CLIPping the Deception: Adapting Vision-Language Models for Universal Deepfake Detection

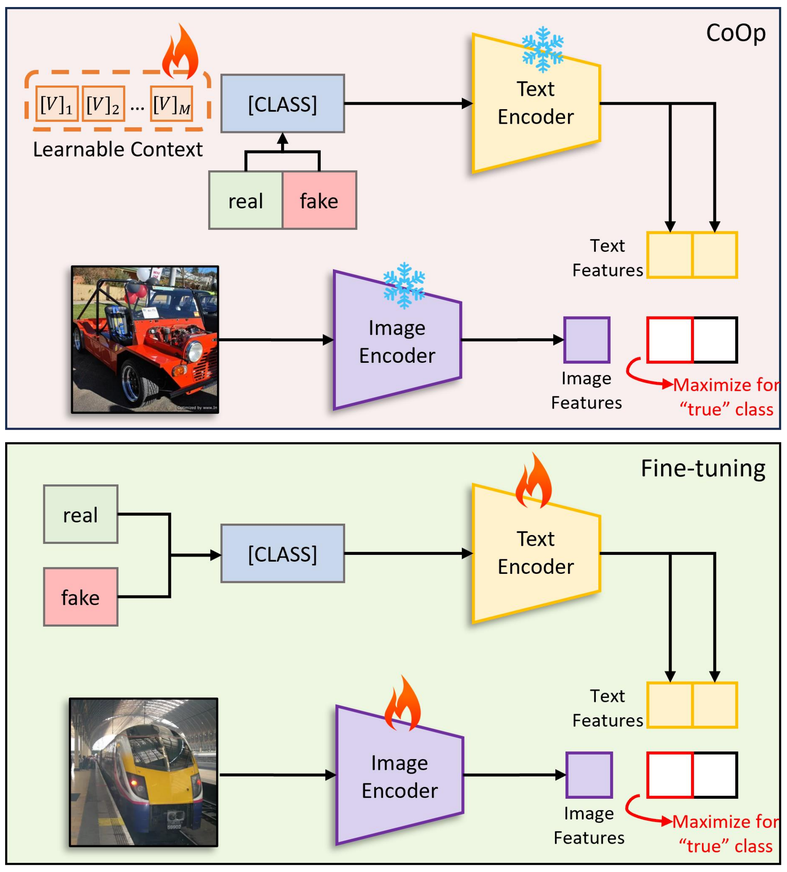
ICMR 2024

Sohail Ahmed Khan and Duc-Tien Dang-Nguyen

Media Futures

Contributions

1. Investigated four transfer learning strategies to adapt CLIP for deepfake detection, inspired by recent VLM research.
2. Our strategies, especially Prompt Tuning, outperform the current state-of-the-art.
3. Few-shot experiments show excellent performance with only 32 real/fake samples per LSUN object category.
4. Robustness analysis against post-processing operations such as, JPEG compression and Gaussian blurring.
5. Demonstrated solid performance of CLIP-based detectors with smaller training sets (20k real/fake images).
6. Open-source code and trained models released.



Why CLIP?

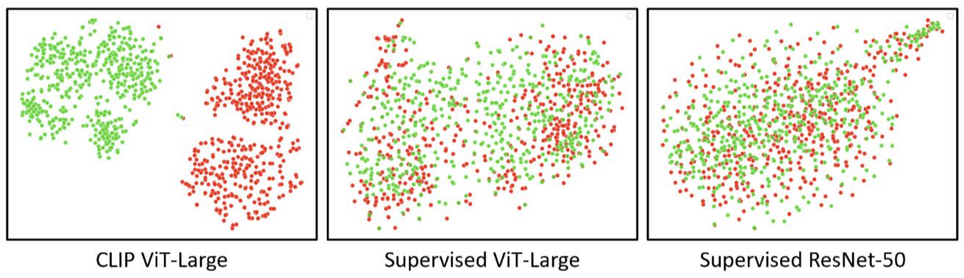


Figure 1: Visualization of real (in red) and a subset of StarGAN [6] generated fake (in green) images utilizing t-SNE in the feature space of three different image encoders. CLIP's feature space demonstrates superior separation of real and fake images as compared to other two supervised models.

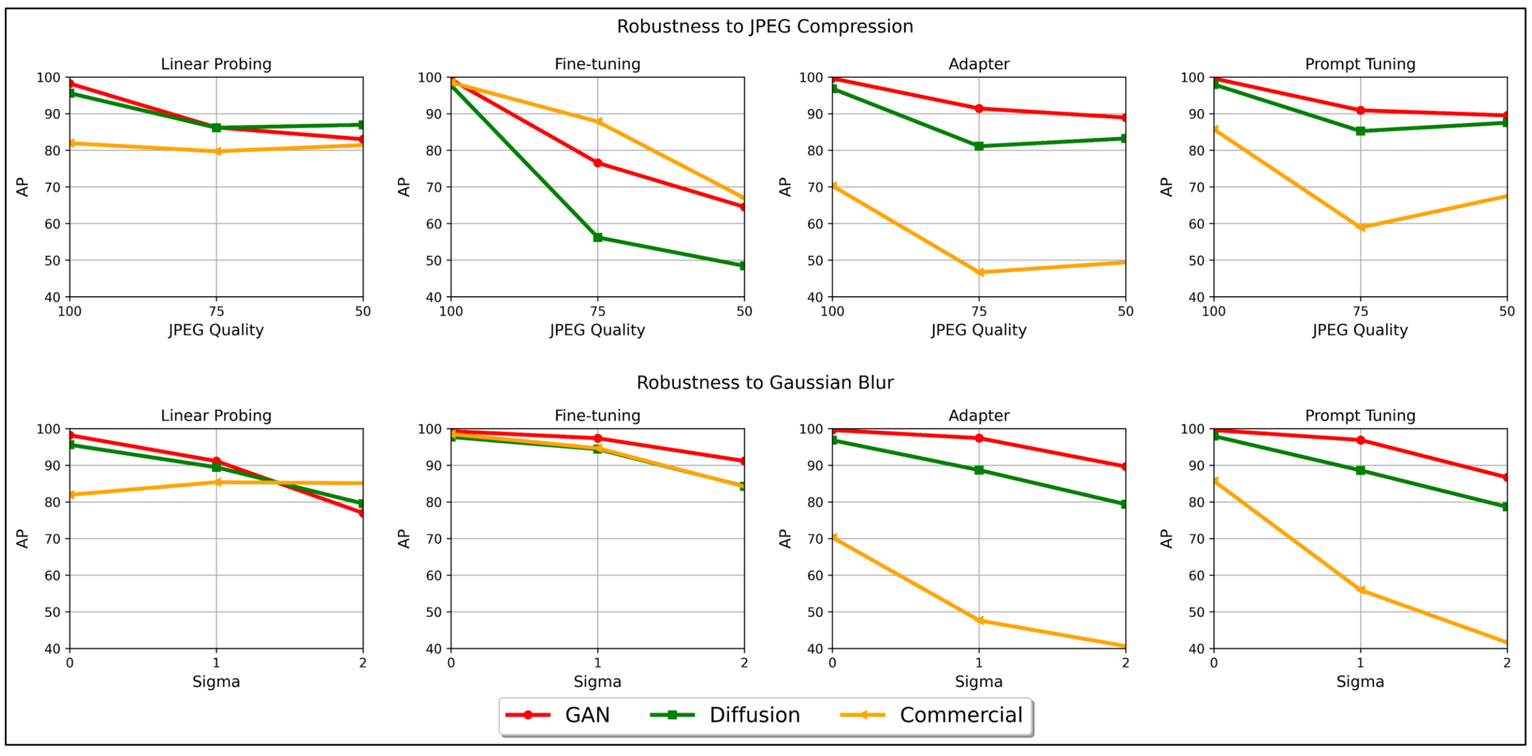
Results

Table 3: Generalization performance. This table compares the accuracy (Acc) scores attained by our proposed techniques with various previous studies. The proposed CLIP adaptation strategies show noteworthy performance gains compared to previous baselines and SOTA techniques.

Method	Variant	Generative Adversarial Networks										DALL-E	Denoising Diffusion Models					FF++		Avg. Acc
		Pro GAN	Big GAN	Cycle GAN	EG3D	Gau GAN	Star GAN	Style GAN	Style GAN-2	Style GAN-3	Taming-T		Glide	Guided	LDM	SD	SDXL	Deep Fakes	Face Swap	
Wang et al. (CVPR'20)	Blur+JPEG (0.1)	99.90	67.65	79.50	72.65	76.63	89.72	82.10	77.05	80.68	56.45	55.05	61.15	62.90	54.03	52.50	53.40	52.67	49.68	66.09
	Blur+JPEG (0.5)	99.65	58.13	77.80	50.30	75.56	79.99	69.80	62.30	53.42	51.05	51.90	54.33	52.35	51.35	50.15	51.00	51.46	50.02	59.18
Gragn. et al. (ICME'21)	ResNet-50 No Downsample	**100.00**	93.27	91.75	97.55	94.13	99.65	97.25	89.75	97.47	67.45	60.65	69.38	67.30	62.33	59.70	57.75	65.31	50.02	76.59
Corvi et al. (ICASSP'23)	ProGAN/LSUN	**100.00**	**95.85**	90.35	**98.40**	92.46	99.00	**97.65**	84.90	82.79	65.30	69.30	58.98	53.10	58.83	55.70	52.10	59.38	50.11	72.72
	Latent/LSUN	50.94	51.82	46.20	49.25	50.86	48.02	59.40	50.95	50.05	77.65	87.00	59.83	50.95	**99.25**	**99.25**	93.10	69.87	48.14	66.40
Ojha et al. (CVPR'23)	CLIP Linear Probing	98.94	94.48	94.20	57.75	94.65	87.49	85.55	83.40	75.42	89.45	89.20	82.15	79.00	87.80	81.90	74.15	62.71	64.30	82.84
Ours	Linear Probing	98.50	91.75	91.00	98.20	88.08	94.42	81.40	71.70	94.11	91.05	85.80	90.55	79.05	87.42	77.30	83.85	69.37	68.30	86.26
	Fine Tuning	99.60	77.38	71.55	**98.40**	65.70	**100.00**	94.85	**95.30**	**99.89**	94.40	**93.20**	88.78	**92.35**	95.17	91.75	**97.35**	76.46	52.11	88.74
	Adapter	99.88	94.75	**97.45**	95.30	**95.47**	99.12	93.35	78.35	93.11	94.55	92.00	**94.27**	81.65	89.18	67.70	71.60	77.11	70.16	88.72
	Prompt Tuning	99.83	93.80	95.60	93.50	93.43	99.15	95.25	82.95	93.11	**94.95**	91.50	92.88	84.3	88.16	76.45	77.80	**78.45**	**74.66**	**89.45**

Table 5: We present the results from our few-shot (32-shot) experiments, wherein we train CLIP using various transfer learning strategies on *real/fake* images from the ProGAN dataset. We then evaluate the trained models on images generated by GANs, diffusion models and commercial image generators.

Method	Family of Generators			Average AP/Acc
	GAN AP/Acc	Diffusion AP/Acc	Comm. Tools AP/Acc	
Linear Probing	94.39 / 83.62	89.67 / 80.47	76.78 / **69.72**	86.95 / 77.94
Fine-tuning	97.09 / 85.23	90.14 / 77.18	71.35 / 65.90	86.19 / 76.11
Adapter	97.40 / 87.27	90.53 / 81.12	61.69 / 53.93	83.21 / 74.11
Prompt Tuning	**98.61** / **89.88**	**95.97** / **84.76**	**87.23** / 66.38	**93.94** / **80.34**



PARTNERS

amedia Schibsted 2 NRK Universitas Oslo Bergens Tidende nb.no VIZRT WOLFTECH NORCE UiS Faktisk.

HOST

UNIVERSITY OF BERGEN

FUNDED BY

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

Forskningsrådet

Norwegian Centre for Research-based Innovation