Tutor-Student Reinforcement Learning: A Dynamic Curriculum for Robust Deepfake Detection

Zhanhe Lei, Zhongyuan Wang*, Jikang Cheng, Baojin Huang,
Yuhong Yang, Zhen Han*, Chao Liang, Dengpan Ye
School of Computer Science, Wuhan University

Tutor-Student Reinforcement Learning: A Dynamic Curriculum for Robust Deepfake Detection

Motivation: The Generalization Gap & Dynamic Curricula

Suboptimal Uniform Weighting: Standard supervised training applies a uniform loss gradient across all samples, which is suboptimal for learning robust, generalizable features.

Persistent Hard Samples: Conventional baselines fundamentally struggle to reduce difficult instances over time, and this inability directly correlates with compromised generalization.

Static Curriculum Limits: Prior alternative approaches rely on static weightings based on predetermined properties or rigid pacing schedules that blindly ignore the real-time status of the model.

Dynamic Solution: Sample difficulty is not an intrinsic constant but is entirely relative to the detector's instantaneous learning state, requiring an actively adapting curriculum.

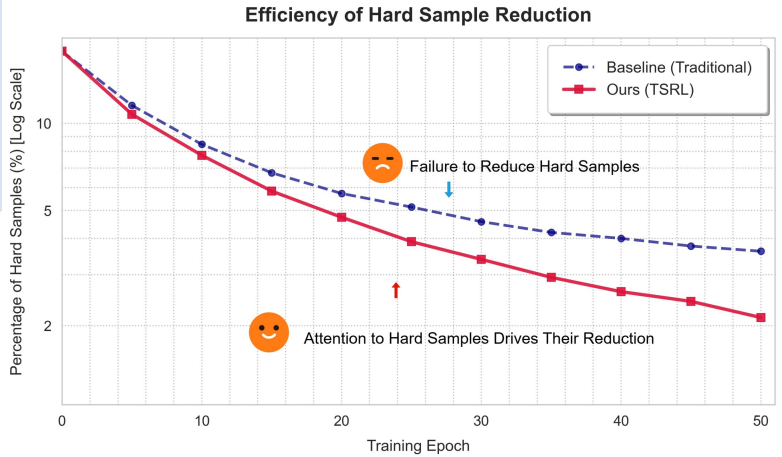
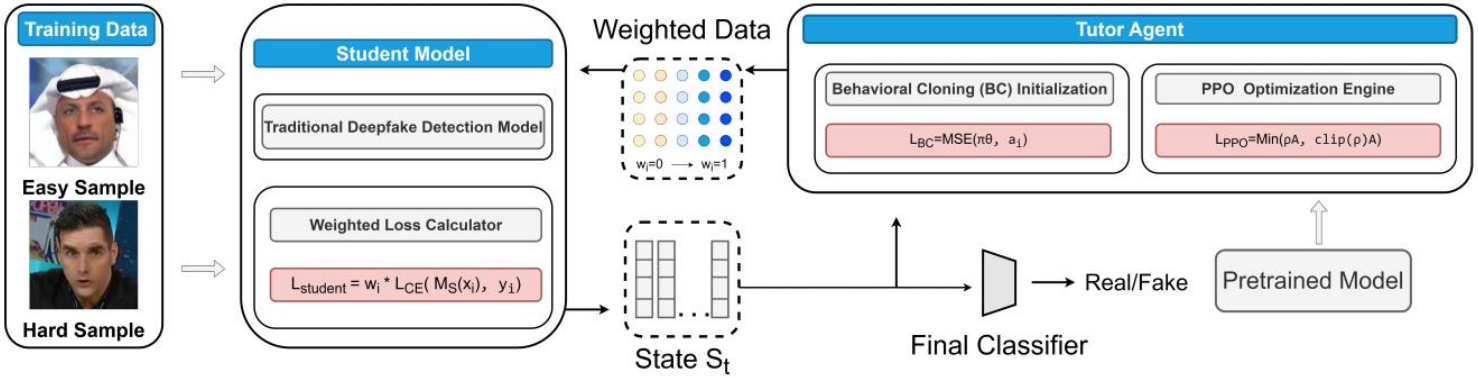


Figure1. Traditional baselines leave persistently high volumes of hard samples unresolved, whereas TSRL drives rapid mastery.

Tutor-Student Reinforcement Learning: A Dynamic Curriculum for Robust Deepfake Detection

Framework: Tutor-Student Reinforcement Learning (TSRL)



Our Tutor-Student Reinforcement Learning (TSRL) framework models detector training as a Markov Decision Process. A PPO-based "Tutor" observes each sample's history-aware state—including visual features, EMA loss, and forgetting counts—to assess long-term difficulty and instability. Based on this, the Tutor assigns a dynamic weight between 0 and 1 to reshape the Student's loss $L_{student} = w_i * L_{CE}(M_s(x_i), y_i)$in real-time. Rather than using delayed validation metrics, the Tutor is driven by an immediate state-change reward that highly incentivizes correcting predictions. Stably initialized via Behavioral Cloning and Student warmup, the framework continuously optimizes data pacing to maximize generalization.

Tutor-Student Reinforcement Learning: A Dynamic Curriculum for Robust Deepfake Detection

Cross-dataset and Cross-method Evaluation:

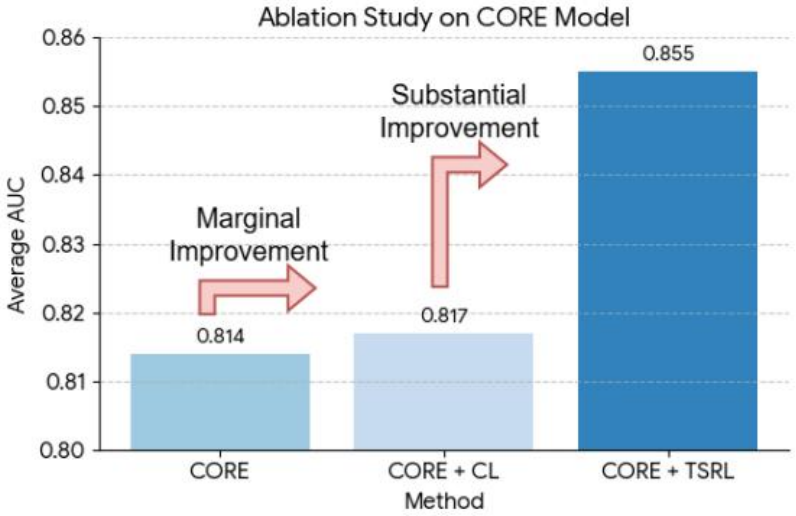
Trained on FF++(HQ); Evaluated on CDF-v2, DFD, DFDC, DFDCP (Cross-dataset)
Evaluated on DF40(Cross-method);

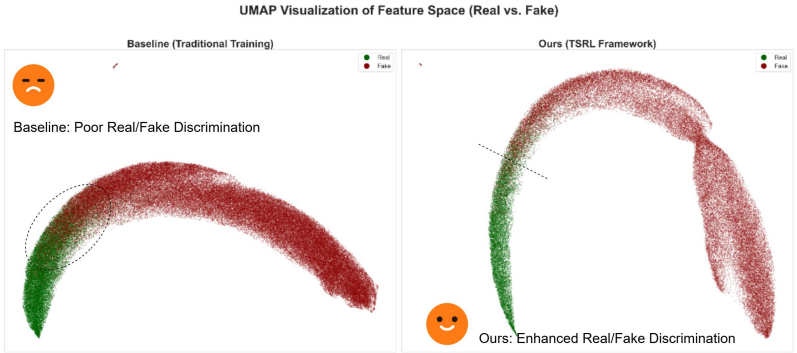
Methods	Cross-dataset Evaluation					Cross-method Evaluation								
	CDF-v2	DFD	DFDC	DFDCP	Avg.	UniFace	BleFace	MobSwap	e4s	FaceDan	FSGAN	InSwap	SimSwap	Avg.
F3Net [23]	0.789	0.844	0.718	0.749	0.775	0.809	0.808	0.867	0.494	0.717	0.845	0.757	0.674	0.746
SPSL [18]	0.799	0.871	0.724	0.770	0.791	0.747	0.748	0.885	0.514	0.666	0.812	0.643	0.665	0.710
SRM [19]	0.840	0.885	0.695	0.728	0.787	0.749	0.704	0.779	0.704	0.659	0.772	0.793	0.694	0.732
RECCE [4]	0.823	0.891	0.696	0.734	0.786	0.898	0.832	0.925	0.683	0.848	0.949	0.848	0.768	0.844
SLADD [5]	0.837	0.904	0.772	0.756	0.817	0.878	0.882	0.954	0.765	0.825	0.943	0.879	0.794	0.865
SBI [27]	0.886	0.827	0.717	0.848	0.820	0.724	0.891	0.952	0.750	0.594	0.803	0.712	0.701	0.766
TALL [33]	0.831	0.833	0.693	0.739	0.774	0.714	0.699	0.805	0.651	0.768	0.863	0.762	0.616	0.735
LSDA [36]	0.875	0.881	0.701	0.812	0.817	0.872	0.875	0.930	0.694	0.721	0.939	0.855	0.793	0.835
CDFA [17]	0.938	0.954	0.830	0.881	0.901	0.762	0.756	0.823	0.631	0.803	0.942	0.772	0.757	0.781
IID† [11]	0.776	0.876	0.711	0.706	0.767	0.830	0.796	0.945	0.675	0.787	0.928	0.800	0.685	0.806
IID [11] + TSRL	**0.791**	**0.889**	**0.704**	**0.752**	**0.784**	**0.833**	**0.812**	**0.927**	**0.651**	**0.827**	**0.903**	**0.825**	**0.704**	**0.810**
CLIP† [24]	0.751	0.752	0.759	0.667	0.732	0.597	0.672	0.798	0.571	0.700	0.755	0.631	0.498	0.653
CLIP [24] + TSRL	**0.849**	**0.732**	**0.768**	**0.724**	**0.768**	**0.598**	**0.633**	**0.827**	**0.598**	**0.728**	**0.828**	**0.663**	**0.514**	**0.674**
CORE† [22]	0.697	0.868	0.692	0.759	0.754	0.855	0.843	0.936	0.658	0.741	0.942	0.834	0.700	0.814
CORE [22] + TSRL	**0.798**	**0.863**	**0.713**	**0.724**	**0.775**	**0.926**	**0.867**	**0.930**	**0.537**	**0.855**	**0.923**	**0.941**	**0.858**	**0.855**
UCF† [34]	0.807	0.845	0.740	0.690	0.771	0.823	0.814	0.930	0.710	0.820	0.919	0.796	0.634	0.806
UCF [34] + TSRL	**0.843**	**0.850**	**0.738**	**0.698**	**0.782**	**0.830**	**0.821**	**0.947**	**0.718**	**0.836**	**0.910**	**0.778**	**0.652**	**0.812**
ProDet† [6]	0.884	0.878	0.711	0.801	0.819	0.869	0.908	0.959	0.754	0.699	0.901	0.809	0.814	0.839
ProDet [6] + TSRL	**0.907**	**0.861**	**0.706**	**0.845**	**0.830**	**0.883**	**0.916**	**0.956**	**0.789**	**0.706**	**0.928**	**0.819**	**0.801**	**0.850**
Effort† [38]	0.871	0.910	0.863	0.899	0.886	0.946	0.839	0.905	0.987	0.910	0.967	0.937	0.870	0.920
Effort [38] + TSRL	**0.901**	**0.904**	**0.882**	**0.924**	**0.903**	**0.954**	**0.902**	**0.933**	**0.983**	**0.948**	**0.975**	**0.937**	**0.904**	**0.942**

Tutor-Student Reinforcement Learning: A Dynamic Curriculum for Robust Deepfake Detection

Dynamic vs. Static: Our RL-based dynamic policy provides a +3.8% gain over static curriculum learning, proving the necessity of real-time adaptation.

Feature Disentanglement: TSRL effectively separates Real from Fake and isolates "Easy Fakes," forcing the student to master the critical decision boundary.





Tutor-Student Reinforcement Learning: A Dynamic Curriculum for Robust Deepfake Detection

Conclusion

Key Takeaways

Rethinking Optimization: Demonstrates that traditional static, uniform sample weighting is a primary bottleneck limiting the out-of-distribution generalization of deepfake detectors.

Dynamic Curriculum (TSRL): Proposes the novel TSRL framework, modeling curriculum generation as a Markov Decision Process (MDP). A PPO-based Tutor observes history-aware states (EMA loss and forgetting counts) to dynamically re-weight training batches in real-time.

Consistent SOTA Gains: Validates that TSRL acts as an orthogonal, highly effective module, delivering consistent improvements in cross-dataset and cross-method generalization across 6 diverse baseline architectures to establish a new SOTA.

Resources & Contact

Open-Source Code:
https://github.com/wannac1/TSRL

Contact Email: **zhanhelei@whu.edu.cn**

Paper

Code