Face Morphing Attacks Detection & Fingerprinting

Na Zhang

Morphing Defense – Morphing Attack Detection (MAD)
– Morphing Attack Fingerprinting (MAF)

- Aims at detecting morphing attacks
- Since a malicious person can successfully pass the system's check as the morphed face resembles the face enrolled in the FRS
- the detection of face morphing attack is becoming an urgent problem

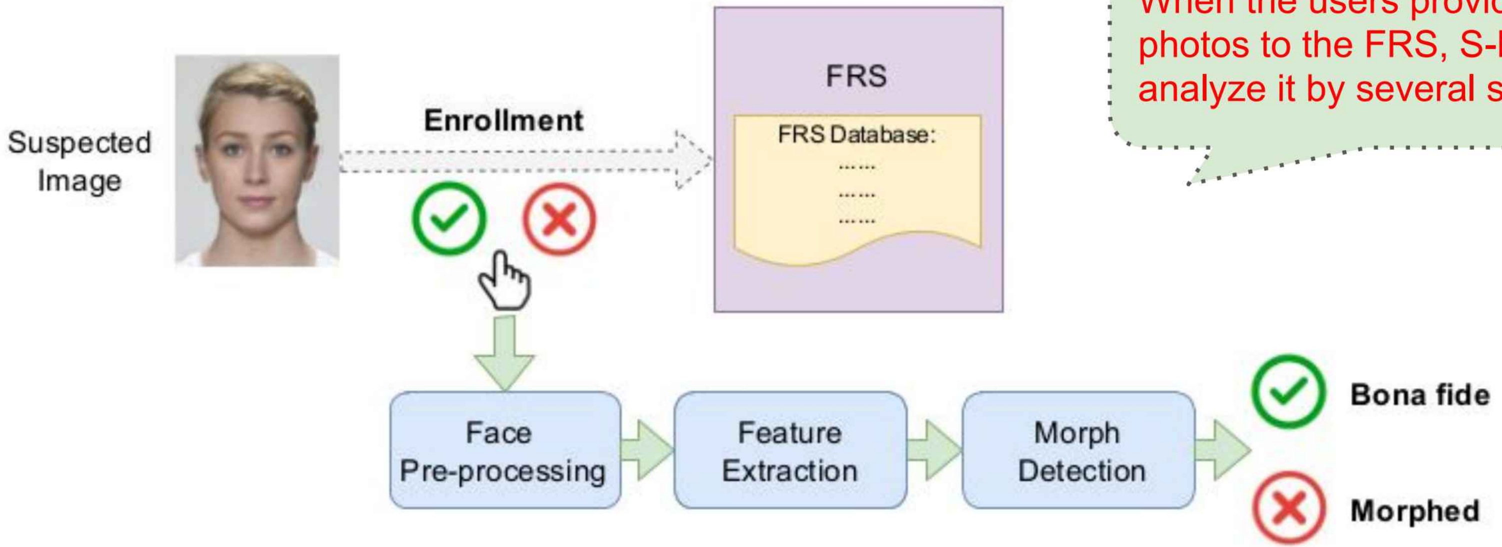
Existing Detection Methods

- A number of morphing attack detection (MAD) approaches have been proposed
- Can be coarsely categorized in two types with respect to the considered morphing detection scenario
  - Single image based MAD (S-MAD)
    - i.e. no-reference
  - Differential image based MAD (D-MAD)
    - reference -based

S-MAD

- Focuses on a single potentially morphed image
- The detection action occurs during enrollment
    - e.g. the passport application process

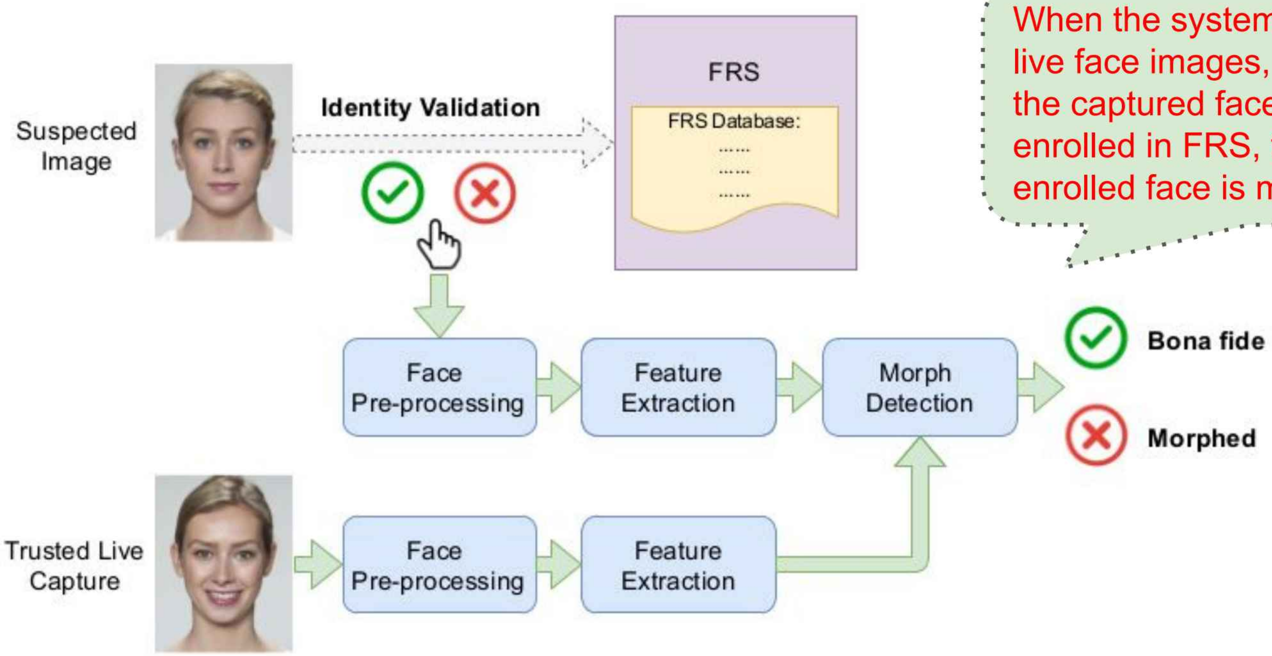
When the users provide their photos to the FRS, S-MAD will analyze it by several steps.



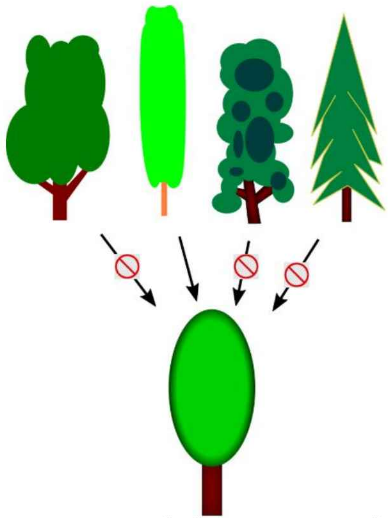
D-MAD

- With a corresponding face image captured in a trusted environment
- The detection action occurs at the time of identity validation
  - e.g. passing through an Automated Border Control (ABC) gates at borders

When the system captures a trusted live face images, D-MAD will analyze the captured face and the face enrolled in FRS, to check if the enrolled face is morphed



Problems of Existing MAD

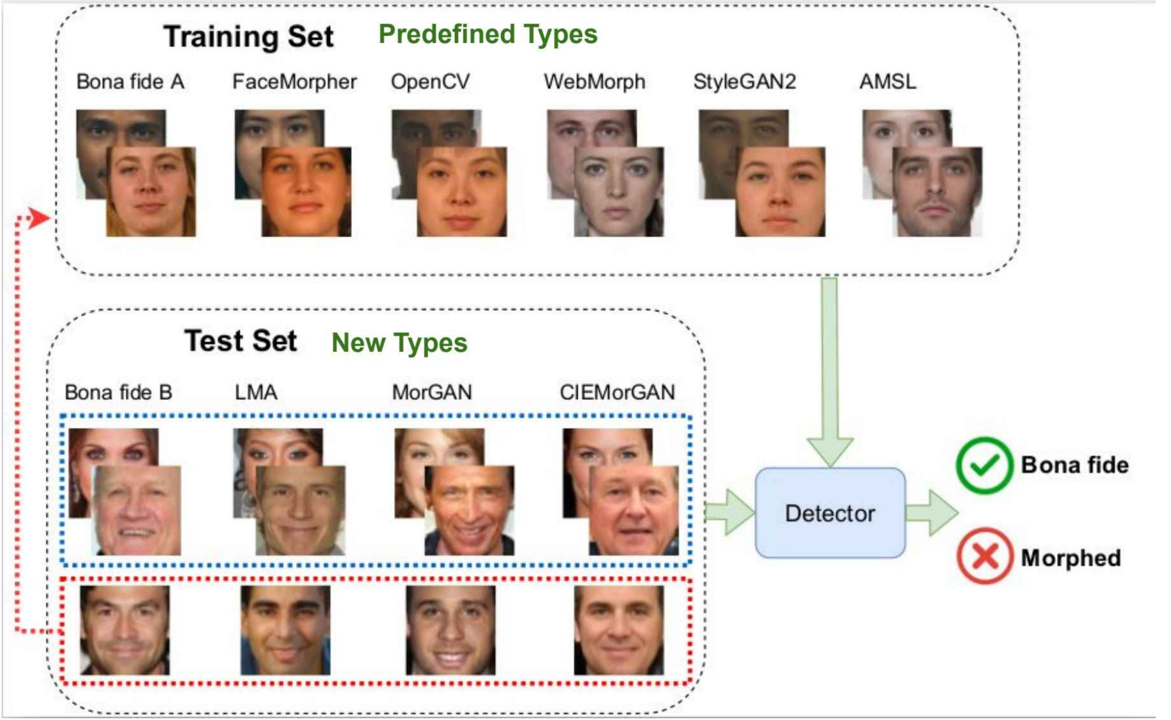


- Low generalization ability
  - Small training dataset
  - Single modality
- Degrades rapidly when facing newly evolved attacks
- Possible solution: fine-tuning existing MAD models
- However, the cost of collecting labeled data for every new morphing attack is often formidable
- Moreover
  - MAD (binary detection) alone is not sufficient to meet the demand of increased security risk
  - need a more aggressive countermeasure to formulate morphing attack fingerprinting (MAF) problem
    - multiclass classification of morphing attack models

Single image based detection

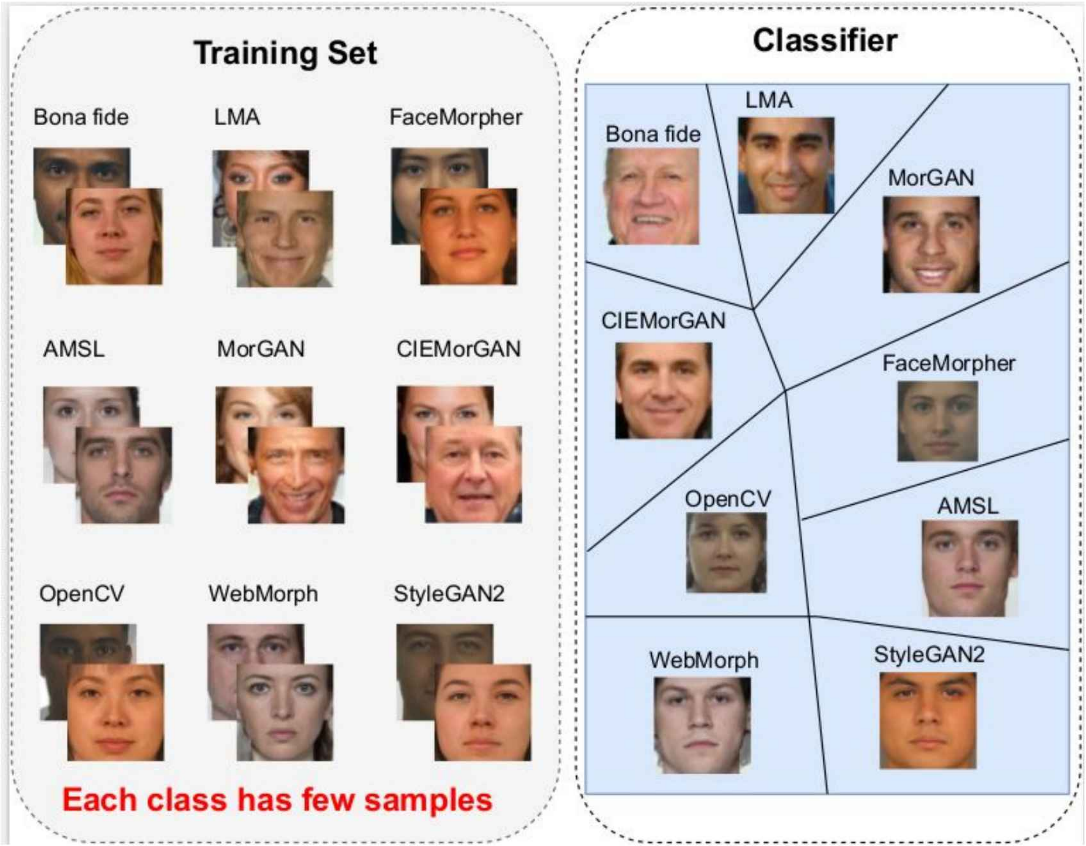
- Formulate MAD/MAF as **few-shot learning (FSL)** problems
- **FS-MAD**
  - **train** the detector using data from both **predefined** models and **new** attack models (only a few samples are required)
  - to predict **unknown test** samples

Few-shot MAD (FS-MAD)

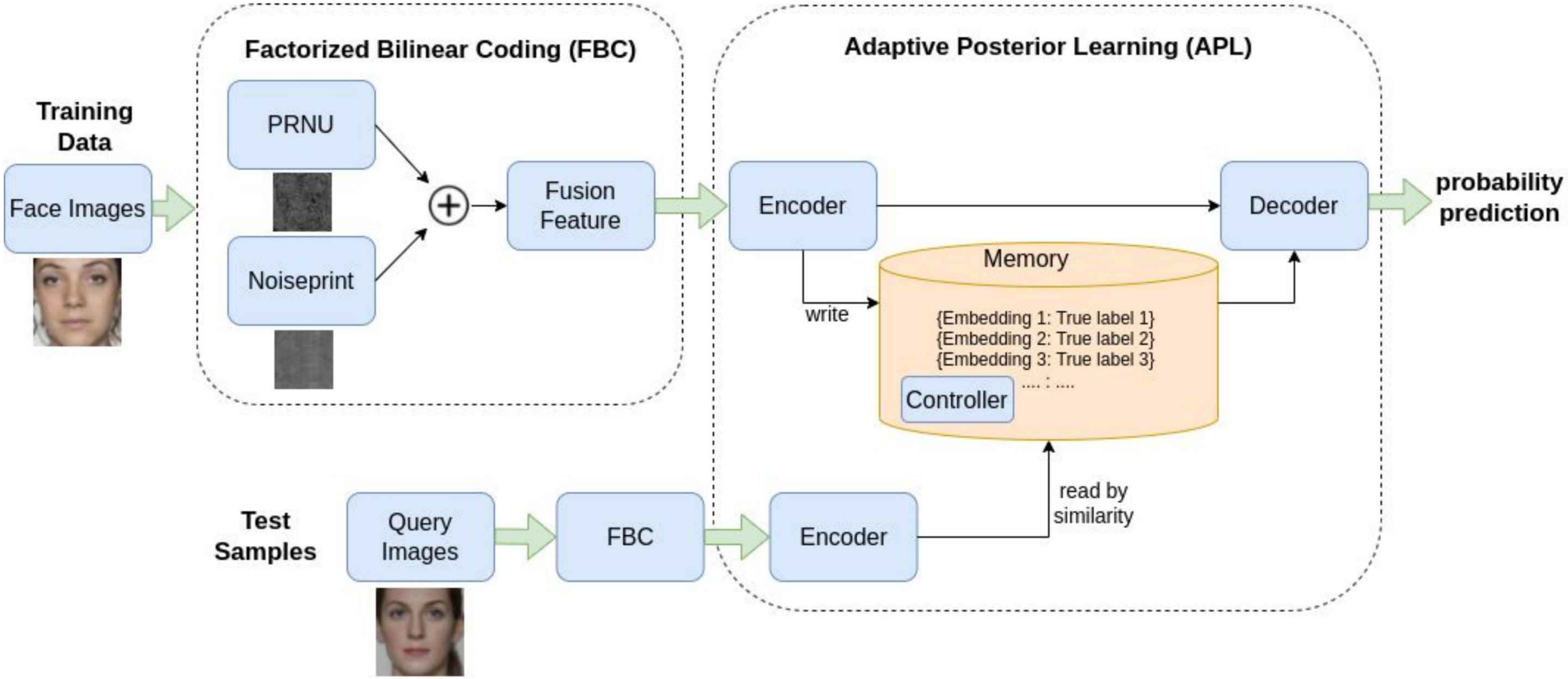


- **FS-MAF**
  - finer-granularity classification
  - multi-class problem
  - classify different types of attacks based on a few samples
  - closely related to
    - camera identification
    - camera model fingerprinting
    - etc.

Few-shot MAF (FS-MAF)

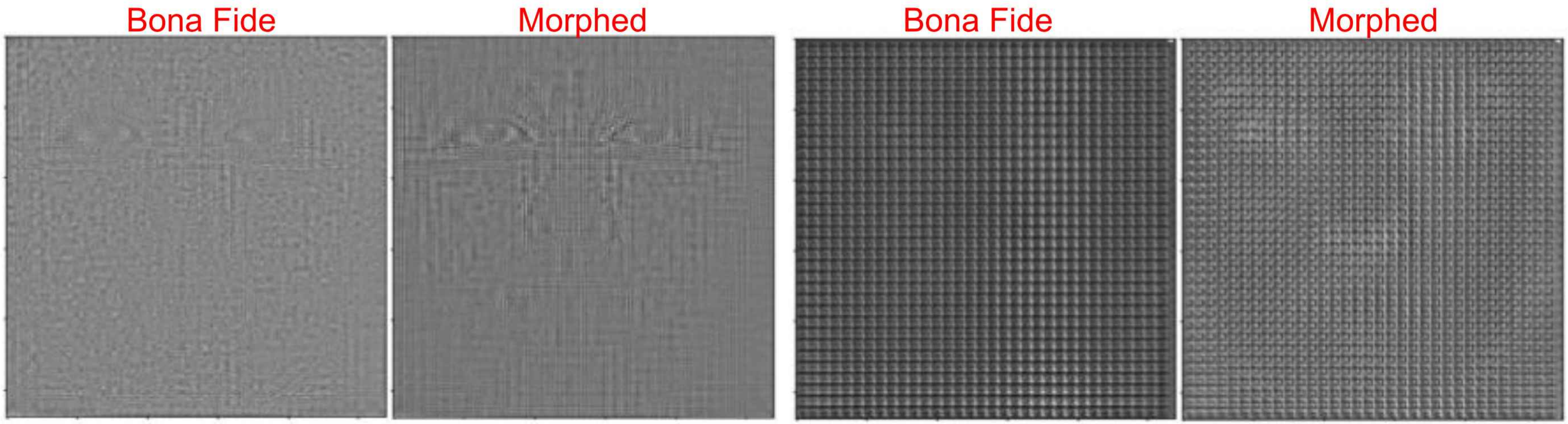


Fusion-based FSL Model



Feature Extraction

- Noise occurs during image manipulation
- Consider two types of sensor noise patterns
  - Photo Response Non-Uniformity (PRNU) [5] — Model-based
  - Noiseprint [6] — Data-driven



(a) PRNU (b) Noiseprint

[5] Jessica Fridrich. Digital image forensics. IEEE Signal Processing Magazine, 26(2):26–37, 2009.
[6] Davide Cozzolino and Luisa Verdoliva. Noiseprint: A cnn-based camera model fingerprint. arXiv preprint arXiv:1808.08396, 2018.

Feature Fusion

- Factorized Bilinear Coding (FBC) [7]
- A sparse coding formulation
  - generate a compact /discriminative representation
  - by learning a dictionary [capture structure of the whole data space]

➔ Let $\boldsymbol{x}_i$: PRNU, $\mathbf{y}_j$: Noiseprint

➔ FBC encodes the two input feature (x_i , y_j) into FBC code c_v (final fusion feature) by solving the following optimization problem:

$$\min_{\boldsymbol{c}_v} \underbrace{\left\|\boldsymbol{x}_i\boldsymbol{y}_j^\top - \sum_{l=1}^{k} c_v^l \boldsymbol{U}_l \boldsymbol{V}_l^\top\right\|^2}_{\text{Reconstruction Error}} + \lambda \underbrace{\|\boldsymbol{c}_v\|_1}_{\text{Sparsity}}$$

λ : a trade-off parameter

$B = \{b_1, b_2, \ldots, b_l, \ldots b_k\}$, Dictionary

k : # atoms of B

$b_l = U_l V_l^T$, U_l, V_l : low-rank matrices

c_v^l : l-th element of c_v

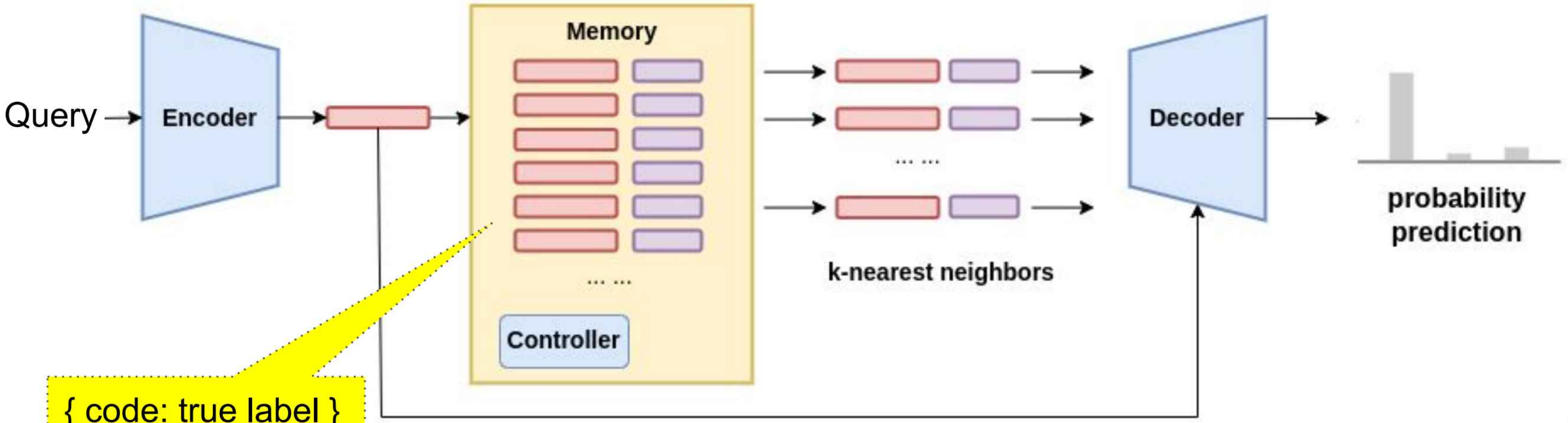
➔ In essence, the bilinear feature $x_i y_j^T$ is reconstructed by $\sum_{l=1}^{k} c_v^l \boldsymbol{U}_l \boldsymbol{V}_l^\top$

[7] Zhi Gao, Yuwei Wu, Xiaoxun Zhang, Jindou Dai, Yunde Jia, and Mehrtash Harandi. Revisiting bilinear pooling: A coding perspective. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 3954–3961, 2020.

Few-shot Learning (FSL)

- Inspired by adaptive posterior learning (APL) [8]
- The key idea
  - to predict the probability by remembering the most surprising observations it has encountered [stored in memory]

The higher the probability the model assigns to true class correctly, the less surprised it will be.

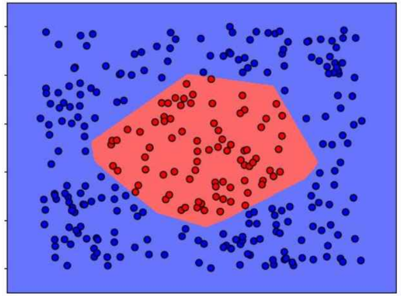


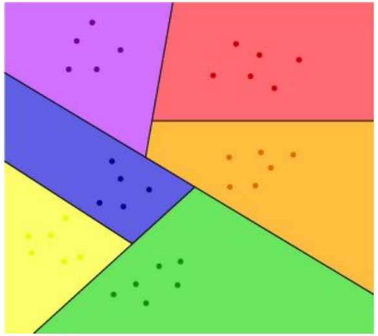
{ code: true label }

[8] *T. Ramalho and M. Garnelo, "Adaptive posterior learning: few-shot learning with a surprise-based memory module," arXiv preprint arXiv:1902.02527, 2019.*

Binary/Multiclass Classification

- APL module easily leads itself to the generalization
  - from binary (FS-MAD) to multiclass (FS-MAF) classification
  - by resetting the hyperparameters, like
    - the number of classes
    - data path for each class, etc.





Database

- Combined 5 datasets for evaluation
  - 4 public
  - 1 self-collected
- A total of over 20K images
  - Bona fide: 6,869
  - Morphed: 15,764
- 8 morphing algorithms
  - 5 landmark based
    - OpenCV
    - FaceMorpher
    - LMA
    - WebMorph
    - AMSL
  - 3 GAN based
    - MorGAN
    - CIEMorGAN
    - StyleGAN2

Table 1. The newly constructed face morphing database consists of five image sources and 3-6 different morphing methods.

Database	Subset	#Number	Resolution
FERET-Morphs	bona fide [12]	576	512x768
	FaceMorpher [13]	529	512x768
	OpenCV [13]	529	512x768
	StyleGAN2 [13]	529	1024x1024
FRGC-Morphs	bona fide [11]	964	1704x2272
	FaceMorpher [13]	964	512x768
	OpenCV [13]	964	512x768
	StyleGAN2 [13]	964	1024x1024
FRLL-Morphs	bona fide [14]	102+1932	413x531
	AMSL [10]	2175	413x531
	FaceMorpher [13]	1222	431x513
	OpenCV [13]	1221	431x513
	LMA	768	413x531
	WebMorph [13]	1221	413x531
	StyleGAN2 [13]	1222	1024x1024
CelebA-Morphs*	bona fide [7]	2989	128x128
	MorGAN [3]	1000	64x64
	CIEMorGAN [2]	1000	128x128
	LMA [3]	1000	128x128
Doppelgänger	bona fide	306	1024x1024
	FaceMorpher	150	1024x1024
	OpenCV	153	1024x1024
	StyleGAN2	153	1024x1024

FS-MAD

- Binary detection
- Training data: predefined types + 1 (for 1-shot) or 5 (5-shot) samples per new type
- Test data: new types

Performance (%) comparison of few-shot MAD

	1-shot			5-shot		
Method	Accu.	D-EER	ACER	Accu.	D-EER	ACER
Xception [31]	66.5	32.5	33.5	73.25	27	26.75
MobileNetV2 [188]	67	36.5	33	71.25	29	28.75
NasNetMobile [262]	59	40.5	41	66.25	35	33.75
DenseNet121 [87]	68.25	31.5	31.75	73.5	24.5	26.5
FaceNet [198]	66.75	30	33.25	66.75	30.5	33.25
ArcFace [49]	58	41	42	62.25	37.5	37.75
Meta-Baseline [29]	60.45	-	-	71.38	-	-
COSOC [141]	66.89	-	-	74.54	-	-
FBC-APL	**99.25**	**1.5**	**0.75**	**99.75**	**0.5**	**0.25**

FS-MAF

- Multiclass
- Each morphing type and the bona fide type are treated as different classes
- Training data
  - 1 and 5 images per class for 1-shot and 5-shot learning, respectively.
- Test data
  - non-overlapping data with training set

Accuracy(%) of 1-shot MAF classification on single and hybrid datasets

Method	FERET-Morphs 4-class	FRGC-Morphs 4-class	FRLL-Morphs 7-class	CelebA-Morphs 4-class	Doppelgänger 4-class	Hybrid 9-class
Xception [31]	29.47	25.26	17.68	16.67	21.05	15.11
MobileNetV2 [188]	31.58	33.68	31.3	55.19	25.26	17.33
NasNetMobile [262]	32.63	27.37	22.61	19.26	23.16	12.88
DenseNet121 [87]	46.32	26.32	22.03	47.04	23.16	19.33
FaceNet [198]	26.79	27.98	16.48	33.67	31.15	15.67
ArcFace [49]	29.33	39.64	26.12	28.33	18.03	15.22
Meta-Baseline [29]	51.05	51.44	34.77	61.43	33.43	53.46
COSOC [141]	54.58	64.37	35.22	63.19	34.3	59.55
FBC	96.93	98.83	94.06	99.5	56.67	96.11
FBC-all	98.11	99.48	98.42	100	84.17	96.78
FBC-APL	**98.82**	**99.61**	**98.24**	**99.67**	**91.67**	**98.11**

Accuracy(%) of 5-shot MAF classification on single and hybrid datasets

Method	FERET-Morphs 4-class	FRGC-Morphs 4-class	FRLL-Morphs 7-class	CelebA-Morphs 4-class	Doppelgänger 4-class	Hybrid 9-class
Xception [31]	46.32	43.16	31.01	73.7	28.42	43.67
MobileNetV2 [188]	55.79	53.68	40	89.26	26.32	54.56
NasNetMobile [262]	48.42	40	24.35	67.41	27.37	37.33
DenseNet121 [87]	54.74	55.79	36.23	89.26	25.26	53.33
FaceNet [198]	23.16	35.79	15.94	40	30.53	18.11
ArcFace [49]	44.34	50.91	33.81	39.67	20.49	29.11
Meta-Baseline [29]	60.6	64.72	50.74	81.42	36.8	61.98
COSOC [141]	65.98	75.04	54.9	89.6	41.81	72.62
FBC	97.64	99.09	96.94	99.5	65.83	96.22
FBC-all	98.11	99.48	98.42	100	84.17	96.78
FBC-APL	**98.82**	**99.61**	**98.24**	**99.67**	**96.67**	**98.22**