



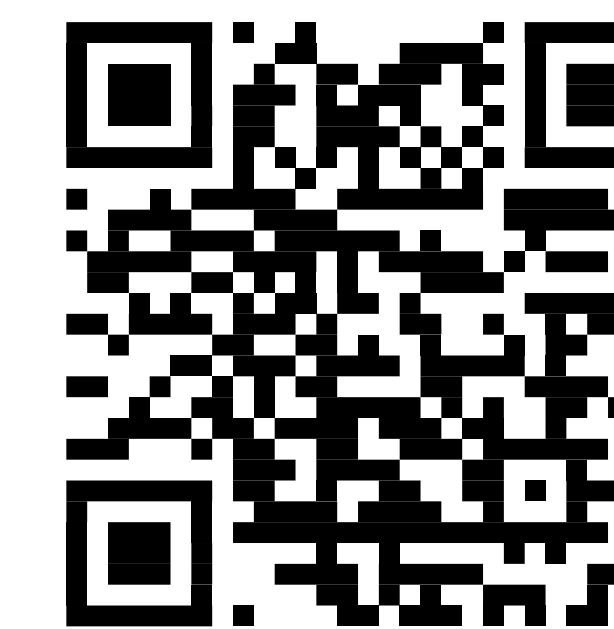
RiSE: Residual Subspace Expert for Generalizable Text-Centric Image Forgery Localization

Kahim Wong¹ · Kemou Li¹ · Yiming Chen¹ · Haiwei Wu² · Jiantao Zhou^{1*}

¹ State Key Laboratory of Internet of Things for Smart City, University of Macau

² School of Information and Software Engineering, University of Electronic Science and Technology of China

* Corresponding author: jtzhou@um.edu.mo



[Paper](#)



[Code](#)



[Live demo](#)

01 Why generalization fails

- Text edits leave **weak forensic traces** that change across editing pipelines and acquisition conditions.
- Full-parameter fine-tuning can collapse features** into a few directions and overfit synthetic artifacts.
- A single representation also **entangles** generation, copy-move, and splicing traces.

02 Preserve + specialize

- Freeze principal SVD components of each ViT linear weight.

$$W = UV^T = W_{n-r} + \Delta W$$

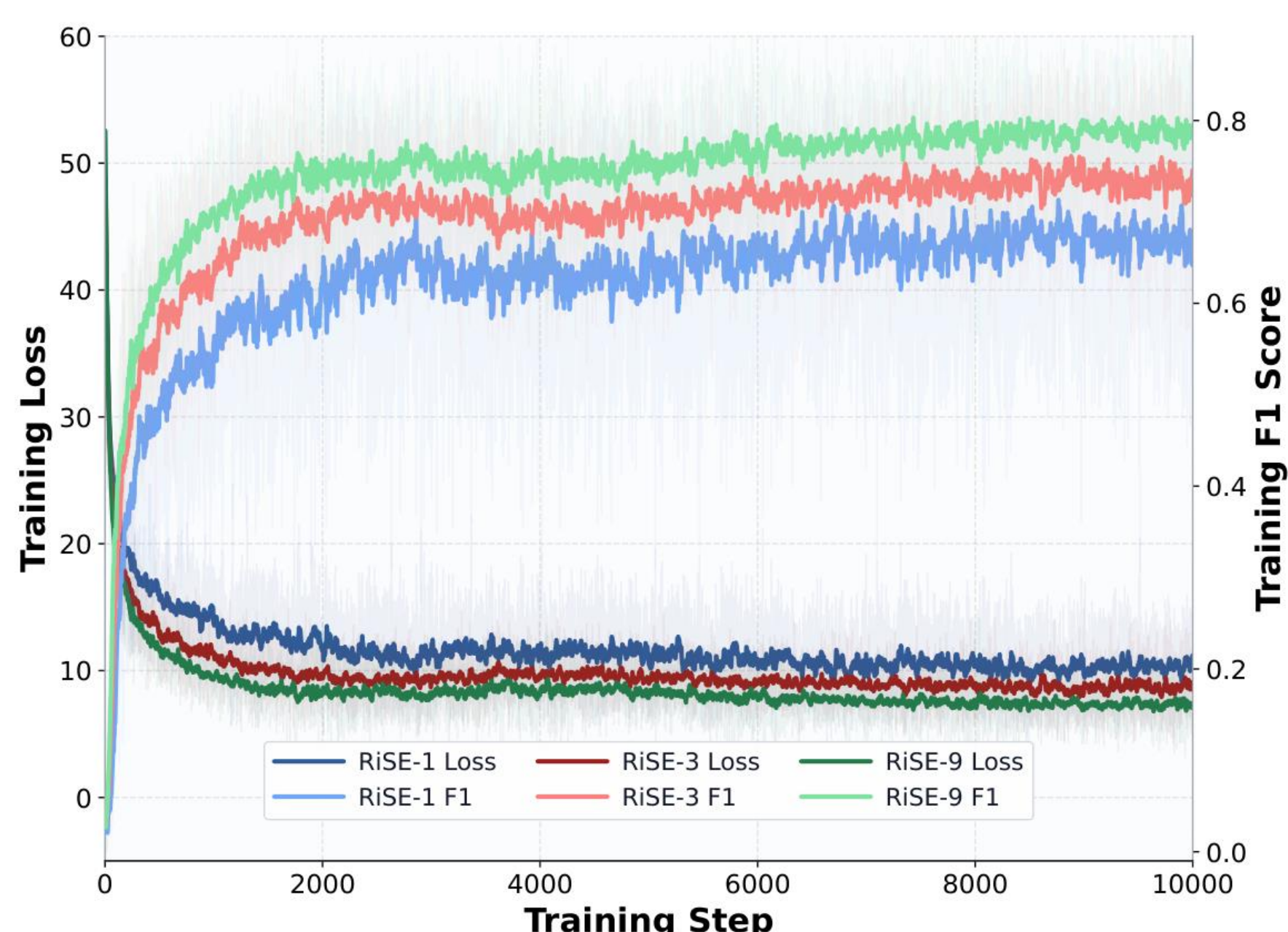
$$\Delta W = U_r \Sigma_r V_r^T$$

- Adapt expert-specific residuals; share principal weights and common modules.

- Fuse DCT features with RGB tokens.

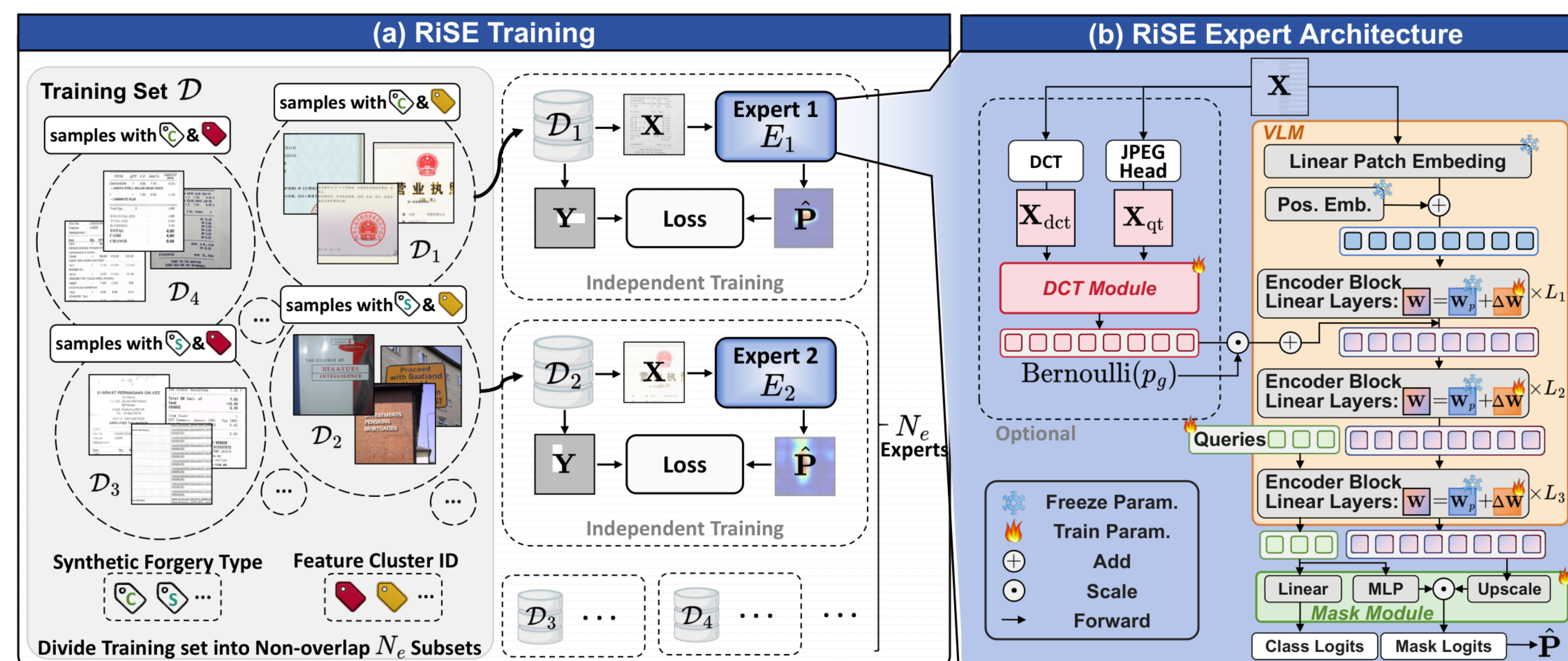
$$g \sim \text{Bernoulli}(0.5)$$

03 Why more experts?



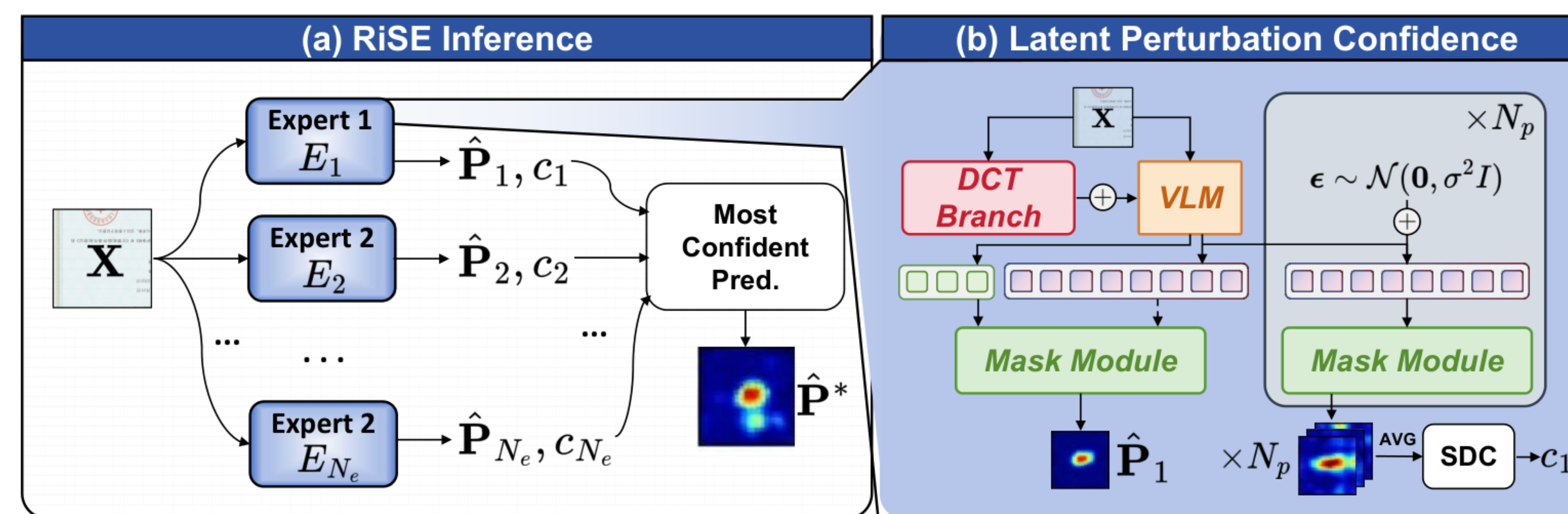
- More experts ease optimization (Fig. 2a).

04 Train experts in the residual subspace



- Partition each of 3 forgery types into 3 balanced clusters: 9 expert subsets.
- Share a DINOv3 ViT-L/16 backbone; train residual experts on separate subsets.
- Train 10,000 steps with weighted CE + BCE + Dice; optional DCT fusion.

05 Select best expert



$$LPC_j = SDC(\mathbb{E}_\epsilon[p_j(Z + \epsilon)]), \quad \epsilon \sim \mathcal{N}(0, I)$$

$$SDC(p) = \frac{2p^T \hat{y}}{\|p\|_1 + \|\hat{y}\|_1}, \quad \hat{y} = 1[p > \tau]$$

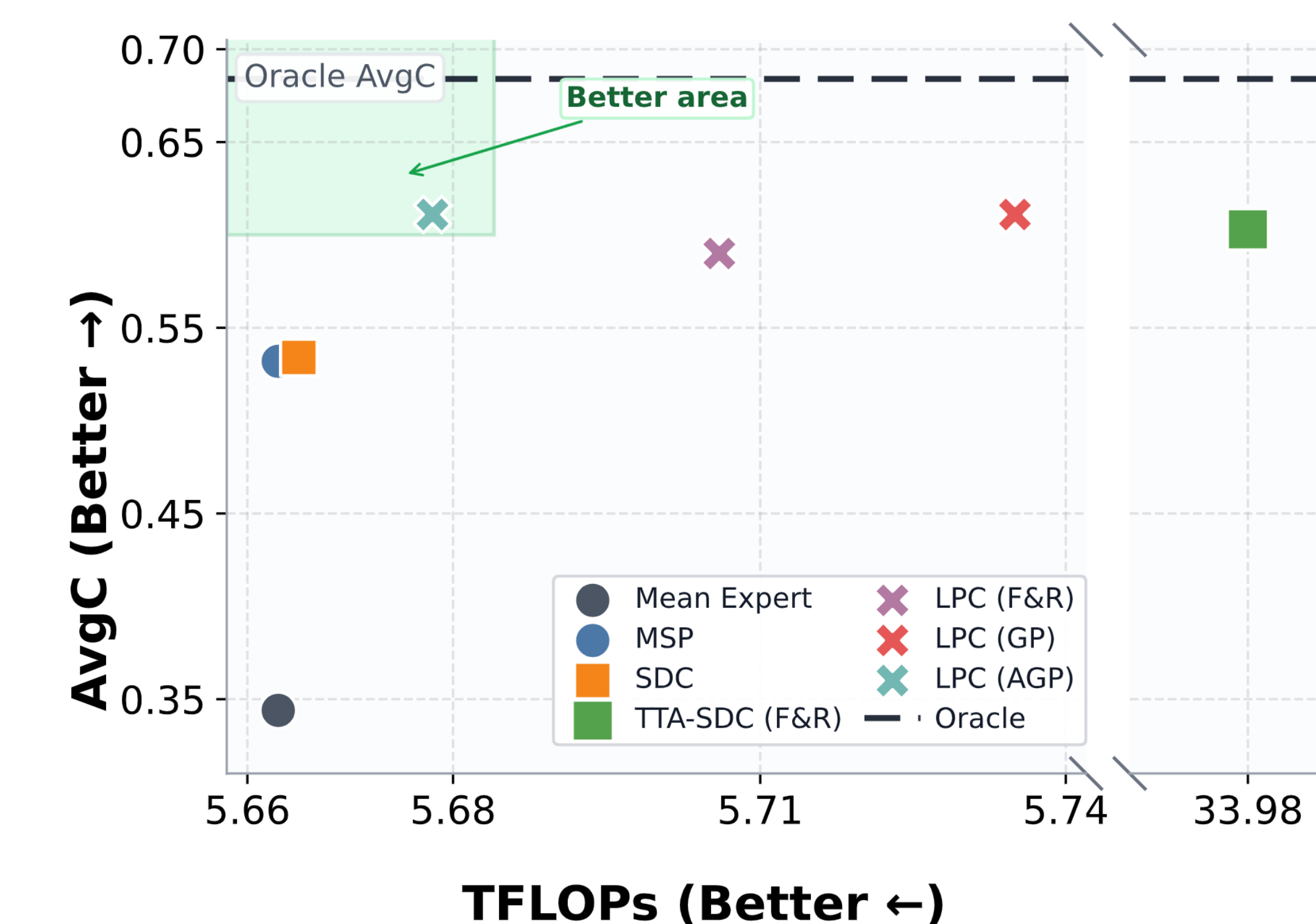
- ADF propagates latent moments analytically and selects the expert with highest LPC.
- All experts run without ground-truth masks or repeated noisy backbone passes.

06 Doc Protocol benchmark

Method	AvgI	AvgC	Avg
PSCC-Net	0.425	0.408	0.416
Cat-Net	0.722	0.298	0.480
DTD	0.743	0.247	0.460
FFDN	0.813	0.300	0.520
ADCD-Net	0.771	0.450	0.588
RiSE-1	0.662	0.542	0.594
RiSE-3	0.676	0.575	0.618
RiSE-9	0.687	0.611	0.644
RiSE-9-DCT	0.767	0.616	0.681

- Train: DocTamer.
- AvgI: FCD/SCD/Test.
- AvgC: T-SROIE/OSTF/TPIC-13/RTM.
- Avg: all 7.

07 Cross-domain F1 versus TFLOPs



08 Takeaways

- FPFT causes low-rank feature and poor generalization
- RiSE freezes the dominant subspace, adapts small residual component, and scales capacity through independently trained forgery-aware experts
- LPC, confidence estimator, latent augments via a single-pass