



# A Probabilistic Framework for Spoofing Aware Speaker Verification

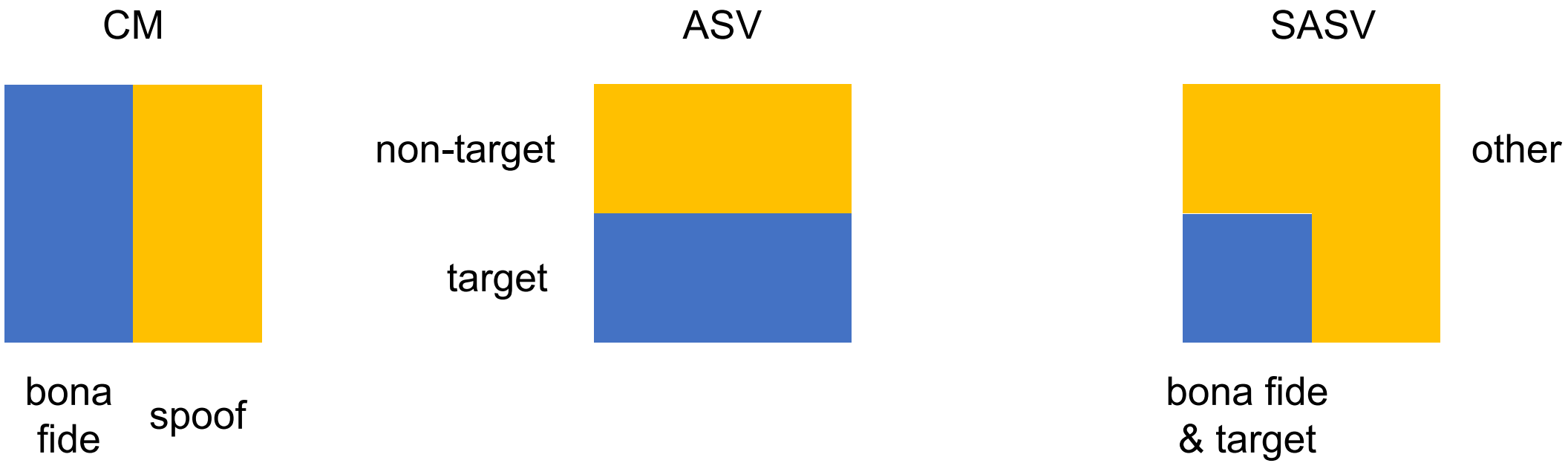
***You Zhang***, *Ge Zhu, Zhiyao Duan*

University of Rochester, USA


UNIVERSITY of ROCHESTER

# Spoofing aware speaker verification

- Automatic Speaker Verification (ASV)
- Anti-spoofing / Spoofing Countermeasure (CM)

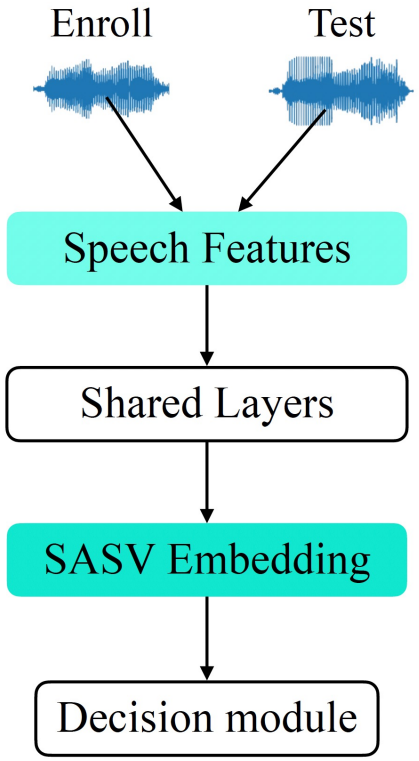




You (Neil) Zhang you.zhang@rochester.edu

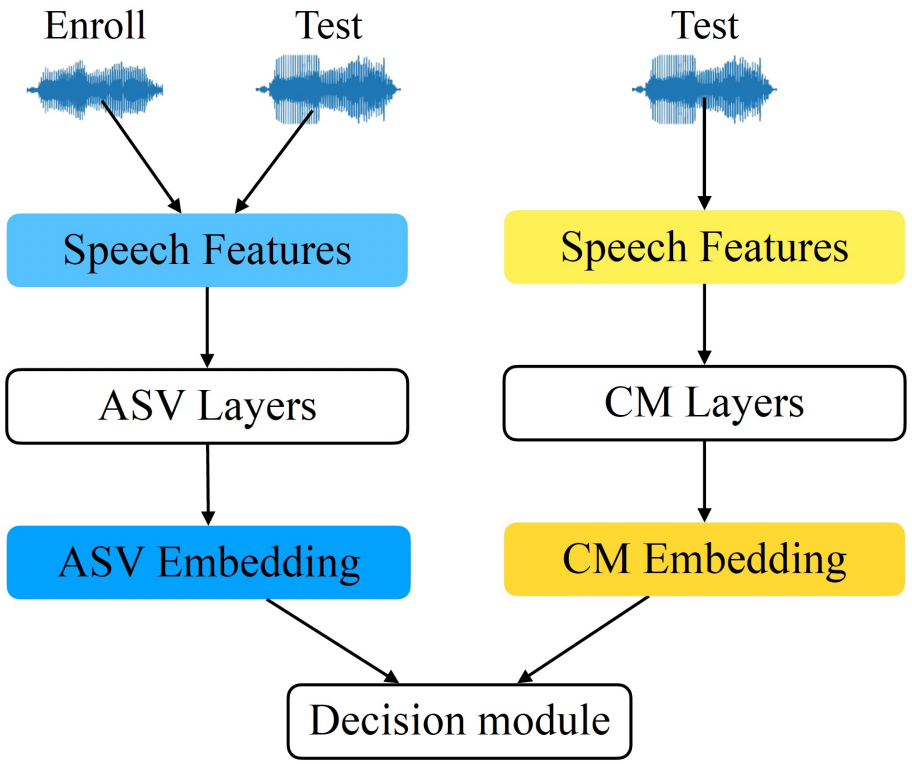
# Existing methods

- Multi-task learning-based methods
  - Require both speaker and spoofing labels (ASVspoof dataset)
  - Might overfit the seen speakers



# Existing methods

- Fusion-based methods
  - Score-level fusion
  - Embedding-level fusion



# SASV challenge 2022

- Dataset

Table 1: Summary of the ASVspoof 2019 LA dataset.

| Partition | #speakers | Bona fide | Spoofing attacks | |
|---|---|---|---|---|
| | | #utterances | #utterances | Attacks type |
| Train | 20 | 2,580 | 22,800 | A01 - A06 |
| Dev | 20 | 2,548 | 22,296 | A01 - A06 |
| Eval | 67 | 7,355 | 63,882 | A07 - A19 |

- Official protocols
  - For each test trial, multiple corresponding enrollment utterances
  - Listing the target, non-target, spoof trials

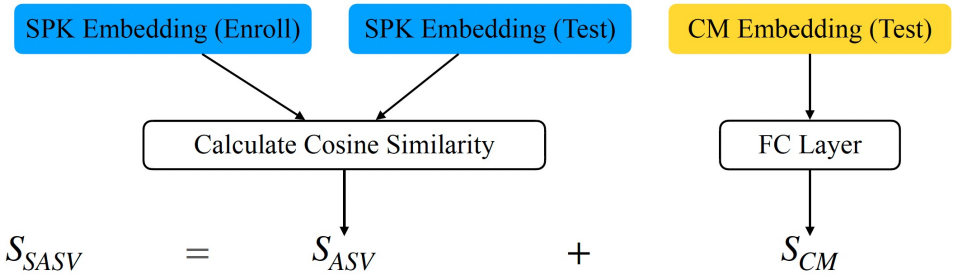
# SASV challenge 2022

- Evaluation metrics

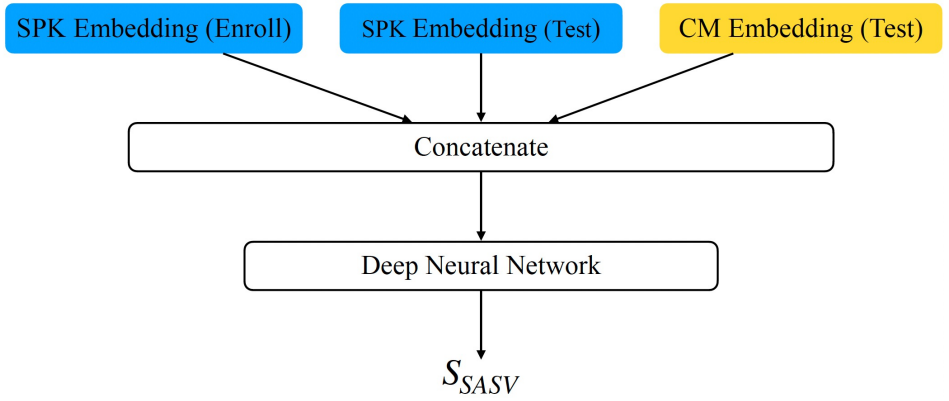
Table 2: Three kinds of EERs for evaluation (Adapted from [15]). "+" denotes the positive class and "-" denotes the negative class. A blank entry denotes classes not used in the metric. SASV-EER is the primary metric for the SASV challenge.

| Evaluation metrics | Target | Non-target | Spoof |
|---|---|---|---|
| SASV-EER | + | - | - |
| SV-EER | + | - | |
| SPF-EER | + | | - |

- Baseline methods

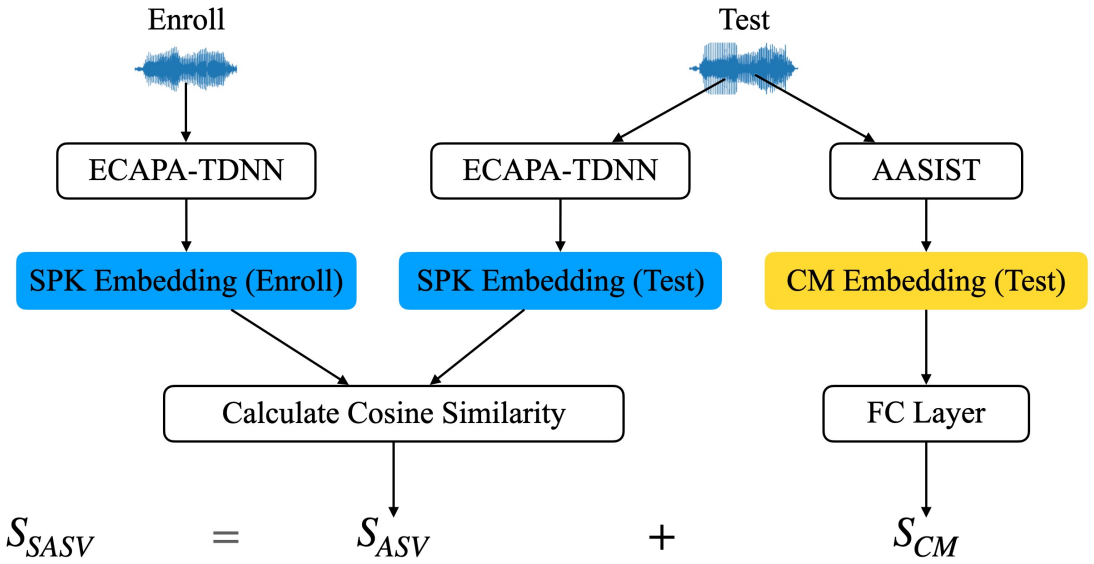



(a) Baseline1 (Score-level fusion)



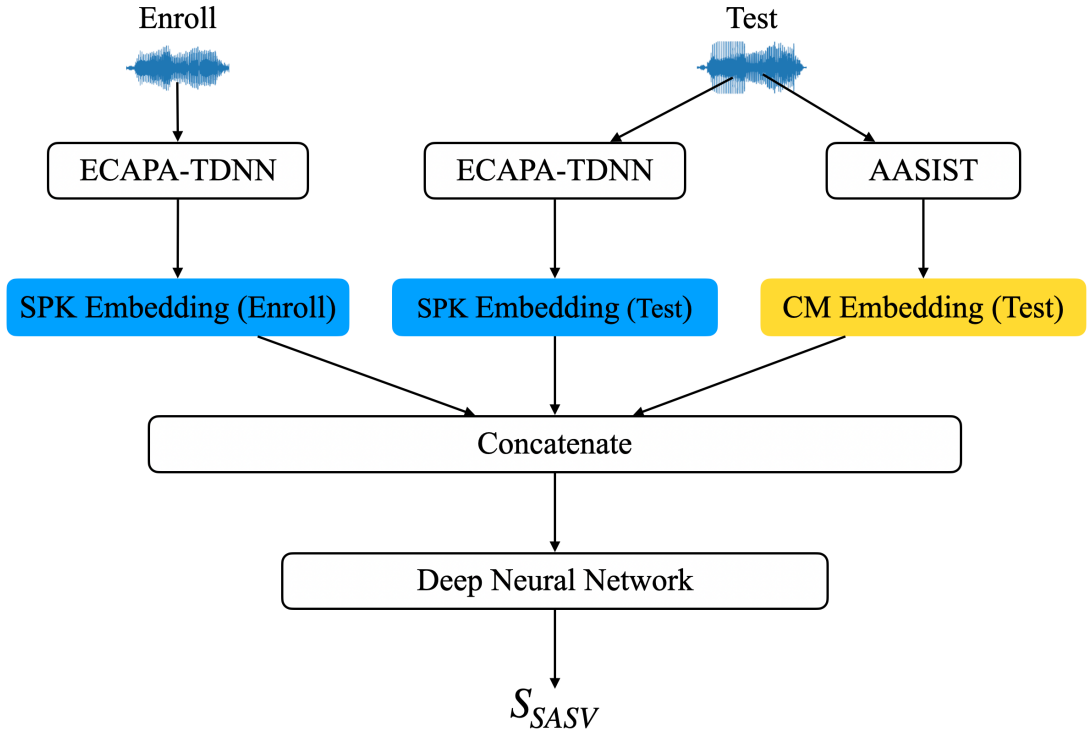

(b) Baseline2 (Embedding-level fusion)

# Challenge baseline methods




Baseline1:

Score-level fusion

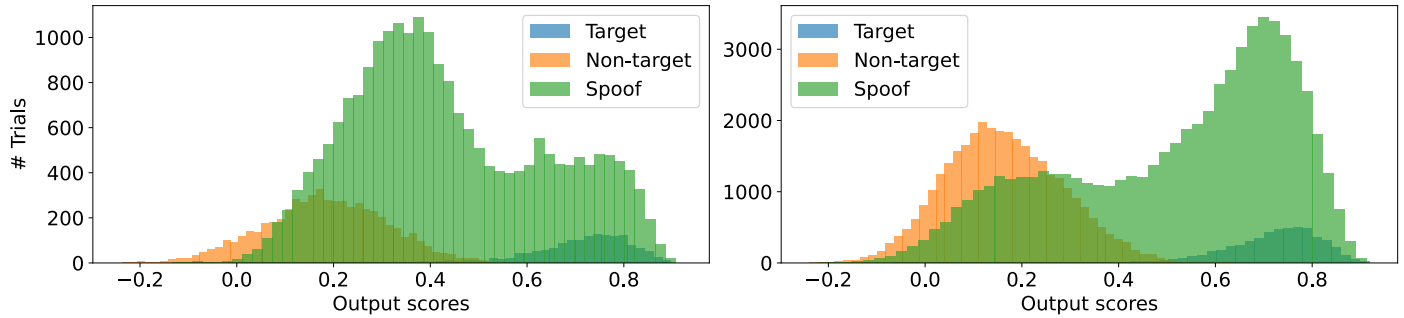



Baseline2:

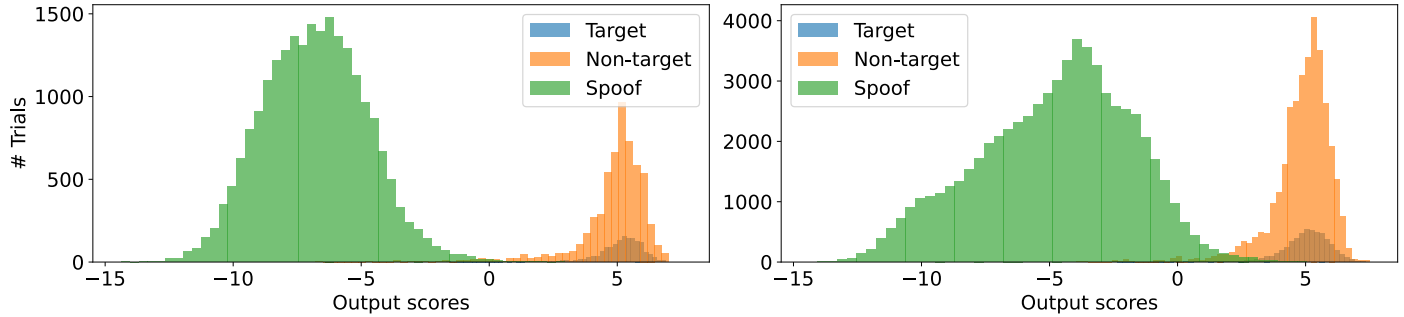
Embedding-level fusion

# Score distribution analysis

- Separate systems



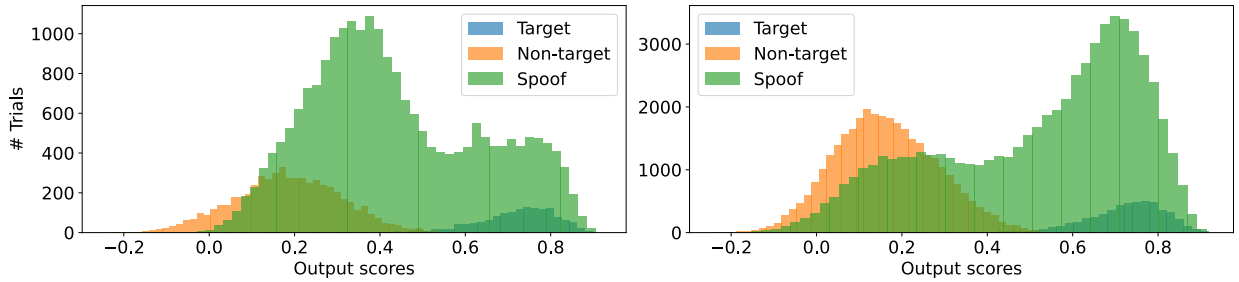

ECAPA-TDNN (ASV system)



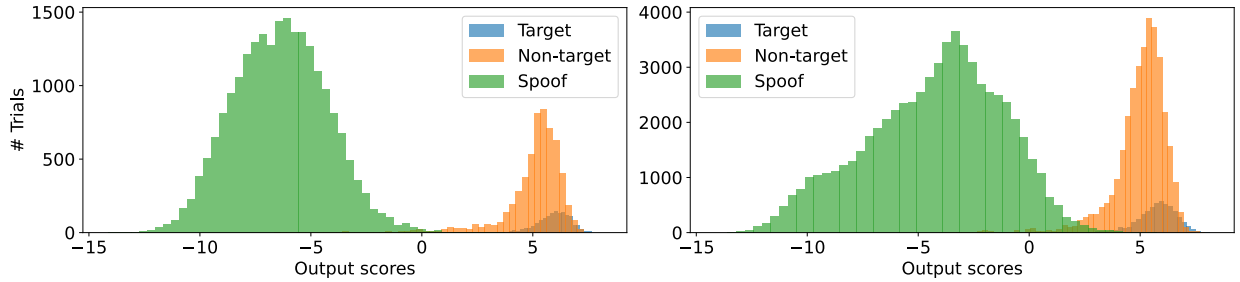

AASIST (CM system)

# Score distribution analysis

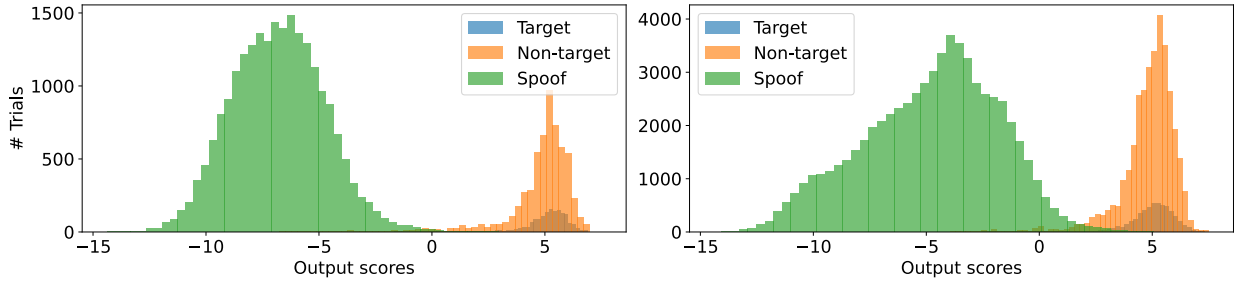
- Baseline systems



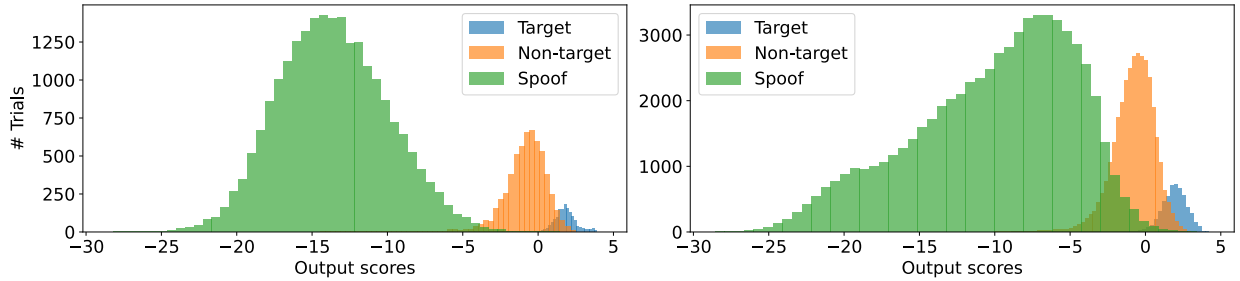

ECAPA-TDNN (ASV system)




Baseline1

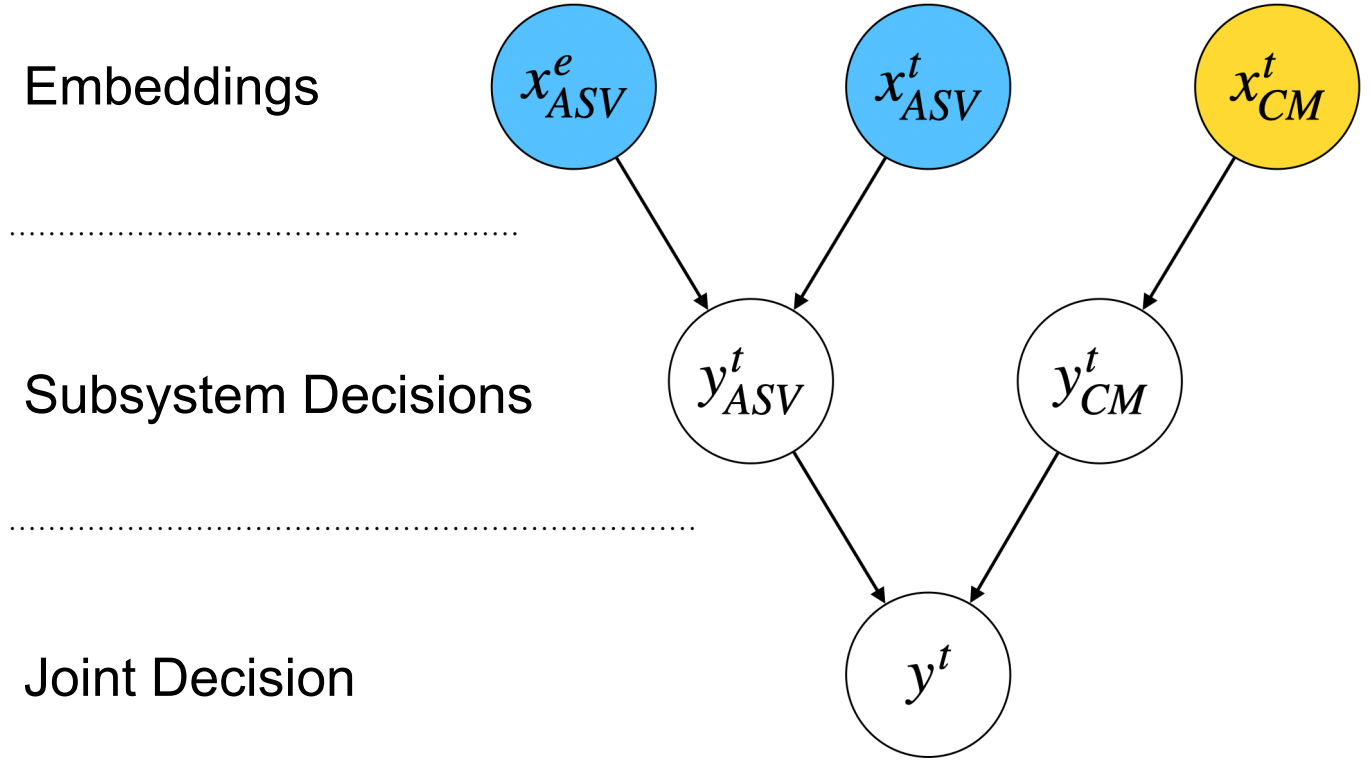



AASIST (CM system)




Baseline2

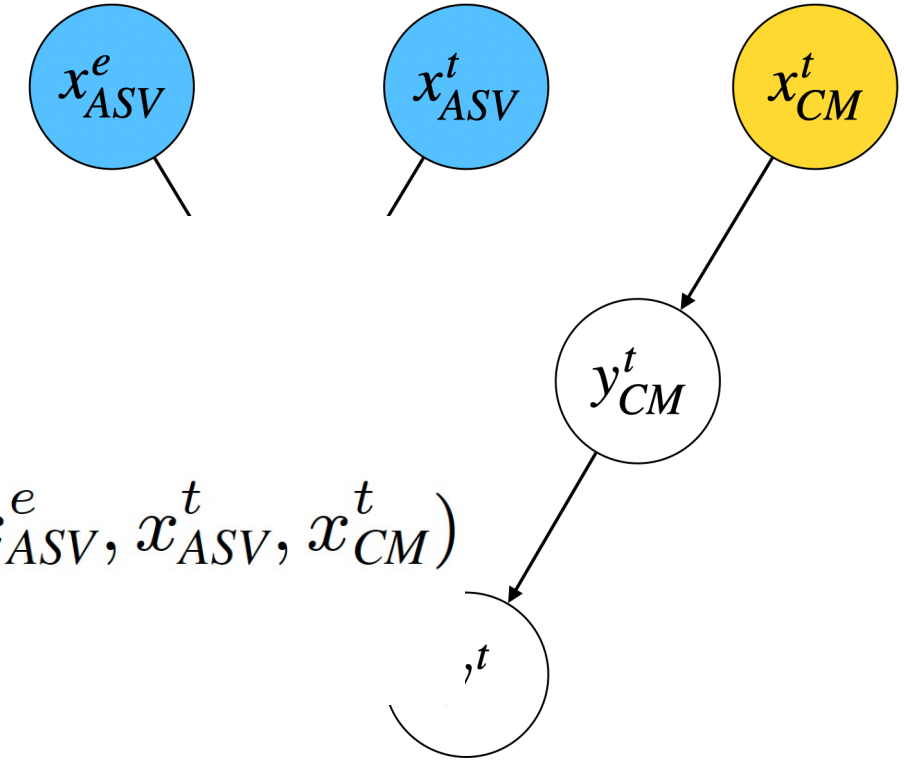
# Probabilistic fusion framework



# Probabilistic fusion framework

- Product rule (PR)

Embeddings




$$
\begin{aligned}
& P(y^t = 1 | x^e_{ASV}, x^t_{ASV}, x^t_{CM}) \\
=& P(y^t_{ASV} = 1, y^t_{CM} = 1 | x^e_{ASV}, x^t_{ASV}, x^t_{CM}) \\
=& P(y^t_{ASV} = 1 | x^e_{ASV}, x^t_{ASV}, x^t_{CM}) P(y^t_{CM} = 1 | x^e_{ASV}, x^t_{ASV}, x^t_{CM}) \\
=& P(y^t_{ASV} = 1 | x^e_{ASV}, x^t_{ASV}) P(y^t_{CM} = 1 | x^t_{CM}).
\end{aligned}
$$

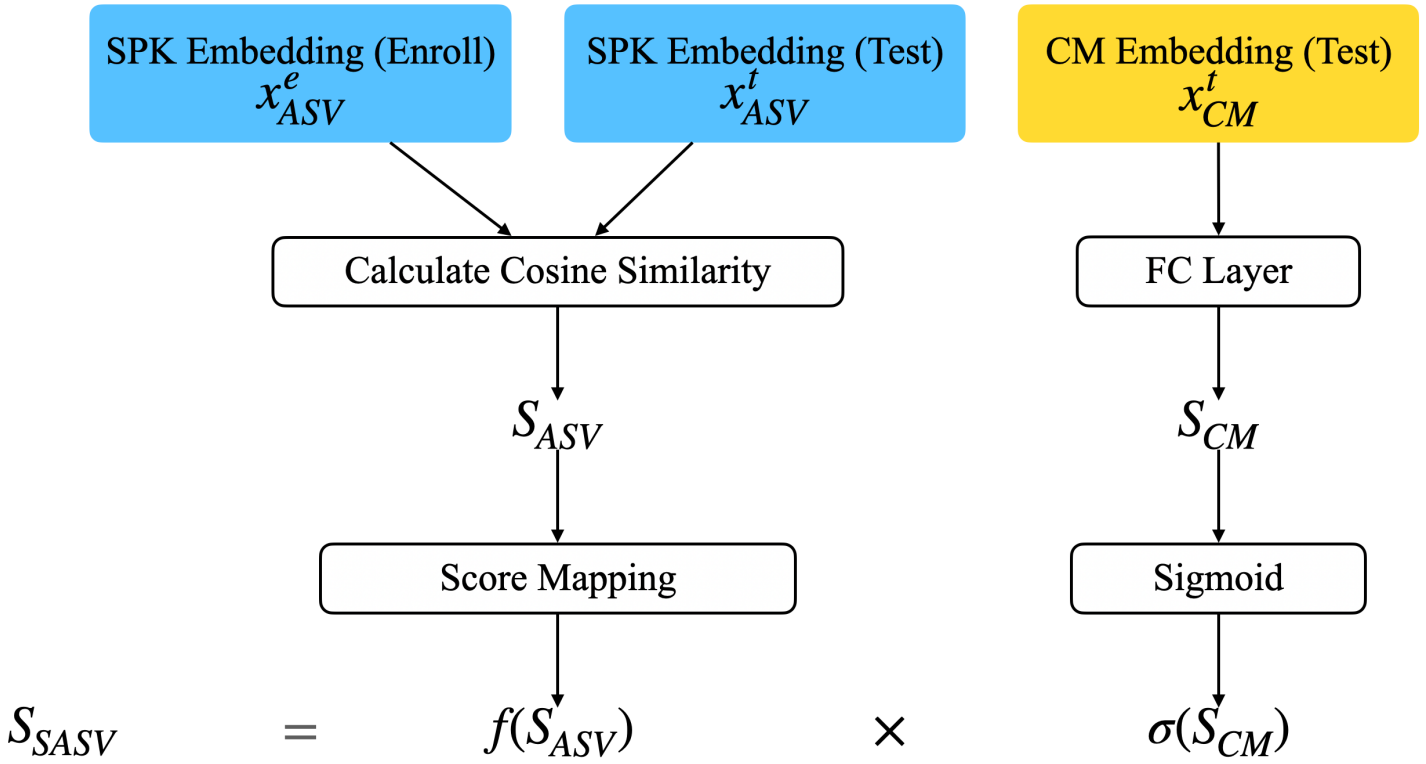
# Proposed strategies

- Direct inference strategy

$$P(y^t_{CM} = 1|x^t_{CM}) = \sigma(\mathcal{S}_{CM}).$$

$$P(y^t_{ASV} = 1|x^e_{ASV}, x^t_{ASV}) = f(\mathcal{S}_{ASV}),$$

$$\mathcal{S}_{SASV} = \sigma(\mathcal{S}_{CM}) \times f(\mathcal{S}_{ASV}).$$



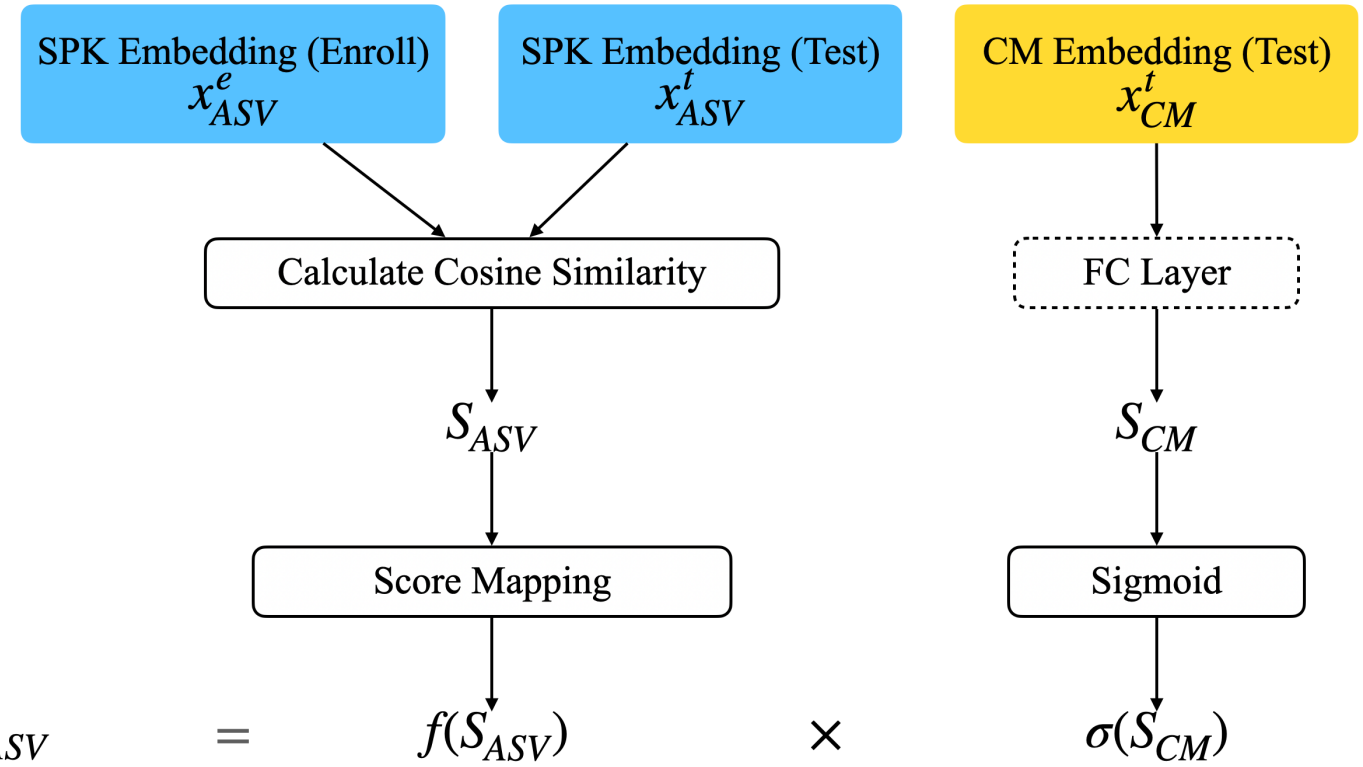
# Proposed strategies

- Fine-tuning strategy

$$
\begin{aligned}
&P(y^t = 1 | x^e_{ASV}, x^t_{ASV}, x^t_{CM}) \\
=&P(y^t_{ASV} = 1, y^t_{CM} = 1 | x^e_{ASV}, x^t_{ASV}, x^t_{CM}) \\
=&P(y^t_{ASV} = 1 | x^e_{ASV}, x^t_{ASV}, x^t_{CM}) P(y^t_{CM} = 1 | y^t_{ASV}, x^e_{ASV}, x^t_{ASV}, x^t_{CM}) \\
=&P(y^t_{ASV} = 1 | x^e_{ASV}, x^t_{ASV}) P(y^t_{CM} = 1 | y^t_{ASV}, x^t_{CM}).
\end{aligned}
$$

# Proposed strategies

- Fine-tuning strategy
  - Re-train the FC layer by optimizing the $S_{SASV}$ based on the cross-entropy loss
  - Speaker-aware anti-spoofing



# Results


Table 3: Comparison of our proposed methods with separate systems and SASV challenge baselines.

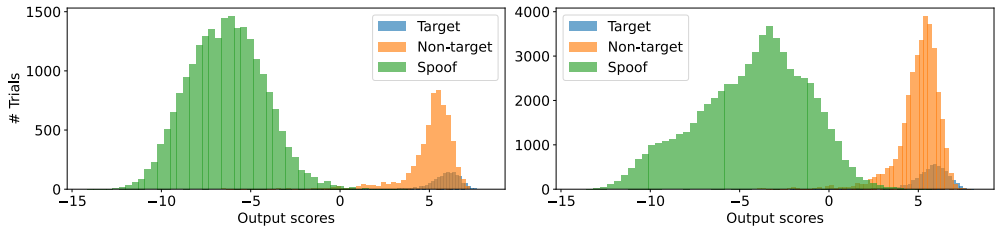
| Systems | SV-EER↓ | | SPF-EER↓ | | SASV-EER↓ | |
|---|---|---|---|---|---|---|
| | Dev | Eval | Dev | Eval | Dev | Eval |
| ECAPA-TDNN | 1.86 | 1.64 | 20.28 | 30.75 | 17.31 | 23.84 |
| AASIST | 46.01 | 49.24 | 0.07 | 0.67 | 15.86 | 24.38 |
| Baseline1 | 32.89 | 35.33 | 0.07 | 0.67 | 13.06 | 19.31 |
| Baseline2 | 7.94 | 9.29 | 0.07 | 0.80 | 3.10 | 5.23 |
| PR-L-I (Ours) | 2.13 | 2.14 | 0.11 | 0.86 | 1.21 | 1.68 |
| PR-S-I (Ours) | 2.43 | 2.57 | 0.07 | 0.78 | 1.34 | 1.94 |
| PR-C-I (Ours) | 1.95 | 1.64 | 0.97 | 2.94 | **1.08** | 2.70 |
| **PR-L-F (Ours)** | 2.02 | 1.92 | 0.07 | 0.80 | 1.10 | **1.54** |
| **PR-S-F (Ours)** | 2.02 | 1.94 | 0.07 | 0.80 | 1.10 | **1.53** |

PR: Product Rule

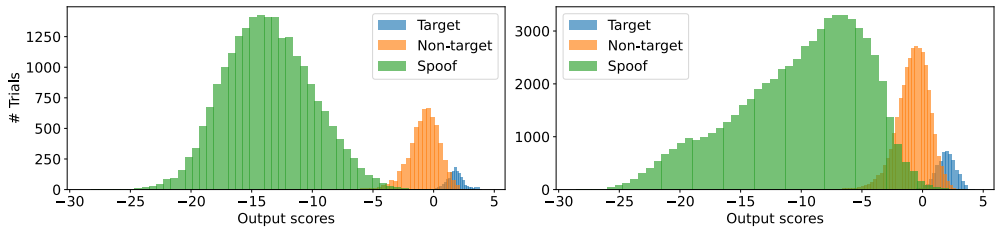
L: linear
S: Sigmoid
C: Calibrated

I: Inference
F: Fine-tuning

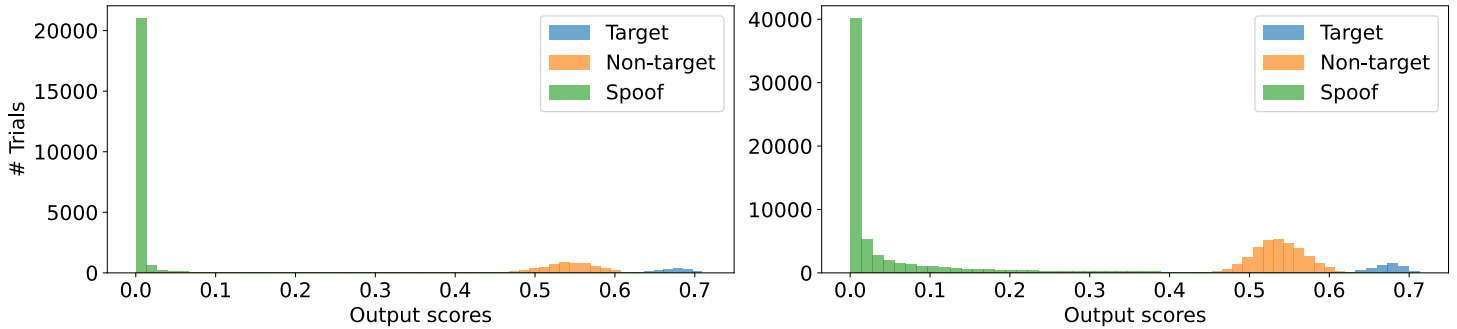
# Score distribution analysis




Baseline1




Baseline2



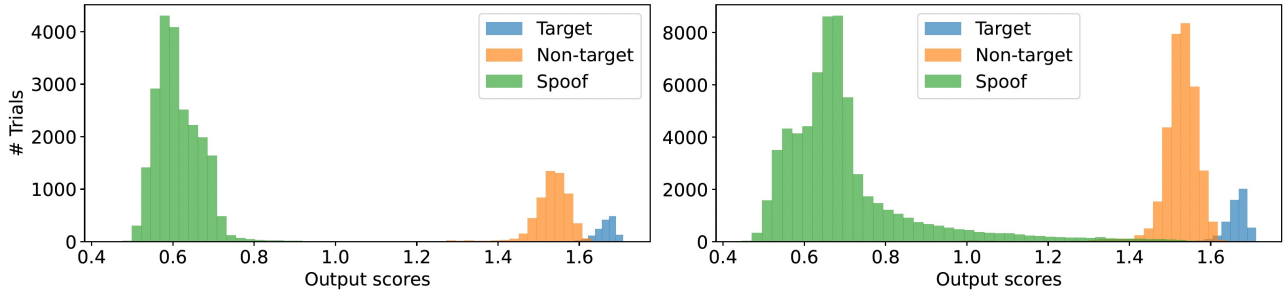

PR-S-F (Ours)


You (Neil) Zhang you.zhang@rochester.edu

# Ablation study

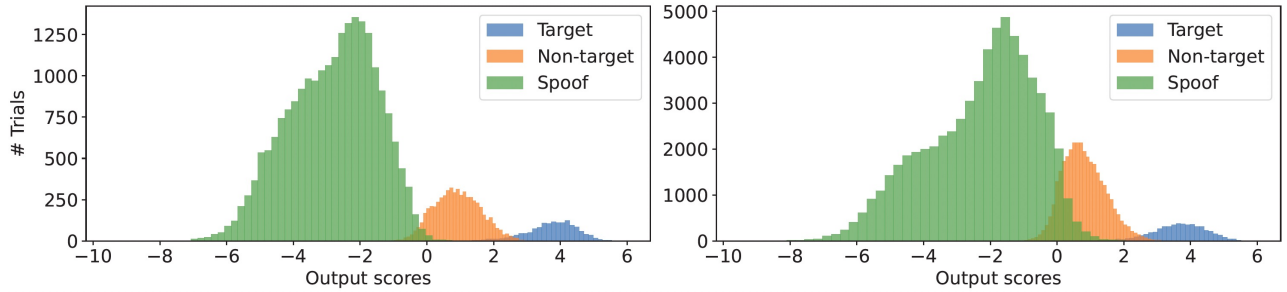
Table 4: Results of ablation study from our proposed best performing system PR−S−F to Baseline1.

| Systems | SASV-EER | |
|---|---|---|
| | Dev | Eval |
| PR−S−F (Ours) | 1.10 | 1.53 |
| PR−S−I (Ours) | 1.34 | 1.94 |
| Restore multiplication to sum (Baseline1 + score mapping) | 1.69 | 2.45 |
| Remove score mapping (Baseline1 + score multiplication) | 2.16 | 2.89 |
| Restore both (Baseline1) | 13.06 | 19.31 |




Baseline1 + score mapping

$$\mathcal{S}_{SASV} = \sigma(\mathcal{S}_{CM}) + \sigma(\mathcal{S}_{ASV})$$




Baseline1 + score multiplication

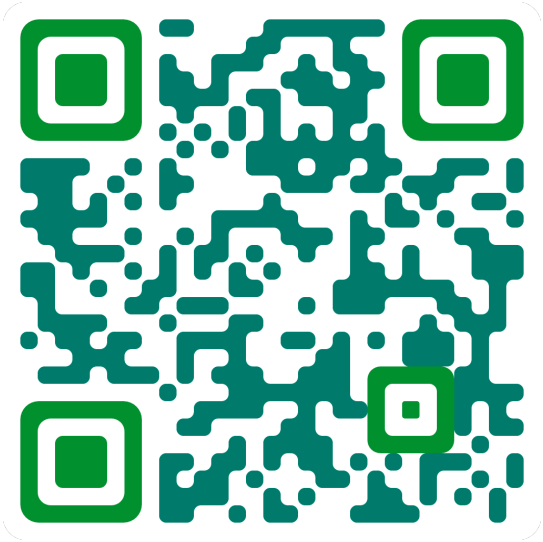
$$\mathcal{S}_{SASV} = \mathcal{S}_{CM} \times \mathcal{S}_{ASV}$$

# Takeaways

- Our proposed probabilistic framework is effective to spoofing aware speaker verification.
- With the product rule and fine-tuning strategy, we achieved 1.53% SASV-EER, which significantly improved the baseline methods.

# Resources

Full Paper



Code

Challenge Submission

# Thank you !

# Q & A