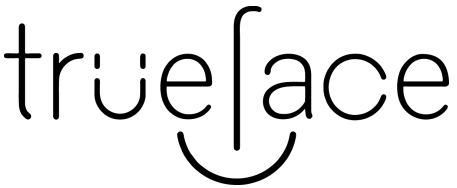
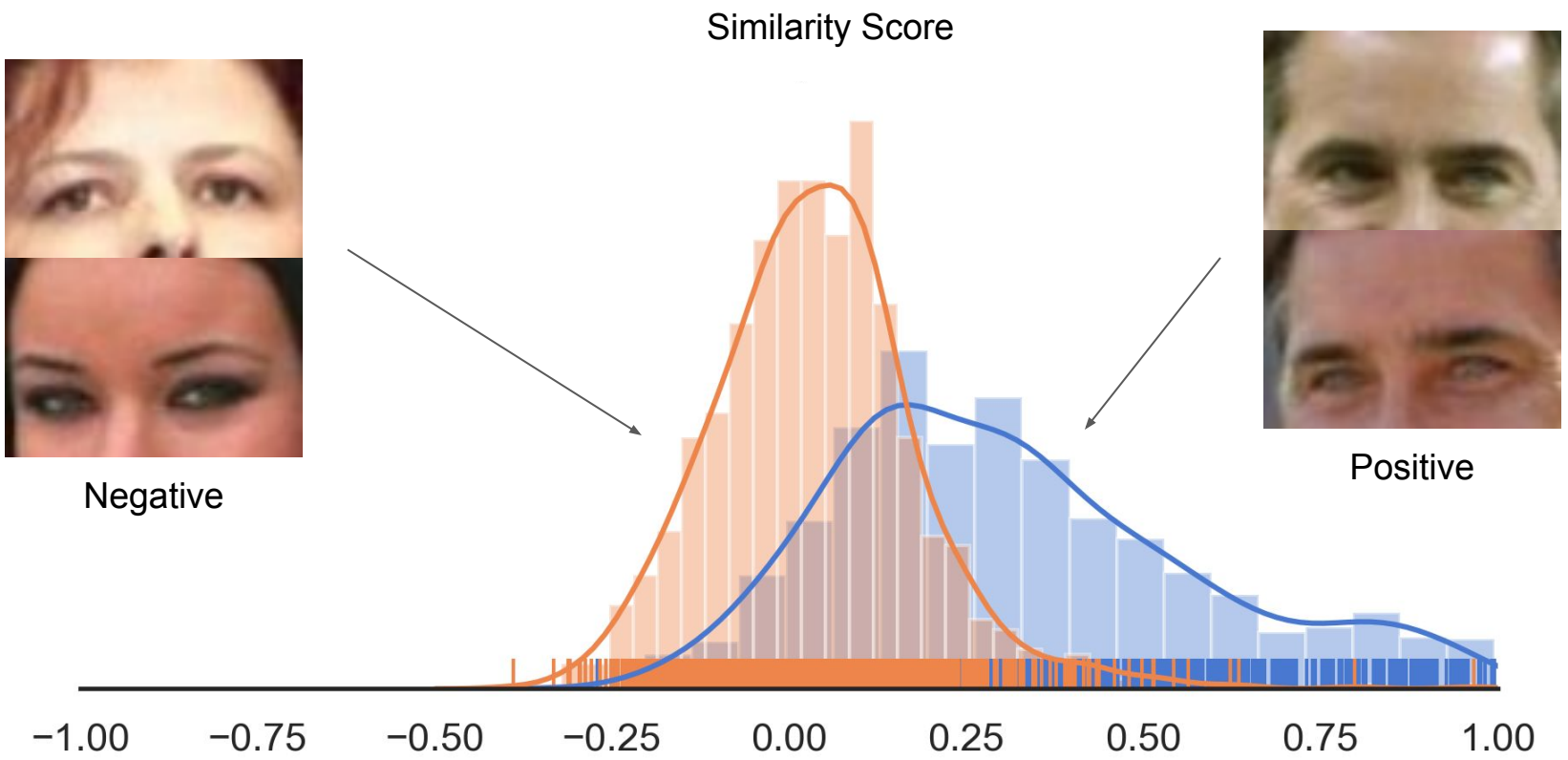
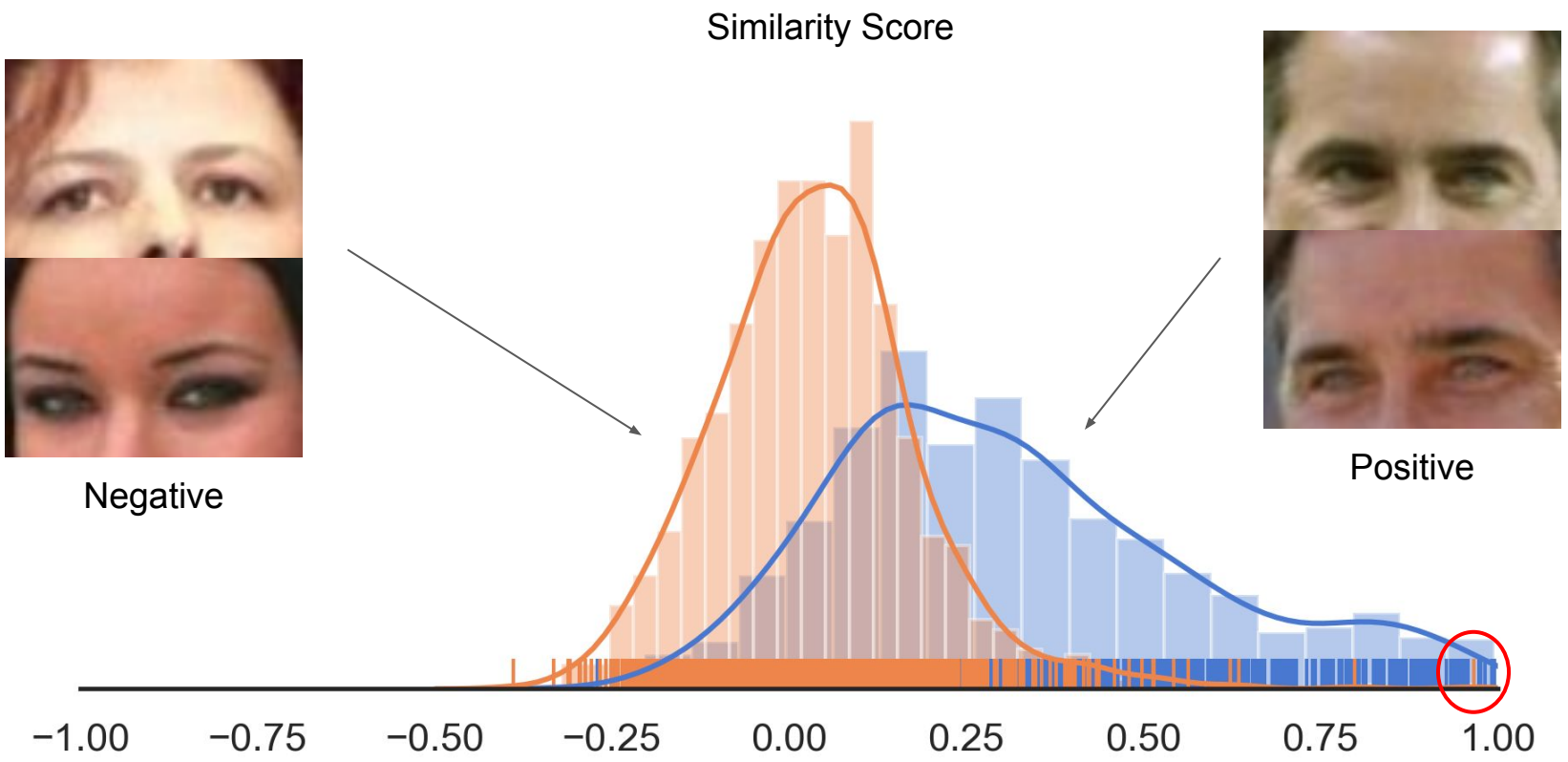
A bias mitigation strategy: overcoming the problem of overly confident false matches

Mosalam Ebrahimi, Trueface
October 29, 2020





(The LFW images are used to test this experimental model.)


Similarity Score
Negative
Positive
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

Trueface Introduction

At Trueface, we teach computers to see like humans; interpreting the data they ingest. Once trained to understand the visual data in question, computers help businesses and agencies make instant decisions, allowing businesses to cut costs and agencies to redistribute human capital to higher-functioning tasks.

Our clients choose Trueface because we are committed to the responsible deployment of computer vision technology and they want to ensure their businesses and their customers are benefiting from the advancements of artificial intelligence.

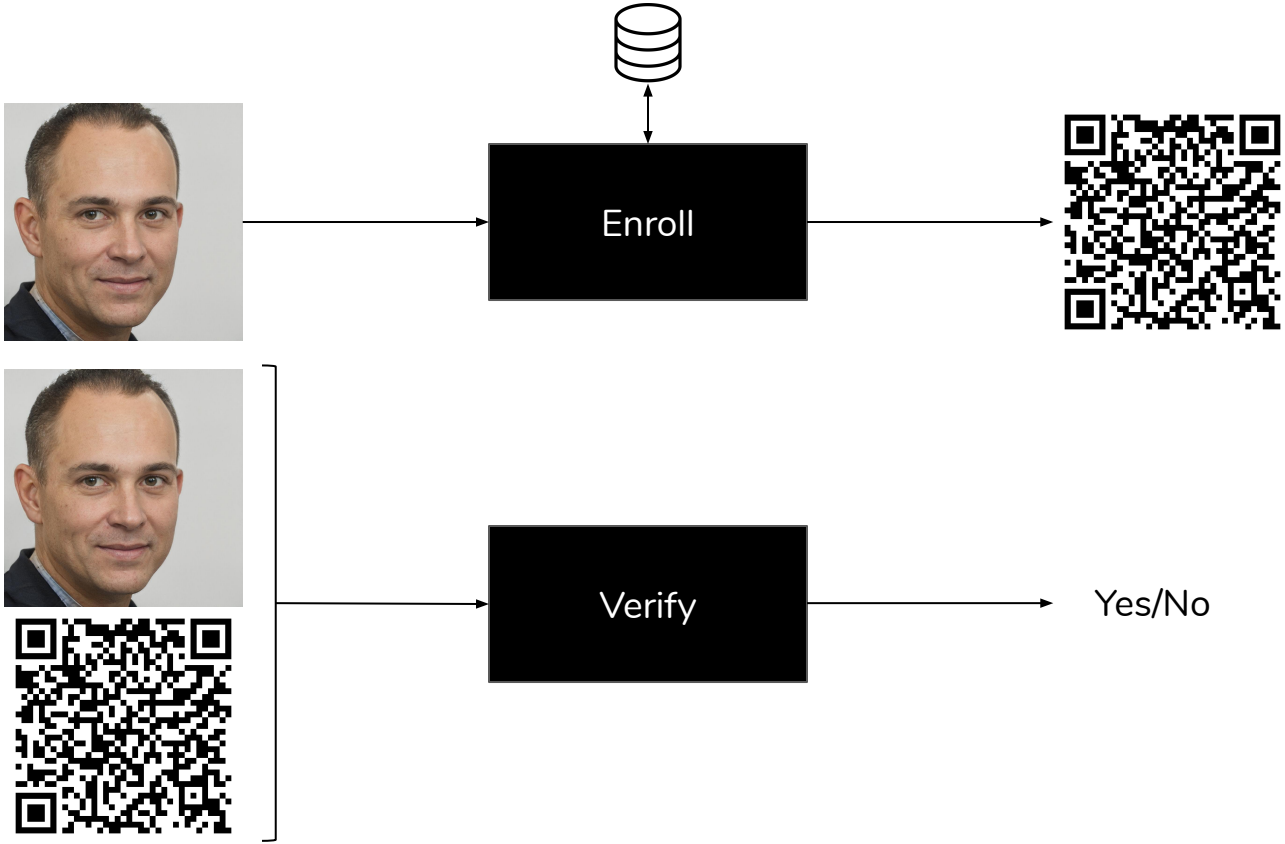
Trueface scores in the top 10 on the NIST Face Recognition Vendor Test for genuine match speed, which is paramount in mission critical deployments. The team has been working on face recognition in constrained environments since 2012 and has partnered with some of the world's most prestigious and innovative companies along the way.

Trueface is proud to be built in the USA and to support our Department of Defense, having been awarded 3 contracts in the last two years.

HQ: Venice, CA

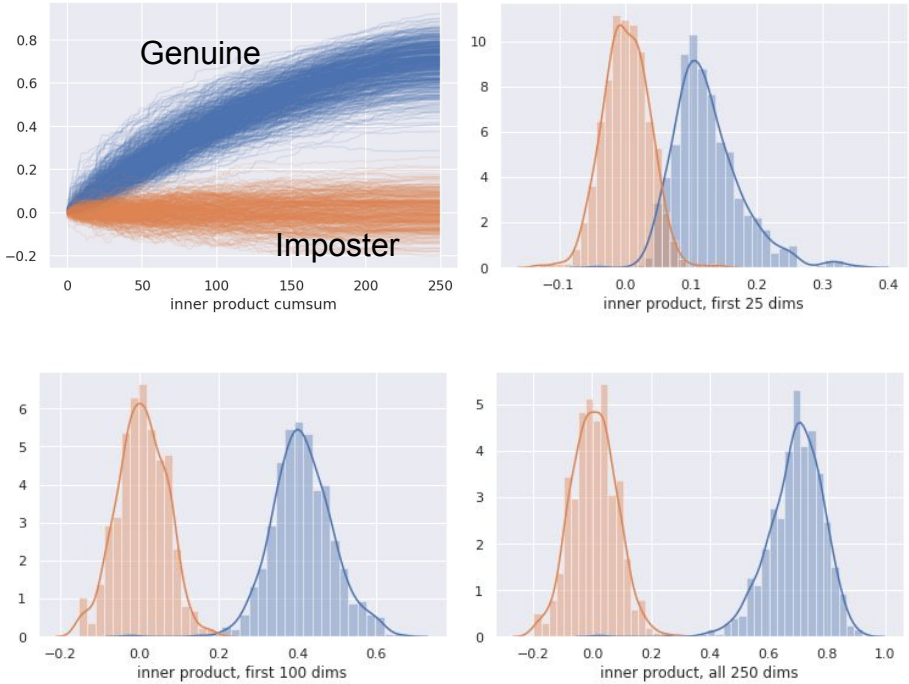
Some of our research projects

Databaseless face verification

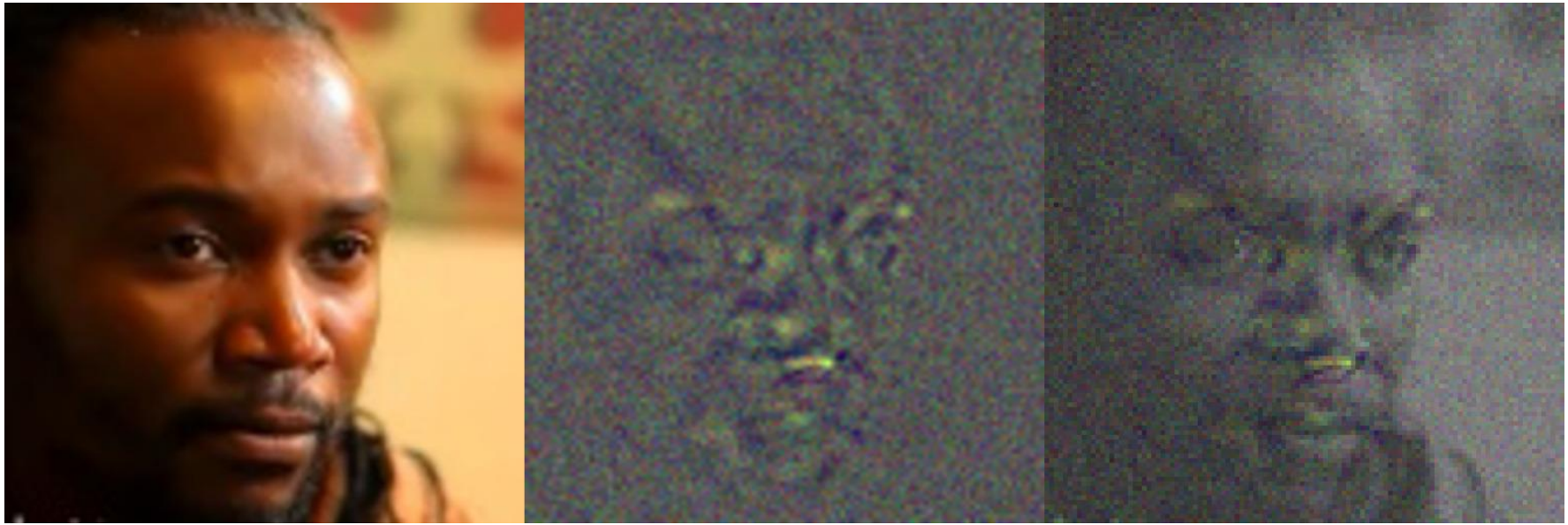


Fast Template Matching

4x faster than the our own method in the FRVT 1:1 report on the same hardware; a non-approximate method.



Detailed Saliency Maps



217 KB

80% JPEG

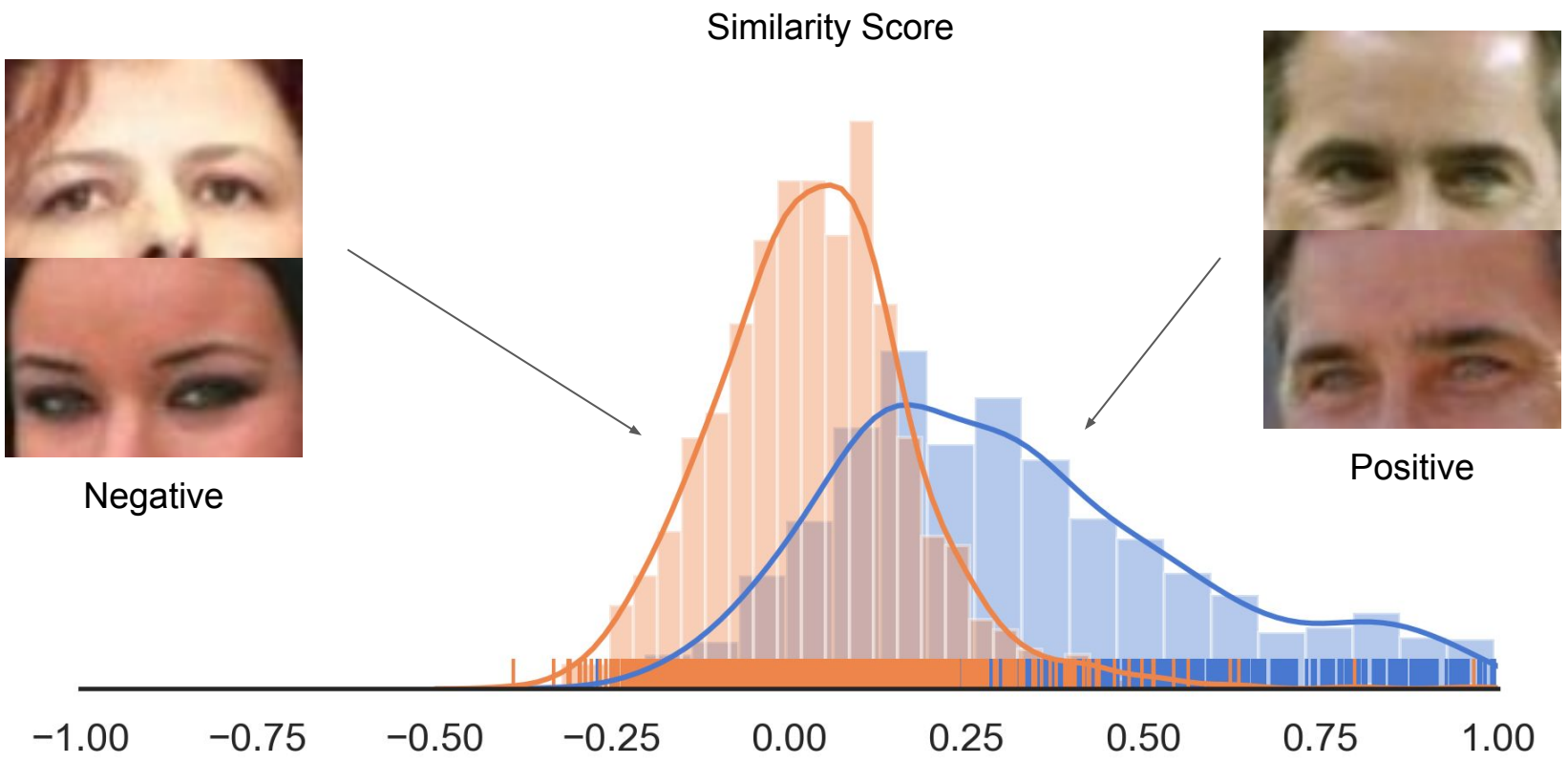
14 KB

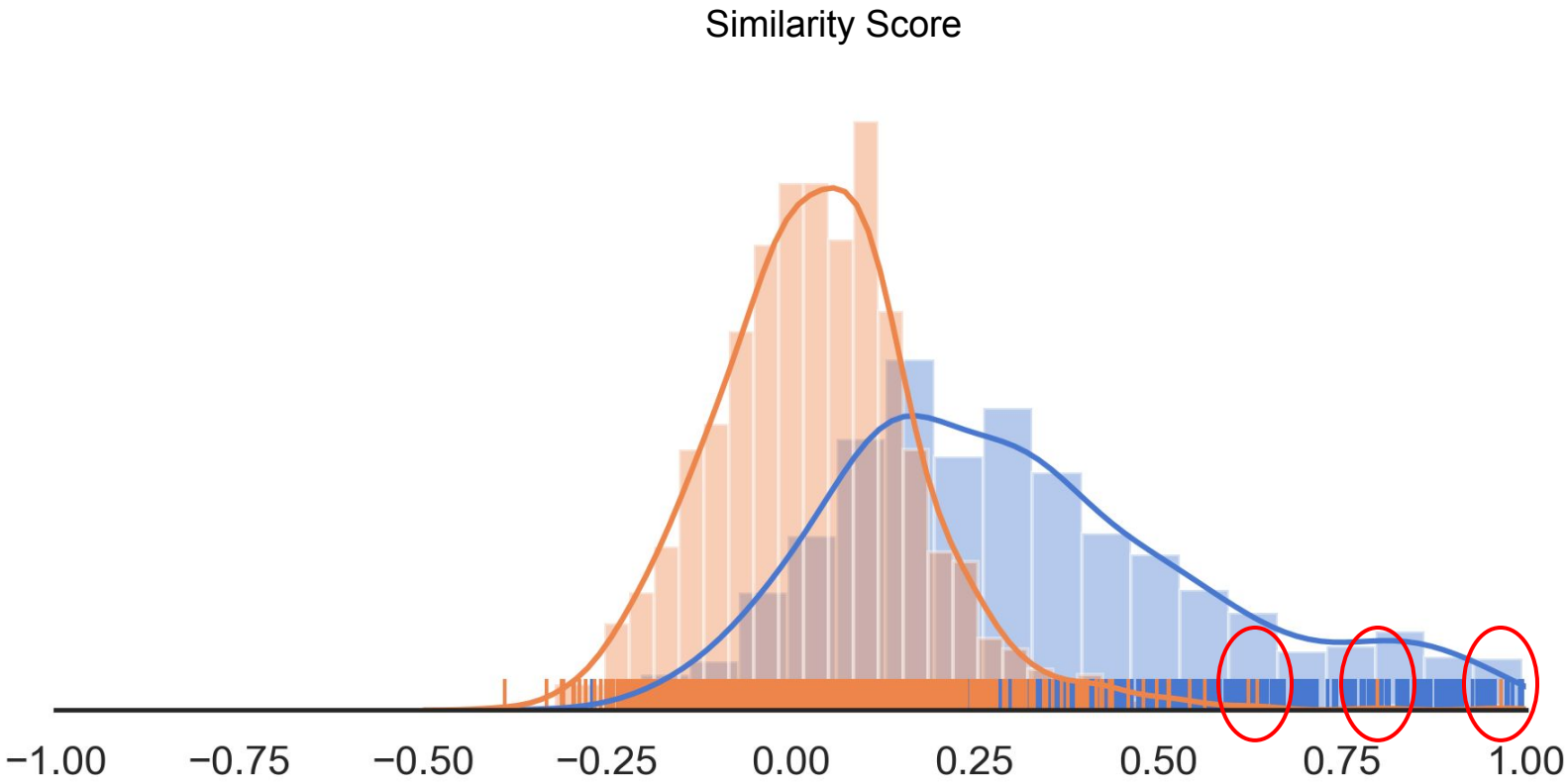
Quality: 10%

26 KB

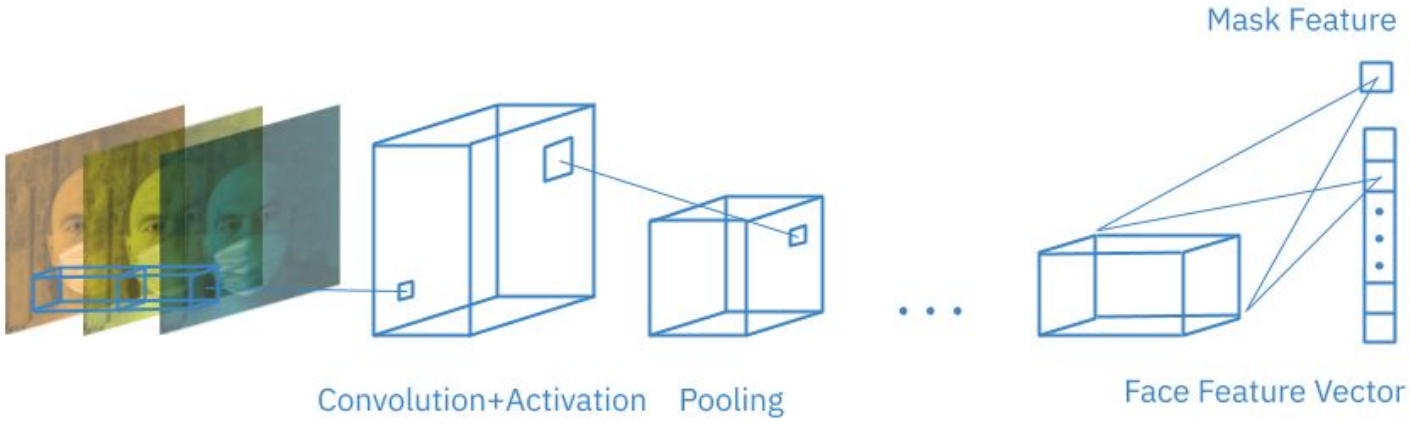
Our Method

Detecting overly confident false matches


Similarity Score
Negative
Positive
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00


Similarity Score
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

The root cause of overly confident false matches


Mask Feature
Convolution+Activation
Pooling
Face Feature Vector

Why the distance (similarity score) is a bad confidence measure?

1. The embedding is non-isometric with **high distortion**.

2. The embedding maps the **out-of-distribution points** to random points in the destination metric.

High distortion: distance in the embedding space is translation variant

0.41

0.55

0.80

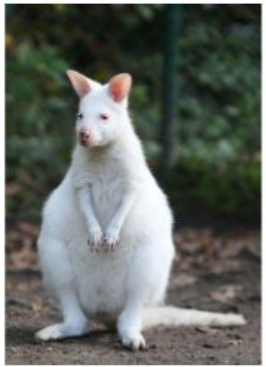
0.82

0.81

0.75

Out-of-distribution samples
are mapped to random points

Input

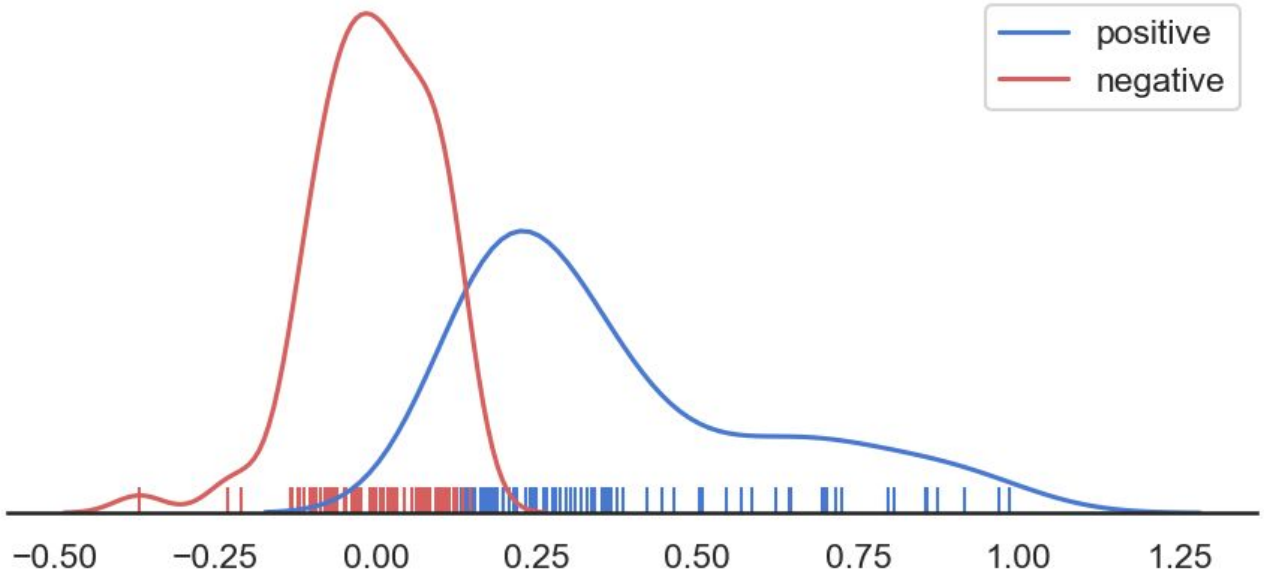
Output



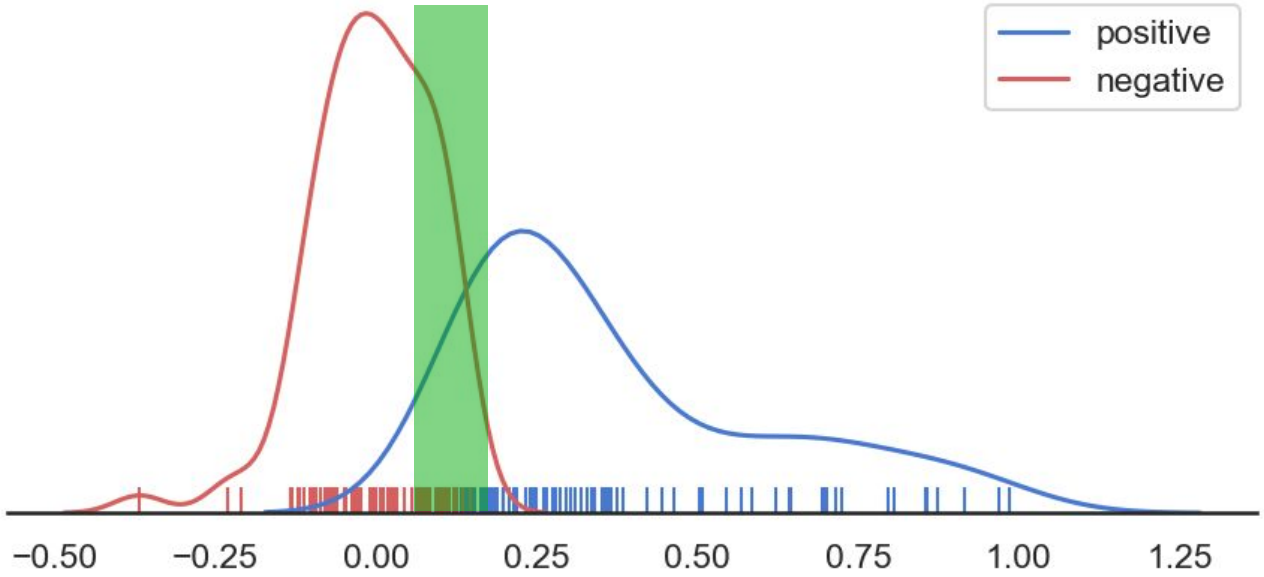

Model: https://tfhub.dev/inaturalist/vision/embedder/inaturalist_V2/1

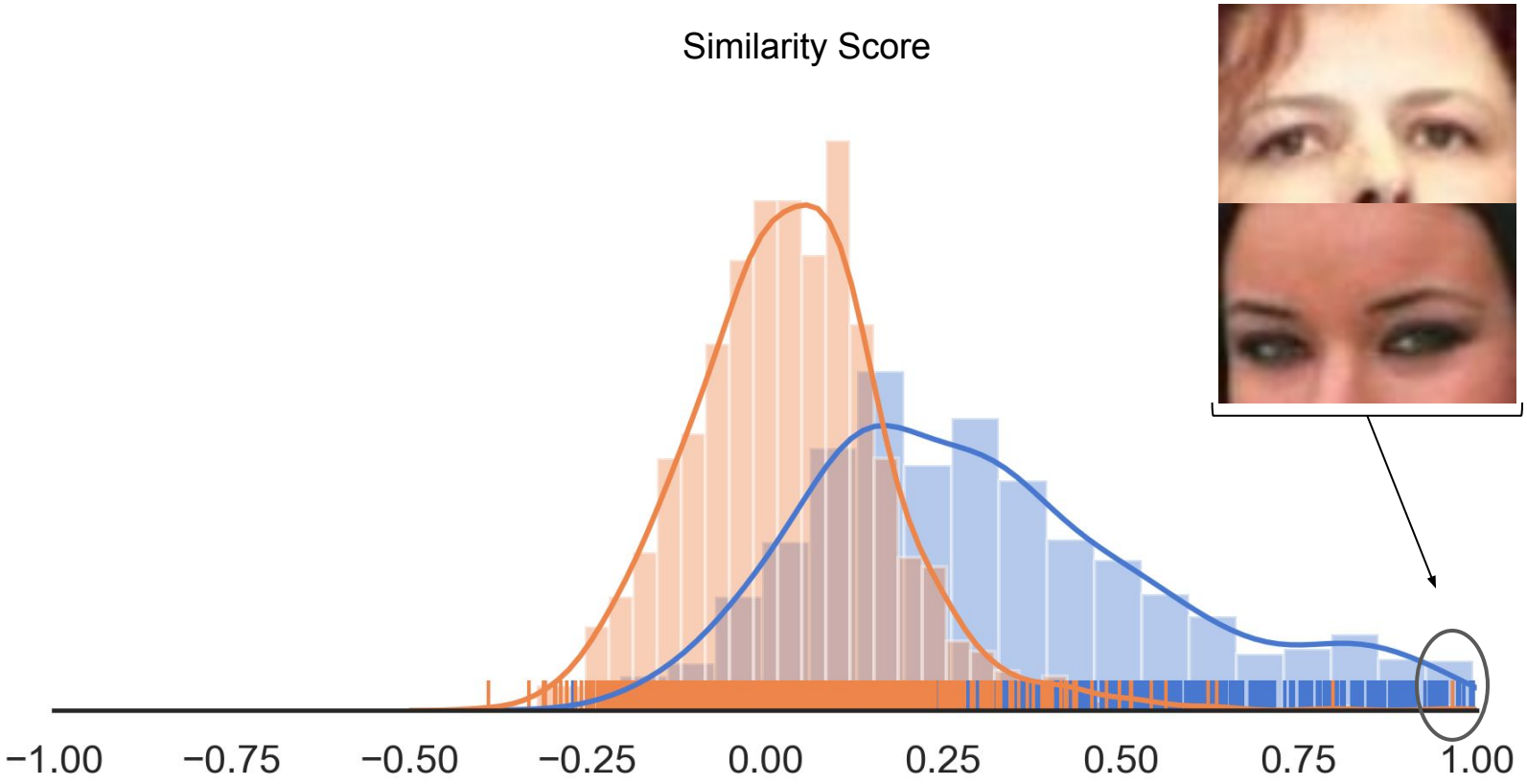
Zero false positive face recognition

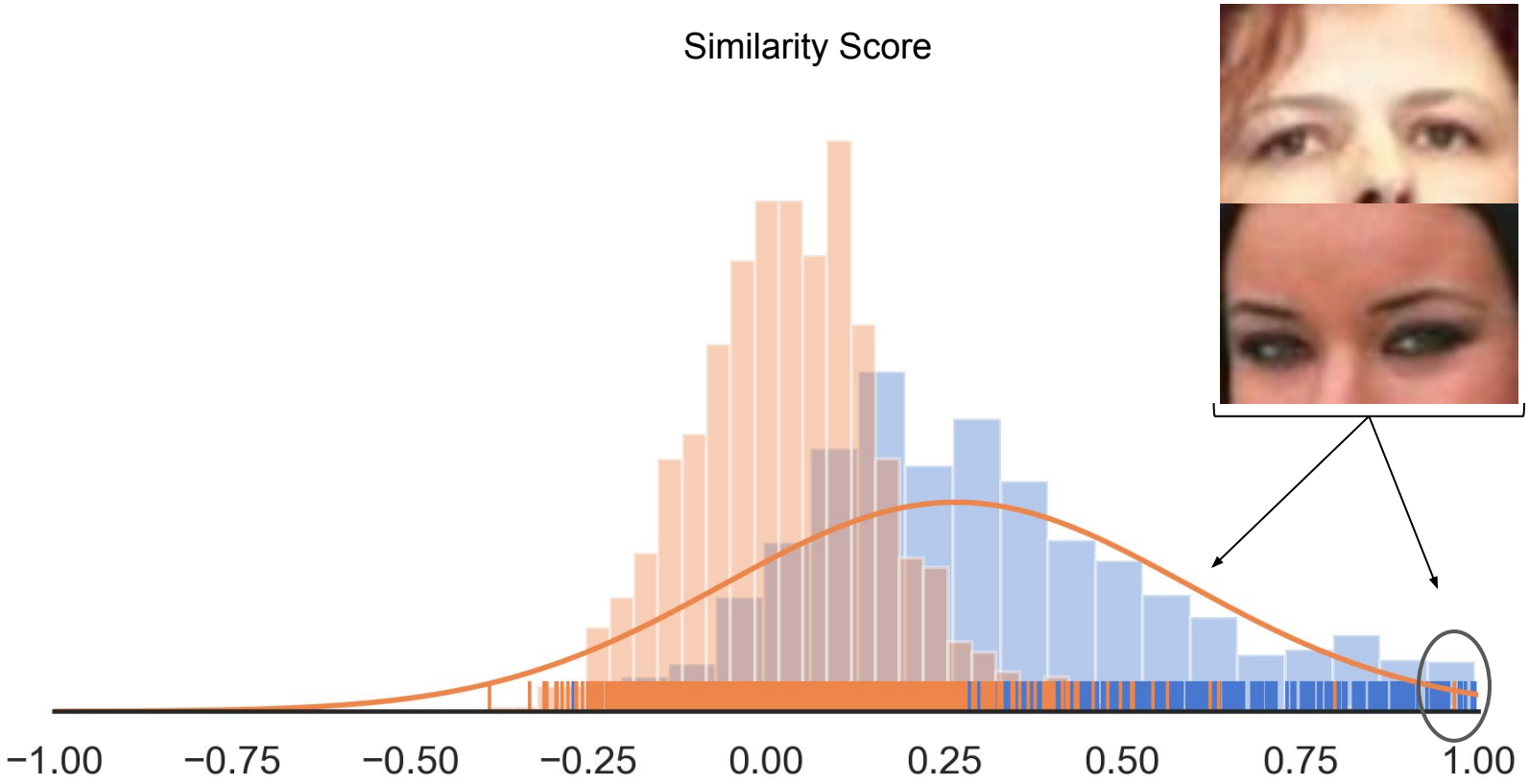
Similarity Score

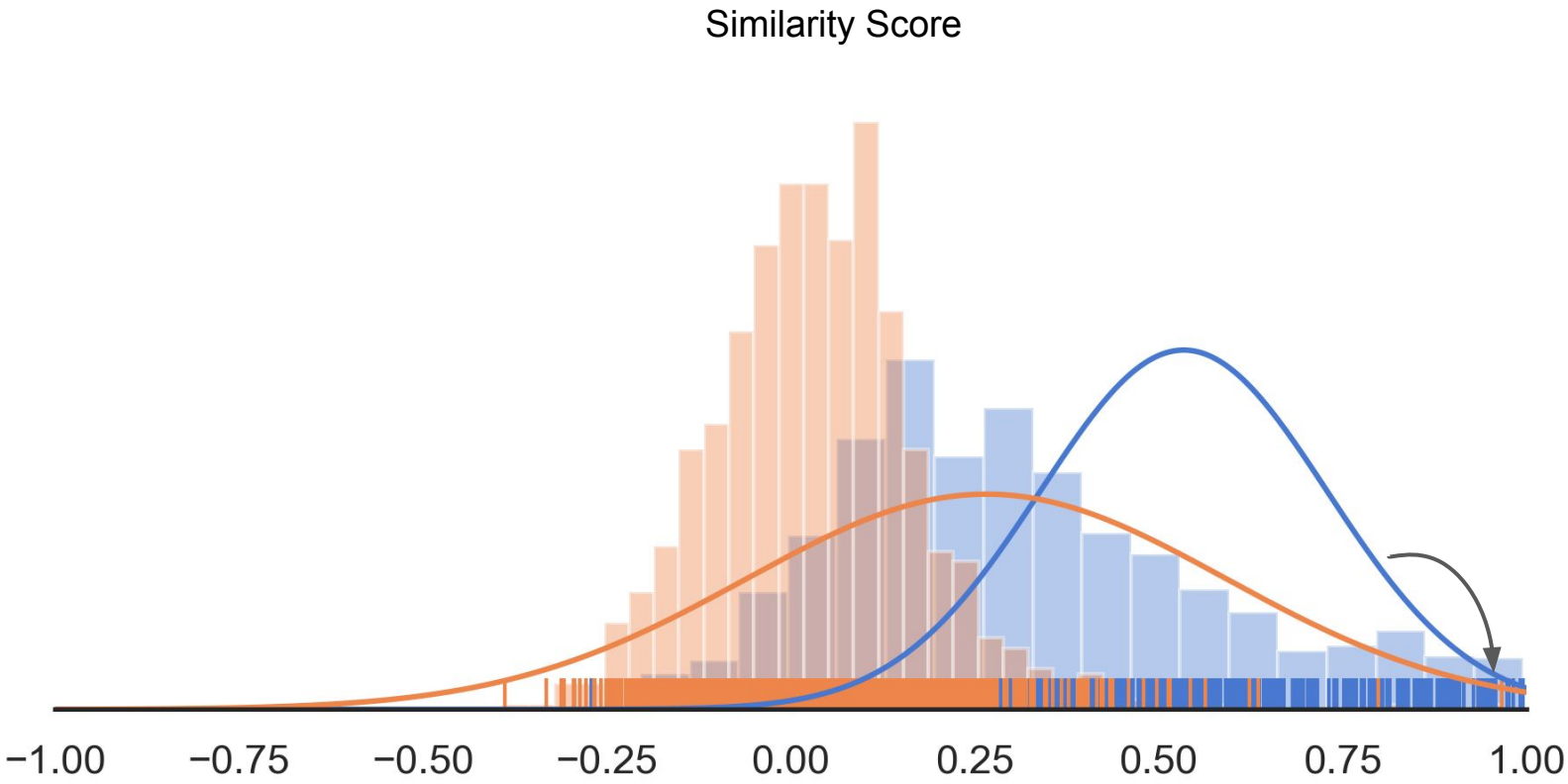


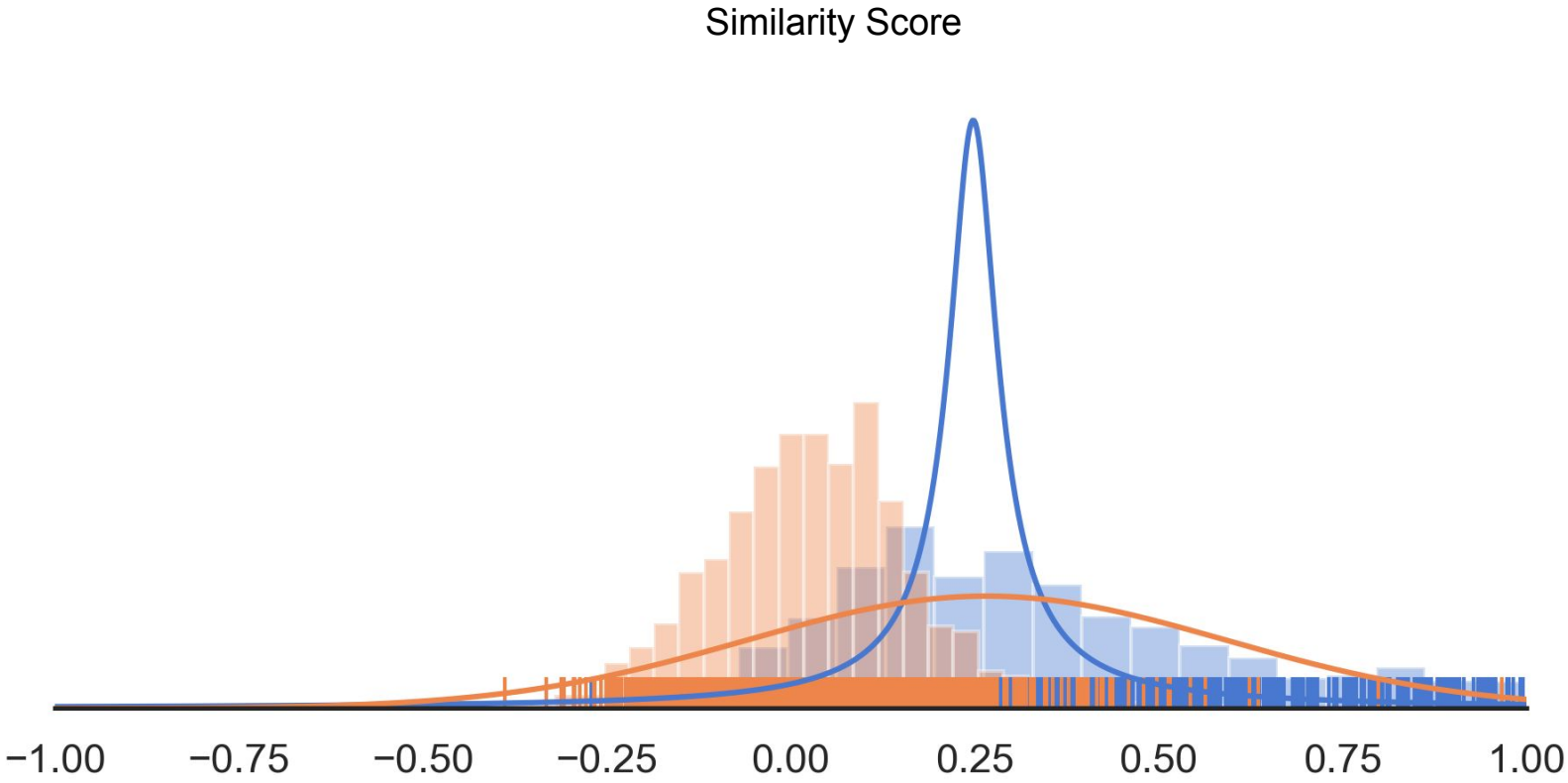
Similarity Score

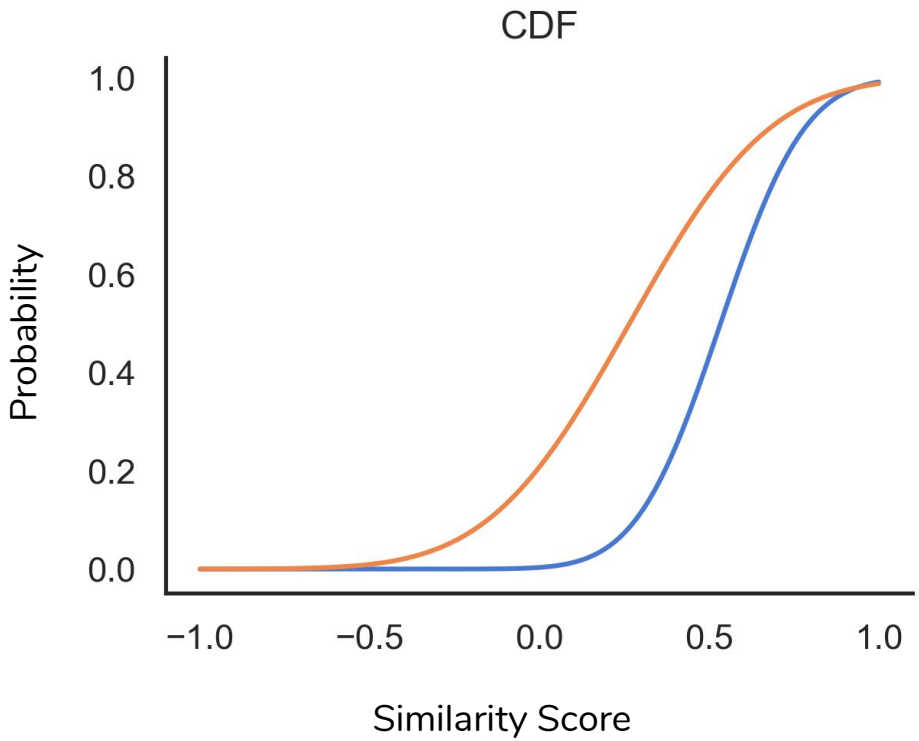


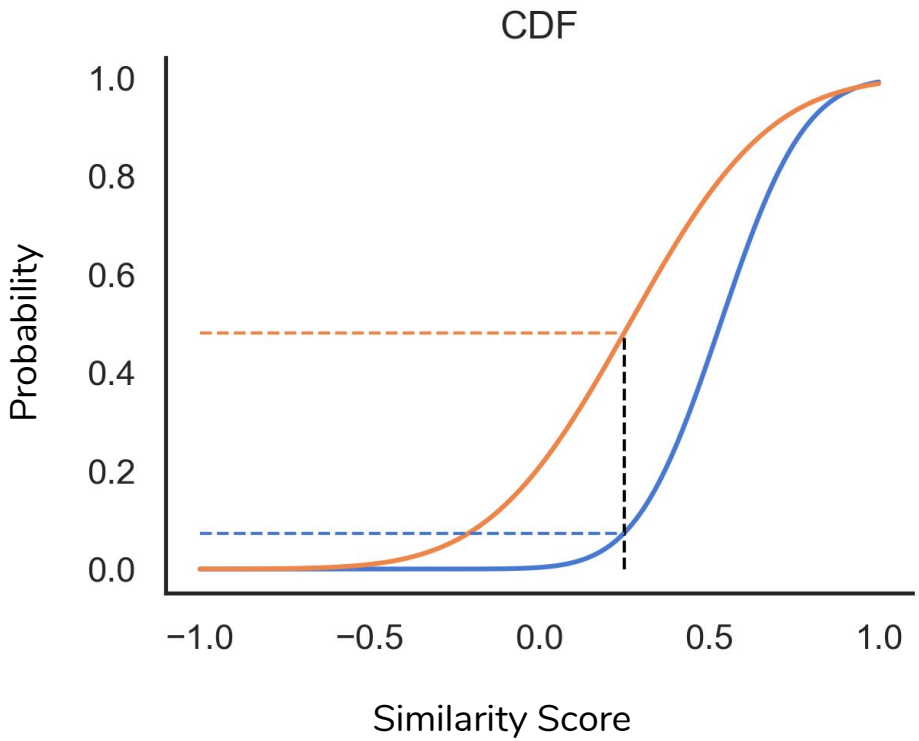

Similarity Score
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

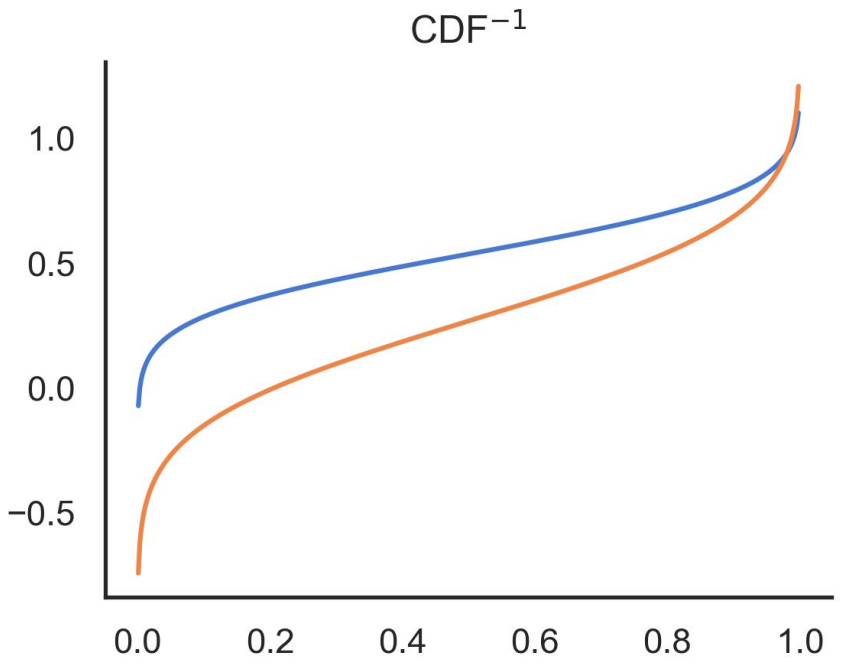

Similarity Score
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

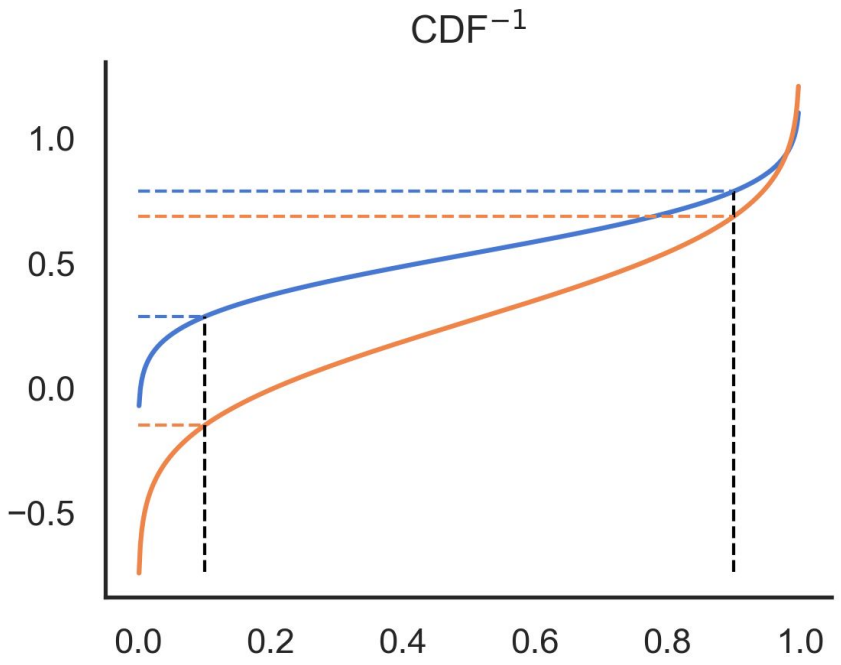

Similarity Score
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

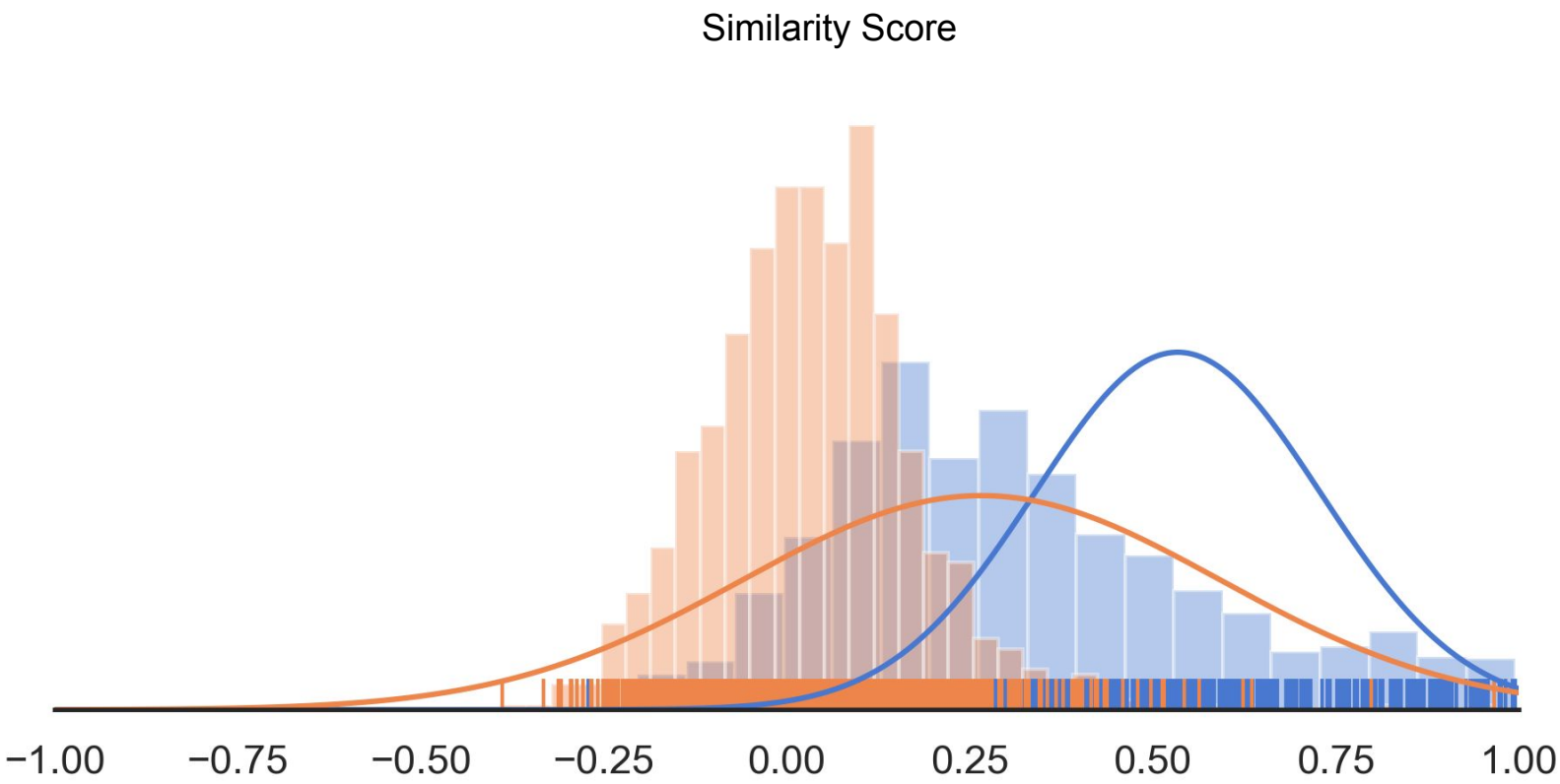

Similarity Score
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

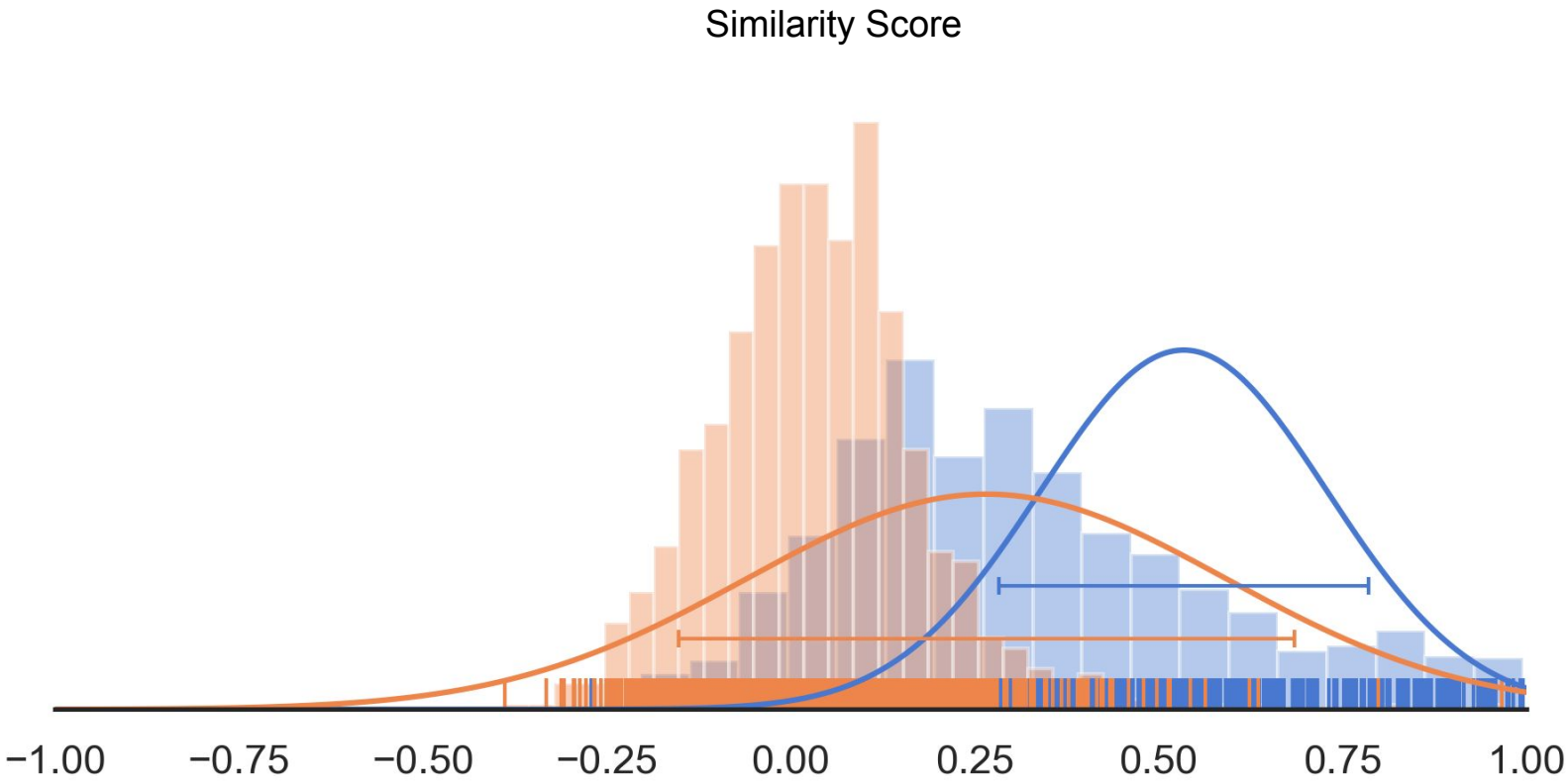

CDF
1.0
0.8
0.6
0.4
0.2
0.0
Probability
−1.0
−0.5
0.0
0.5
1.0
Similarity Score


CDF
1.0
0.8
0.6
0.4
0.2
0.0
Probability
−1.0
−0.5
0.0
0.5
1.0
Similarity Score


CDF^{-1}
1.0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0


CDF^{-1}
1.0
0.5
0.0
−0.5
0.0
0.2
0.4
0.6
0.8
1.0


Similarity Score
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00


Similarity Score
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

Probability density estimation

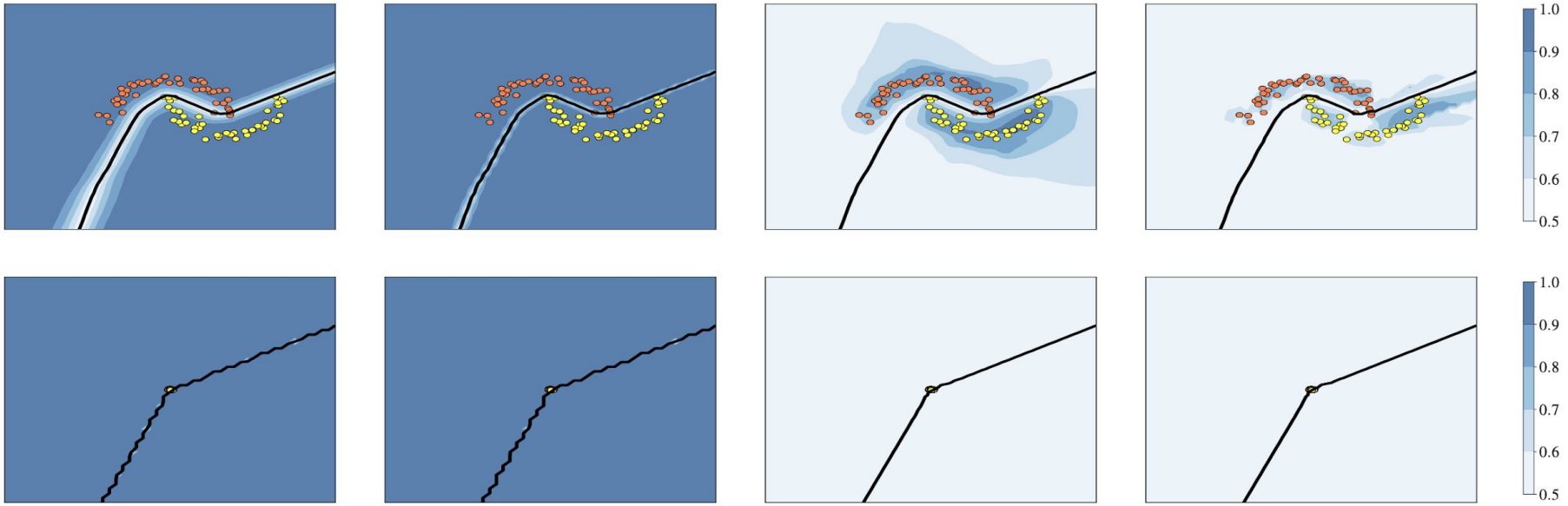


Figure 1. Binary classification on a toy dataset using a MAP estimate, temperature scaling, and both last-layer and all-layer Gaussian approximations over the weights which are obtained via Laplace approximations. Background color and black line represent confidence and decision boundary, respectively. Bottom row shows a zoomed-out view of the top row. The Bayesian approximations—even in the last-layer case—give desirable uncertainty estimates: confident close to the training data and uncertain otherwise. MAP and temperature scaling yield overconfident predictions. The optimal temperature is picked as in Guo et al. (2017).

Being Bayesian, Even Just a Bit, Fixes Overconfidence in ReLU Networks,
A. Kristiadi et al., ICML 2020

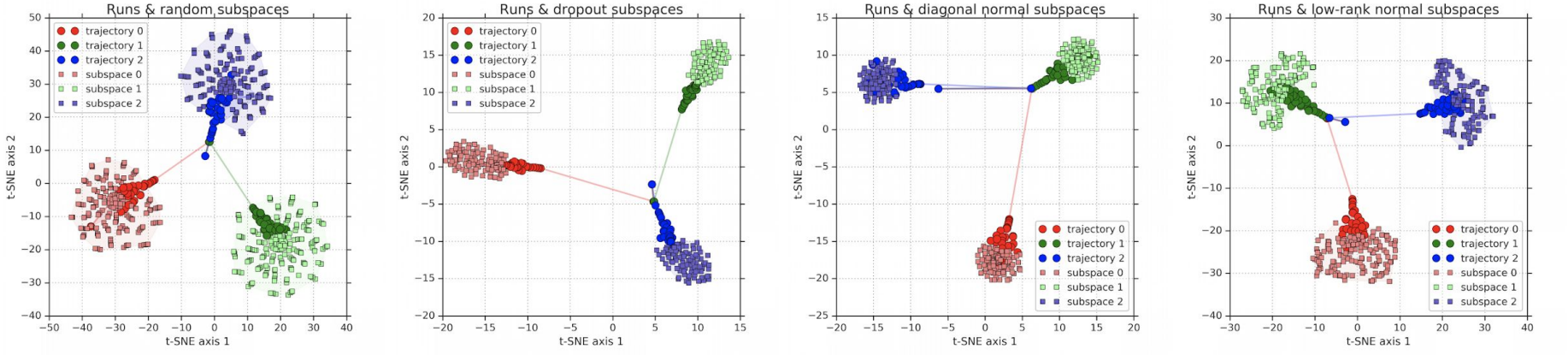


Figure 4: *Results using SimpleCNN on CIFAR-10*: t-SNE plots of validation set predictions for each trajectory along with four different subspace generation methods (showed by squares), in addition to 3 independently initialized and trained runs (different colors). As visible in the plot, the subspace-sampled functions stay in the prediction-space neighborhood of the run around which they were constructed, demonstrating that truly different functions are not sampled.

Deep Ensembles: A Loss Landscape Perspective,
S. Fort et al., ICML 2020

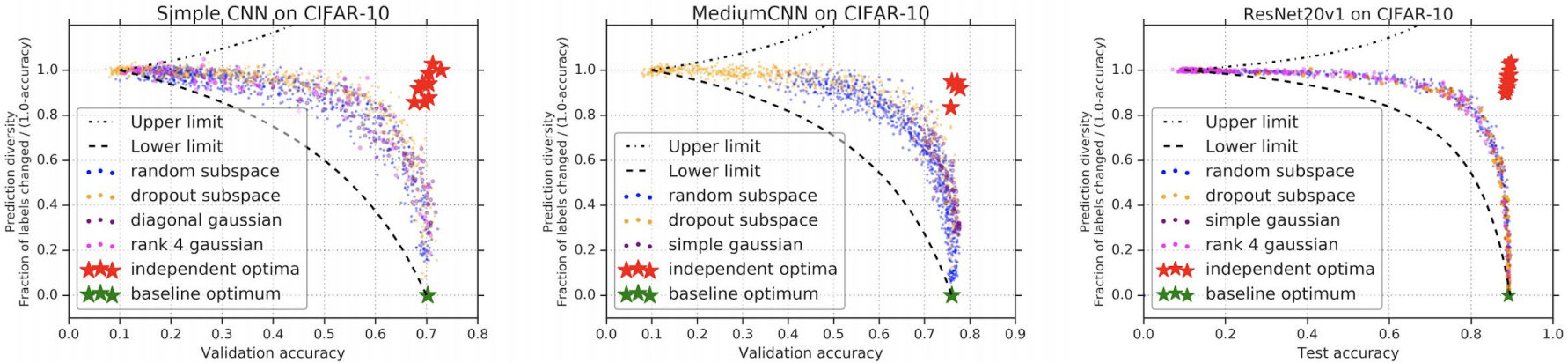


Figure 5: *Diversity versus accuracy plots for 3 models trained on CIFAR-10: SmallCNN, MediumCNN and a ResNet20v1.* The clear separation between the subspace sampling populations (for 4 different subspace sampling methods) and the population of independently initialized and optimized solutions (red stars) is visible. The 2 limiting curves correspond to solution generated by perturbing the reference solution's predictions (bottom curve) and completely random predictions at a given accuracy (upper curve).

Deep Ensembles: A Loss Landscape Perspective,
S. Fort et al., ICML 2020

Summary

Probability density estimation instead of point estimation of the similarity scores

Overly confident false matches can be prevented

Better evaluation metrics for fail-safe face recognition?

mosalam@trueface.ai

nchafni@trueface.ai - CTO

shaunpmoore@trueface.ai - CEO