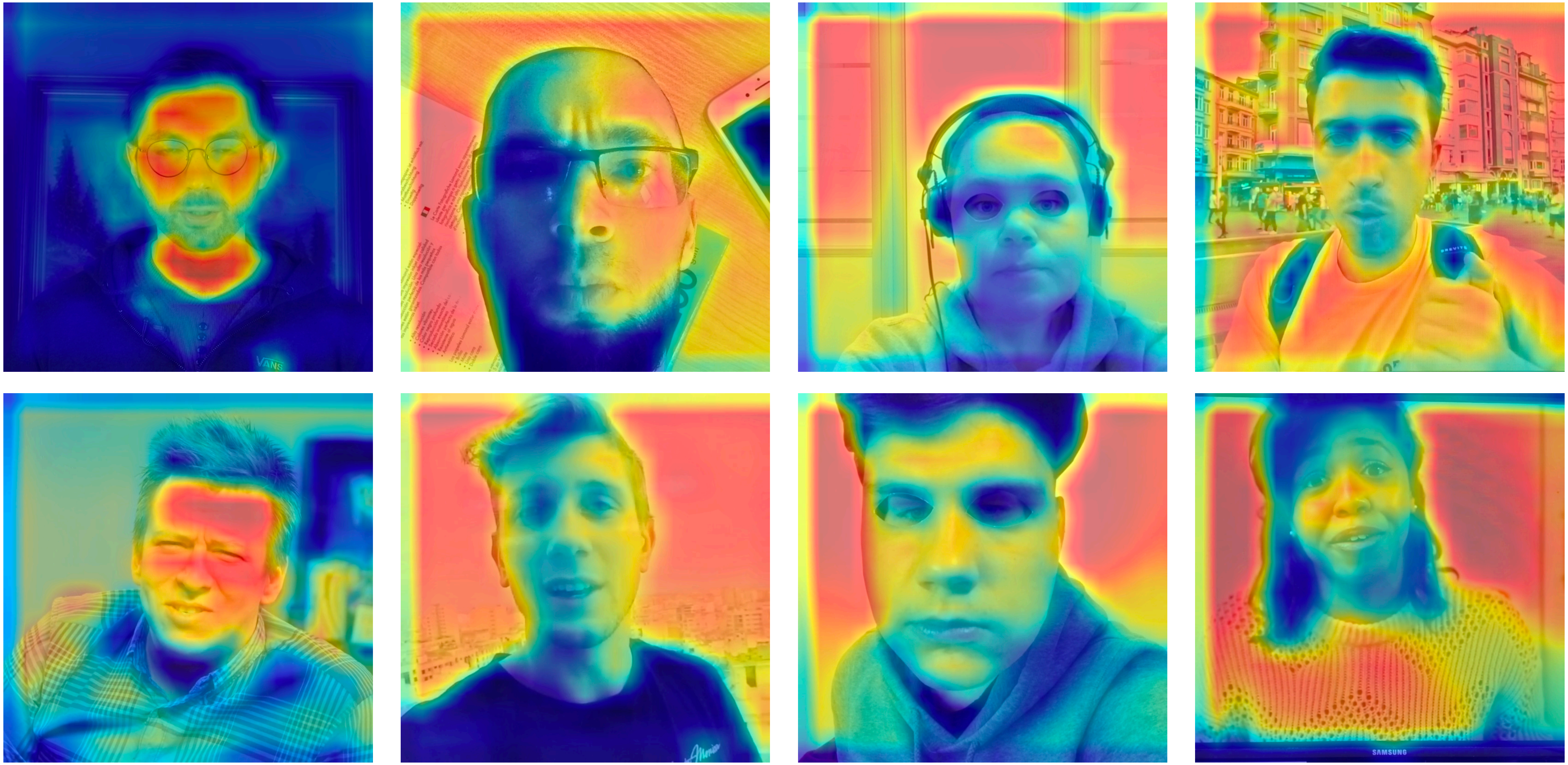
replay

Dataset: live and spoof images from 11500 subjects. For each subject, it has 5 frames captured with in interval of 200ms and 1080P HD resolution. The dataset are collected from the wild and has different variation of poses and backgrounds. There are three type of attack in dataset: 2d-mask, replay, printed paper.
Test: ~ 1000 subjects (public/private - 50/50)

Metric: min((P(false alarm) + 19 · P(miss)) / 20) by threshold == \
min((FP / (FP + TN) + 19 · FN / (FN + TP)) / 20) by threshold

https://datasouls.com/c/idrnd-antispoof/description

Activation Heatmap



live

printed paper

2d-mask

replay

Thanks!