Deepfake Detection for Mobile Devices

組員：徐仁瓏、曾文海、陳亦宥、蕭合亭

Outline

Motivation

Goal

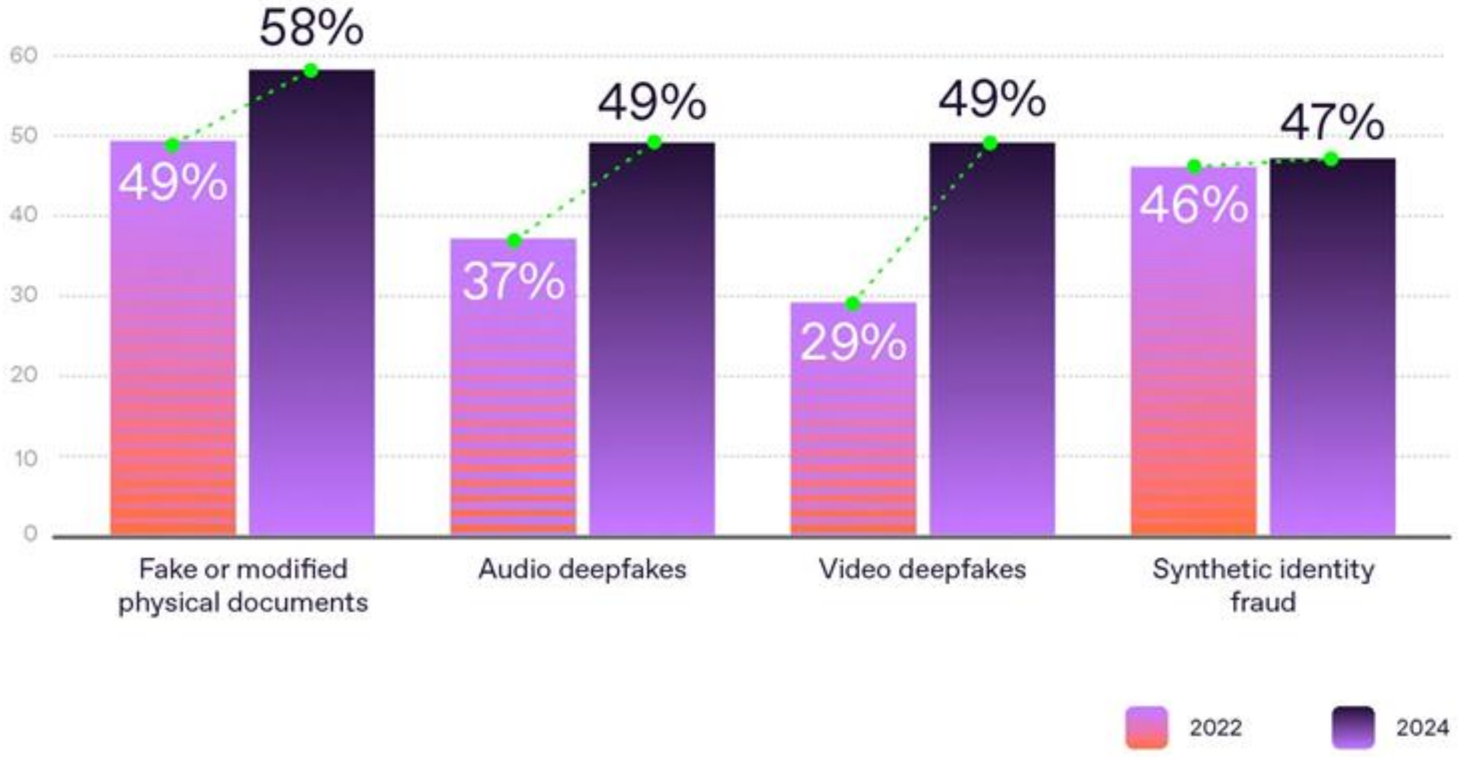
Challenge & Solution

Dataset

Conclusion

Motivation

Motivation



https://regulaforensics.com/news/deepfake-fraud-doubles-down/

Motivation

1. 小玉「Deepfake換臉A片」、韓國N號房2.0
2. Real time deepfake

https://www.thenewslens.com/article/168524
https://www.gvm.com.tw/article/115603
https://youtu.be/w0Wkhz4G6OA?si=lXWMS2x-8MTvu52w

Motivation

在目前手機普及的時代，手機也成為現代人獲得訊息最常見的工具之一，但是這些訊息常常參雜著偽造的訊息，例如透過生成式 AI 偽造影像。

為了讓人們能夠發現偽造影像我們將開發輕量化的 deepfake 偵測模型並部署到手機裝置，讓用戶能即時檢測影像真偽，減少受虛假資訊影響的機會。

手機裝置的算力和記憶體資源有限，因此開發輕量化模型是為了在有限資源下提供高效能的偵測能力。

Goal

Goal

開發一款可在手機平台上運行的 Deepfake 偵測應用，主要針對圖像進行偵測，並涵蓋目前所有主流的Deepfake生成技術。

- ➢ 記憶體限制：確保模型在現有的記憶體下可以穩定運行，並將內存使用量降到最低，以適應不同型號的手機。

- ➢ 處理能力要求：盡管即時性不是本次的強制需求，推理速度仍是影響用戶體驗的重要指標，因此會在模型設計中考量優化運算效率，使得模型在有限的算力下能提供相對快速的反應。

Challenge & Solution

Challenge

❑ 模型壓縮與性能權衡

❑ 適應性與泛化能力

❑ 部署到手機平台的兼容性

Challenge

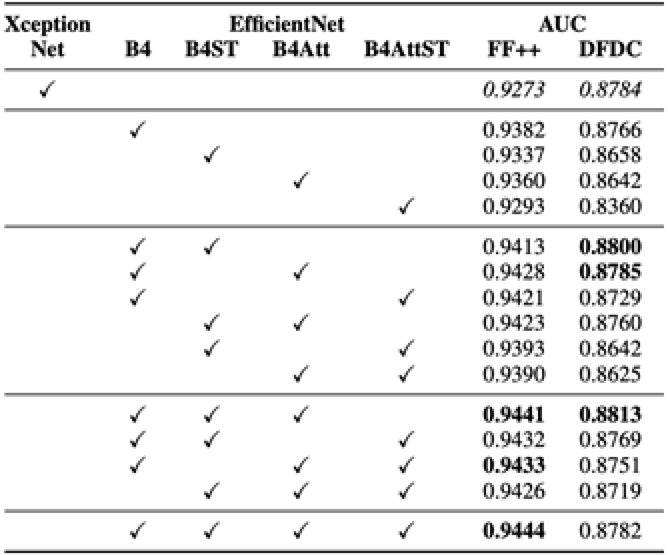
❑ 模型壓縮與性能權衡

❑ 適應性與泛化能力

❑ 部署到手機平台的兼容性

Challenge

現有的Deepfake偵測模型，如基於CNN的架構（例如Xception、EfficientNet等），通常具有高準確率，但這些模型的計算複雜，無法直接在資源受限的裝置上運行。

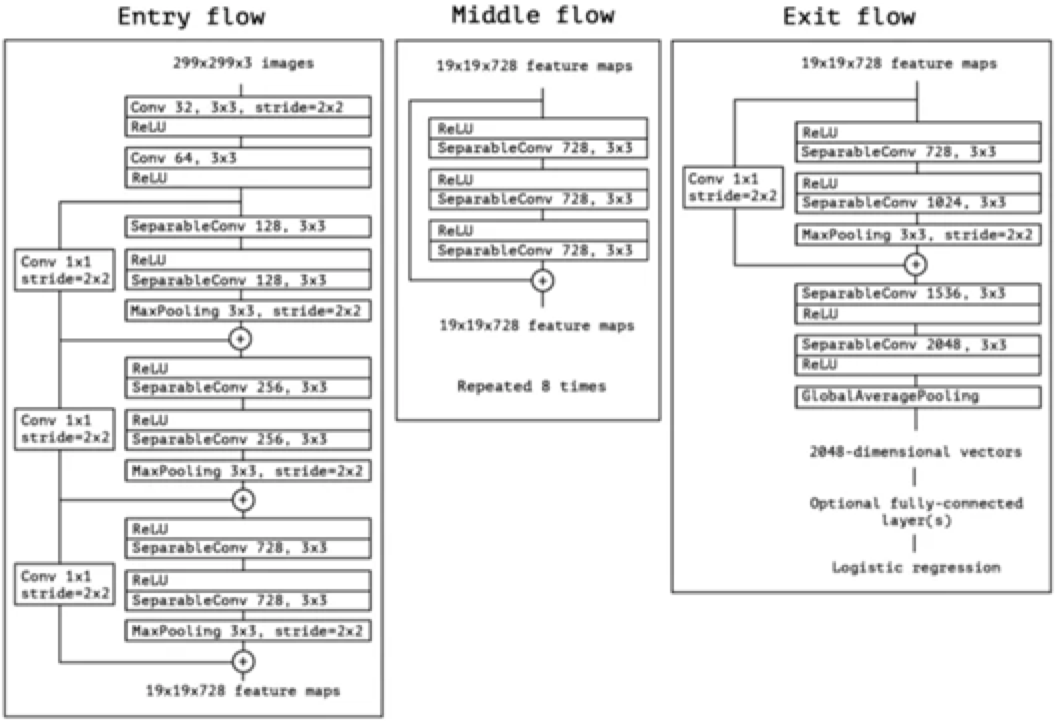


Xception	EfficientNet				AUC	
Net	B4	B4ST	B4Att	B4AttST	FF++	DFDC
✓					*0.9273*	*0.8784*
	✓				0.9382	0.8766
		✓			0.9337	0.8658
			✓		0.9360	0.8642
				✓	0.9293	0.8360
	✓	✓			0.9413	**0.8800**
	✓		✓		0.9428	**0.8785**
	✓			✓	0.9421	0.8729
		✓	✓		0.9423	0.8760
		✓		✓	0.9393	0.8642
			✓	✓	0.9390	0.8625
	✓	✓	✓		**0.9441**	**0.8813**
	✓	✓		✓	0.9432	0.8769
	✓		✓	✓	**0.9433**	0.8751
		✓	✓	✓	0.9426	0.8719
	✓	✓	✓	✓	**0.9444**	0.8782

Video Face Manipulation Detection Through Ensemble of CNNs

Challenge

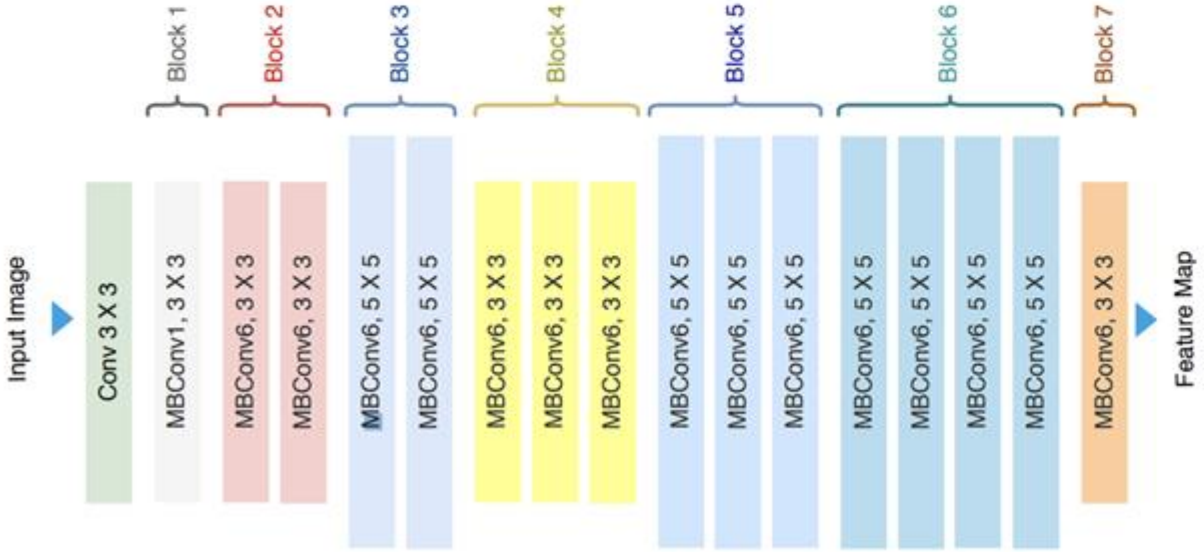
Xception



Xception Model Architecture

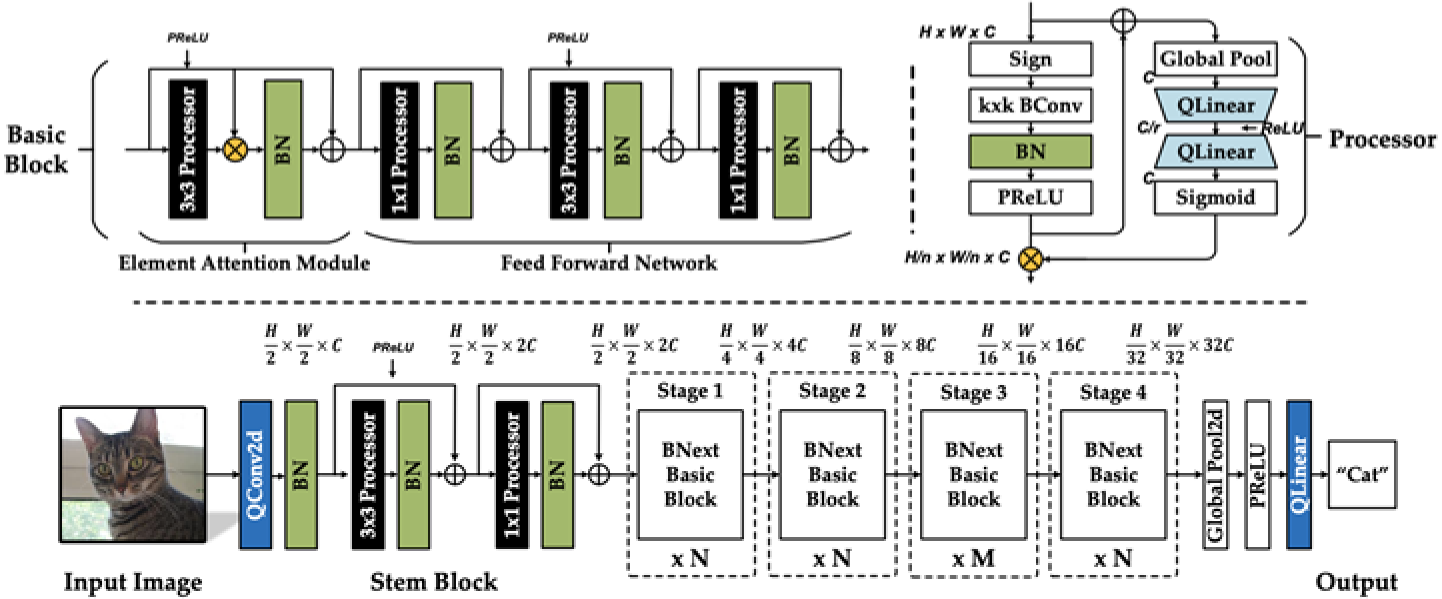
Challenge

EfficientNet



Solution

BNext



Solution

Quantization

將 BNext model 中藍色部分 quantize 成 INT8 或 INT4。

Pruning

Remove unsignificant weights

Fake-buster: a lightweight solution for deepfake detection

Quantization

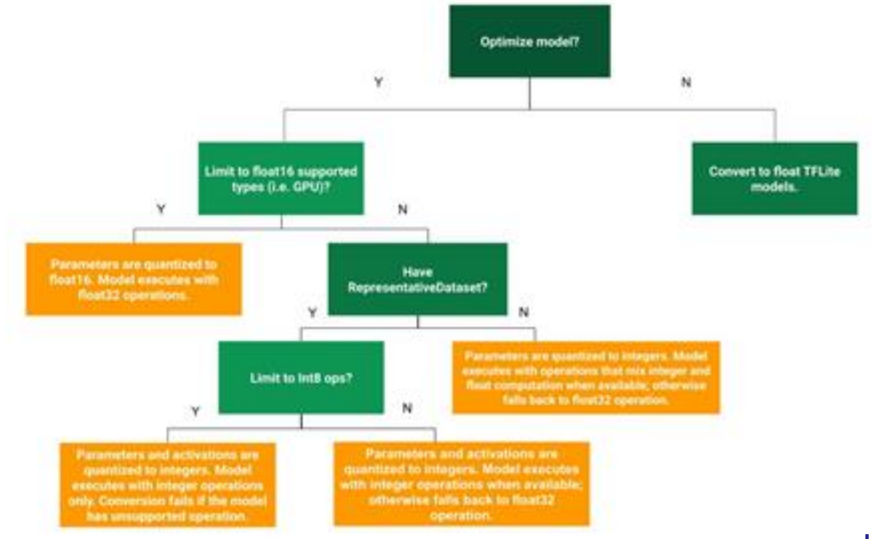
訓練模型需要部署在硬體較為受限的智慧型裝置，模型運算在吃緊的硬體資源中顯得笨重，此時可以採取模型優化策略改進。

Pros:

- 神經網路的參數需要過多空間
- 減少檔案的大小
- 減少運算資源
- 達到更快更小的優化成果

Post Training Quantization

訓練後量化 Post Training Quantization 是一種轉換技術，可以減少模型大小，同時還可以改善 CPU 和硬件加速器的延遲，模型精度幾乎沒有下降。



Optimize model?

Limit to float16 supported types (i.e. GPU)?

Convert to float TFLite models.

Parameters are quantized to float16. Model executes with float32 operations.

Have RepresentativeDataset?

Limit to Int8 ops?

Parameters are quantized to integers. Model executes with operations that mix integer and float computation when available; otherwise falls back to float32 operation.

Parameters and activations are quantized to integers. Model executes with integer operations only. Conversion fails if the model has unsupported operation.

Parameters and activations are quantized to integers. Model executes with integer operations when available; otherwise falls back to float32 operation.

https://ai.google.dev/edge/litert/models/post_training_quantization?hl=zh-tw

Post Training Quantization

訓練流程

模型準備：

獲取已訓練好的 `float32` 模型。

選擇量化框架：

選擇支持 PTQ 的框架。

校準數據集：

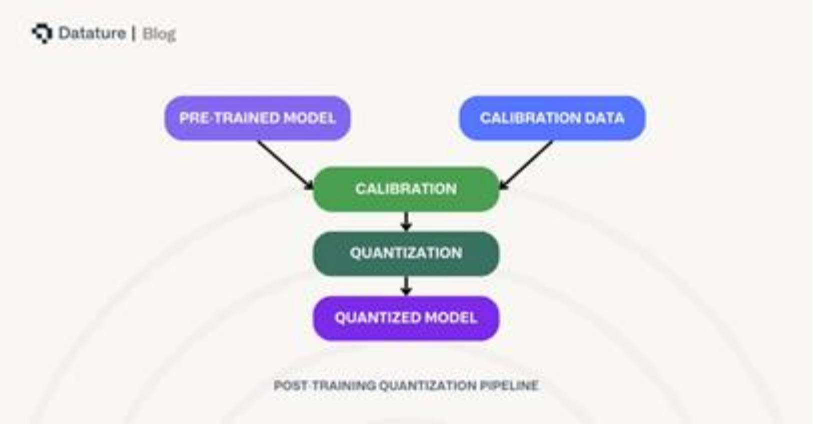
使用一小部分未標註的數據作為校準數據集，計算量化過程中的縮放比例（scale）和零點（zero-point）。

模型量化：

使用框架提供的工具對模型進行量化，將權重和激活轉換為低精度。

驗證與性能測試：

驗證量化後的模型性能是否符合需求，並進行精度和推理速度測試。



https://www.datature.io/blog/introducing-post-training-quantization-feature-and-mechanics-explained

Pruning

Using Unstructure Pruning to achieve higher sparsity

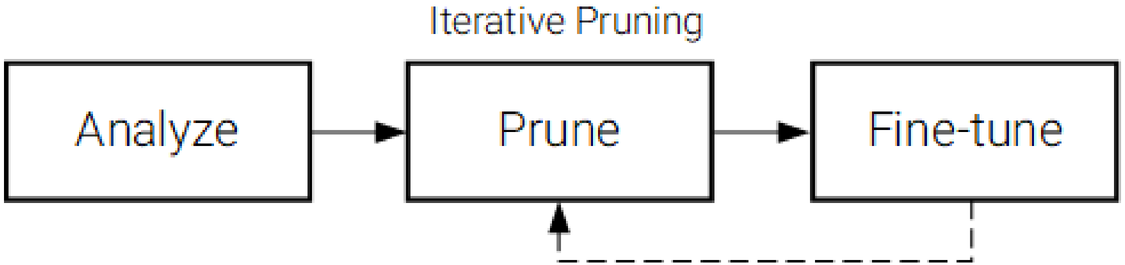
- **TensorFlow Lite**
- Using **XNNPACK library** to apply pruning wrapper only to the parts that can be accelerated
- **XNNPACK library** is included with the pre-built TensorFlow Lite binaries for Android and iOS

Build fast, sparse on-device models with the new TF MOT Pruning API
Accelerating TensorFlow Lite with XNNPACK Integration
TensorFlow Lite Core ML delegate enables faster inference on iPhones and iPads

Pruning

Using Iteritive Pruning and Fintuning

- Leveraging the process of iterative pruning, higher pruning rates can be achieved without any significant loss of model performance.



https://docs.amd.com/r/2.5-English/ug1333-ai-optimizer/Overview

Challenge

❑ 模型壓縮與性能權衡

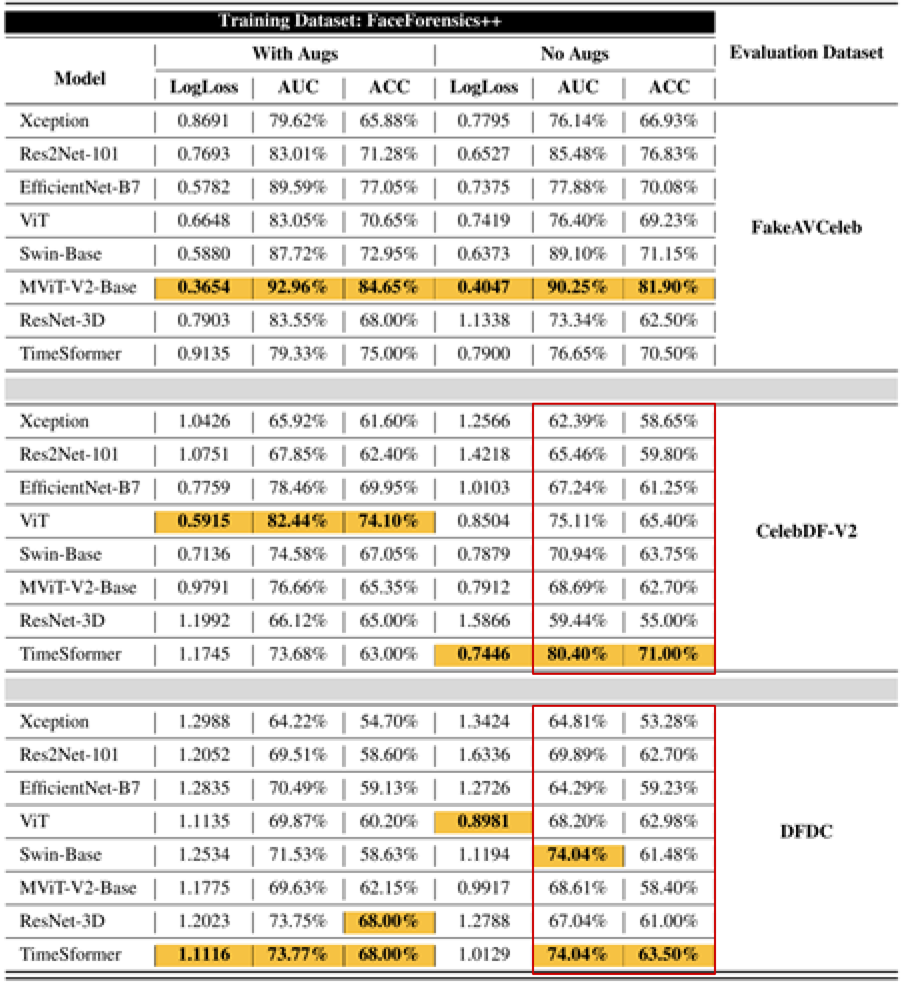
❑ 適應性與泛化能力

❑ 部署到手機平台的兼容性

Challenge

適應性與泛化能力

挑戰：Deepfake技術多樣化且不斷演進，期望該模型在其他測試集上也能維持良好的偵測準確率，展現強大的泛化能力。考慮到Deepfake技術的多樣化和不斷演進，我們將致力於找到有效的方法來提升模型在不同Deepfake生成技術下的適應性。

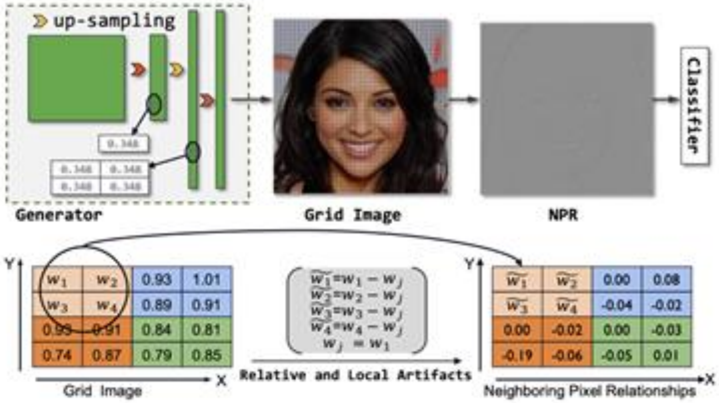


	Training Dataset: FaceForensics++						
Model	With Augs			No Augs			Evaluation Dataset
	LogLoss	AUC	ACC	LogLoss	AUC	ACC	
Xception	0.8691	79.62%	65.88%	0.7795	76.14%	66.93%	FakeAVCeleb
Res2Net-101	0.7693	83.01%	71.28%	0.6527	85.48%	76.83%	
EfficientNet-B7	0.5782	89.59%	77.05%	0.7375	77.88%	70.08%	
ViT	0.6648	83.05%	70.65%	0.7419	76.40%	69.23%	
Swin-Base	0.5880	87.72%	72.95%	0.6373	89.10%	71.15%	
MViT-V2-Base	**0.3654**	**92.96%**	**84.65%**	**0.4047**	**90.25%**	**81.90%**	
ResNet-3D	0.7903	83.55%	68.00%	1.1338	73.34%	62.50%	
TimeSformer	0.9135	79.33%	75.00%	0.7900	76.65%	70.50%	
Xception	1.0426	65.92%	61.60%	1.2566	62.39%	58.65%	CelebDF-V2
Res2Net-101	1.0751	67.85%	62.40%	1.4218	65.46%	59.80%	
EfficientNet-B7	0.7759	78.46%	69.95%	1.0103	67.24%	61.25%	
ViT	**0.5915**	**82.44%**	**74.10%**	0.8504	75.11%	65.40%	
Swin-Base	0.7136	74.58%	67.05%	0.7879	70.94%	63.75%	
MViT-V2-Base	0.9791	76.66%	65.35%	0.7912	68.69%	62.70%	
ResNet-3D	1.1992	66.12%	65.00%	1.5866	59.44%	55.00%	
TimeSformer	1.1745	73.68%	63.00%	**0.7446**	**80.40%**	**71.00%**	
Xception	1.2988	64.22%	54.70%	1.3424	64.81%	53.28%	DFDC
Res2Net-101	1.2052	69.51%	58.60%	1.6336	69.89%	62.70%	
EfficientNet-B7	1.2835	70.49%	59.13%	1.2726	64.29%	59.23%	
ViT	1.1135	69.87%	60.20%	**0.8981**	68.20%	62.98%	
Swin-Base	1.2534	71.53%	58.63%	1.1194	**74.04%**	61.48%	
MViT-V2-Base	1.1775	69.63%	62.15%	0.9917	68.61%	58.40%	
ResNet-3D	1.2023	73.75%	**68.00%**	1.2788	67.04%	61.00%	
TimeSformer	**1.1116**	**73.77%**	**68.00%**	1.0129	**74.04%**	**63.50%**	

Deepfake Detection: Analyzing Model Generalization Across Architectures, Datasets, and Pre-Training Paradigms

Solution

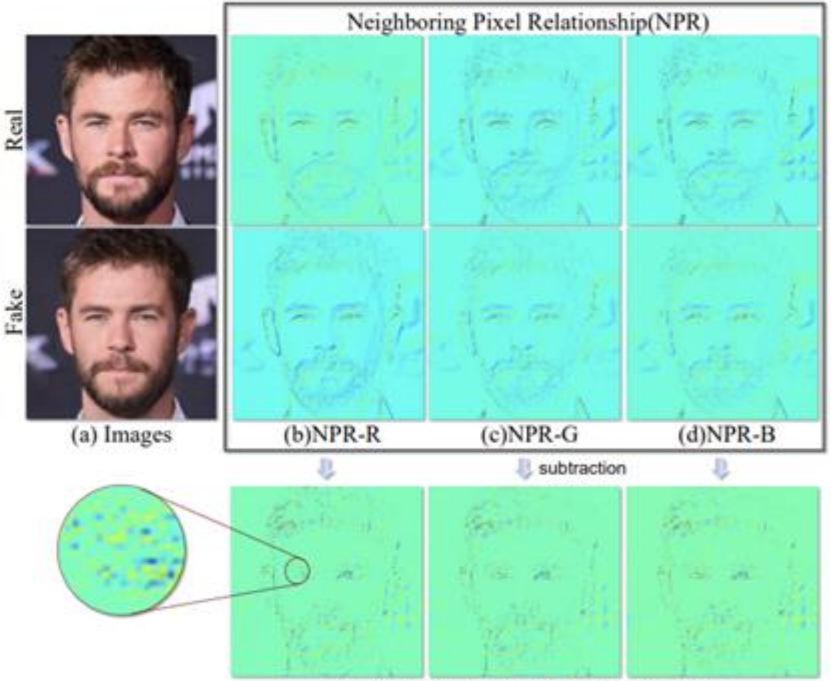
Neighboring Pixel Relationship



w_1	w_2	0.93	1.01
w_3	w_4	0.89	0.91
0.93	0.91	0.84	0.81
0.74	0.87	0.79	0.85

$$\widetilde{w_1} = w_1 - w_j$$
$$\widetilde{w_2} = w_2 - w_j$$
$$\widetilde{w_3} = w_3 - w_j$$
$$\widetilde{w_4} = w_4 - w_j$$
$$w_j = w_1$$

$\widetilde{w_1}$	$\widetilde{w_2}$	0.00	0.08
$\widetilde{w_3}$	$\widetilde{w_4}$	-0.04	-0.02
0.00	-0.02	0.00	-0.03
-0.19	-0.06	-0.05	0.01

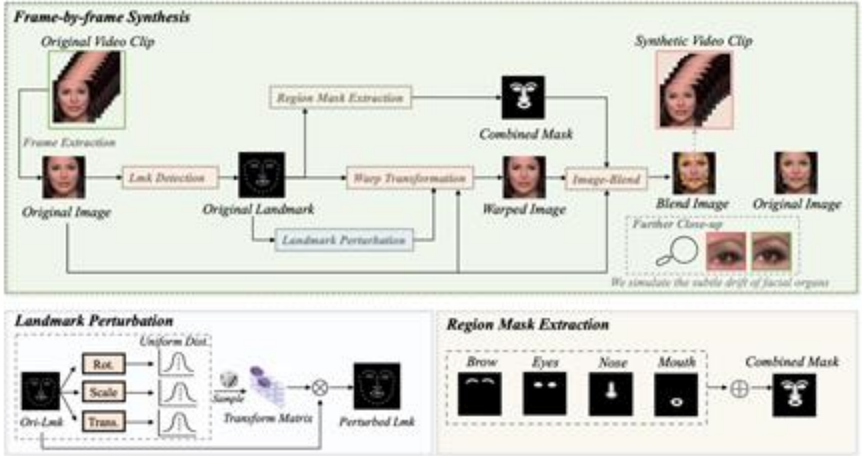


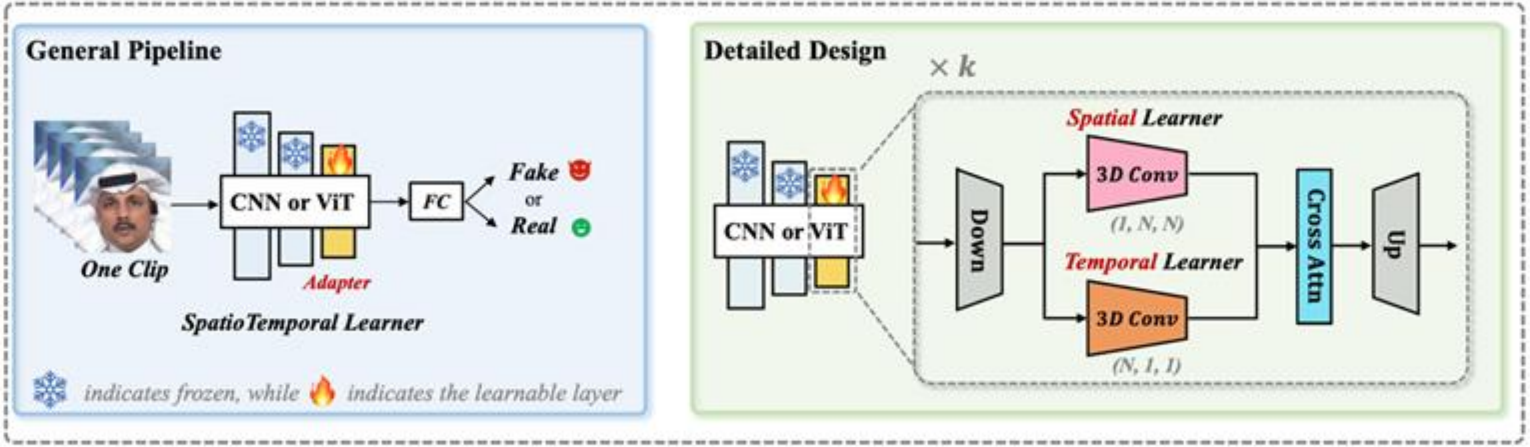
(e)Different NPRs bewteen Real and Fake

Rethinking the Up-Sampling Operations in CNN-based Generative Network for Generalizable Deepfake Detection

Solution

Video-Level Blending and Spatiotemporal Adapter

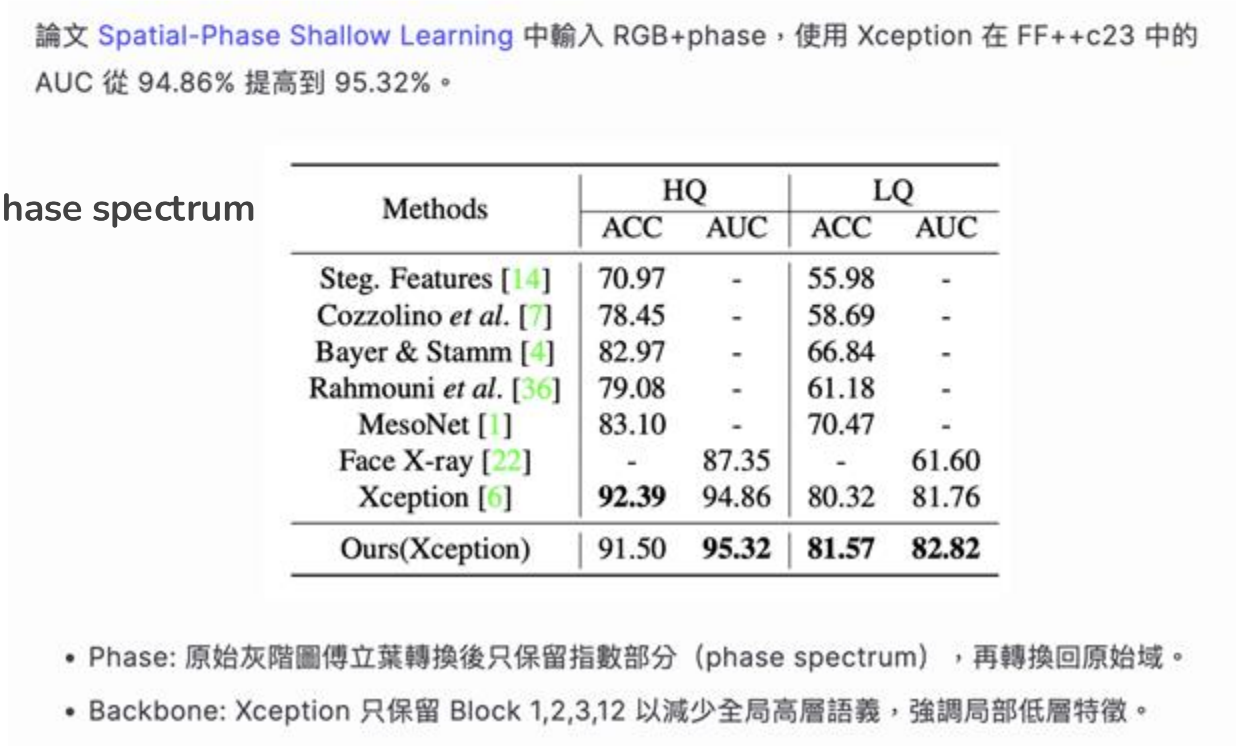




Generalizing Deepfake Video Detection with Plug-and-Play: Video-Level Blending and Spatiotemporal Adapter

Solution

Magnitude & Phase spectrum



論文 Spatial-Phase Shallow Learning 中輸入 RGB+phase，使用 Xception 在 FF++c23 中的 AUC 從 94.86% 提高到 95.32%。

Methods	HQ		LQ	
	ACC	AUC	ACC	AUC
Steg. Features [14]	70.97	-	55.98	-
Cozzolino *et al*. [7]	78.45	-	58.69	-
Bayer & Stamm [4]	82.97	-	66.84	-
Rahmouni *et al*. [36]	79.08	-	61.18	-
MesoNet [1]	83.10	-	70.47	-
Face X-ray [22]	-	87.35	-	61.60
Xception [6]	**92.39**	94.86	80.32	81.76
Ours(Xception)	91.50	**95.32**	**81.57**	**82.82**

- Phase: 原始灰階圖傅立葉轉換後只保留指數部分（phase spectrum），再轉換回原始域。
- Backbone: Xception 只保留 Block 1,2,3,12 以減少全局高層語義，強調局部低層特徵。

Spatial-Phase Shallow Learning: Rethinking Face Forgery Detection in Frequency Domain

Challenge

❑ 模型壓縮與性能權衡

❑ 適應性與泛化能力

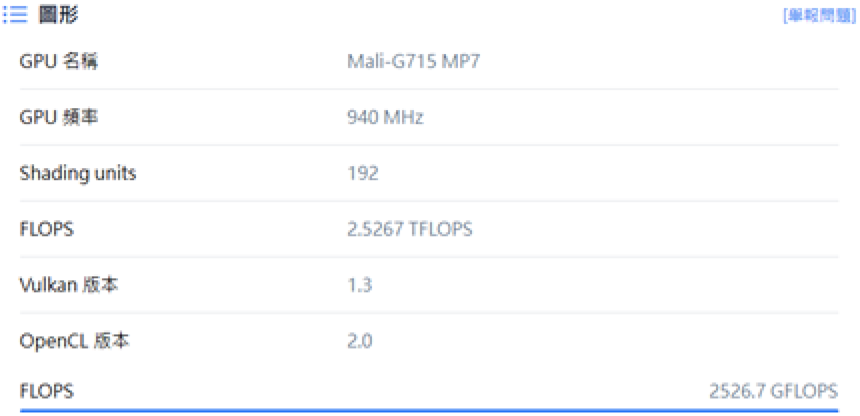
❑ **部署到手機平台的兼容性**

https://pytorch.org/mobile/home/

Challenge

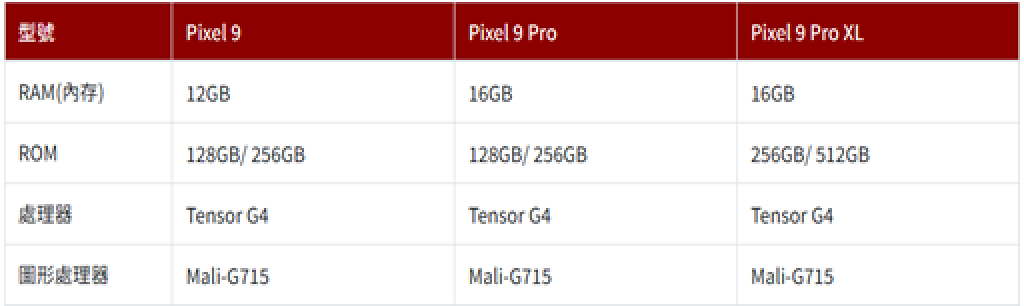
將DeepFake模型成功部署在手機上(Google Pixel 9)進行運作，並且在使用體驗上達到不會讓使用有等待的感覺。

目標: RAM控制在1GB內，即時的推理完成



圖形 [舉報問題]

GPU 名稱	Mali-G715 MP7
GPU 頻率	940 MHz
Shading units	192
FLOPS	2.5267 TFLOPS
Vulkan 版本	1.3
OpenCL 版本	2.0
FLOPS	2526.7 GFLOPS



型號	Pixel 9	Pixel 9 Pro	Pixel 9 Pro XL
RAM(內存)	12GB	16GB	16GB
ROM	128GB/ 256GB	128GB/ 256GB	256GB/ 512GB
處理器	Tensor G4	Tensor G4	Tensor G4
圖形處理器	Mali-G715	Mali-G715	Mali-G715

https://www.landtop.com.tw/reviews/430

Dataset

Dataset

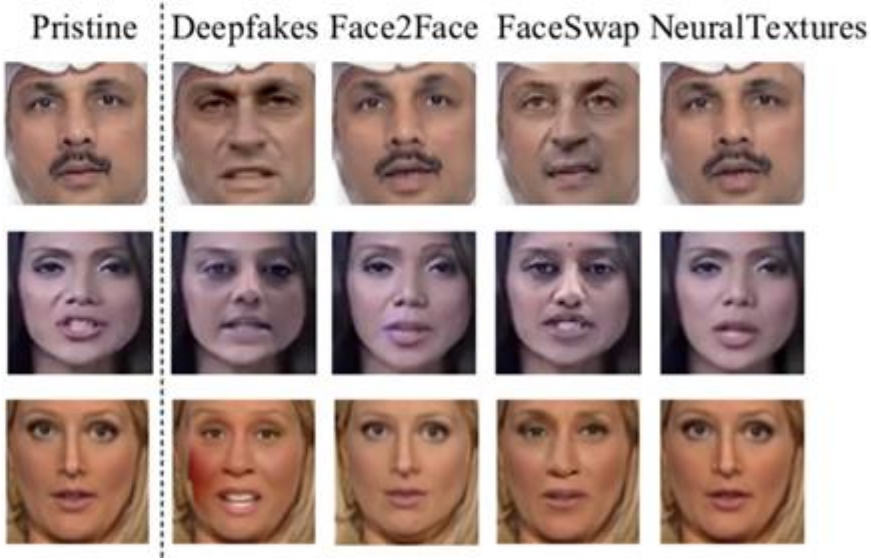
FaceForensics++ (FF++)

描述：FF++資料集是深偽檢測研究中最早且最常用的資料集之一，包含四種不同的Deepfake生成技術（包括DeepFakes、FaceSwap、Face2Face、NeuralTextures），提供不同的影像質量級別（原始、高質量、低質量）。

用途：適合用於評估模型在各種深偽技術和圖像質量下的偵測性能。

總量：FF++包含1000個真實影片和4000個偽造影片（每種Deepfake生成技術生成1000個影片）。

訓練、驗證、測試集劃分：常見劃分為720個影片用於訓練、140個用於驗證、140個用於測試。



Dataset

Celeb-DF

描述：Celeb-DF資料集專注於提升Deepfake影片的真實性，採用了更高級的Deepfake生成技術，使生成影片更加逼真且難以分辨，並包含多位名人視頻。

用途：適合用於測試模型在高仿真度Deepfake影像下的準確性，特別是對於高質量偽造影像的檢測。

總量：Celeb-DF v2版本包含590個真實影片和5639個偽造影片。

訓練、驗證、測試集劃分：通常使用70%的數據（約413部真實影片和3950部偽造影片）進行訓練，15%用於驗證，15%用於測試。

Dataset

DeepFake Detection Challenge (DFDC)

描述：DFDC資料集由Facebook提供，規模龐大，包含來自不同年齡、性別和種族的人物影像，並結合了多種Deepfake生成技術。

用途：適合用於測試模型在多樣化的人群、場景和生成技術上的泛化能力，是針對Deepfake模型訓練和測試的全面性資料集。

總量：DFDC資料集包含超過5000個真實影片和超過10萬個偽造影片。

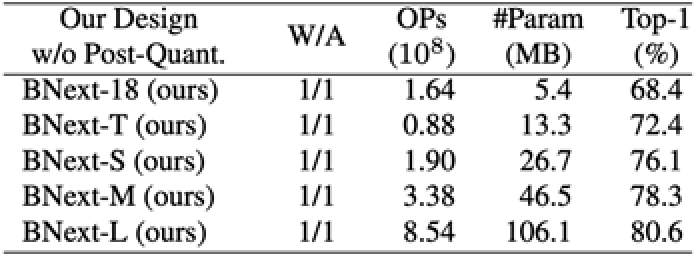
訓練、驗證、測試集劃分：官方劃分約80%數據用於訓練，10%用於驗證，10%用於測試。

Results

報告

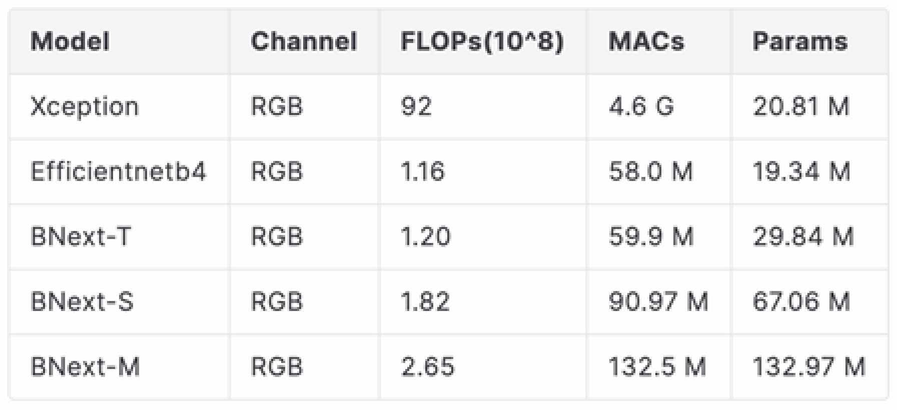
- 三個模型的參數比較(Xception / EfficientNetB4 / BNext)
- 三個模型在三個資料集的表現(FF++c23 / Celeb-DF / DFDC)
- 加入剪枝後的表現
- 加入量化後的表現
- 加入NPR的表現
- 加入VB的表現
- 加入其他方法（Maginitude, Phase）的表現

參數比較



Our Design w/o Post-Quant.	W/A	OPs (10^8)	#Param (MB)	Top-1 (%)
BNext-18 (ours)	1/1	1.64	5.4	68.4
BNext-T (ours)	1/1	0.88	13.3	72.4
BNext-S (ours)	1/1	1.90	26.7	76.1
BNext-M (ours)	1/1	3.38	46.5	78.3
BNext-L (ours)	1/1	8.54	106.1	80.6

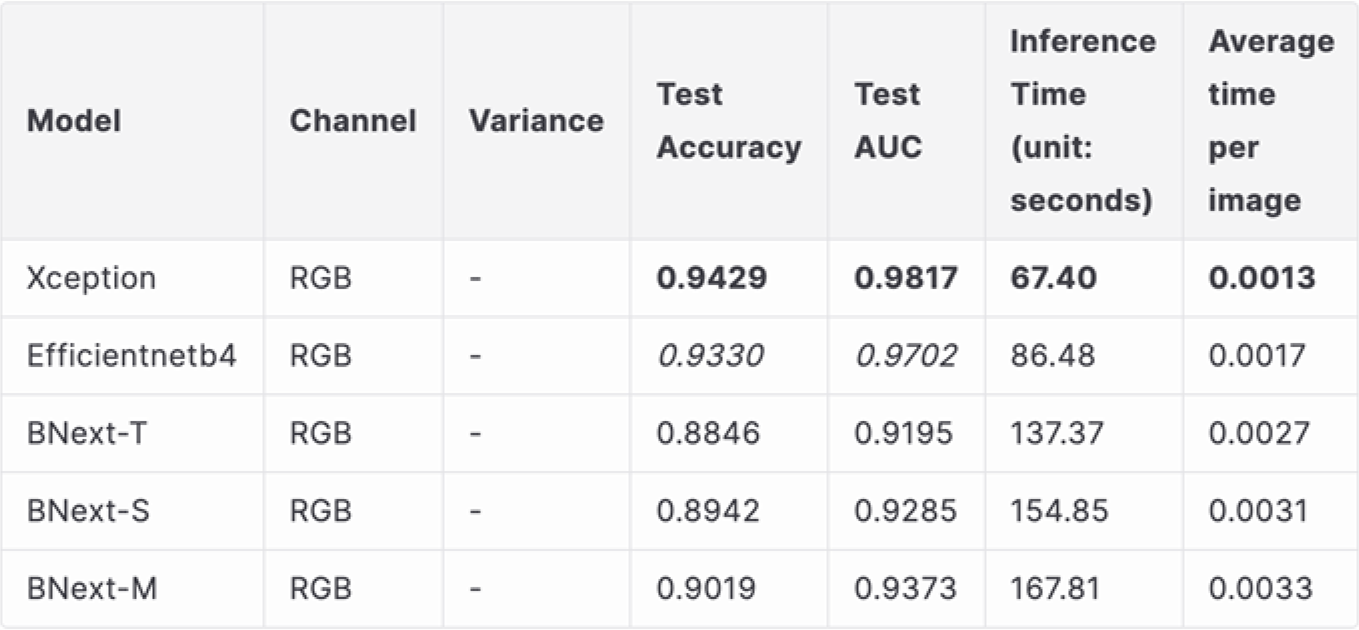
論文中的參數結果



Model	Channel	FLOPs(10^8)	MACs	Params
Xception	RGB	92	4.6 G	20.81 M
Efficientnetb4	RGB	1.16	58.0 M	19.34 M
BNext-T	RGB	1.20	59.9 M	29.84 M
BNext-S	RGB	1.82	90.97 M	67.06 M
BNext-M	RGB	2.65	132.5 M	132.97 M

實驗得出的參數結果

FF++c23 (In-dataset)

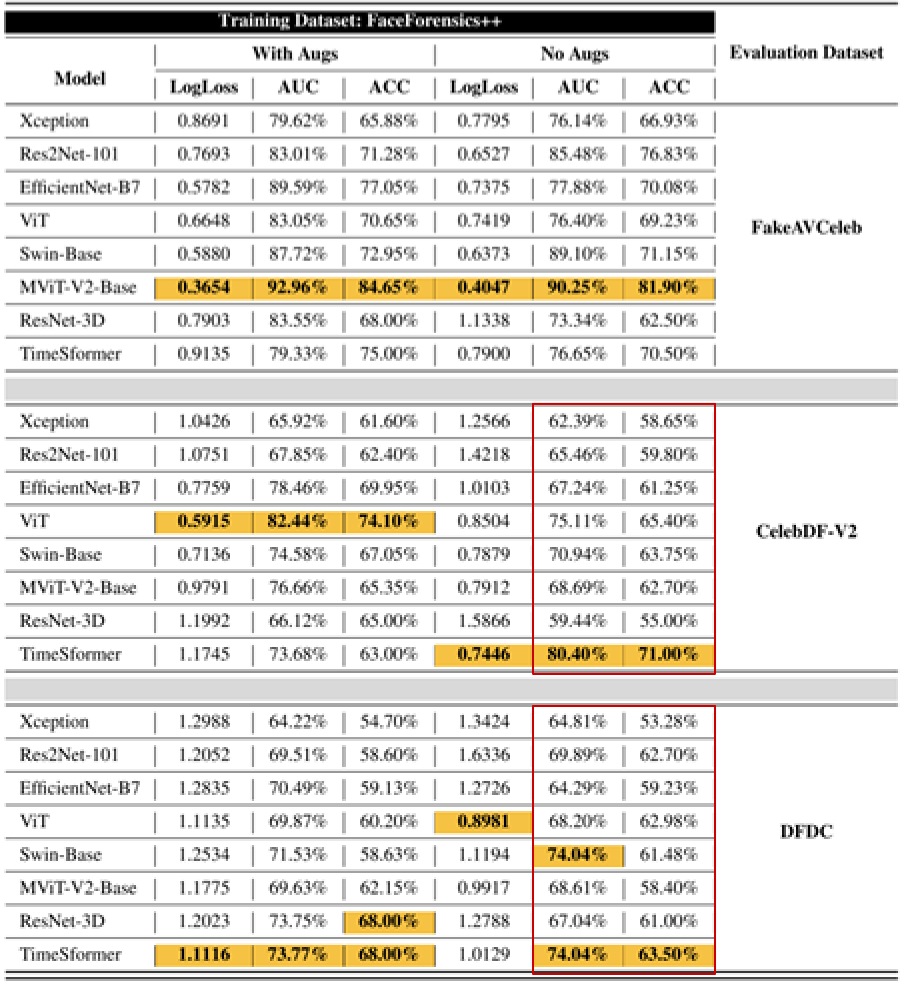


Model	Channel	Variance	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
Xception	RGB	-	**0.9429**	**0.9817**	**67.40**	**0.0013**
Efficientnetb4	RGB	-	*0.9330*	*0.9702*	86.48	0.0017
BNext-T	RGB	-	0.8846	0.9195	137.37	0.0027
BNext-S	RGB	-	0.8942	0.9285	154.85	0.0031
BNext-M	RGB	-	0.9019	0.9373	167.81	0.0033

Challenge

適應性與泛化能力

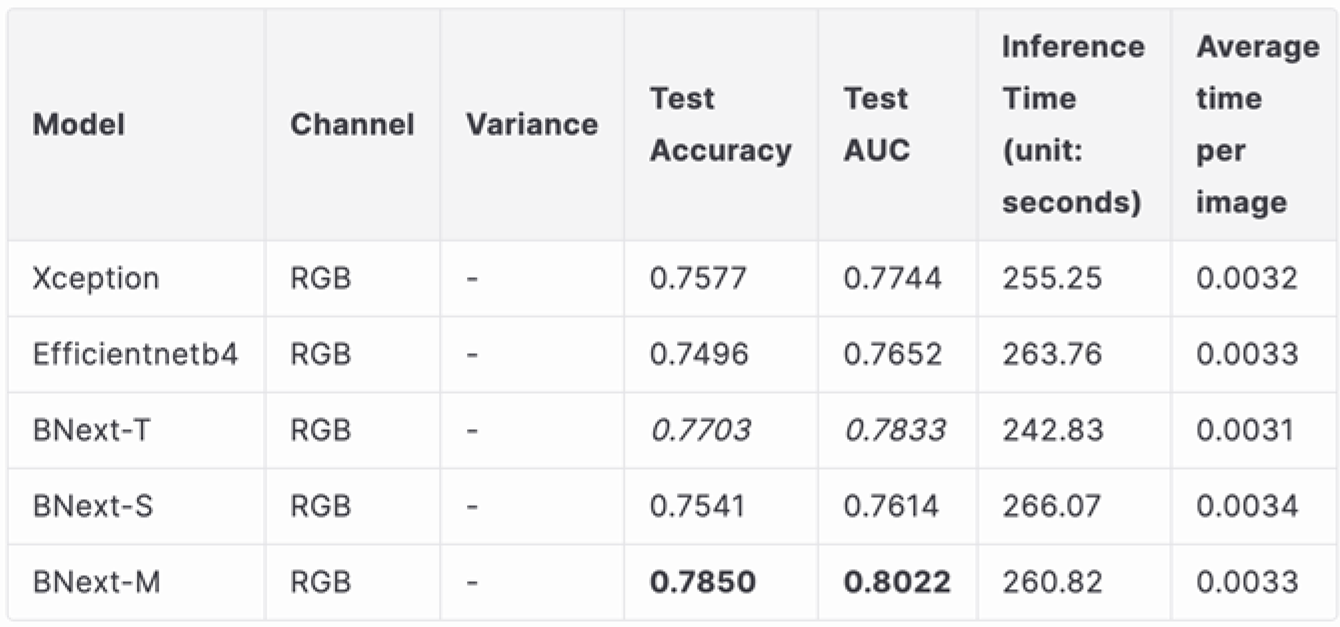
挑戰：Deepfake技術多樣化且不斷演進，期望該模型在其他測試集上也能維持良好的偵測準確率，展現強大的泛化能力。考慮到Deepfake技術的多樣化和不斷演進，我們將致力於找到有效的方法來提升模型在不同Deepfake生成技術下的適應性。



Training Dataset: FaceForensics++							
Model	With Augs			No Augs			Evaluation Dataset
	LogLoss	AUC	ACC	LogLoss	AUC	ACC	
Xception	0.8691	79.62%	65.88%	0.7795	76.14%	66.93%	FakeAVCeleb
Res2Net-101	0.7693	83.01%	71.28%	0.6527	85.48%	76.83%	
EfficientNet-B7	0.5782	89.59%	77.05%	0.7375	77.88%	70.08%	
ViT	0.6648	83.05%	70.65%	0.7419	76.40%	69.23%	
Swin-Base	0.5880	87.72%	72.95%	0.6373	89.10%	71.15%	
MViT-V2-Base	**0.3654**	**92.96%**	**84.65%**	**0.4047**	**90.25%**	**81.90%**	
ResNet-3D	0.7903	83.55%	68.00%	1.1338	73.34%	62.50%	
TimeSformer	0.9135	79.33%	75.00%	0.7900	76.65%	70.50%	
Xception	1.0426	65.92%	61.60%	1.2566	62.39%	58.65%	CelebDF-V2
Res2Net-101	1.0751	67.85%	62.40%	1.4218	65.46%	59.80%	
EfficientNet-B7	0.7759	78.46%	69.95%	1.0103	67.24%	61.25%	
ViT	**0.5915**	**82.44%**	**74.10%**	0.8504	75.11%	65.40%	
Swin-Base	0.7136	74.58%	67.05%	0.7879	70.94%	63.75%	
MViT-V2-Base	0.9791	76.66%	65.35%	0.7912	68.69%	62.70%	
ResNet-3D	1.1992	66.12%	65.00%	1.5866	59.44%	55.00%	
TimeSformer	1.1745	73.68%	63.00%	**0.7446**	**80.40%**	**71.00%**	
Xception	1.2988	64.22%	54.70%	1.3424	64.81%	53.28%	DFDC
Res2Net-101	1.2052	69.51%	58.60%	1.6336	69.89%	62.70%	
EfficientNet-B7	1.2835	70.49%	59.13%	1.2726	64.29%	59.23%	
ViT	1.1135	69.87%	60.20%	**0.8981**	68.20%	62.98%	
Swin-Base	1.2534	71.53%	58.63%	1.1194	**74.04%**	61.48%	
MViT-V2-Base	1.1775	69.63%	62.15%	0.9917	68.61%	58.40%	
ResNet-3D	1.2023	73.75%	**68.00%**	1.2788	67.04%	61.00%	
TimeSformer	**1.1116**	**73.77%**	**68.00%**	1.0129	**74.04%**	**63.50%**	

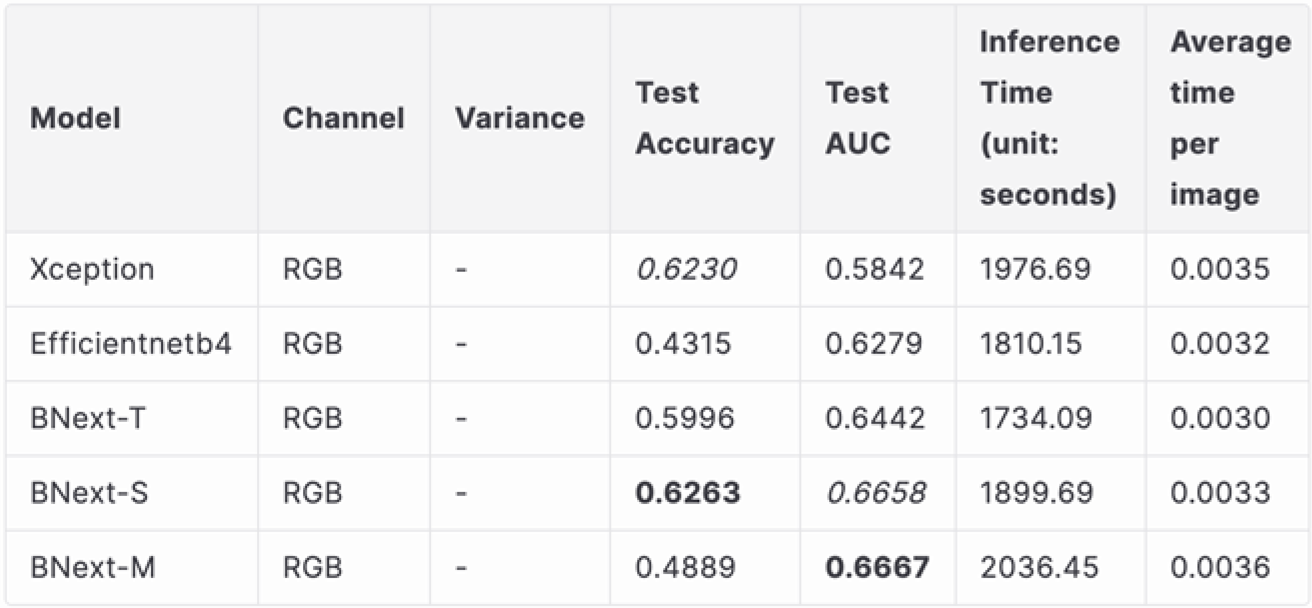
Deepfake Detection: Analyzing Model Generalization Across Architectures, Datasets, and Pre-Training Paradigms

Celed-DF (Cross-dataset)



Model	Channel	Variance	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
Xception	RGB	-	0.7577	0.7744	255.25	0.0032
Efficientnetb4	RGB	-	0.7496	0.7652	263.76	0.0033
BNext-T	RGB	-	*0.7703*	*0.7833*	242.83	0.0031
BNext-S	RGB	-	0.7541	0.7614	266.07	0.0034
BNext-M	RGB	-	**0.7850**	**0.8022**	260.82	0.0033

DFDC (Cross-dataset)



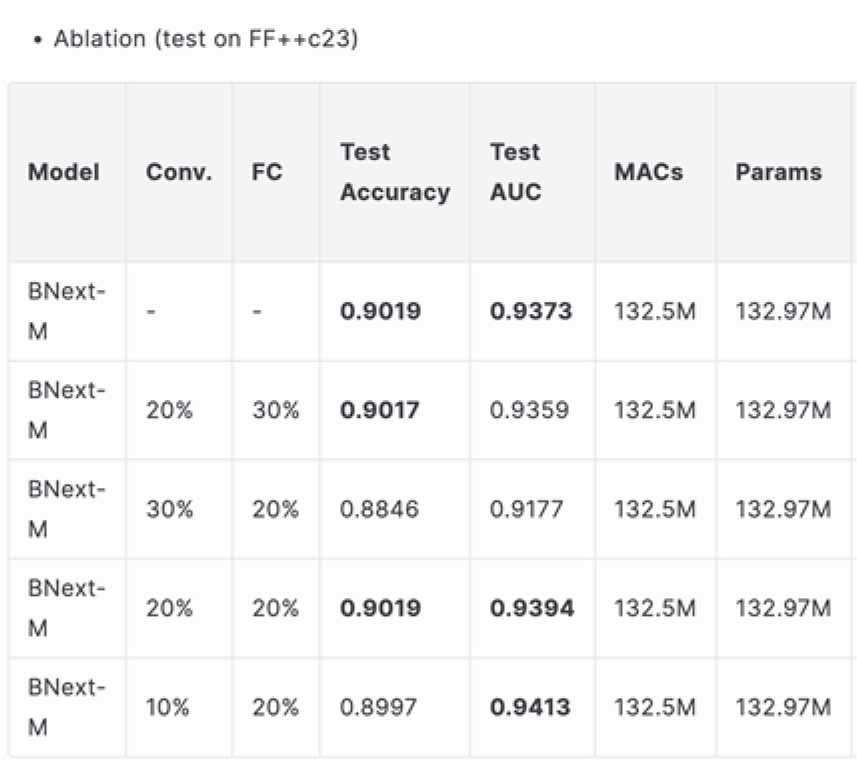
Model	Channel	Variance	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
Xception	RGB	-	*0.6230*	0.5842	1976.69	0.0035
Efficientnetb4	RGB	-	0.4315	0.6279	1810.15	0.0032
BNext-T	RGB	-	0.5996	0.6442	1734.09	0.0030
BNext-S	RGB	-	**0.6263**	*0.6658*	1899.69	0.0033
BNext-M	RGB	-	0.4889	**0.6667**	2036.45	0.0036

Pruning

比例：20%

方法：L1非結構化剪枝

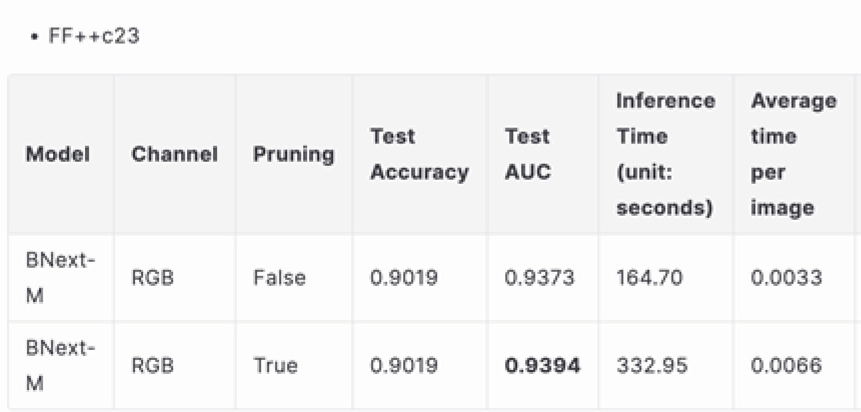
-> 選擇 Conv20%, FC20%



- Ablation (test on FF++c23)

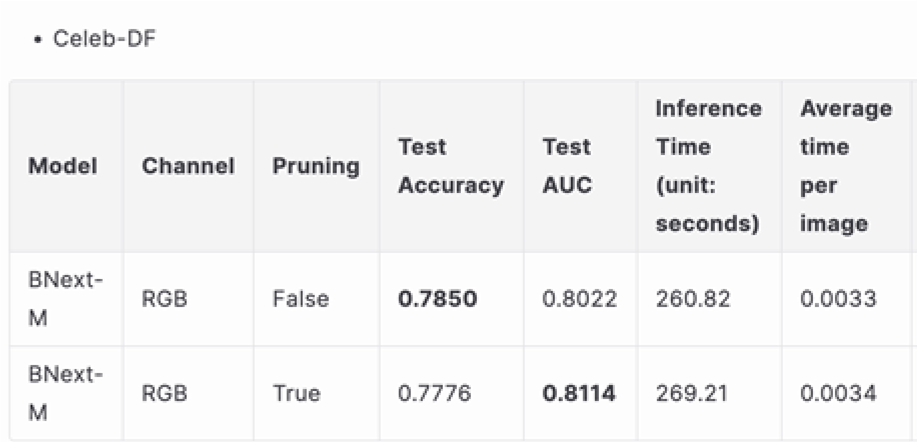
Model	Conv.	FC	Test Accuracy	Test AUC	MACs	Params
BNext-M	-	-	**0.9019**	**0.9373**	132.5M	132.97M
BNext-M	20%	30%	**0.9017**	0.9359	132.5M	132.97M
BNext-M	30%	20%	0.8846	0.9177	132.5M	132.97M
BNext-M	20%	20%	**0.9019**	**0.9394**	132.5M	132.97M
BNext-M	10%	20%	0.8997	**0.9413**	132.5M	132.97M

Pruning



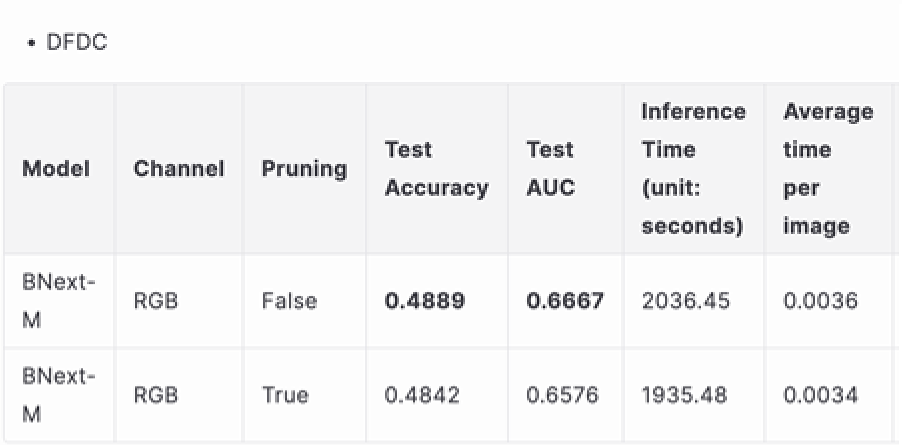
- FF++c23

Model	Channel	Pruning	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	False	0.9019	0.9373	164.70	0.0033
BNext-M	RGB	True	0.9019	**0.9394**	332.95	0.0066



- Celeb-DF

Model	Channel	Pruning	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	False	**0.7850**	0.8022	260.82	0.0033
BNext-M	RGB	True	0.7776	**0.8114**	269.21	0.0034

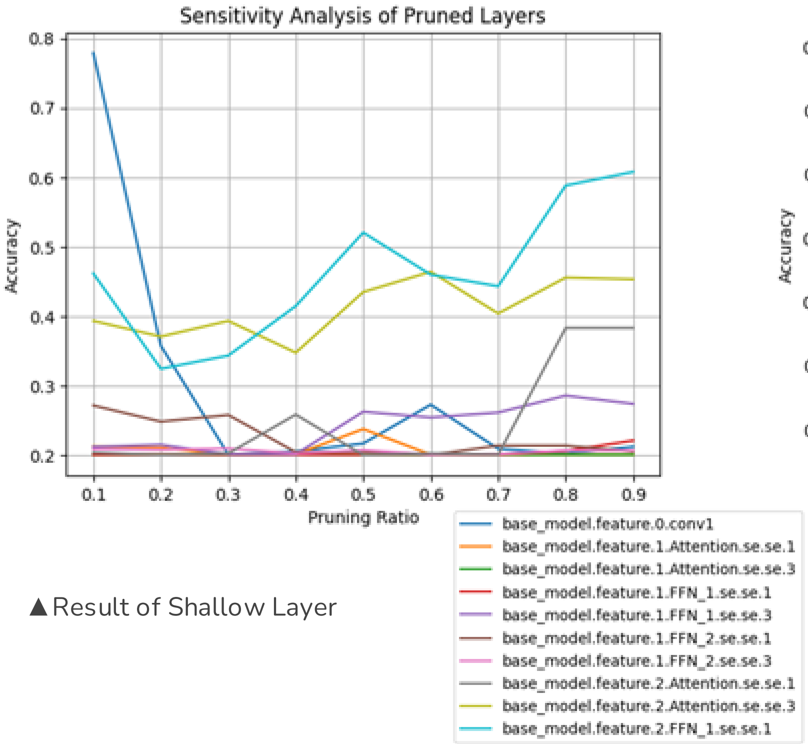


- DFDC

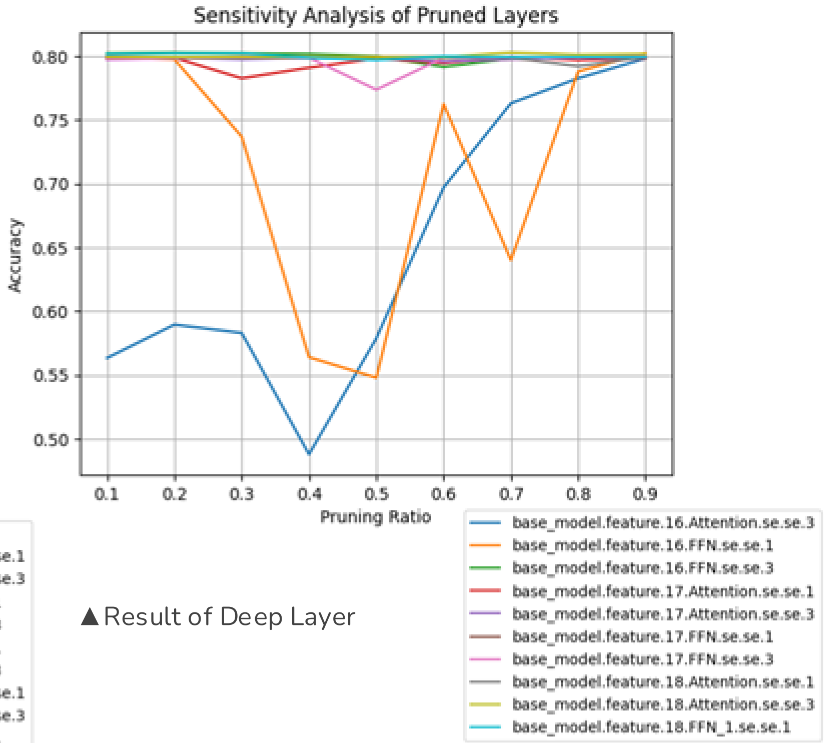
Model	Channel	Pruning	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	False	**0.4889**	**0.6667**	2036.45	0.0036
BNext-M	RGB	True	0.4842	0.6576	1935.48	0.0034

Pruning 2 - Sensitivity Analysis

Prune each layer and test the model



▲Result of Shallow Layer



▲Result of Deep Layer

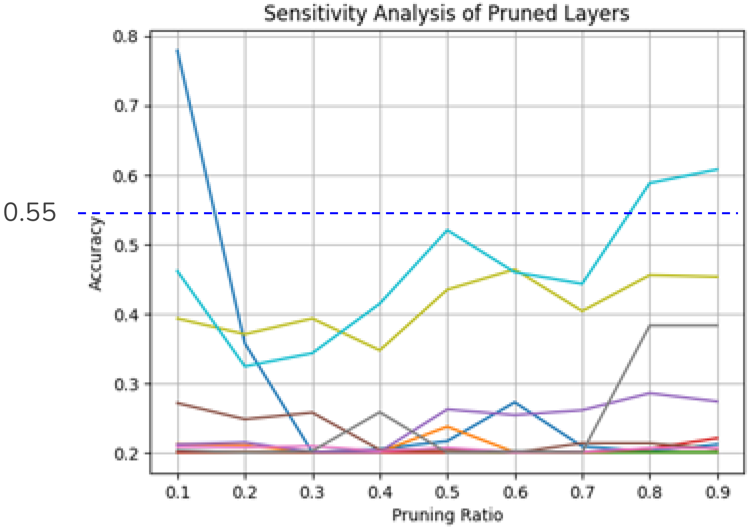
Pruning 2 - Sensitivity Analysis

Base ratio: 0.2

If accuracy in analysis > 0.55, set ratio higher

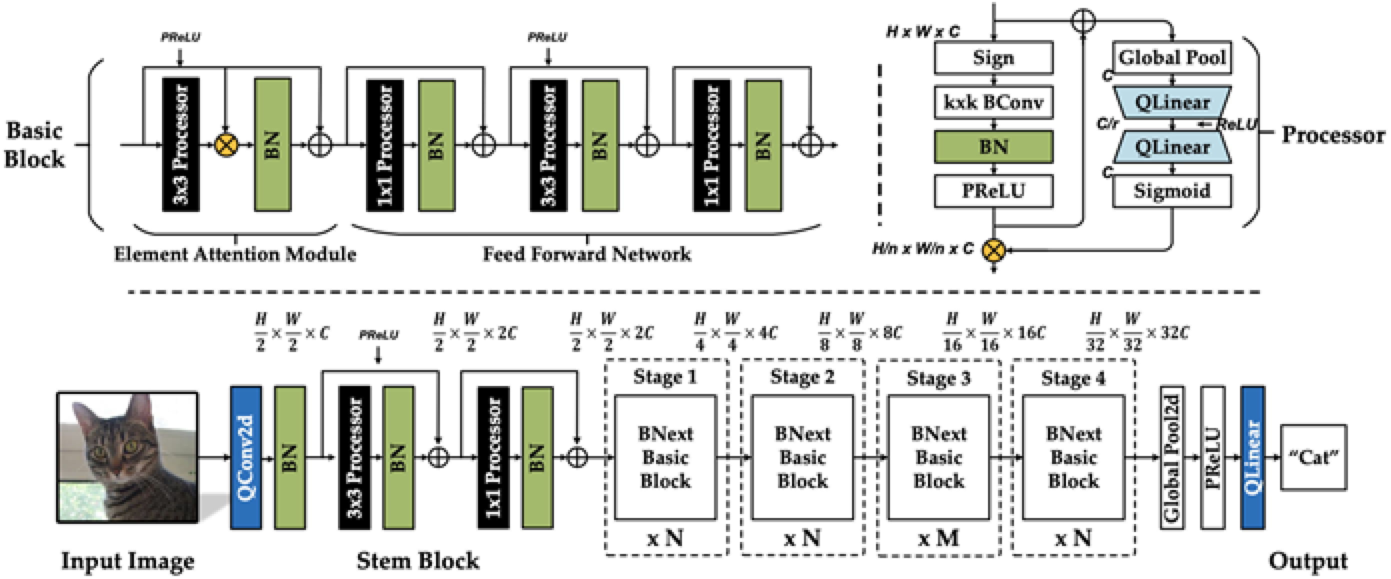
- Ablation (test on FF++c23)

Model	Test Acc	Test AUC	MACs	Params
BNext-M	0.9019	0.9373	132.5M	132.97M
BNext-M	0.8930	0.9102	132.5 M	132.97M

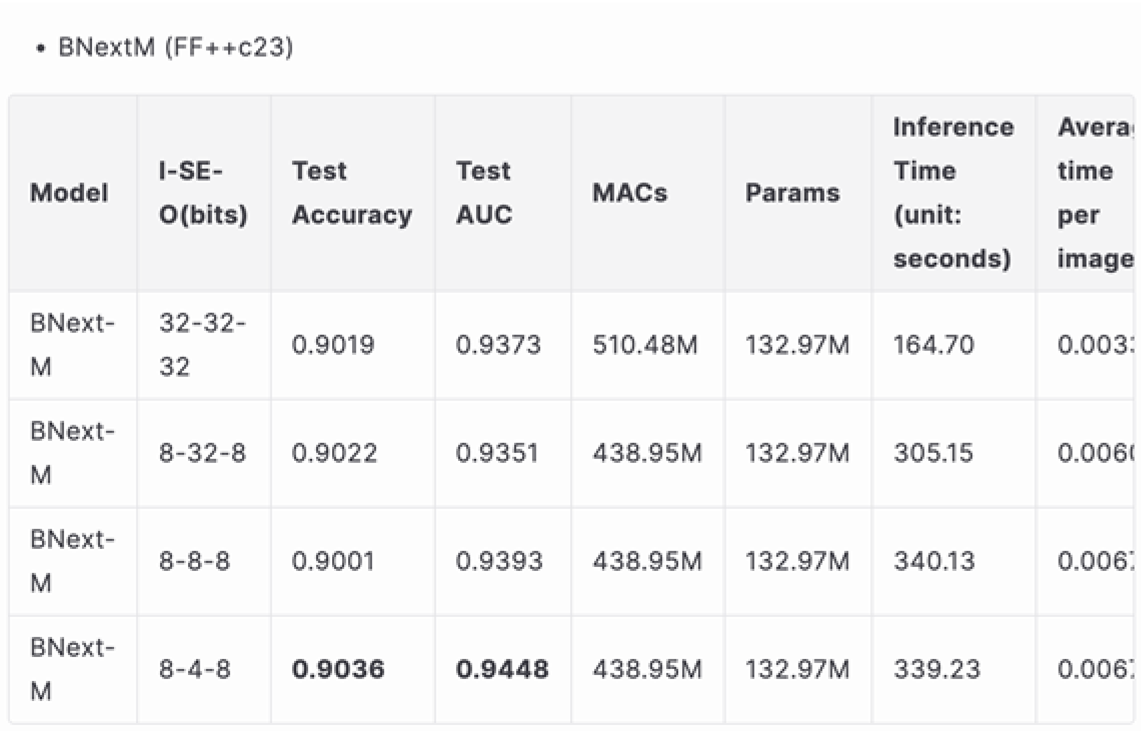


Quantization

BNext



Quantization

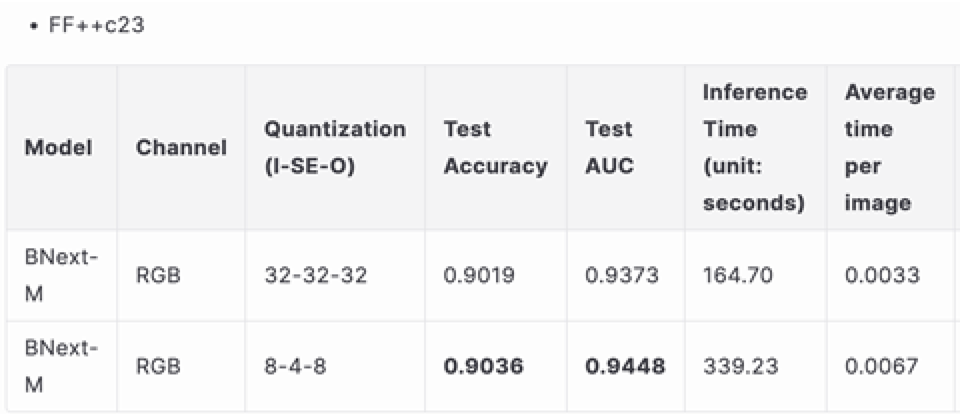


- BNextM (FF++c23)

Model	I-SE-O(bits)	Test Accuracy	Test AUC	MACs	Params	Inference Time (unit: seconds)	Avera time per image
BNext-M	32-32-32	0.9019	0.9373	510.48M	132.97M	164.70	0.003
BNext-M	8-32-8	0.9022	0.9351	438.95M	132.97M	305.15	0.006
BNext-M	8-8-8	0.9001	0.9393	438.95M	132.97M	340.13	0.006
BNext-M	8-4-8	**0.9036**	**0.9448**	438.95M	132.97M	339.23	0.006

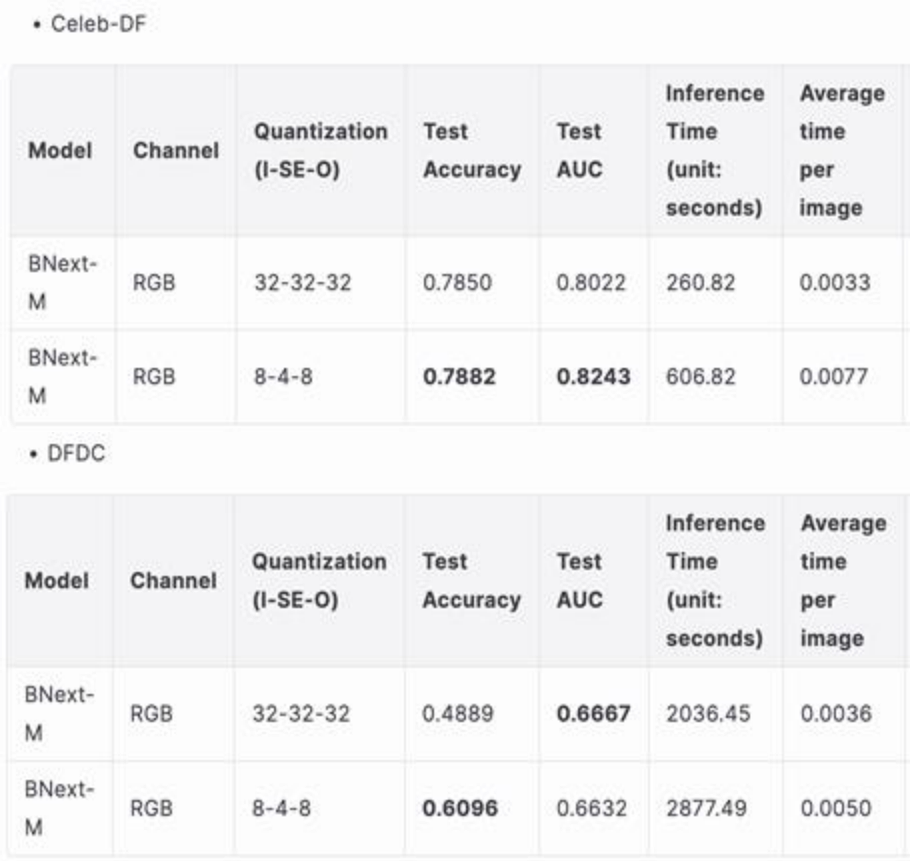
-> 選擇 8-4-8 的 model

Quantization



- FF++c23

Model	Channel	Quantization (I-SE-O)	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	32-32-32	0.9019	0.9373	164.70	0.0033
BNext-M	RGB	8-4-8	**0.9036**	**0.9448**	339.23	0.0067



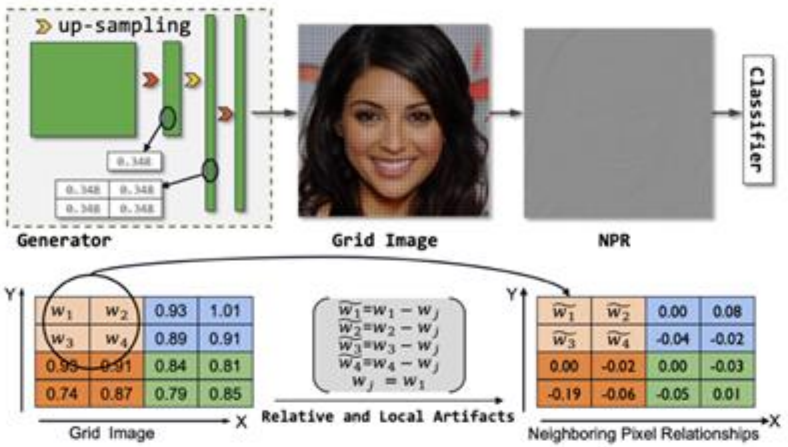
- Celeb-DF

Model	Channel	Quantization (I-SE-O)	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	32-32-32	0.7850	0.8022	260.82	0.0033
BNext-M	RGB	8-4-8	**0.7882**	**0.8243**	606.82	0.0077

- DFDC

Model	Channel	Quantization (I-SE-O)	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	32-32-32	0.4889	**0.6667**	2036.45	0.0036
BNext-M	RGB	8-4-8	**0.6096**	0.6632	2877.49	0.0050

NPR

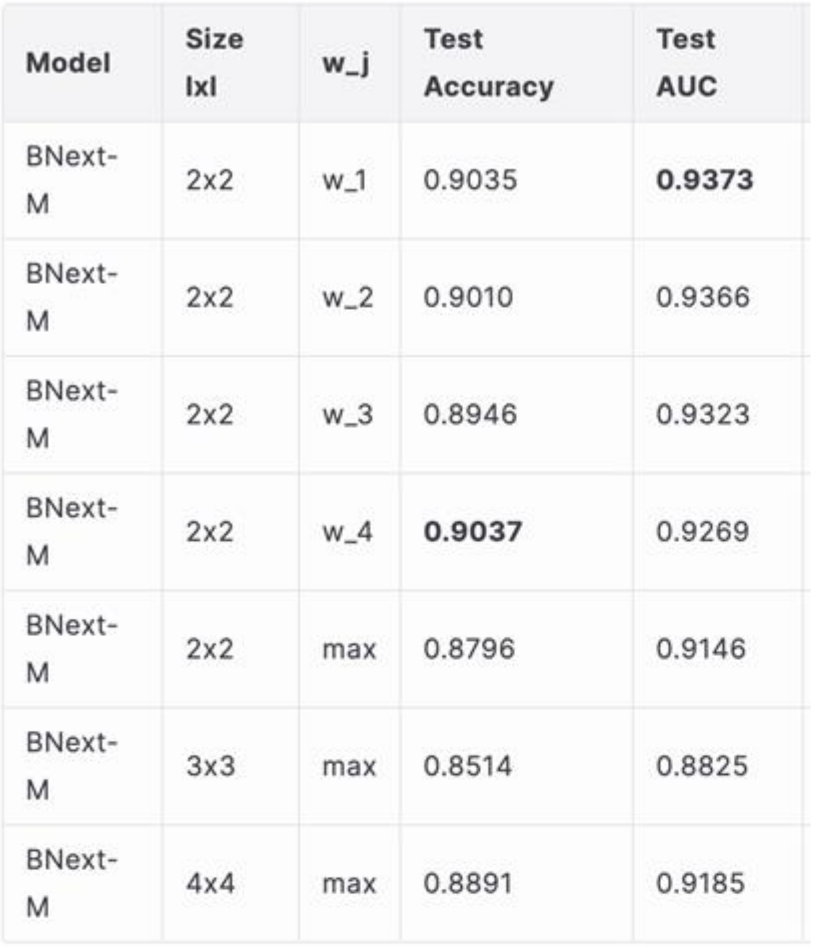


w_1	w_2	0.93	1.01
w_3	w_4	0.89	0.91
0.93	0.91	0.84	0.81
0.74	0.87	0.79	0.85

$$\widetilde{w_1}=w_1-w_j$$
$$\widetilde{w_2}=w_2-w_j$$
$$\widetilde{w_3}=w_3-w_j$$
$$\widetilde{w_4}=w_4-w_j$$
$$w_j = w_1$$

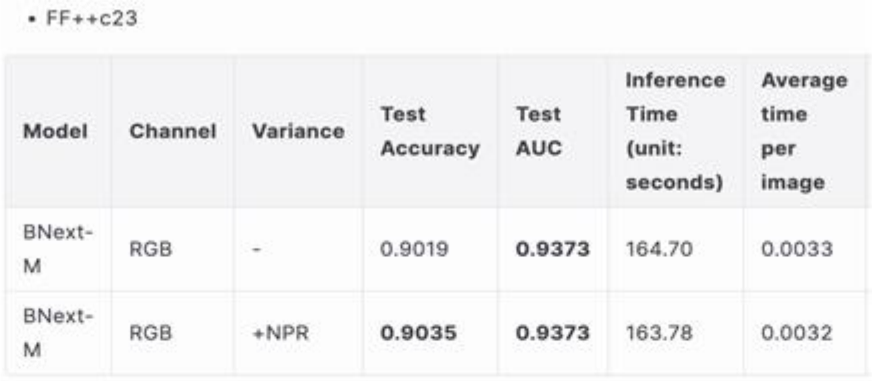
$\widetilde{w_1}$	$\widetilde{w_2}$	0.00	0.08
$\widetilde{w_3}$	$\widetilde{w_4}$	-0.04	-0.02
0.00	-0.02	0.00	-0.03
-0.19	-0.06	-0.05	0.01

-> 選擇參數為 w_1 的 model



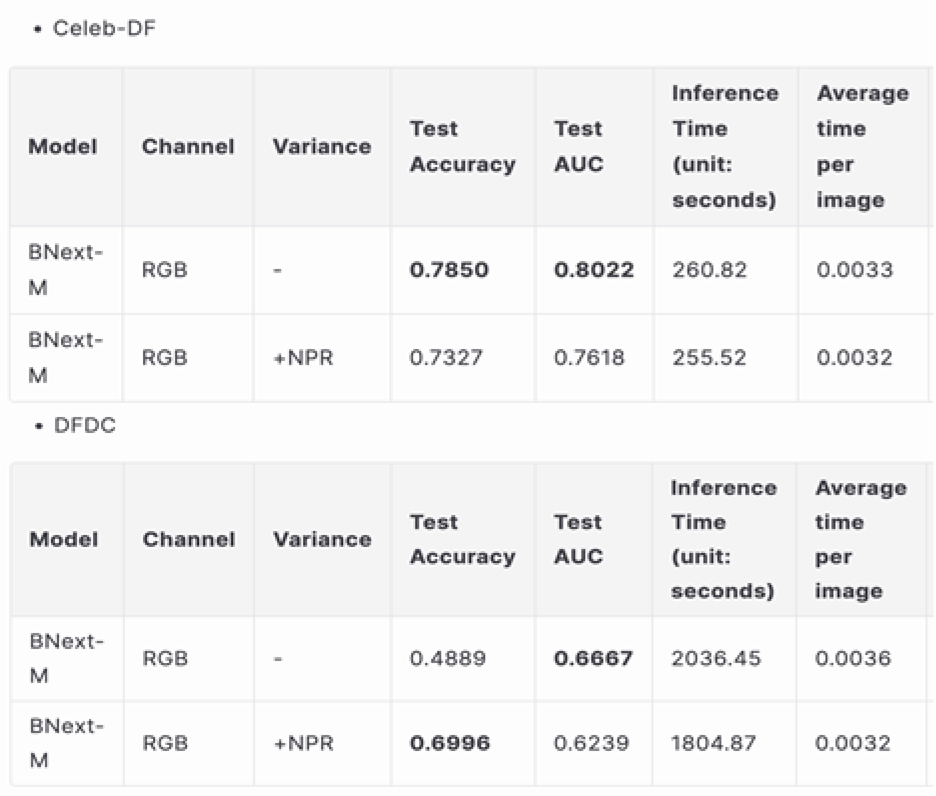
Model	Size lxl	w_j	Test Accuracy	Test AUC
BNext-M	2x2	w_1	0.9035	**0.9373**
BNext-M	2x2	w_2	0.9010	0.9366
BNext-M	2x2	w_3	0.8946	0.9323
BNext-M	2x2	w_4	**0.9037**	0.9269
BNext-M	2x2	max	0.8796	0.9146
BNext-M	3x3	max	0.8514	0.8825
BNext-M	4x4	max	0.8891	0.9185

NPR



- FF++c23

Model	Channel	Variance	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	-	0.9019	**0.9373**	164.70	0.0033
BNext-M	RGB	+NPR	**0.9035**	**0.9373**	163.78	0.0032



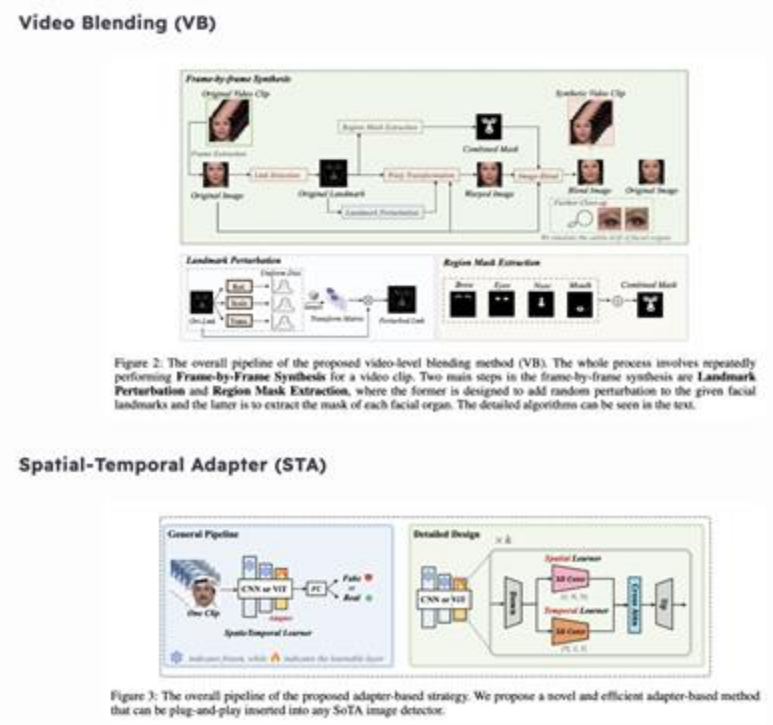
- Celeb-DF

Model	Channel	Variance	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	-	**0.7850**	**0.8022**	260.82	0.0033
BNext-M	RGB	+NPR	0.7327	0.7618	255.52	0.0032

- DFDC

Model	Channel	Variance	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	-	0.4889	**0.6667**	2036.45	0.0036
BNext-M	RGB	+NPR	**0.6996**	0.6239	1804.87	0.0032

VB&
StA

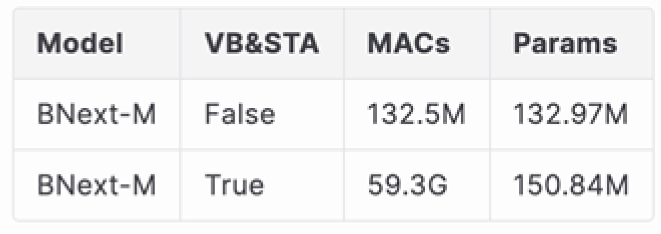


Video Blending (VB)

Figure 2: The overall pipeline of the proposed video-level blending method (VB). The whole process involves repeatedly performing **Frame-by-Frame Synthesis** for a video clip. Two main steps in the frame-by-frame synthesis are **Landmark Perturbation** and **Region Mask Extraction**, where the former is designed to add random perturbation to the given facial landmarks and the latter is to extract the mask of each facial organ. The detailed algorithms can be seen in the text.

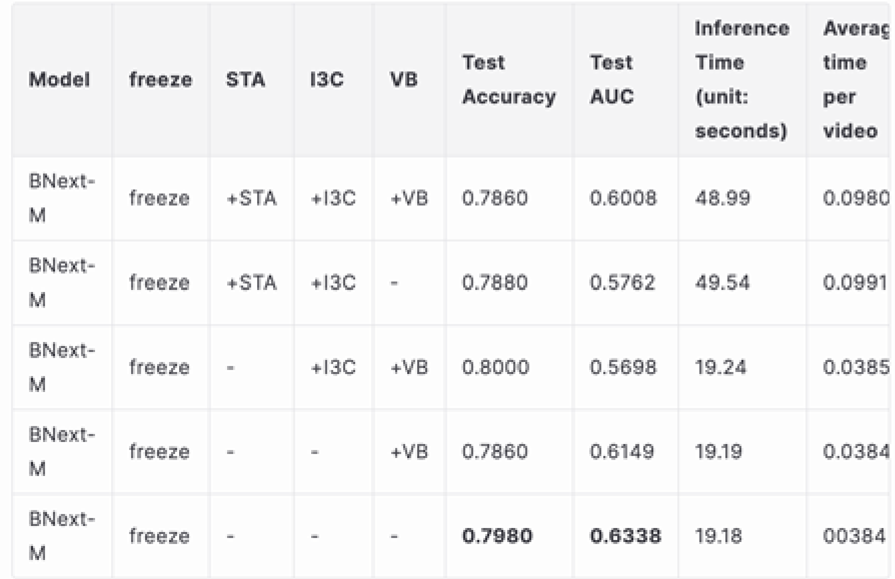
Spatial-Temporal Adapter (STA)

Figure 3: The overall pipeline of the proposed adapter-based strategy. We propose a novel and efficient adapter-based method that can be plug-and-play inserted into any SoTA image detector.



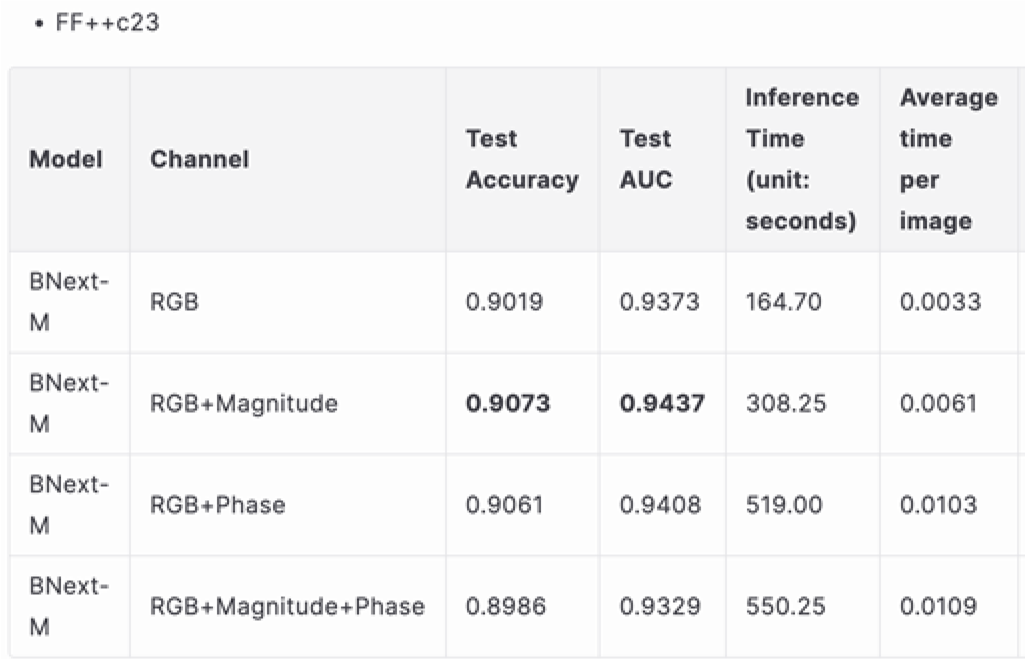
Model	VB&STA	MACs	Params
BNext-M	False	132.5M	132.97M
BNext-M	True	59.3G	150.84M

計算量會變超級大！！！



Model	freeze	STA	I3C	VB	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Averag time per video
BNext-M	freeze	+STA	+I3C	+VB	0.7860	0.6008	48.99	0.0980
BNext-M	freeze	+STA	+I3C	-	0.7880	0.5762	49.54	0.0991
BNext-M	freeze	-	+I3C	+VB	0.8000	0.5698	19.24	0.0385
BNext-M	freeze	-	-	+VB	0.7860	0.6149	19.19	0.0384
BNext-M	freeze	-	-	-	**0.7980**	**0.6338**	19.18	00384

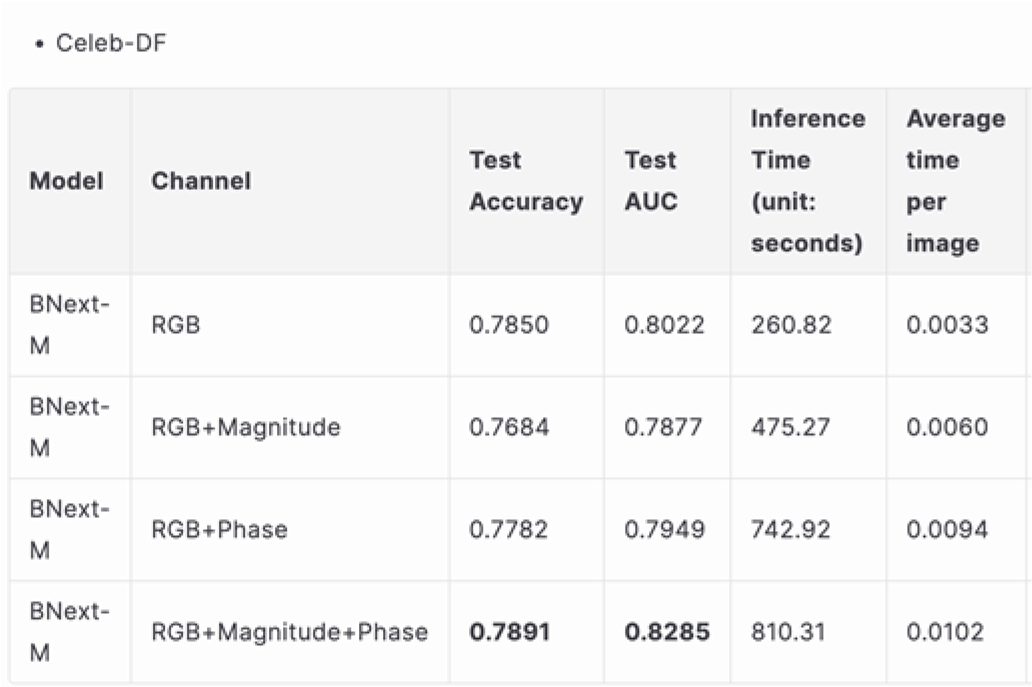
Magnitude & Phase Spectrum



- FF++c23

Model	Channel	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	0.9019	0.9373	164.70	0.0033
BNext-M	RGB+Magnitude	**0.9073**	**0.9437**	308.25	0.0061
BNext-M	RGB+Phase	0.9061	0.9408	519.00	0.0103
BNext-M	RGB+Magnitude+Phase	0.8986	0.9329	550.25	0.0109

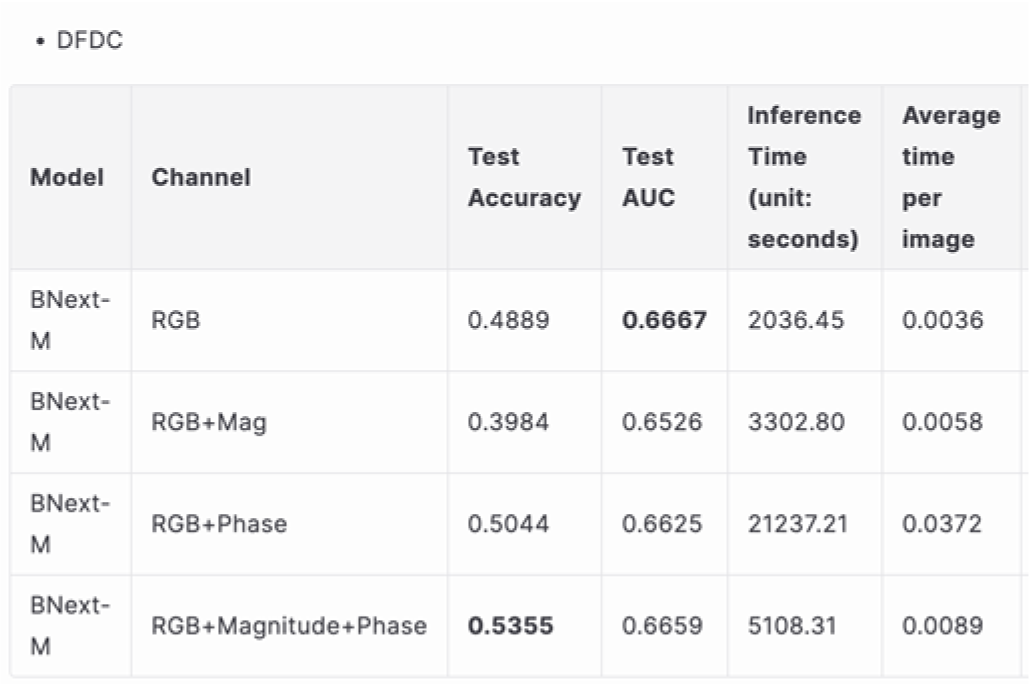
Magnitude & Phase Spectrum



- Celeb-DF

Model	Channel	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	0.7850	0.8022	260.82	0.0033
BNext-M	RGB+Magnitude	0.7684	0.7877	475.27	0.0060
BNext-M	RGB+Phase	0.7782	0.7949	742.92	0.0094
BNext-M	RGB+Magnitude+Phase	**0.7891**	**0.8285**	810.31	0.0102

Magnitude & Phase Spectrum

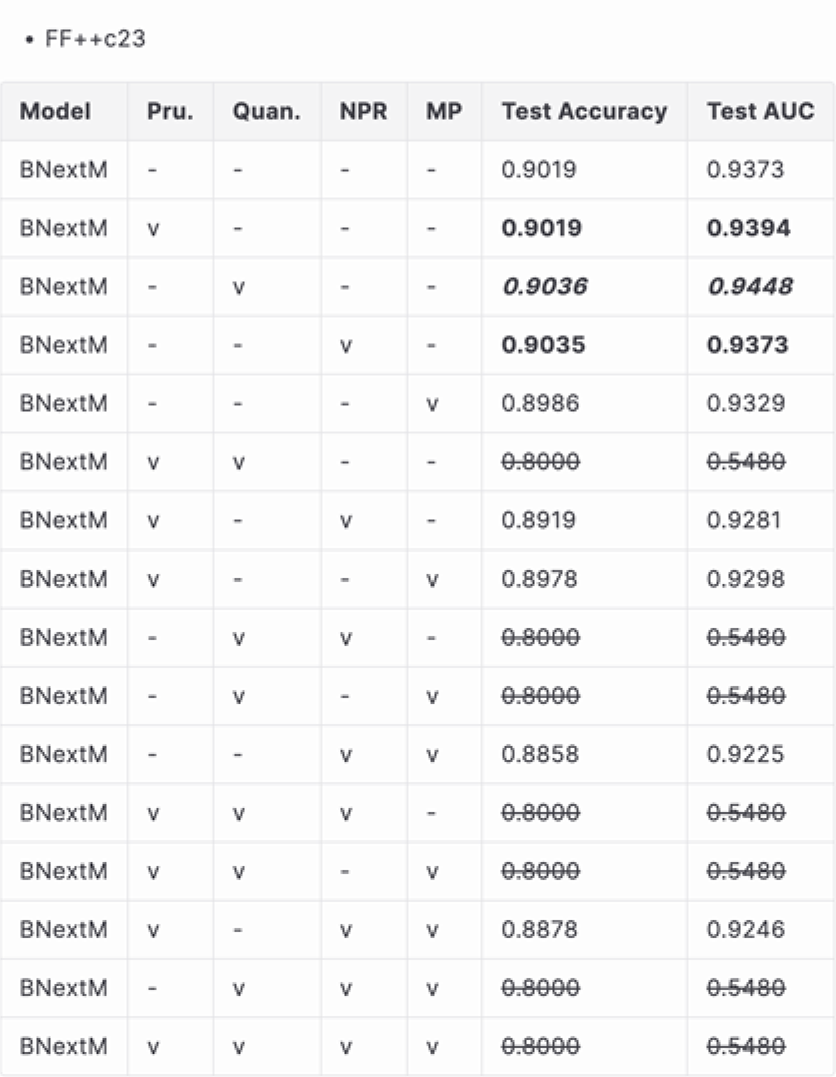


- DFDC

Model	Channel	Test Accuracy	Test AUC	Inference Time (unit: seconds)	Average time per image
BNext-M	RGB	0.4889	**0.6667**	2036.45	0.0036
BNext-M	RGB+Mag	0.3984	0.6526	3302.80	0.0058
BNext-M	RGB+Phase	0.5044	0.6625	21237.21	0.0372
BNext-M	RGB+Magnitude+Phase	**0.5355**	0.6659	5108.31	0.0089

-> 選擇兩者皆加入

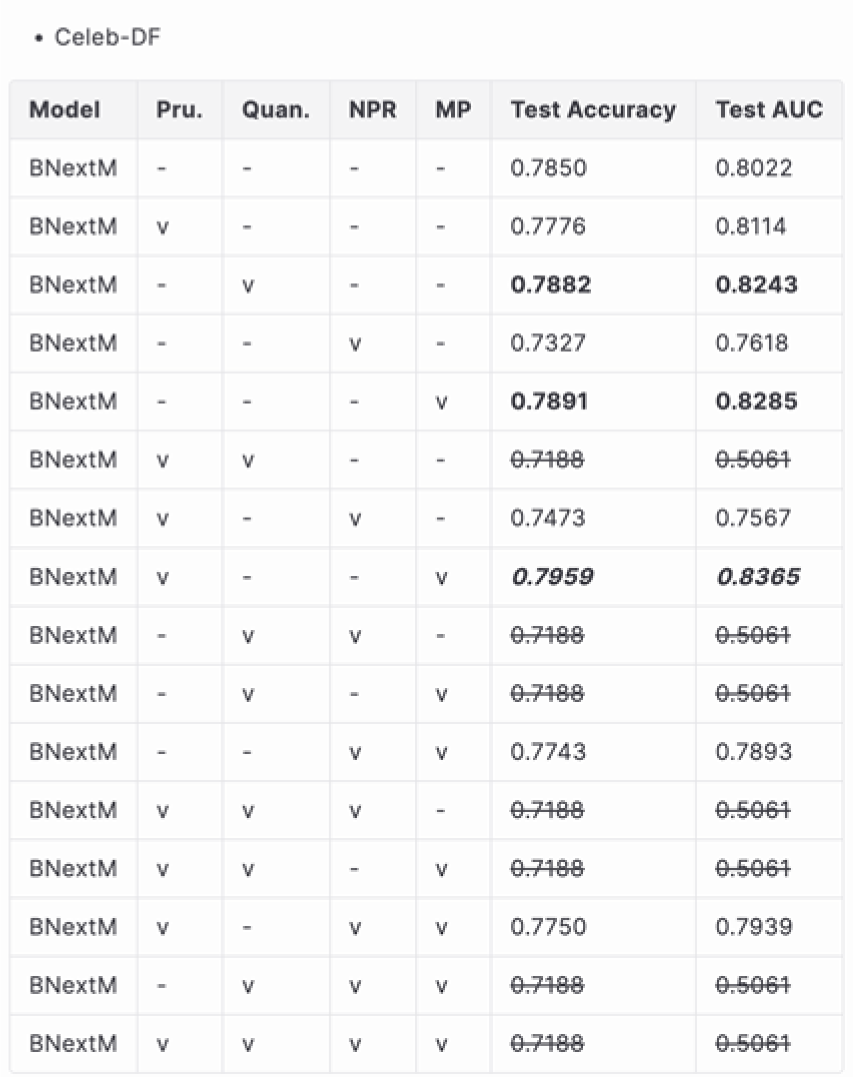
Ablation Study



- FF++c23

Model	Pru.	Quan.	NPR	MP	Test Accuracy	Test AUC
BNextM	-	-	-	-	0.9019	0.9373
BNextM	v	-	-	-	**0.9019**	**0.9394**
BNextM	-	v	-	-	***0.9036***	***0.9448***
BNextM	-	-	v	-	**0.9035**	**0.9373**
BNextM	-	-	-	v	0.8986	0.9329
BNextM	v	v	-	-	~~0.8000~~	~~0.5480~~
BNextM	v	-	v	-	0.8919	0.9281
BNextM	v	-	-	v	0.8978	0.9298
BNextM	-	v	v	-	~~0.8000~~	~~0.5480~~
BNextM	-	v	-	v	~~0.8000~~	~~0.5480~~
BNextM	-	-	v	v	0.8858	0.9225
BNextM	v	v	v	-	~~0.8000~~	~~0.5480~~
BNextM	v	v	-	v	~~0.8000~~	~~0.5480~~
BNextM	v	-	v	v	0.8878	0.9246
BNextM	-	v	v	v	~~0.8000~~	~~0.5480~~
BNextM	v	v	v	v	~~0.8000~~	~~0.5480~~

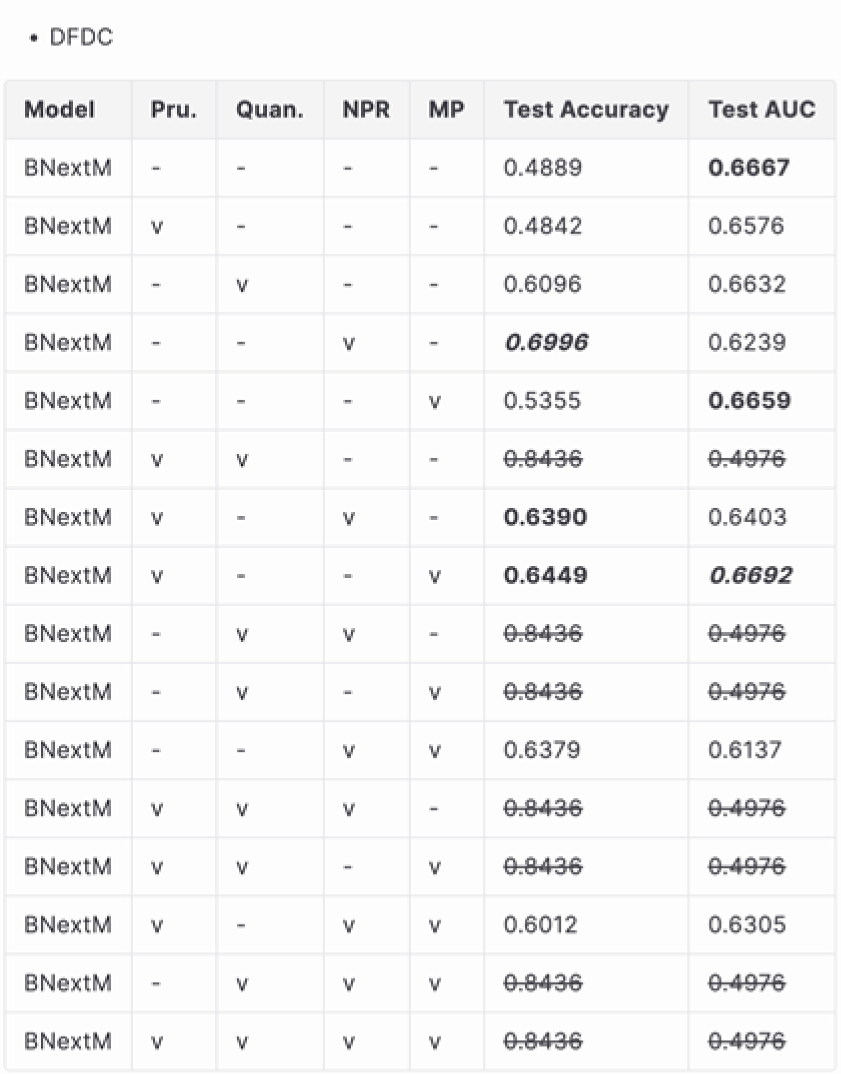
Ablation Study



- Celeb-DF

Model	Pru.	Quan.	NPR	MP	Test Accuracy	Test AUC
BNextM	-	-	-	-	0.7850	0.8022
BNextM	v	-	-	-	0.7776	0.8114
BNextM	-	v	-	-	**0.7882**	**0.8243**
BNextM	-	-	v	-	0.7327	0.7618
BNextM	-	-	-	v	**0.7891**	**0.8285**
BNextM	v	v	-	-	~~0.7188~~	~~0.5061~~
BNextM	v	-	v	-	0.7473	0.7567
BNextM	v	-	-	v	***0.7959***	***0.8365***
BNextM	-	v	v	-	~~0.7188~~	~~0.5061~~
BNextM	-	v	-	v	~~0.7188~~	~~0.5061~~
BNextM	-	-	v	v	0.7743	0.7893
BNextM	v	v	v	-	~~0.7188~~	~~0.5061~~
BNextM	v	v	-	v	~~0.7188~~	~~0.5061~~
BNextM	v	-	v	v	0.7750	0.7939
BNextM	-	v	v	v	~~0.7188~~	~~0.5061~~
BNextM	v	v	v	v	~~0.7188~~	~~0.5061~~

Ablation Study



- DFDC

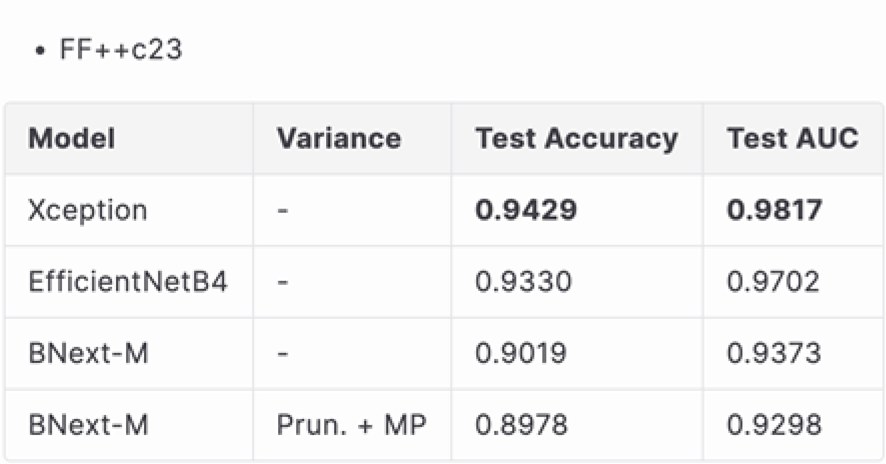
Model	Pru.	Quan.	NPR	MP	Test Accuracy	Test AUC
BNextM	-	-	-	-	0.4889	**0.6667**
BNextM	v	-	-	-	0.4842	0.6576
BNextM	-	v	-	-	0.6096	0.6632
BNextM	-	-	v	-	***0.6996***	0.6239
BNextM	-	-	-	v	0.5355	**0.6659**
BNextM	v	v	-	-	~~0.8436~~	~~0.4976~~
BNextM	v	-	v	-	**0.6390**	0.6403
BNextM	v	-	-	v	**0.6449**	***0.6692***
BNextM	-	v	v	-	~~0.8436~~	~~0.4976~~
BNextM	-	v	-	v	~~0.8436~~	~~0.4976~~
BNextM	-	-	v	v	0.6379	0.6137
BNextM	v	v	v	-	~~0.8436~~	~~0.4976~~
BNextM	v	v	-	v	~~0.8436~~	~~0.4976~~
BNextM	v	-	v	v	0.6012	0.6305
BNextM	-	v	v	v	~~0.8436~~	~~0.4976~~
BNextM	v	v	v	v	~~0.8436~~	~~0.4976~~

Conclusion

Conclusion

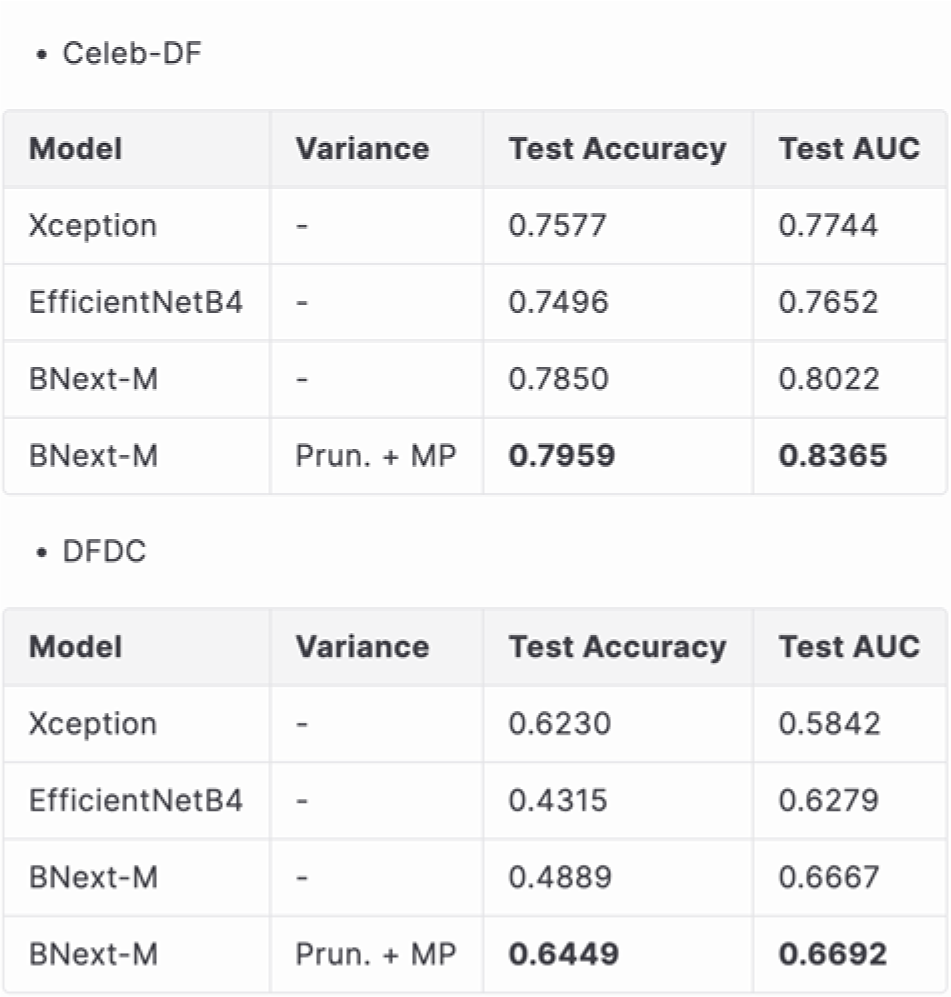
- ➢ 將模型壓縮至適合手機部署的規格。
- ➢ 保持準確度以確保良好的測試效果。
- ➢ 設法提升模型在跨資料集上的表現，以實現更強的泛化能力。

Conclusion



- FF++c23

Model	Variance	Test Accuracy	Test AUC
Xception	-	**0.9429**	**0.9817**
EfficientNetB4	-	0.9330	0.9702
BNext-M	-	0.9019	0.9373
BNext-M	Prun. + MP	0.8978	0.9298



- Celeb-DF

Model	Variance	Test Accuracy	Test AUC
Xception	-	0.7577	0.7744
EfficientNetB4	-	0.7496	0.7652
BNext-M	-	0.7850	0.8022
BNext-M	Prun. + MP	**0.7959**	**0.8365**

- DFDC

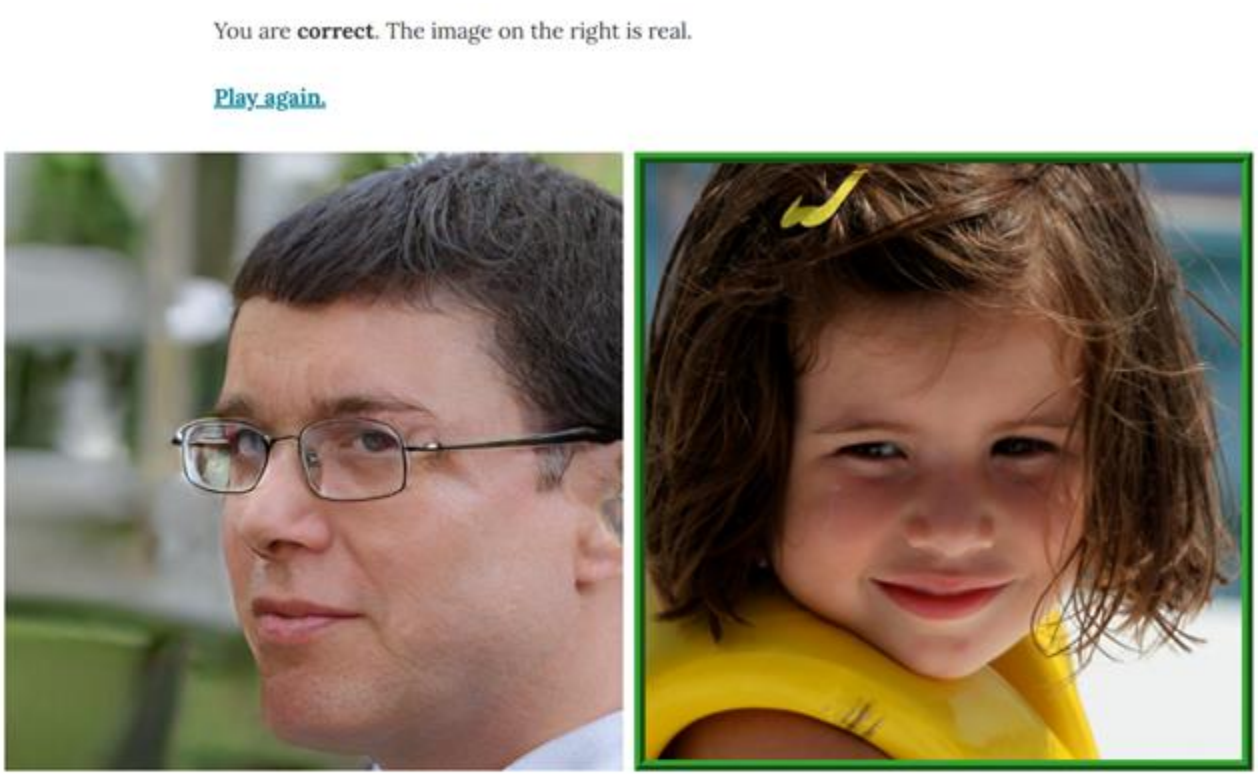
Model	Variance	Test Accuracy	Test AUC
Xception	-	0.6230	0.5842
EfficientNetB4	-	0.4315	0.6279
BNext-M	-	0.4889	0.6667
BNext-M	Prun. + MP	**0.6449**	**0.6692**

Challenges & Future work

- **實驗驗證**：需先執行一次實驗，確認結果是否與論文一致。
- **模型穩定性**：BNext 為 BNN 模型，進行模型壓縮或加入方法可能導致不穩定或無法收斂。
- **超參數篩選**：建議針對每種方法的超參數進行全面實驗，以尋求最佳解。
- **Pruning**：應先匯入權重再進行剪枝。應採用**結構化剪枝**方法來減小模型。
- **轉換模型**：模型中的一些 Module 是自訂義，因此TorchScript並不支援。可以尋找 TorchScript 的相同功能 Module 替代或是在 Android Studio 使用 Java 寫過。

Demo

Demo

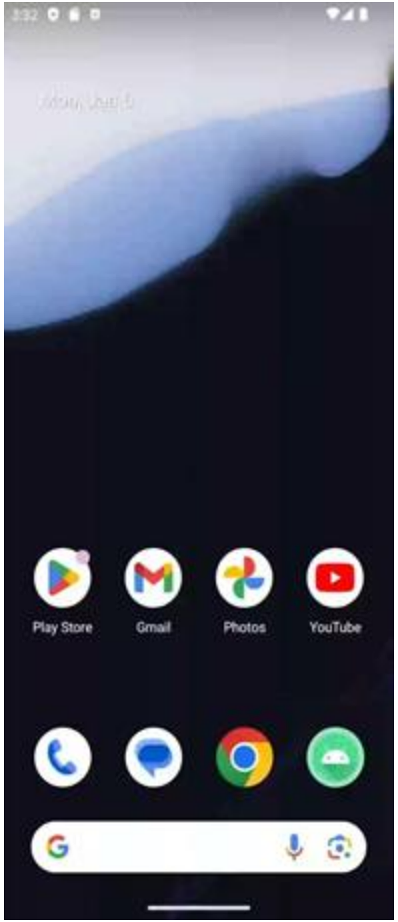


You are **correct**. The image on the right is real.

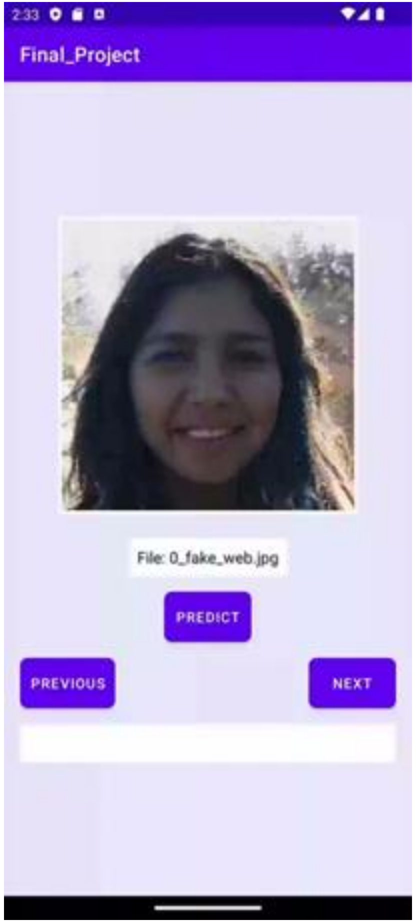
Play again.

https://www.whichfaceisreal.com/results.php?r=1&p=1&i1=image-2019-02-18_215504.jpeg&i2=21923.jpeg

Demo



FF++資料集測試

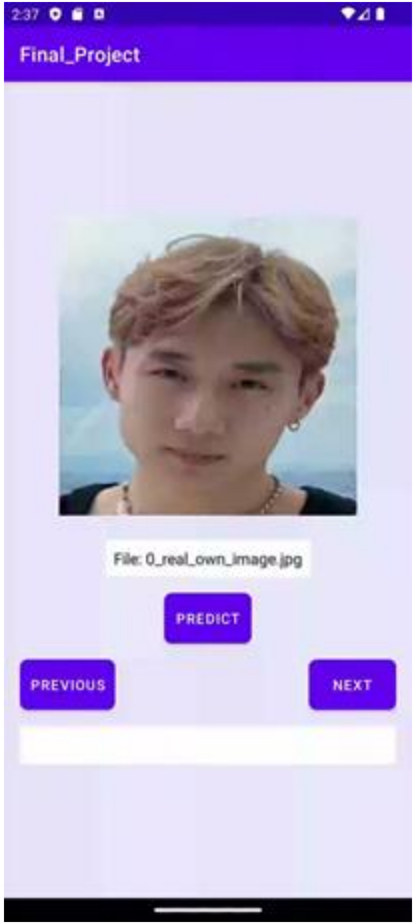


Final_Project

File: 0_fake_web.jpg

whichfaceisreal測試

Demo

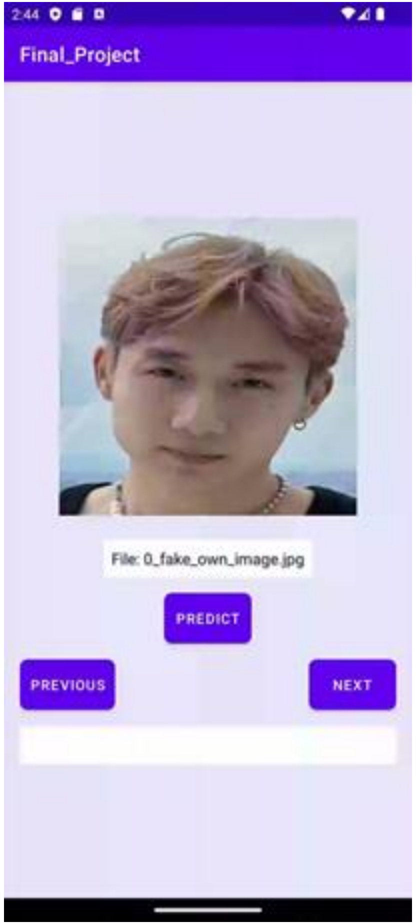


2:37

Final_Project

File: 0_real_own_image.jpg

自己的照片測試(Generative AI)



2:44

Final_Project

File: 0_fake_own_image.jpg

自己的照片測試(濾鏡)