Faster Than Lies: Real-time Deepfake Detection using Binary Neural Networks

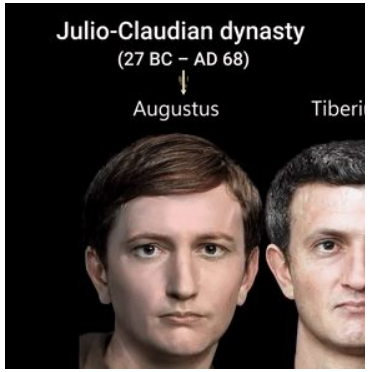
Romeo Lanzino, Federico Fontana, Anxhelo Diko,
Marco Raoul Marini, Luigi Cinque

{lanzino, fontana.f, diko, marini, cinque}@di.uniroma1.it

SAPIENZA
UNIVERSITÀ DI ROMA

The (Real) World on the Brink of Ruin

- Why We Want Deepfakes:
  - Entertainment and media creation
  - Enhanced educational tools (e.g., historical figures)
  - Personalization in marketing
- Problematic Use-Cases of Deepfakes:
  - Politics: A deep fake of a world leader announcing false policies
  - Security: Impersonation for fraudulent activities
  - Personality Substitution: Identity theft and social engineering
  - Fabricated Celebrity Videos: Tom Cruise performing stunts he never did
  - Manipulated News Broadcasts: News anchors appearing to report events that never happened



🖊 YouTube, @panagiotisconstantinou

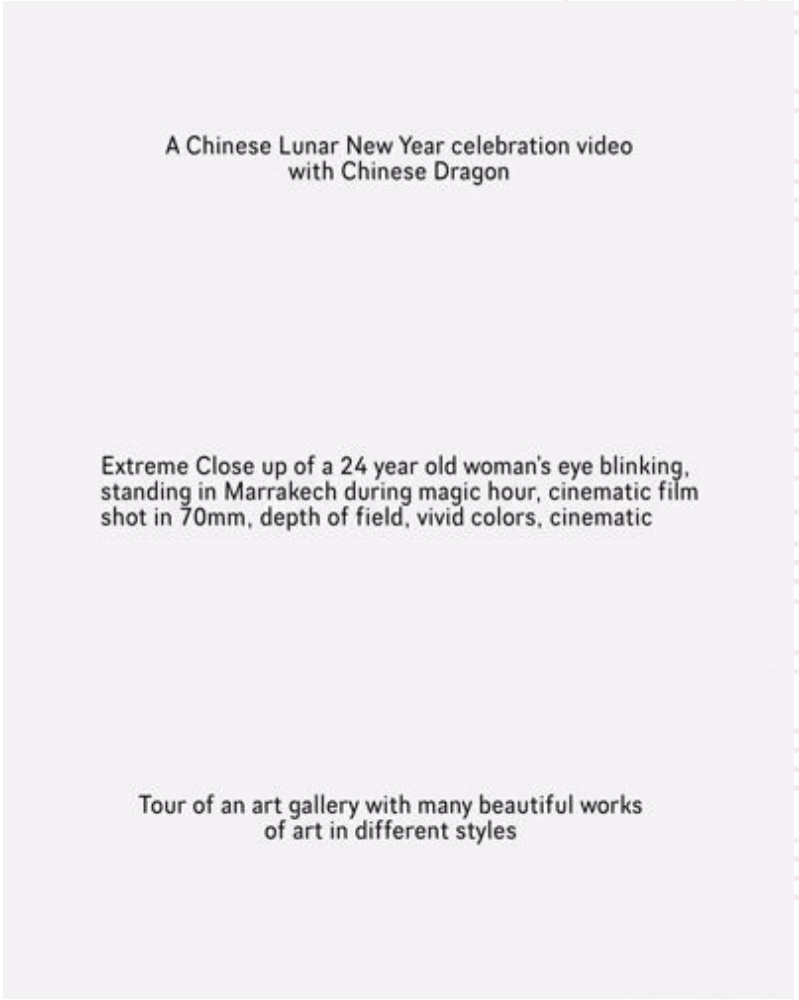
🖊 YouTube, @Shamook

🖊 BBC News, Kayleen Devlin and Joshua Cheetham

🖊 YouTube, @Vecanoi

The Current State of Deep-Fake Detection

- More and More Powerful Generative Models:
  - Rapid advancements in Diffusion models and other generative architectures
- Prediction vs. Generation:
  - Generating a fake image takes seconds
  - Detecting a fake image needs to be faster than the generation process
- Challenges in Real-Time Detection:
  - **Scalability**: High computational demands
  - **Latency**: Immediate detection is crucial in live scenarios
  - **Accuracy**: Balancing speed and detection precision



🖊 YouTube, @OpenAI

Our Vision

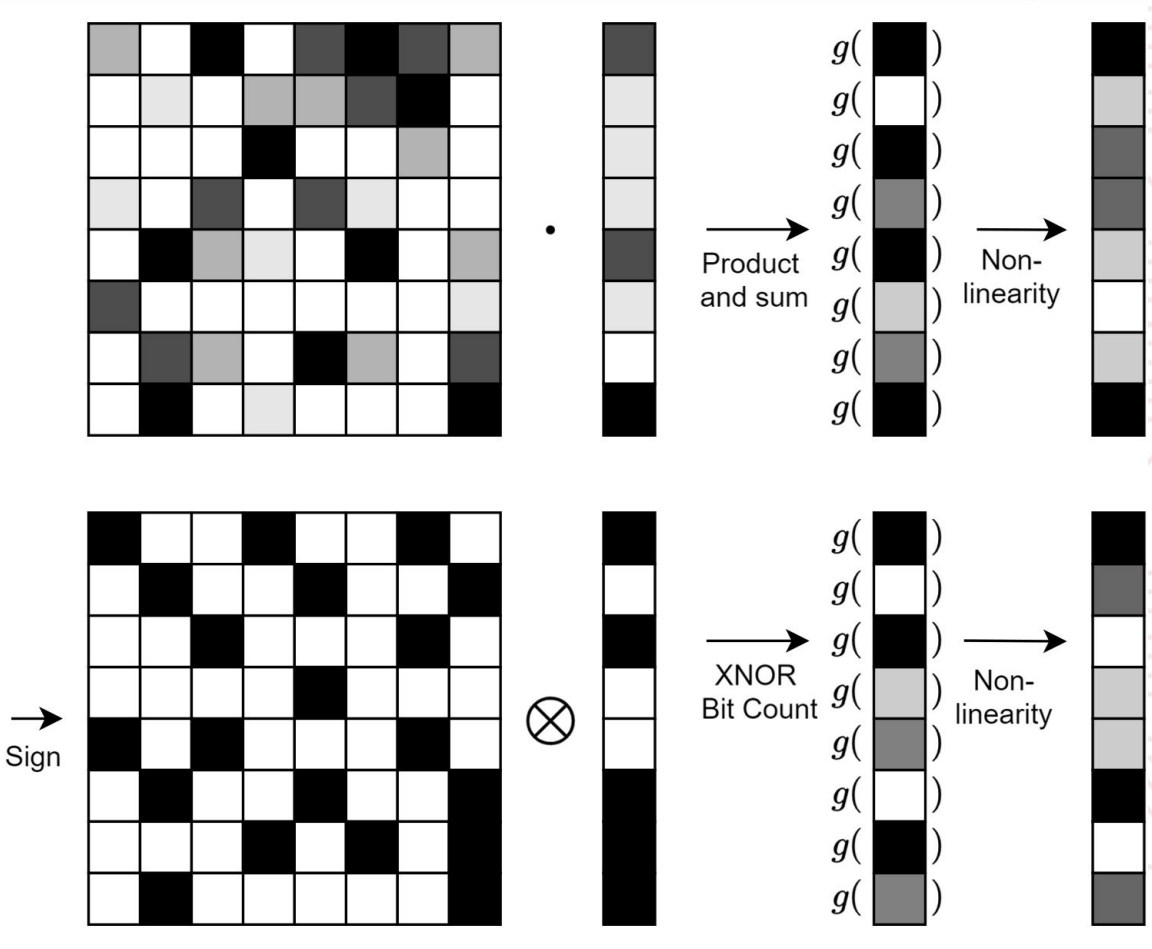
- Future Prospects:
  - Envisioning a future where **every device** is **capable of real-time deepfake detection**
  - Real-time detection is essential for maintaining the integrity and trustworthiness of digital content
  - Aim to significantly reduce the time and computational resources required for accurate deep fake identification

- Our Contribution:
  - Proposing **an efficient network** specifically designed **for image deepfake detection**
  - Introducing **a Binary Neural Network (BNN)** to enhance efficiency and speed
  - BNNs provide a streamlined approach to detect deep fakes with **minimal resource usage while maintaining high accuracy**

Method (BNNs)

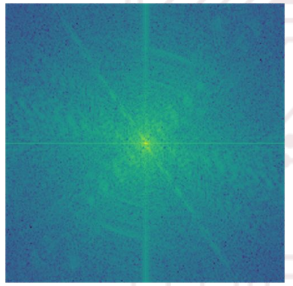
- Binary Neural Networks (BNNs):
  - Utilization of BNNs for **efficiency**

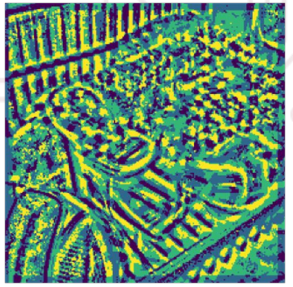
- Pros and Cons of BNNs:
  - Advantages:
    - Up to 58x more efficient
    - 32x less memory usage
  - Challenges:
    - Theoretical efficiency not fully realized due to hardware limitations



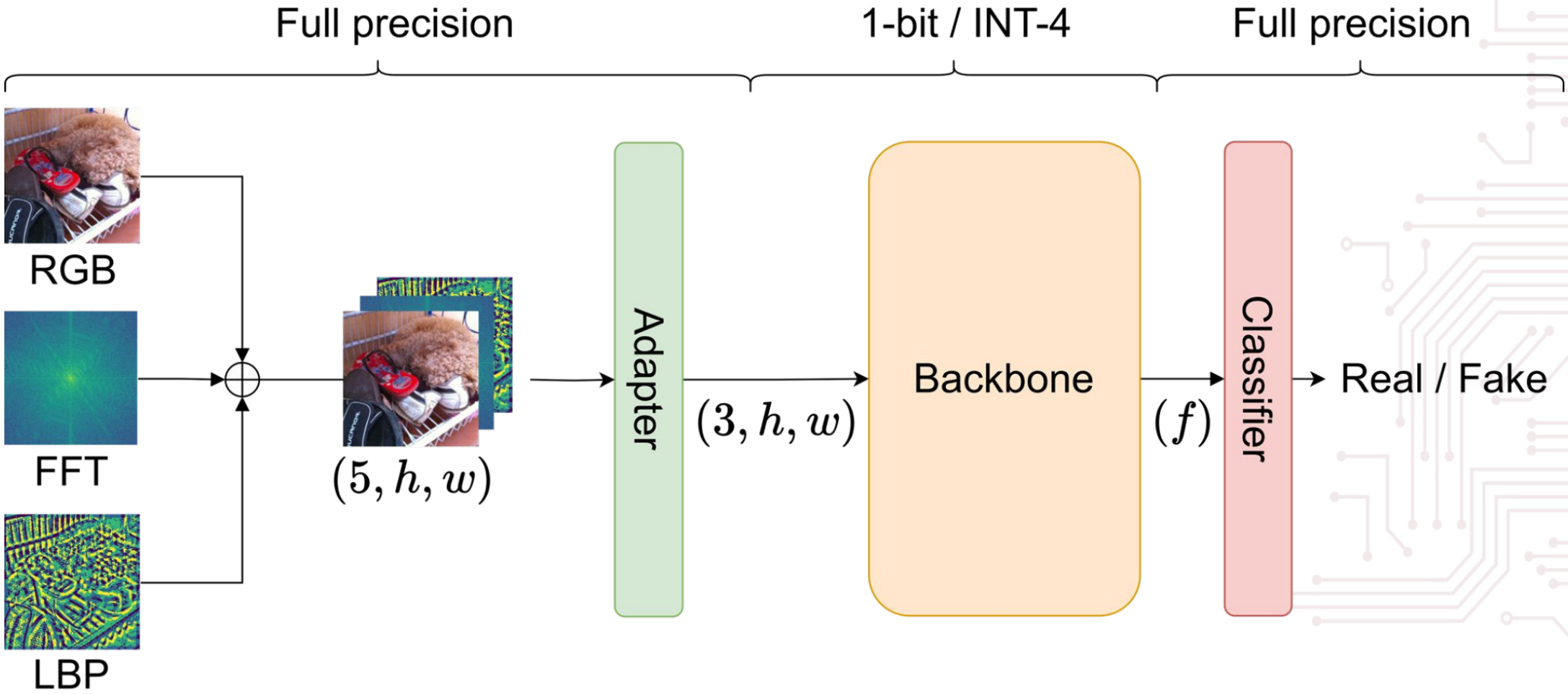
Method (Additional Channels)

- Feature Selection:
  - Use of Fast Fourier Transform (FFT) and Local Binary Patterns (LBP) in addition to RGB
- Intuition Behind Features:
  - FFT:
    - Captures **unusual frequency peaks** in the images
    - Effective in identifying artificial textures and anomalies in frequency components
  - LBP:
    - Highlights **unusual intensities in neighboring pixels**
    - Useful for detecting inconsistencies in local patterns and textures
- Motivation:
  - Combining FFT and LBP with RGB **provides a** comprehensive **representation of both frequency and spatial anomalies**, enhancing deepfake detection accuracy





Method (Proposed Model)



Ablation

- Tested various feature combinations for deepfake detection

- Adding the magnitude of the Sobel filter resulted in worse performance compared to RGB, FFT, and LBP
- The Sobel filter failed to capture essential deepfake characteristics due to its focus on edge detection, which misses subtle frequency and texture anomalies present in deepfakes

- RGB, FFT, and LBP provided complementary features that enhanced the detection accuracy across different datasets

Ablation	Variations	Accuracy (%)
Baseline	-	90.35
Features added to the learned ones	Magnitude	82.36
	FFT	88.18
	LBP	88.42
	Magnitude and FFT	81.20
	Magnitude and LBP	81.67
	FFT and LBP	**91.60**
	Magnitude, FFT and LBP	81.56

Results

- We evaluated our model on three datasets: COCOFake, DFFD, and CIFAKE, showing consistent and robust detection performance across diverse scenarios.

Method	Model	Pre-training dataset	Accuracy	AUC	Parameters (M)	FLOPs (G)
[2]	ResNet50	ImageNet	90.31	-	25.6	4.8
	ViT-B/32	ImageNet	87.64	-	88.3	8.56
	CLIP-ResNet50	OpenAI WIT	99.07	-	25.6	4.8
	CLIP-ViT-B/32	OpenAI WIT	99.11	-	88.3	8.56
	OpenCLIP-ViT-B/32	LAION-400M	97.88	-	88.3	8.56
	OpenCLIP-ViT-B/32	LAION-2B	**99.68**	-	88.3	8.56
Ours	BNext-T with frozen backbone	ImageNet	83.65	81.98	29.8	**0.89**
	BNext-S with frozen backbone	ImageNet	93.15	95.19	67.1	1.91
	BNext-M with frozen backbone	ImageNet	84.59	82.11	133	3.39
	BNext-T	ImageNet	99.25	99.86	29.8	**0.89**
	BNext-S	ImageNet	99.28	99.89	67.1	1.91
	BNext-M	ImageNet	99.18	**99.91**	133	3.39

Results

- We evaluated our model on three datasets: COCOFake, DFFD, and CIFAKE, showing consistent and robust detection performance across diverse scenarios.

Method	Model	Accuracy	AUC	Parameters (M)	FLOPs (G)
[10]	Xception	-	99.64	40.0	18.0
	VGG16	-	99.67	138.4	15.5
Ours	BNext-T with frozen backbone	89.56	87.65	29.8	**0.89**
	BNext-S with frozen backbone	89.69	88.58	67.1	1.91
	BNext-M with frozen backbone	89.61	86.64	133	3.39
	BNext-T	98.95	**99.94**	29.8	**0.89**
	BNext-S	**99.01**	**99.94**	67.1	1.91
	BNext-M	98.75	99.92	133	3.39

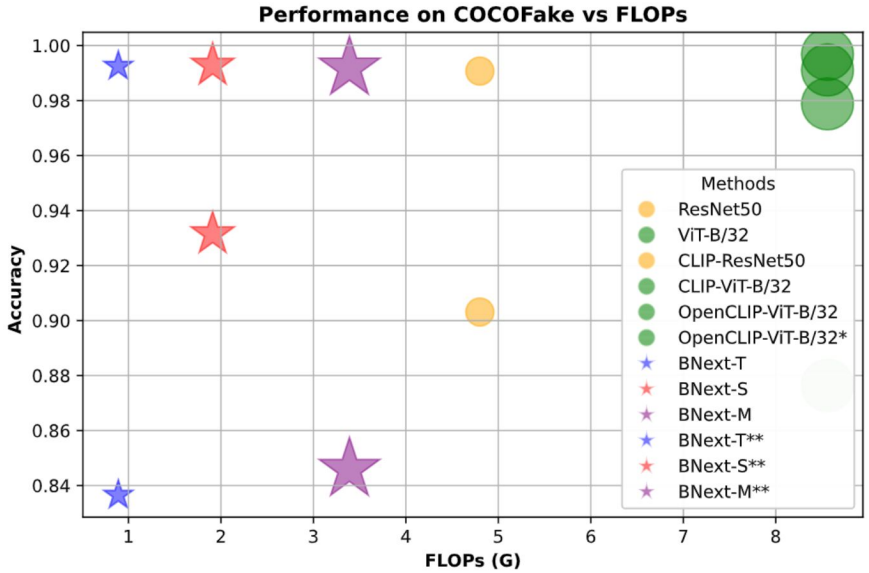
VisionLab

Results

- We evaluated our model on three datasets: COCOFake, DFFD, and CIFAKE, showing consistent and robust detection performance across diverse scenarios.

Method	Model	Accuracy	AUC	Parameters (M)	FLOPs (G)
[53]	ResNet-50	95.00	99.00	25.6	4.8
	VGG	96.00	99.00	133	7.63
	DenseNet	**98.00**	99.00	7.9	5.6
Ours	BNext-T with frozen backbone	83.89	91.70	29.8	**0.89**
	BNext-S with frozen backbone	80.71	89.25	67.1	1.91
	BNext-M with frozen backbone	82.77	90.73	133	3.39
	BNext-T	97.29	**99.65**	29.8	**0.89**
	BNext-S	96.96	99.55	67.1	1.91
	BNext-M	97.35	99.62	133	3.39

Conclusion



- Summary:
  - **Focused on efficiency** in a hardware-agnostic framework
  - **Achieved high accuracy** in deepfake detection

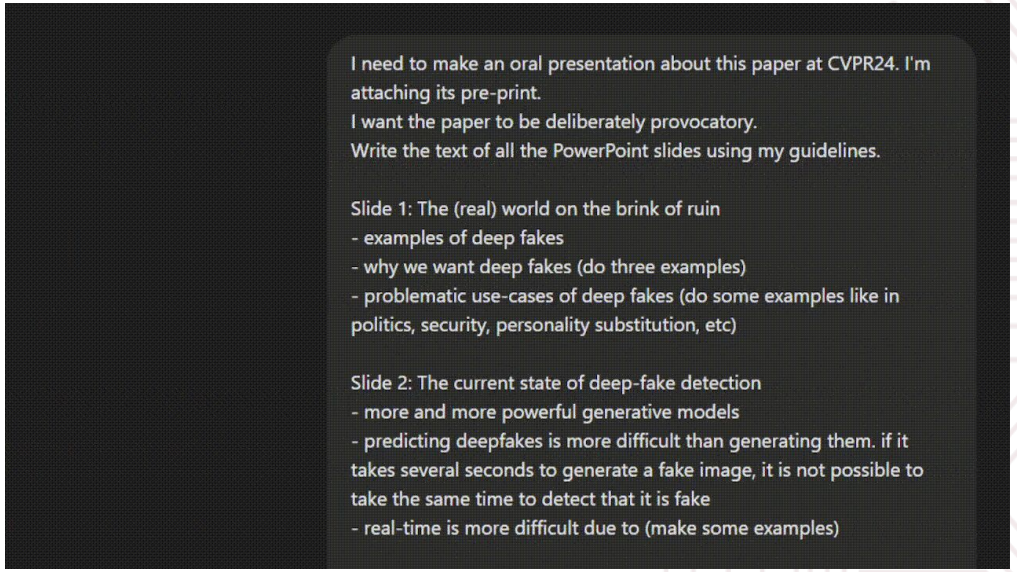
- Limitations:
  - **Emphasis on theoretical FLOP reductions;** real-world application of BNNs requires a specialized framework or accelerator to fully realize benefits
  - Evaluation limited to a network pretrained on ImageNet; **other studies used larger datasets** for enhanced transfer-learning

- Future Works:
  - Potential for **practical implementation** on specialized hardware or within specific computational frameworks to achieve theoretical efficiency gains
  - Exploring alternative pre-training datasets to further **enhance transfer-learning efficacy** and robustness in deep fake detection systems

Conclusion

- Detecting fakes is hard; we took a single step towards making real-time detection feasible

- Did anyone guess that every single word in this presentation was generated by an LLM?



- Find the code and paper at GitHub
  - https://github.com/fedeloper/binary_deepfake_detection

- Come and meet us at the poster session!

VisionLab

Thank you for your attention!