Eunju Park[1] Taekyung Kim[2] Minju Kim[3] Hojun Lee[4] Gil-Jun Lee[5]

[1]Ewha Womans Univ. [2]Dongguk Univ. [3]Chung-Ang Univ. [4]Keimyung Univ. [5]SK On

Interact Virtually

CVPR Nashville JUNE 11-15, 2025

Motivation

Precise anomaly localization is essential for industrial visual inspection.

Limitations of existing reverse distillation methods

- **Feature degradation** from student over-compression leads to poor detection of subtle or small anomalies (e.g., transistor defects).
- **Insufficient semantic transfer** via layer-wise skips limits accuracy on complex parts.

▶ **Our solution**: Asymmetric skip connections enable cross-hierarchical feature transfer, significantly improving anomaly localization and detection performance.

Framework of SK-RD4AD
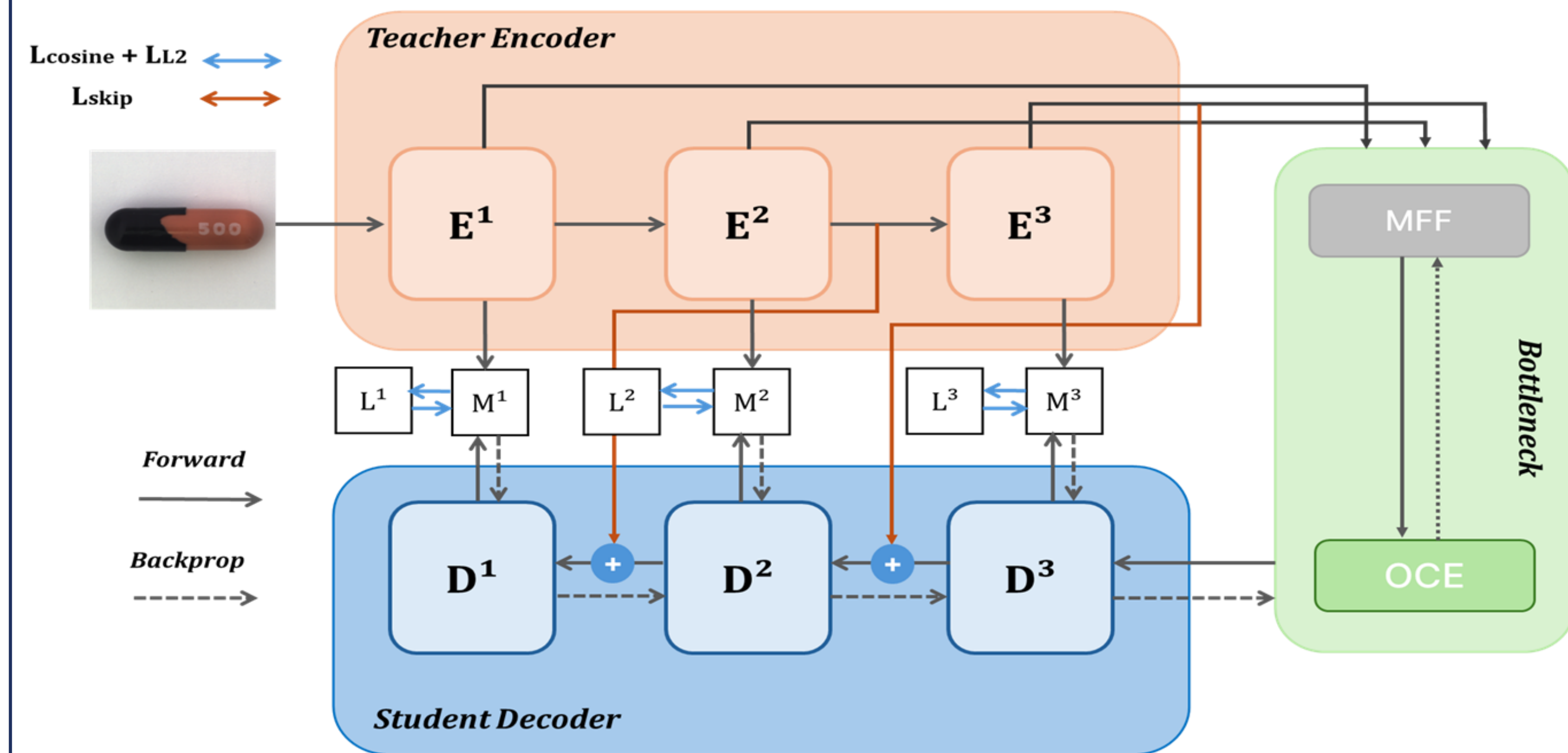
1) **Asymmetric Skip Knowledge Module:**
   Asymmetric skip connections from teacher encoder (E^2, E^3) to student decoder (D^1, D^2) enable cross-hierarchical feature transfer and preserve semantic information for accurate anomaly localization.

2) **Dual-Loss Supervision:**
   Combining cosine similarity and skip consistency losses guides the student to mimic teacher features and align representations, enhancing both global and local details.

Methods

0. Notations

X : input image
T(·) : teacher encoder
S(·) : student decoder
E^1, E^2, E^3 : teacher encoder features
D^1, D^2, D^3 : student decoder features
L_{cosine} : cosine similarity loss
L_{skip} : skip consistency loss
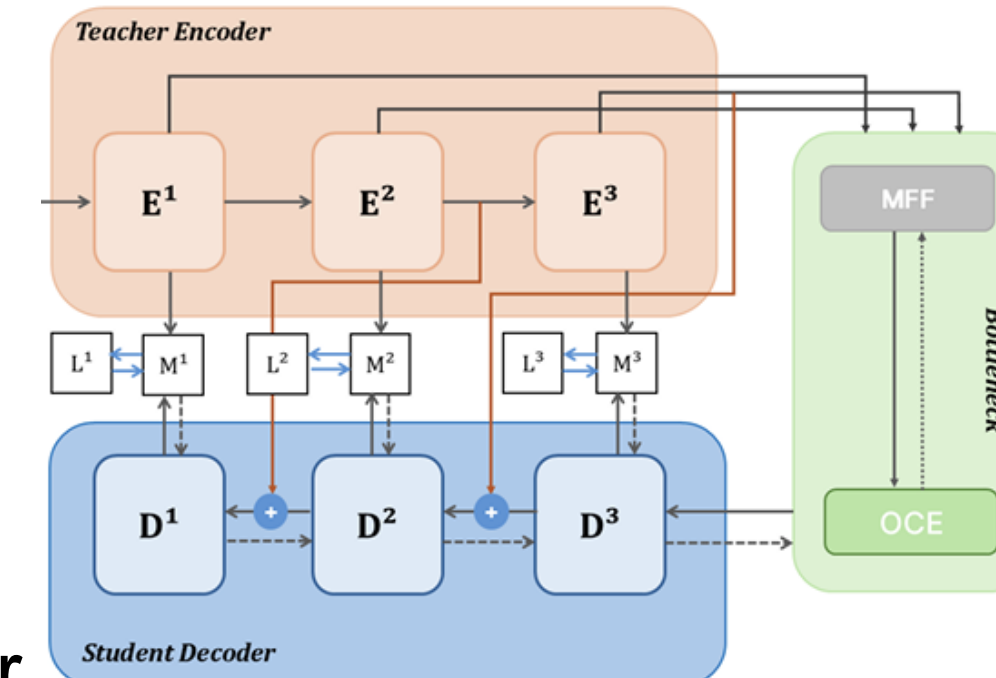
1. Asymmetric Skip Knowledge Module

1) **Non-corresponding Skip Connections**
   - Connects $E^2 \rightarrow D^1$, $E^3 \rightarrow D^2$
   - Enables cross-hierarchical feature transfer
   - **Preserves both low-level texture and high-level semantics**

2) **Feature Fusion**
   : D1 = ReLU(Conv1x1(E^1)) + E^2
   : D2 = ReLU(Conv1x1(D^1)) + E^3
   - Enhances decoder's ability to reconstruct complex features

2. Triple-Loss Supervision

1) **Cosine Similarity Loss**
   : L_{cosine} = 1 - cosine(F_T, F_S)
   - Guides student to mimic teacher's feature direction

2) **Skip Consistency Loss**
   : L_{skip} = 0.5 * (|| E^2 - D^1 ||² + || E^3 - D^2 ||²)
   - Enforces alignment between transferred and reconstructed features

3) **L2 Reconstruction Loss**
   : L_{recon} = ||X − S(T(X))||²
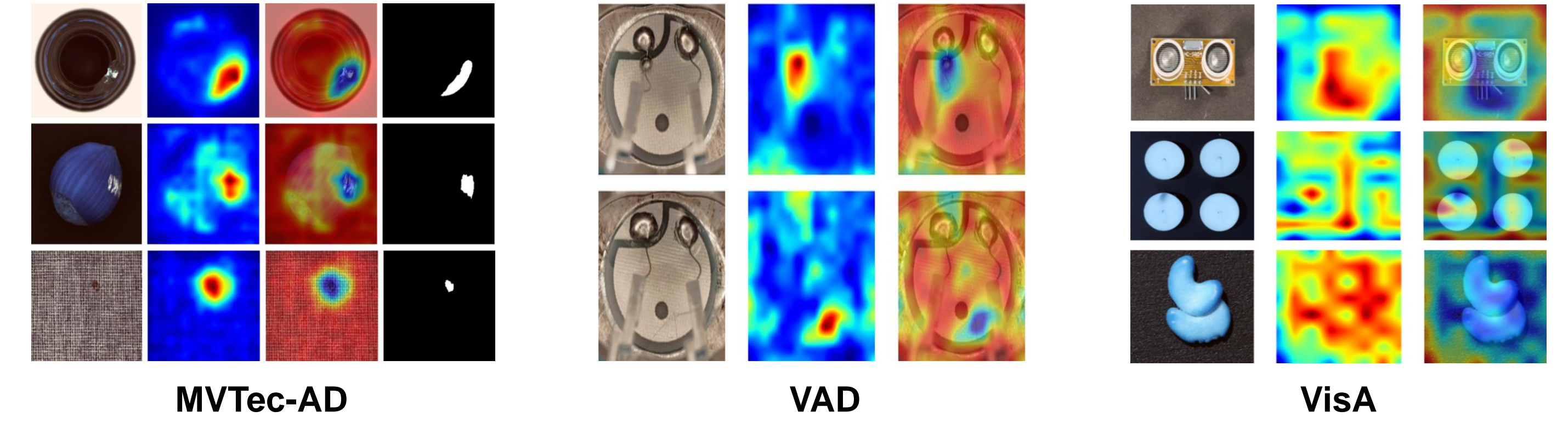   - Ensures pixel-level reconstruction fidelity.

3. Total Objective

$$L_{total} = \alpha * L_{recon} + \beta * L_{cosine} + \gamma * L_{skip}$$

▶The three losses **ensure pixel-level, feature-level, and hierarchical consistency.**

Experiments

▶ **Qualitative Analysis of MVTec-AD, VAD and VisA datasets**

MVTec-AD   **VAD**   **VisA**

▶ **Quantitative Analysis of MVTec-AD, VisA datasets**

Category/Method		US [3]	MF [16]	SPADE [5]	PaDiM [6]	RIAD [17]	CutPaste [12]	RD4AD [7]	Ours
Textures	Carpet	- / 87.9	- / 87.8	97.5 / 94.7	99.1 / 96.2	96.3 / -	98.3 / -	98.9 / 97.0	**99.2 / 97.7**
	Grid	- / 95.2	- / 86.5	93.7 / 86.7	97.3 / 94.6	98.8 / -	- / -	99.3 / 97.6	**99.4 / 99.2**
	Leather	- / 94.5	- / 95.9	97.6 / 97.2	99.2 / 97.8	99.4 / -	99.5 / -	99.4 / 99.1	**99.6 / 99.2**
	Tile	- / 94.6	- / 88.1	87.4 / 75.9	94.1 / 86.0	89.1 / -	90.5 / -	95.6 / 90.6	**96.1 / 91.7**
	Wood	- / 91.1	- / 84.8	88.5 / 87.4	94.9 / 91.1	85.8 / -	95.5 / -	95.3 / 90.9	**95.3 / 90.8**
	Average	- / 92.7	- / 88.6	92.9 / 88.4	96.9 / 93.2	93.9 / -	96.3 / -	97.7 / 95.0	**97.92 / 95.74**
Objects	Bottle	- / 93.1	- / 88.8	98.4 / 95.5	98.3 / 94.8	97.6 / -	97.6 / -	98.7 / 96.6	**98.8 / 96.9**
	Cable	- / 81.8	- / 93.7	97.2 / 90.9	96.7 / 88.8	84.2 / -	90.0 / -	97.4 / 91.0	**98.0 / 92.9**
	Capsule	- / 96.8	- / 87.9	99.0 / 93.7	98.5 / 93.5	92.8 / -	97.4 / -	98.7 / 95.8	**98.7 / 96.2**
	Hazelnut	- / 96.5	- / 88.6	99.1 / 95.4	98.2 / 92.6	96.1 / -	97.3 / -	98.9 / 95.5	**99.1 / 96.2**
	Metal Nut	- / 92.0	- / 86.9	**98.1 / 94.4**	97.2 / 85.6	92.5 / -	93.1 / -	97.3 / 92.3	96.9 / 92.7
	Pill	- / 96.1	- / 93.0	96.5 / 94.6	95.7 / 92.7	95.7 / -	95.7 / -	98.2 / 96.4	**98.4 / 97.2**
	Screw	- / 94.2	- / 95.4	98.9 / 96.0	98.5 / 94.4	98.8 / -	96.7 / -	99.6 / 98.2	**99.6 / 98.5**
	Toothbrush	- / 93.3	- / 87.7	97.9 / 93.5	98.8 / 93.1	98.9 / -	98.1 / -	99.1 / **94.5**	99.1 / 94.3
	Transistor	- / 66.6	- / 92.6	94.1 / **87.4**	**97.5 / 84.5**	87.7 / -	93.0 / -	92.5 / 78.0	93.2 / 80.1
	Zipper	- / 95.1	- / 93.6	96.5 / 92.6	98.5 / 95.9	97.8 / -	99.3 / -	98.2 / 95.4	**98.5 / 96.3**
	Average	- / 90.8	- / 90.8	97.6 / 93.4	97.8 / 91.6	94.3 / -	95.8 / -	97.9 / 93.4	**98.13 / 94.16**
Total Average		- / 91.4	- / 90.1	96.5 / 91.7	97.5 / 92.1	94.2 / -	96.0 / -	97.8 / 93.9	**98.06 / 94.69**

Category/Method	PaDiM [6]	SPADE [5]	PatchCore [14]	RD4AD [7]	Ours
Candle	98.6 / **95.7**	97.9 / 93.2	**99.2** / 94.0	98.9 / 92.2	98.6 / 93.9
Capsule	97.4 / 74.9	60.7 / 36.1	96.5 / 85.5	**99.4** / 56.9	99.1 / **91.9**
Cashew	98.5 / 87.9	86.4 / 57.4	**99.2 / 94.5**	94.4 / 79.0	98.1 / 87.3
Chewing gum	**98.9** / 83.5	98.6 / 93.9	**98.9** / 84.6	97.6 / 92.5	97.7 / **94.3**
Fryum	95.4 / 80.2	**96.7** / 91.3	95.9 / **95.3**	96.4 / 81.0	**96.7** / 90.3
Macaroni1	99.1 / 92.1	96.2 / 61.3	98.5 / 95.4	99.3 / 71.9	**99.3 / 95.5**
Macaroni2	96.5 / 75.4	87.5 / 63.4	93.5 / 94.4	99.1 / 68.0	**99.3 / 95.2**
PCB1	**99.8 / 94.3**	66.9 / 38.4	99.8 / 94.3	99.6 / 43.2	99.6 / 93.7
PCB2	98.7 / 88.7	71.1 / 42.2	98.4 / 89.2	98.3 / 46.4	**98.3 / 91.7**
PCB3	98.7 / 84.9	95.1 / 80.3	98.9 / **90.9**	99.3 / 80.3	98.3 / 90.3
PCB4	97.9 / 81.6	89.0 / 71.6	**99.3 / 90.1**	98.2 / 72.2	98.6 / 89.0
Pipe fryum	**99.3** / 92.5	81.8 / 61.7	**99.3 / 95.7**	99.1 / 68.3	99.1 / 94.8
Average	98.1 / 85.9	85.6 / 65.9	98.1 / 91.2	98.3 / 70.9	**98.5 / 92.1**

▶ **Ablation Study**

Model	AUROC (%)	AUPRO (%)
RD4AD (Baseline)	97.3	92.3
SK-RD4AD (Ours)	**98.06**	**94.69**
Corresponding Skip Connections	97.72	93.50
No Skip Connections	96.3	91.4
Bottleneck Removal	97.91	94.33
Additional Skip Connection	97.90	94.17

Skip Connection Strategy Comparison

Factor	RD4AD	SK-RD4AD
Inference Time (s)	0.31	0.37 (+19%)
Memory Usage (MB)	352	401 (+13%)
Performance (AUROC)	97.3	**98.06 (+0.76)**

Computational Trade-Off
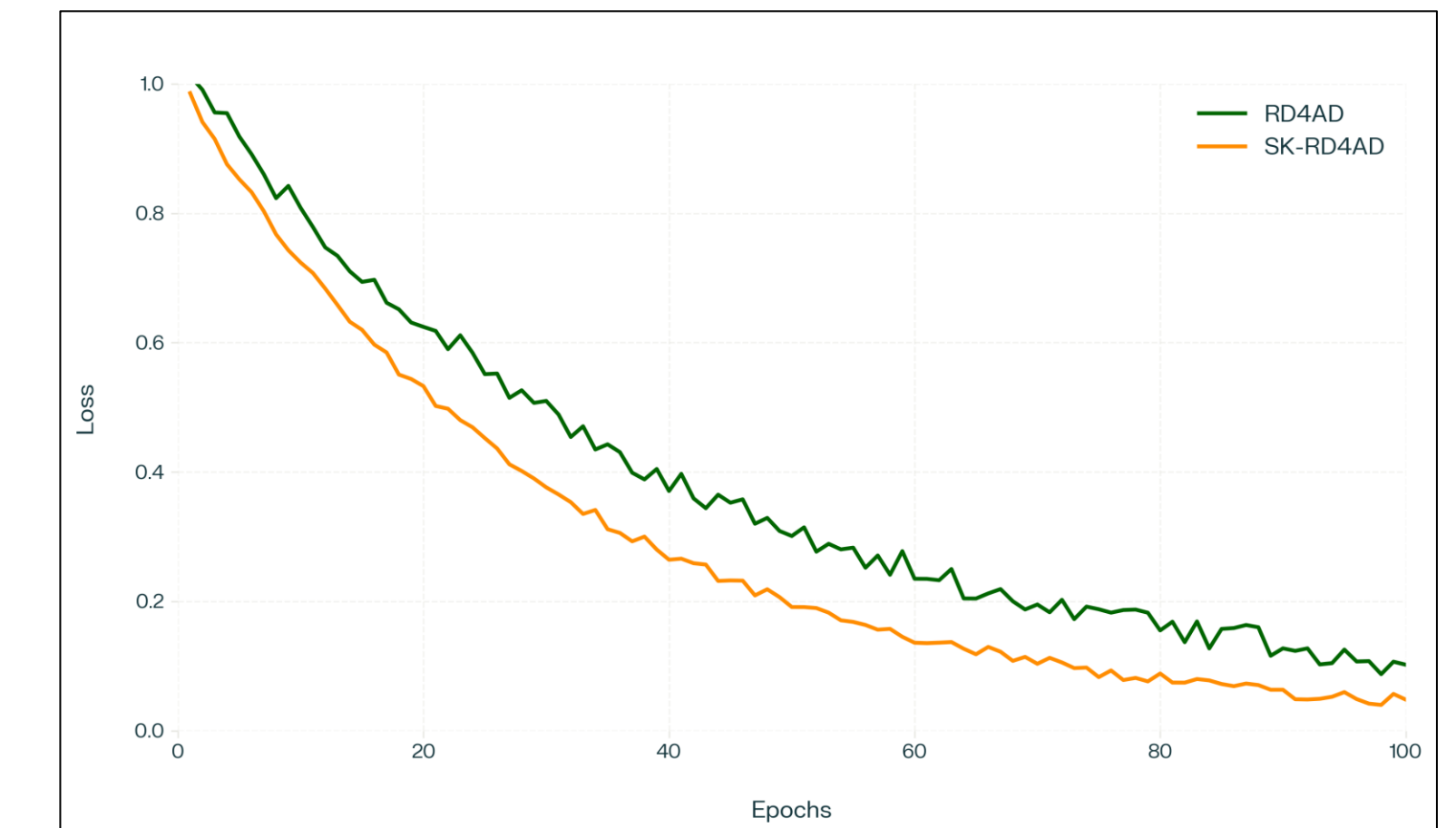
Training Convergence Curve

▶ **Key observation** : SK-RD4AD achieves **the highest AUROC (98.06%) and AUPRO (94.69%), converges faster and more stably than RD4AD**, and does so **with only minor increases in inference time and memory usage** — demonstrating superior accuracy and practical efficiency.