Divide and Conquer: Reliable Multi-View Evidential Learning for Deepfake Detection

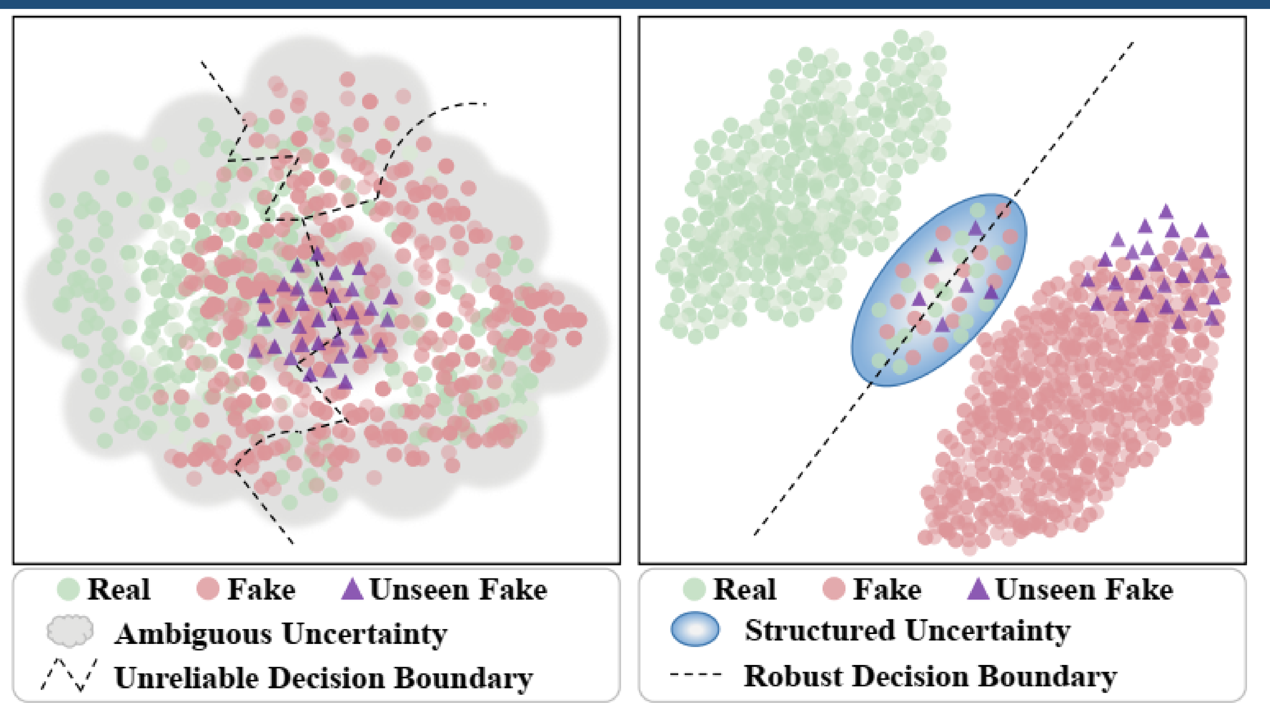
Xiaolu Kang[1], Zhongyuan Wang[1,*], Jikang Cheng[2], Baojin Huang[3], Zhanhe Lei[1], Gang Wu[4], Qin Zou[1], Qian Wang[1]
[1]Wuhan University [2]Peking University [3]Huazhong Agricultural University [4]Tarim University

Background

Generalization remains challenging:

- Forgery techniques evolve faster than detector training data.
- High-fidelity deepfakes hide artifacts behind realistic semantics.
- Existing methods often overfit known patterns and fail on unseen attacks.

Motivation



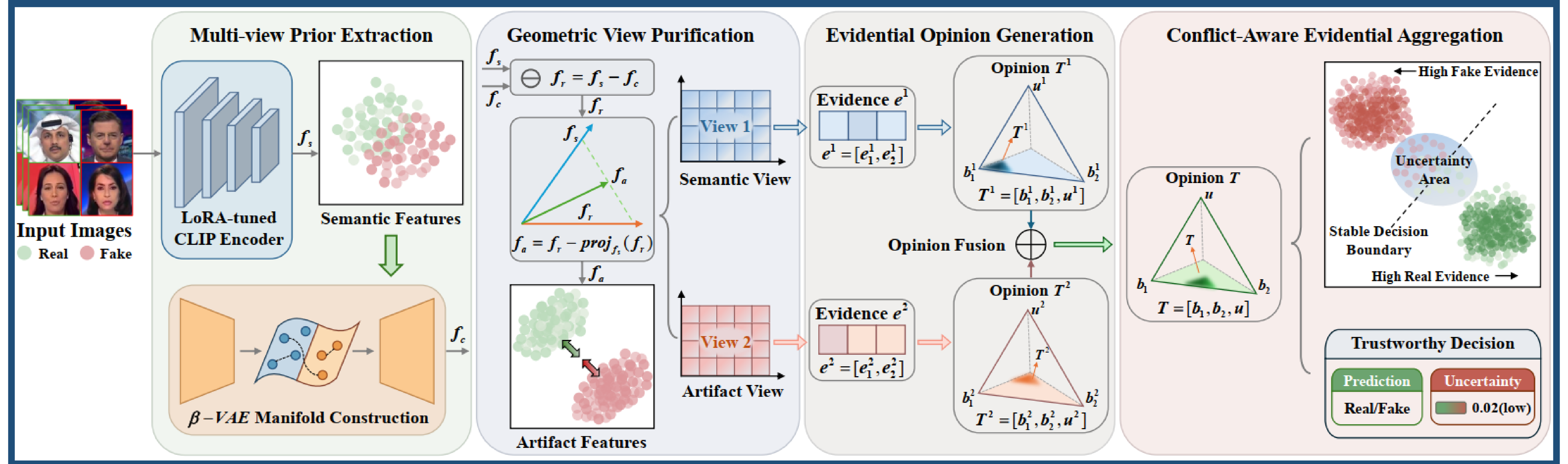
(a) Baseline: Entangled & Brittle (b) Ours: Purified & Reliable

- Existing detectors learn **entangled features**.
- **Semantic Masking Effect** hides subtle artifacts.
- Unseen fakes cause **ambiguous uncertainty** and **unreliable boundaries**.

Contributions

- **Divide:** decouple **Semantic View** and **Artifact View** via **Geometric View Purification**.
- **Conquer:** fuse both views with **Uncertainty-Aware Evidential Learning**.
- **Benefit:** reduces **Semantic Masking Effect**, captures **epistemic conflict**, and improves generalization.

Method



1. **Multi-view Prior Extraction:** Extract Semantic View with a LoRA-tuned CLIP encoder f_s, a lightweight β-VAE reconstructs a manifold-consistent feature f_c,
2. **Geometric View Purification:** Compute raw residual $f_r = f_s - f_c$, then remove semantic leakage by $f_a = f_r - proj_{f_s}(f_r)$, yielding a purified Artifact View.
3. **Evidential Opinion Generation:** Each view produces evidence $e^v = Softplus(W^v f^v + C^v)$ and opinion $T^v = [b_1^v, b_2^v, u^v]$, where b^v is belief and u^v is uncertainty.
4. **Conflict-Aware Evidential Aggregation:** Fuse opinions by Dempster-Shafer theory $T = T^1 \oplus T^2$; conflicting evidence increases uncertainty for trustworthy Real/Fake decision.

Ablation Study

(a) Architectural mechanisms.

Variant	View Comp.	Ortho.	CDFv2	DFDC	MFS	Avg.
(A)	f_s	-	0.927	0.856	0.933	0.905
(B)	$f_s + f_s$	-	0.954	0.867	0.929	0.917
(C)	$f_s + f_r$	×	0.956	0.860	0.951	0.922
Ours	$f_s + f_a$	✓	**0.977**	**0.882**	**0.956**	**0.938**

(b) Decision logic and fusion strategy.

Variant	Loss	Fusion	Unc.	CDFv2	DFDC	MFS	Avg.
(D)	$\mathcal{L}_{ce}$	Mean	×	0.956	0.874	0.942	0.924
(E)	$\mathcal{L}_{edl}$	Mean	✓	0.961	0.878	0.941	0.927
Ours	$\mathcal{L}_{edl}$	DS-C.	✓	**0.977**	**0.882**	**0.956**	**0.938**

(c) Alignment and KL-divergence losses.

Variant	$\mathcal{L}_{align}$	$\mathcal{L}_{kld}$	CDFv2	DFDC	MFS	Avg.
(F)	×	✓	0.955	0.866	0.944	0.922
(G)	✓	×	0.948	0.872	0.944	0.921
Ours	✓	✓	**0.977**	**0.882**	**0.956**	**0.938**

(d) Backbones and pre-training.

Backbone	Pre-train	Patch	CDFv2	DFDC	MFS	Avg.
CLIP-B (H)	Text-Img	16	0.886	0.811	0.790	0.829
DINO-L (I)	Self-Sup.	14	0.856	0.796	0.773	0.808
CLIP-L	**Text-Img**	**14**	**0.977**	**0.882**	**0.956**	**0.938**

(e) Calibration (ECE) ↓

Method	CDFv2	DFD	MFS	Avg.
Baseline (D)	0.13	0.15	0.19	0.16
Ours	**0.07**	**0.07**	**0.08**	**0.07**

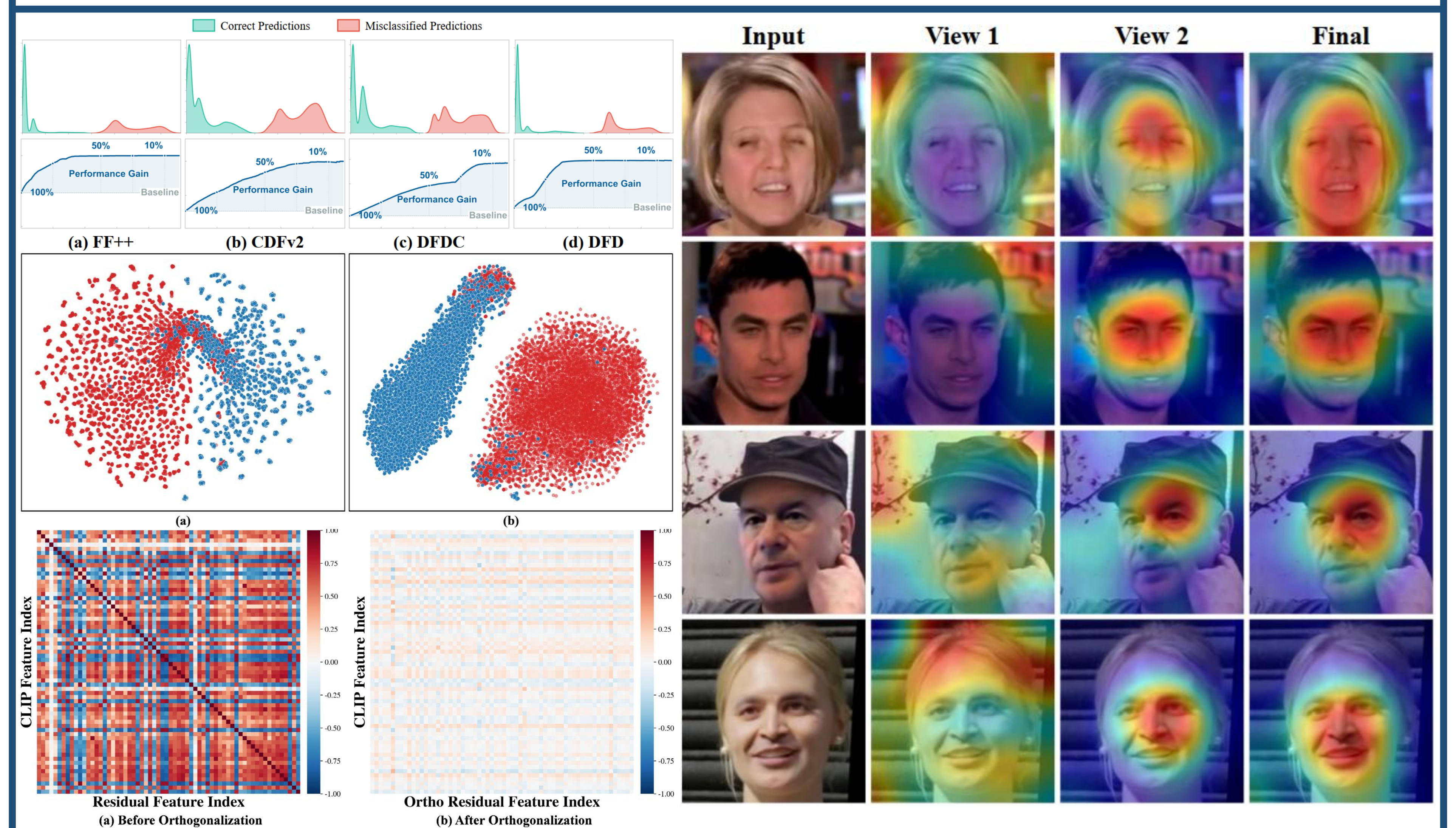
(f) Efficiency comparison

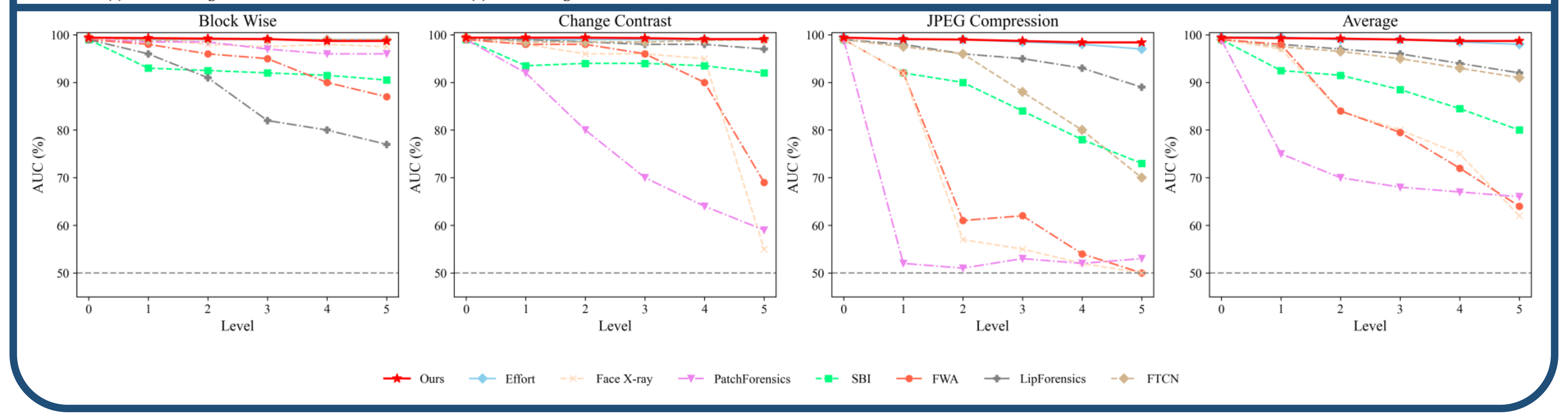
Method	GFLOPs	Time (ms/f)	FPS
Effort	103.893	22.71	44.04
GenD	155.634	17.98	55.61
Ours	**156.080**	**21.11**	**47.37**

All components contribute to stronger generalization and reliability.

Results

Methods	Venue	Cross-dataset Evaluation							Cross-manipulation Evaluation								
		CDFv2	DFD	DFDC	DFo	WDF	CDFv3	Avg.	UniFace	BleFace	MobSwap	e4s	FaceDan	FSGAN	InSwap	SimSwap	Avg.
F3Net	ECCV'20	0.789	0.844	0.718	0.730	0.728	0.736	0.758	0.809	0.808	0.867	0.494	0.717	0.845	0.757	0.674	0.746
SPSL	CVPR'20	0.799	0.871	0.724	0.723	0.702	0.740	0.760	0.747	0.748	0.885	0.514	0.666	0.812	0.643	0.665	0.710
SRM	CVPR'21	0.840	0.885	0.695	0.722	0.702	0.793	0.773	0.749	0.704	0.779	0.704	0.659	0.772	0.793	0.694	0.732
CORE	CVPR'21	0.809	0.882	0.721	0.765	0.724	0.750	0.775	0.871	0.843	0.959	0.679	0.774	0.958	0.855	0.724	0.833
RECCE	CVPR'22	0.823	0.891	0.696	0.784	0.756	0.809	0.793	0.898	0.832	0.925	0.683	0.848	0.949	0.848	0.768	0.844
SBI	CVPR'22	0.886	0.827	0.717	0.899	0.703	0.738	0.795	0.724	0.891	0.952	0.750	0.594	0.803	0.712	0.701	0.766
UCF	ICCV'23	0.837	0.867	0.742	0.808	0.774	0.761	0.798	0.831	0.827	0.950	0.731	0.862	0.937	0.809	0.647	0.824
IID	CVPR'23	0.838	0.939	0.700	0.808	0.666	0.760	0.785	0.839	0.789	0.888	0.766	0.844	0.927	0.789	0.644	0.811
LSDA	CVPR'24	0.875	0.881	0.701	0.768	0.797	0.727	0.792	0.872	0.875	0.930	0.694	0.721	0.939	0.855	0.793	0.835
ProDet	NIPS'24	0.926	0.901	0.707	0.879	0.781	0.732	0.821	0.908	0.929	0.975	0.771	0.747	0.928	0.837	0.844	0.867
Effort	ICML'25	0.956	0.965	0.843	0.977	0.848	0.850	0.907	0.962	0.873	0.953	0.983	0.926	0.957	0.936	0.926	0.940
GenD	WACV'26	0.960	0.970	0.871	0.989	0.890	0.859	0.923	0.974	0.935	0.976	0.983	0.942	0.974	0.943	0.938	0.958
Ours	–	**0.977**	**0.982**	**0.882**	**0.993**	**0.911**	**0.886**	**0.939**	**0.982**	**0.963**	**0.978**	**0.991**	**0.972**	**0.989**	**0.970**	**0.965**	**0.976**





Resources

Paper

Code

Contact

Xiaolu Kang
Wuhan University
Email: xlkang@whu.edu.cn
Website: https://github.com/kxl0825
Phone: +86 18827389221