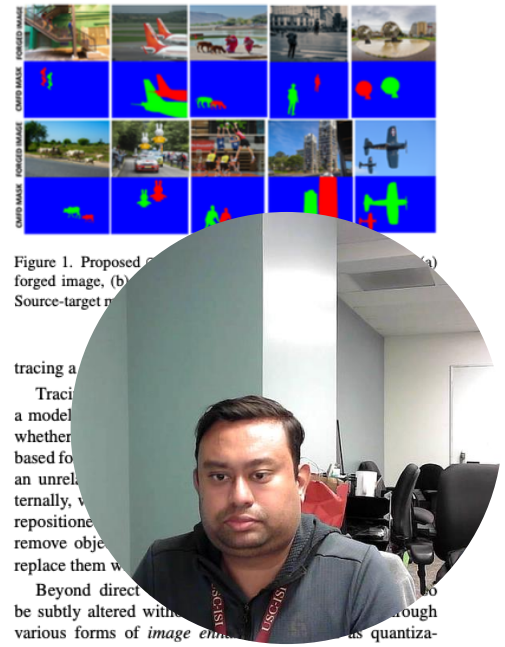
CAN IMAGE SPLICING AND COPY-MOVE FORGERY BE DETECTED BY THE SAME MODEL? FORENSIM: AN ATTENTION-BASED STATE-SPACE APPROACH

SOUMYAROOP NANDI[1,2], **PREM NATARAJAN**[1,2]
[1]USC INFORMATION SCIENCES INSTITUTE, MARINA DEL REY, CA, USA
[2]USC THOMAS LORD CS DEPARTMENT, LOS ANGELES, CA, USA
{SOUMYARN, PREMKUMN}@USC.EDU



Can Image Splicing and Copy-Move Forgery Be Detected by the Same Model? Forensim: An Attention-Based State-Space Approach

Soumyaroop Nandi[1,2] Prem Natarajan[1,2]
[1]USC Information Sciences Institute, Marina del Rey, CA, USA
[2]USC Thomas Lord Department of Computer Science, Los Angeles, CA, USA
{soumyarn, premkumn}@usc.edu

Abstract

We introduce Forensim, an attention-based state-space framework for image forgery detection that jointly localizes both manipulated (target) and source regions. Unlike traditional approaches that rely solely on artifact cues to detect spliced or forged areas, Forensim is designed to capture duplication patterns crucial for understanding context. In scenarios like protest imagery, detecting only the forged region—e.g., a duplicated act of violence inserted into a peaceful crowd—can mislead interpretation, highlighting the need for joint source-target localization. Forensim outputs three-class masks (pristine, source, target) and supports detection of both splicing and copy-move forgeries within a unified architecture. We propose a visual state-space model that leverages normalized attention maps to identify internal similarities, paired with a region-based block-attention module to distinguish manipulated regions. This design enables end-to-end training and precise localization. Forensim achieves state-of-the-art performance on standard benchmarks. We also release CMFD_Anything, a new dataset addressing limitations of existing copy-move forgery datasets. Project page and code.

1. Introduction

The proliferation of generative models has broadened image manipulation far beyond traditional editing tools. The challenge for today's researchers is to detect manipu-

ROADMAP

- **What is Semantic Forgery Detection?**
  - Previous research in Image Forgery Detection
  - What are the Challenges in Image Forgery Detection?
- Solution Proposed – Forensim
  - Forensim Contributions
  - Why Three Class-based Training?
  - State Space Model Overview
  - Forensim Modules
  - CMFD-Anything Dataset
  - Evaluation and Results

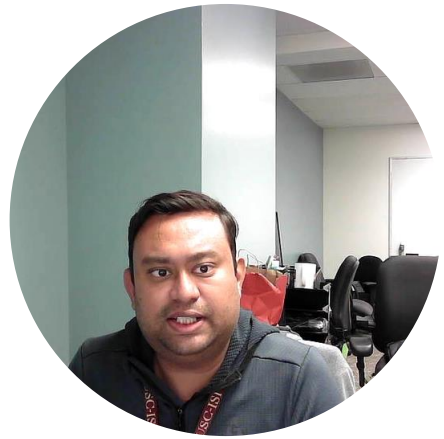
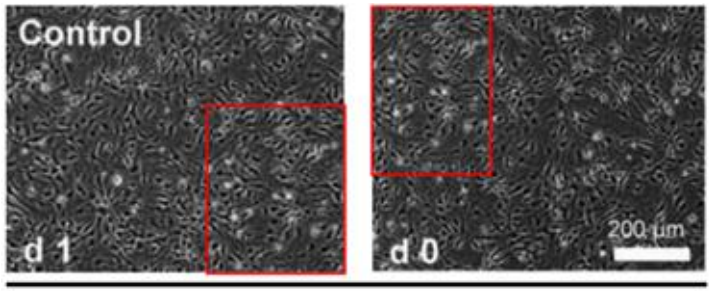


IMAGE MANIPULATION SOURCES

Image Editing
et al. TrainFors, ICCVW 2023

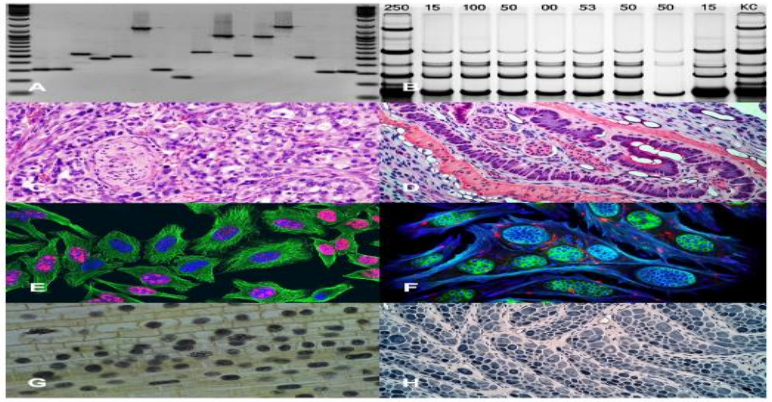


Biomedical Image Tampering
et al. BioFors, ICCV 2021

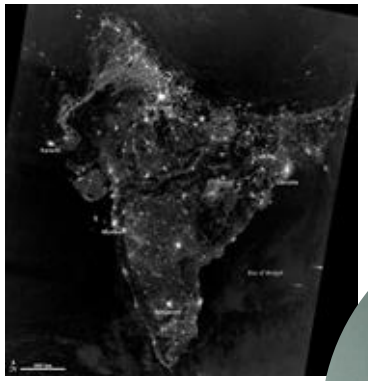
Image Translation
et al. PNP_Diffusion, CVPR 2023

A bird with **black eye rings** and a **black bill**, with a **red crown** and a **red belly**.

Text-Guided Image Tampering
et al. Manigan, CVPR 2020



Laboratory Images Vs DALL-E 3 Generated
Kim et al. AI and Ethics 2024



Satellite Im
et al. Horvá

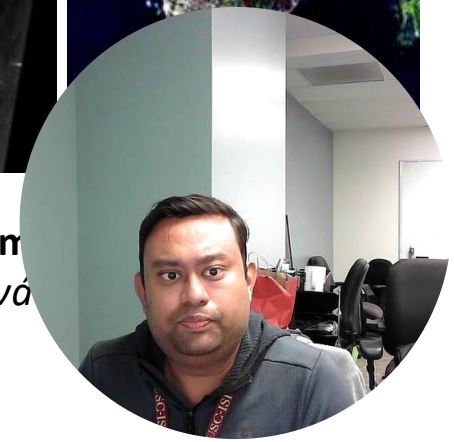


IMAGE MANIPULATION SOURCES

Digital Image Manipulation

GENERATIVE MODEL BASED TAMPERING

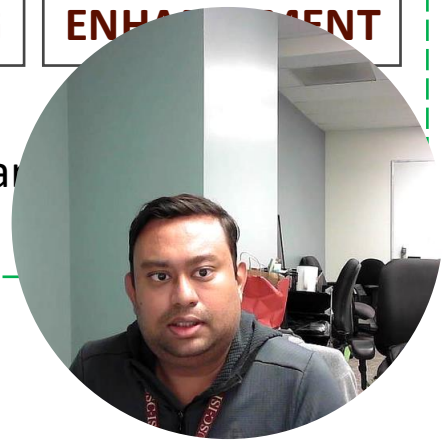
- **LLM/ TEXT-GUIDED**
- **STYLE TRANSFER**
- **DEEP FAKES**

Entire Image Synthetically Generated or
Only Local Regions Generated and Blended

EDITING TOOLS BASED TAMPERING

- **SPLICING**
- **COPYMOVE**
- **INPAINTING**
- **ENH[illegible]ENT**

Only Local Regions of Image Ta[illegible]

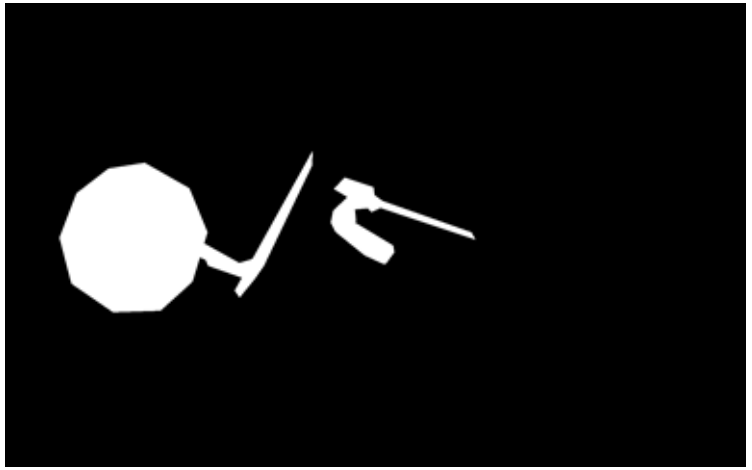


WHAT IS SEMANTIC FORGERY DETECTION?

IMAGE MANIPULATION DETECTION AND LOCALIZATION (IMDL)

PRISTINE

FORGED



GROUNDTRUTH M

LOCAL IMAGE FORGERY TYPES - SPLICING

PRISTINE **FORGED** **GROUNDTRUTH**

Fig 1: Image Splicing Forgery Localization Example

LOCAL IMAGE FORGERY TYPES - COPYMOVE

PRISTINE **FORGED** **GROUNDTRUTH**

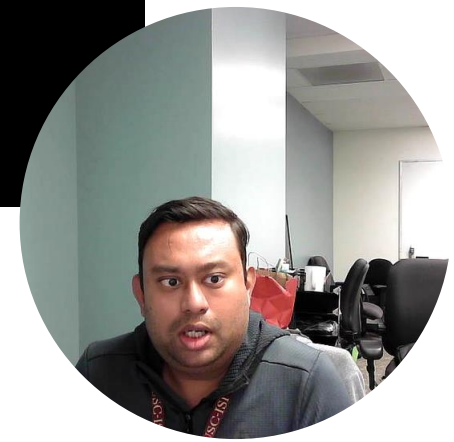


Fig 2: Copymove Image Forgery Localization Example

LOCAL IMAGE FORGERY TYPES - REMOVAL

PRISTINE

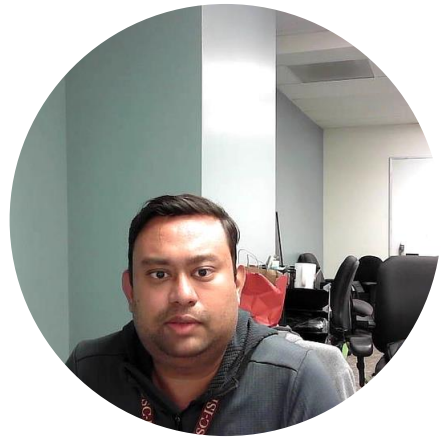
FORGED

GROUNDTRUTH

Fig 3: Image Removal Forgery Localization Example

FORGERY LAUNDERING TECHNIQUES

Fig 4: Image Enhancement Laundering



ROADMAP

- What is Semantic Forgery Detection?
  - **Previous research in Image Forgery Detection**
  - What are the Challenges in Image Forgery Detection?
- Solution Proposed – Forensim
  - Forensim Contributions
  - Why Three Class-based Training?
  - State Space Model Overview
  - Forensim Modules
  - CMFD-Anything Dataset
  - Evaluation and Results

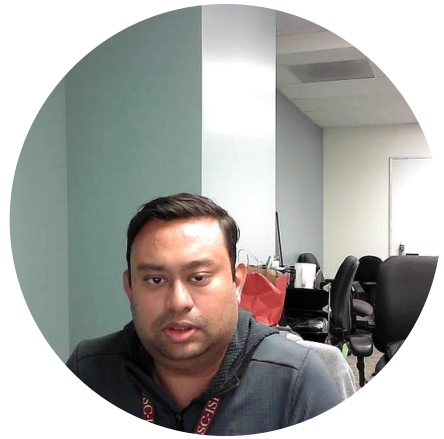


IMAGE TAMPERING ARTIFACTS

- **Noise frequency features** are captured using Bayar Convolution, SRM filters, DCT filters, Sobel filters
- **RGB artifacts** capture edge inconsistency, color consistency, visual similarity, and EXIF consistency
- **Camera model artifacts** capture photo-response non-uniformity noise (PRNU) in the form of Noiseprint

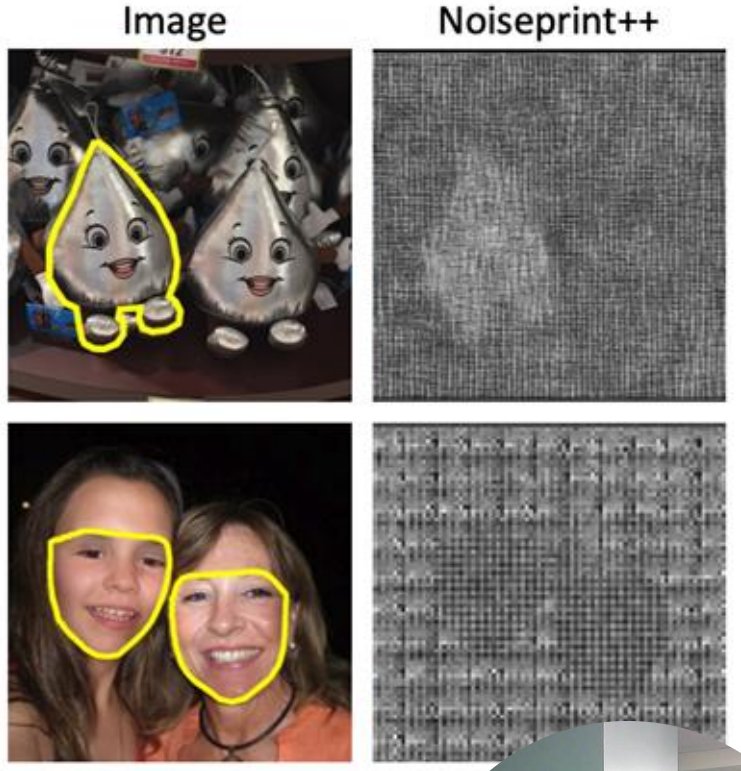
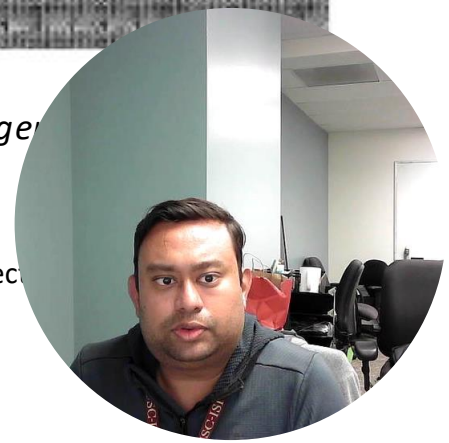


*Fig 5: Noiseprint ge
TruFor



*Fabrizio Guillaro, Davide Cozzolino, Avneesh Sud, Nicholas Dufour, and Luisa Verdoliva. 2023. TruFor: Leveraging all-round clues for trustworthy image forgery detec

EARLIEST HANDCRAFTED FEATURE BASED IMDL

- **Error Level Analysis Artifacts (ELA)** finds the compression error difference between forged regions and pristine regions through different JPEG compression qualities
- **Noise Inconsistency Artifacts (NOI)** modelled local noise using high pass wavelet coefficients
- **Camera Filter Array Artifacts** (**CFA)** approximated the camera filter array patterns by using nearby pixels to generate tampering probability of each pixel

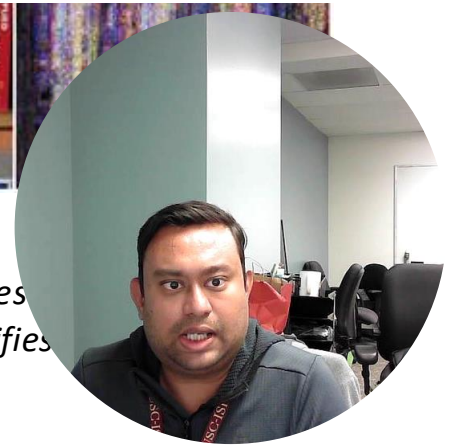


*Fig 6: An original photo res
– 95% ELA identifies

*Neal Krawetz and Hacker Factor Solutions. 2007. A picture's worth. Hacker Factor Solutions 6, 2 (2007), 2.

FIRST GENERATION EARLY DNN TECHNIQUES

- First Generation DNN based methods focused on a specific type of image forgery – **splicing** being most common
- But type of forgery implemented on real-world manipulated images is not known apriori
- General forgery detection algorithms were proposed to detect any type of unknown manipulation

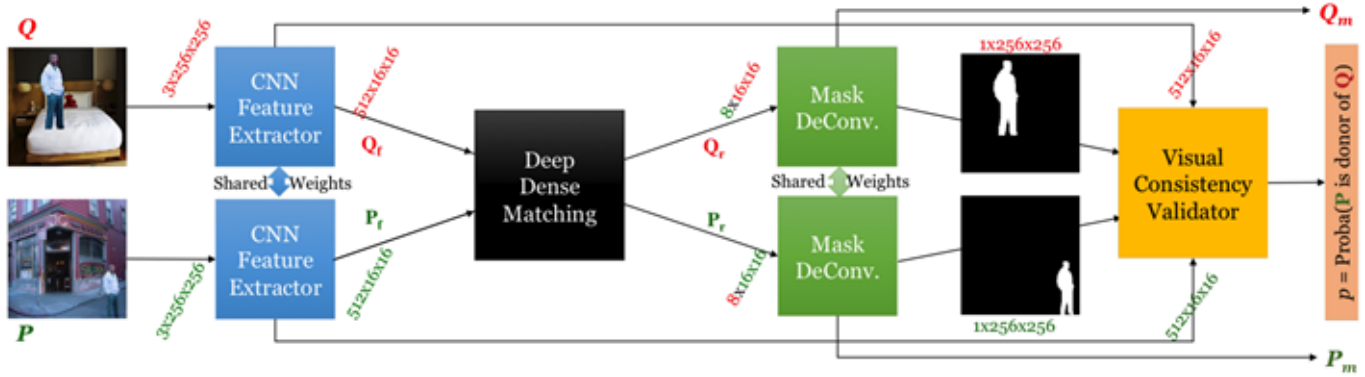
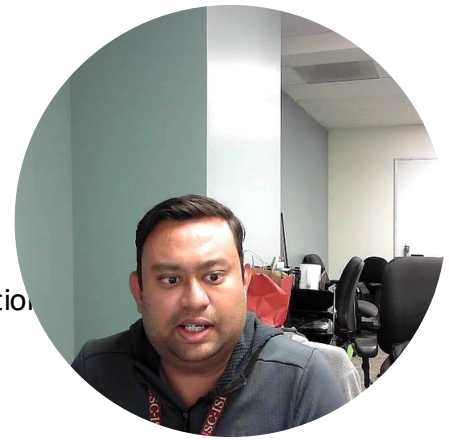


*Fig 7: Given a query and donor image, ***DMVN*** estimated the probability that the donor image has been used to splice the query image*



*Yue Wu, Wael Abd-Almageed, and Prem Natarajan. 2017. Deep matching and validation network: An end-to-end solution to constrained image splicing localizatio

GENERALIZED DNN FORGERY DETECTION

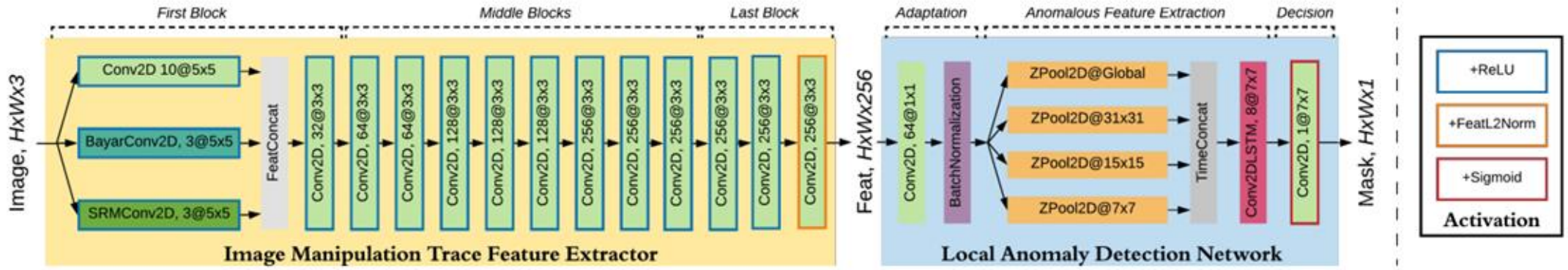
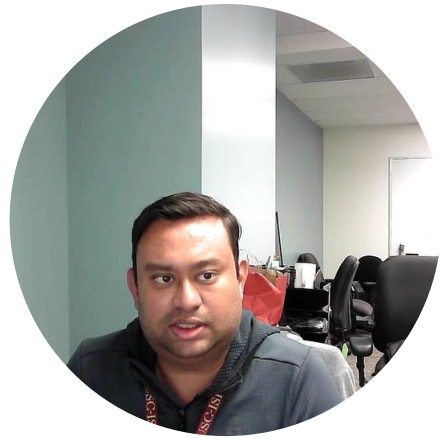


*Fig 8: ***ManTraNet****: Two subnets - image manipulation tracing feature extractor and local anomaly detection network generate a binary forgery localization map*

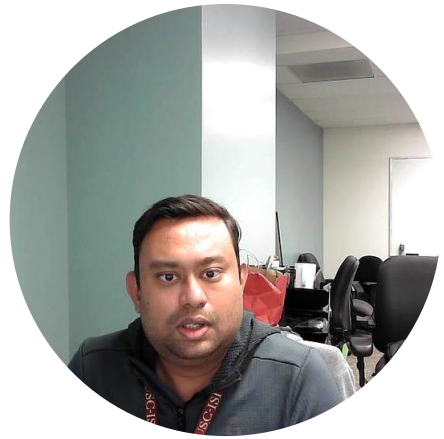
- The general framework used by the generalized DNN based IMDL frameworks consists of
  - Input feature extraction network
  - Feature fusion (if multiple artifacts used)
  - Anomaly detection network
- Anomaly detection network is pre-trained with a synthetic dataset sampled from MS-COCO
- Pre-trained models are fine-tuned on the train-split of the IFDL evaluation datasets



*Yue Wu, Wael Abd-Almageed, and Prem Natarajan. 2019. ManTraNet, CVPR 2019

ROADMAP

- What is Semantic Forgery Detection?
  - Previous research in Image Forgery Detection
  - **What are the Challenges in Image Forgery Detection?**
- Solution Proposed – Forensim
  - Forensim Contributions
  - Why Three Class-based Training?
  - State Space Model Overview
  - Forensim Modules
  - CMFD-Anything Dataset
  - Evaluation and Results



WHAT ARE THE CHALLENGES IN IMDL?

- **Specific Forgery Artifacts** used for training may not be relevant in many cases
  - If the boundary is smooth, edge inconsistency-based methods will fail
  - If multiple jpeg compression is applied, resampling feature-based method will fail
- A **single IFDL model** may not work for all types of manipulations
- Keeping the pristine and tampered **regions well separated** during each phase of training is a big challenge

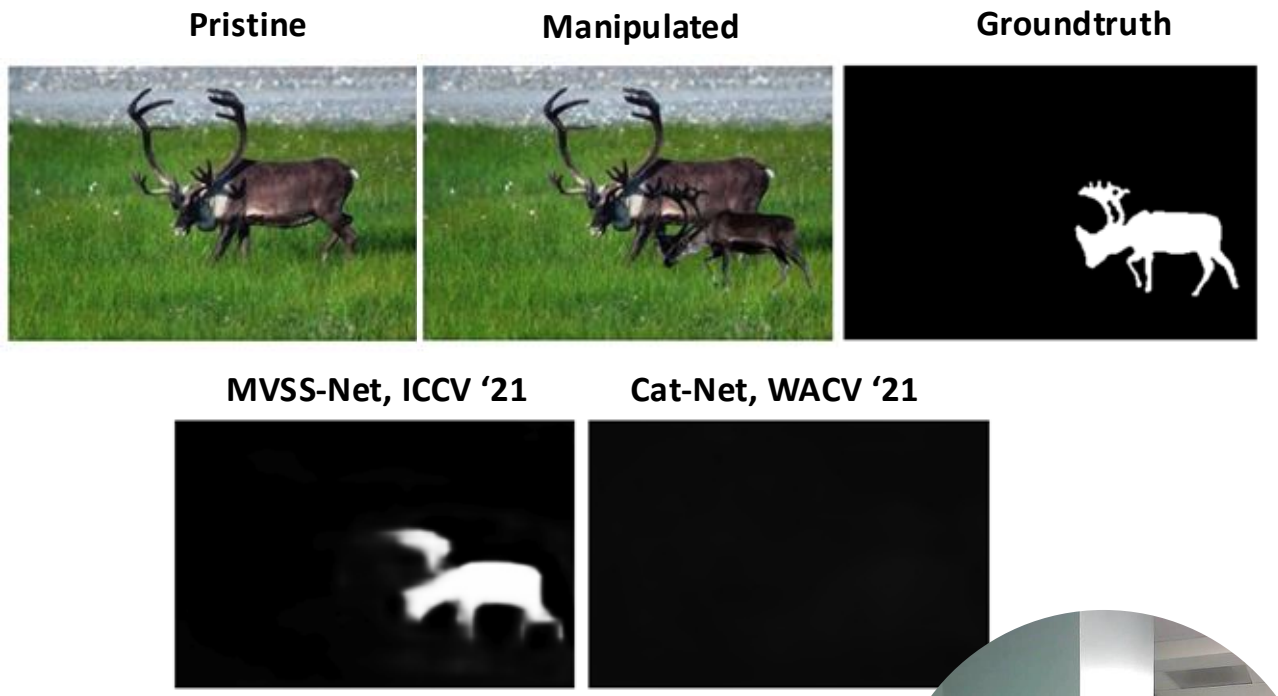
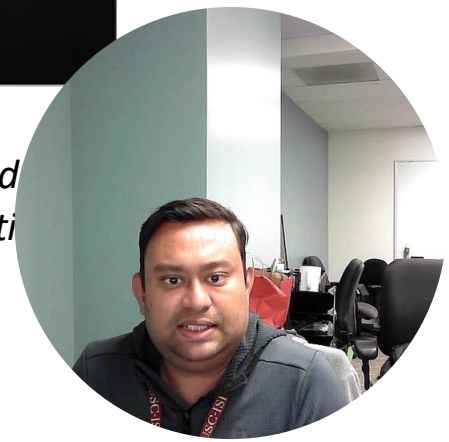


Fig 10: Cat-Net trained with spliced images fails for Copymove manipulati



WHAT ARE THE CHALLENGES IN IMDL?

- Even Generalized Forgery Detection models are not robust for different domains:
  - ManTraNet does not work on Biomedical images
- **Balancing** between **False Alarms** and **Misdetections** is challenging
- **New ways of tampering** generated everyday, hard to keep track and counter

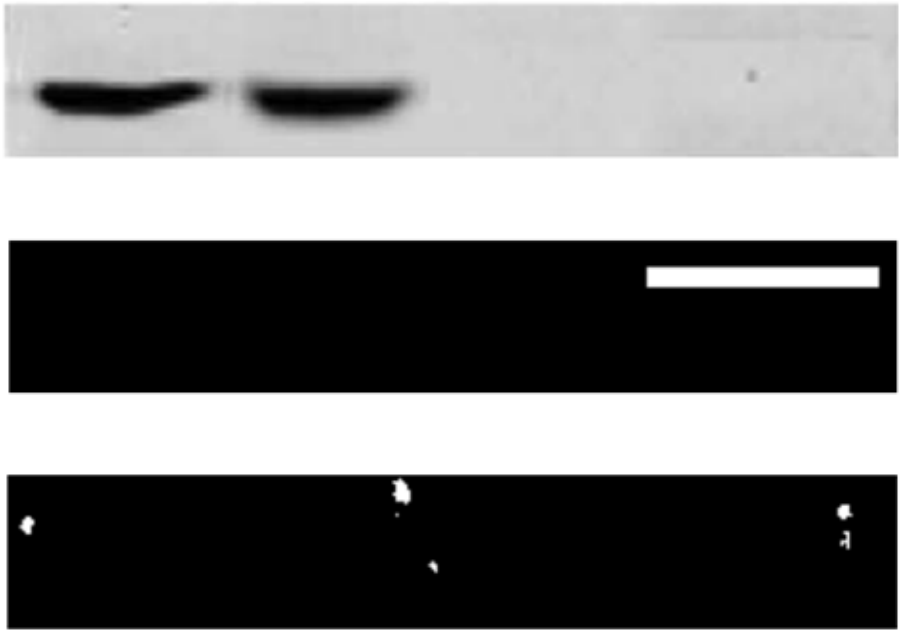
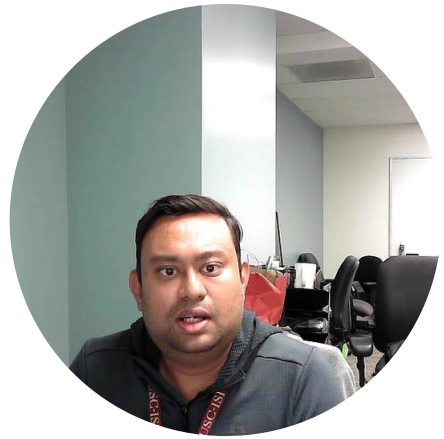


Fig 11: ManTraNet failed to detect biomedic
Et al. BioFors, ICCV 20

ROADMAP

- What is Semantic Forgery Detection?
  - Previous research in Image Forgery Detection
  - What are the Challenges in Image Forgery Detection?
- Solution Proposed – Forensim
  - **Forensim Contributions**
  - Why Three Class-based Training?
  - State Space Model Overview
  - Forensim Modules
  - CMFD-Anything Dataset
  - Evaluation and Results



FORENSIM INTRODUCTION

Problem Statement:

- Generative Models, Editing Tools, Social Media accelerate forgery
- Diverse manipulation strategies increase detection difficulty
- Existing detectors are task-specific – splicing or CMFD
- Existing detectors lack a unified correspondence framework

Key Question:

- *Can a single unified model detect both CMFD and splicing forgeries?*

Can Image Splicing and Copy-Move Forgery Be Detected by the Same Model? Forensim: An Attention-Based State-Space Approach

Soumyaroop Nandi[1,2] Prem Natarajan[1,2]
[1]USC Information Sciences Institute, Marina del Rey, CA, USA
[2]USC Thomas Lord Department of Computer Science, Los Angeles, CA, USA
{soumyarn, premkumn}@usc.edu

Abstract

We introduce Forensim, an attention-based state-space framework for image forgery detection that jointly localizes both manipulated (target) and source regions. Unlike traditional approaches that rely solely on artifact cues to detect spliced or forged areas, Forensim is designed to capture duplication patterns crucial for understanding context. In scenarios like protest imagery, detecting only the forged region—e.g., a duplicated act of violence inserted into a peaceful crowd—can mislead interpretation, highlighting the need for joint source-target localization. Forensim outputs three-class masks (pristine, source, target) and supports detection of both splicing and copy-move forgeries within a unified architecture. We propose a visual state-space model that leverages normalized attention maps to identify internal similarities, paired with a region-based block-attention module to distinguish manipulated regions. This design enables end-to-end training and precise localization. Forensim achieves state-of-the-art performance on standard benchmarks. We also release CMFD_Anything, a new dataset addressing limitations of existing copy-move forgery datasets. Project page and code.

1. Introduction

The proliferation of generative models has broadened image manipulation far beyond traditional editing tools. The challenge for today's researchers is to detect manipulated images, irrespective of the source of manipulation, be

Figure 1. Proposed CMFD_Anything samples- Rows show: (a) forged image, (b) RGB mask, (c) forged image, (d) RGB mask. Source-target mask encodes untampered, source, & target regions.

tracing a tampered

Tracing the s
a model is trai
whether invol
based forgeri
an unrelated
ternally, whe
repositioned
remove objec
replace them w

Beyond direc
be subtly altered w
various forms of *ima*

FORENSIM CONTRIBUTIONS

- **Unified Framework**
  - Jointly models **Copy-Move** and **Image Splicing** detection
- **Novel Attention Design**
  - **Similarity Attention** → Precise copy-move localization
  - **Manipulation Attention** → Tampering region highlighting
- **Efficient Backbone**
  - Built on **State-Space Models (SSM)**
  - Captures long-range dependencies with **linear complexity**
- **CMFD_Anything Dataset**
  - High-resolution, diverse CMFD benchmark
  - Derived from Segment Anything images
  - Real-world complex manipulation scenarios

Can Image Splicing and Copy-Move Forgery Be Detected by the Same Model? Forensim: An Attention-Based State-Space Approach

Soumyaroop Nandi[1,2] Prem Natarajan[1,2]
[1]USC Information Sciences Institute, Marina del Rey, CA, USA
[2]USC Thomas Lord Department of Computer Science, Los Angeles, CA, USA
{soumyarn, premkumn}@usc.edu

Abstract

We introduce Forensim, an attention-based state-space framework for image forgery detection that jointly localizes both manipulated (target) and source regions. Unlike traditional approaches that rely solely on artifact cues to detect spliced or forged areas, Forensim is designed to capture duplication patterns crucial for understanding context. In scenarios like protest imagery, detecting only the forged region—e.g., a duplicated act of violence inserted into a peaceful crowd—can mislead interpretation, highlighting the need for joint source-target localization. Forensim outputs three-class masks (pristine, source, target) and supports detection of both splicing and copy-move forgeries within a unified architecture. We propose a visual state-space model that leverages normalized attention maps to identify internal similarities, paired with a region-based block-attention module to distinguish manipulated regions. This design enables end-to-end training and precise localization. Forensim achieves state-of-the-art performance on standard benchmarks. We also release CMFD_Anything, a new dataset addressing limitations of existing copy-move forgery datasets. Project page and code.

Figure 1. Proposed CMFD_Anything samples- Rows show: (a) forged image, (b) RGB mask, (c) forged image, (d) RGB mask. Source-target mask encodes untampered, source, & target regions.

1. Introduction

The proliferation of generative models has broadened image manipulation far beyond traditional editing tools. The challenge for today's researchers is to detect manipulated images, irrespective of the source of manipulation, be

tracing a tampered

Tracing the s
a model is trai
whether invol
based forgeri
an unrelated
ternally, whe
repositioned
remove objec
replace them w

Beyond direc
be subtly altered w
various forms of *imag*

USC INFORMATION SCIENCES INSTITUTE | USC Viterbi School of Engineering Information Sciences Institute | WACV 2026 TUCSON, AZ 3/6 - 3/10

ROADMAP

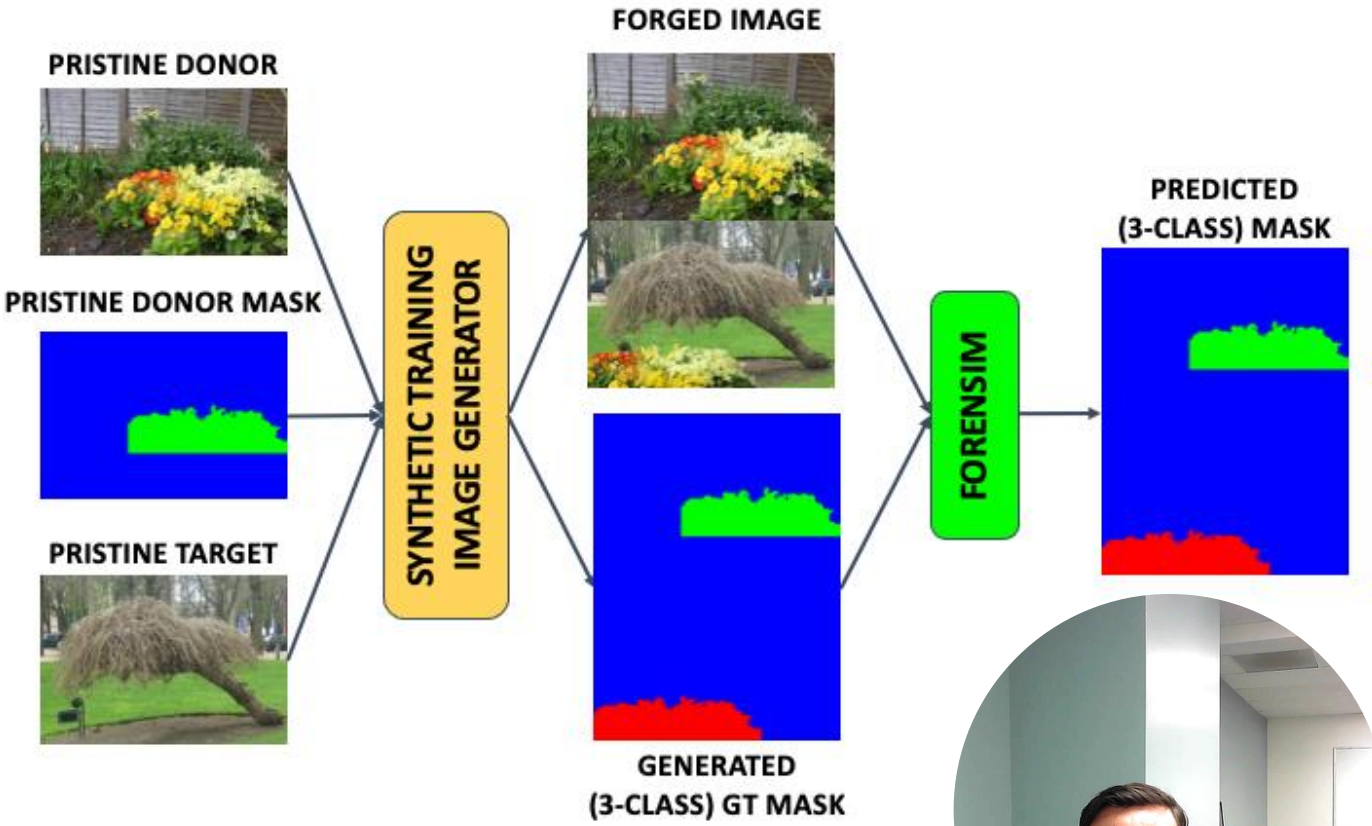
- What is Semantic Forgery Detection?
  - Previous research in Image Forgery Detection
  - What are the Challenges in Image Forgery Detection?
- Solution Proposed – Forensim
  - Forensim Contributions
  - **Why Three Class-based Training?**
  - State Space Model Overview
  - Forensim Modules
  - CMFD-Anything Dataset
  - Evaluation and Results

WHY THREE CLASS-BASED DETECTION?

- **Limitation of Two-Class Splicing Training**
  - Binary supervision: *manipulated vs pristine*
  - Learns appearance inconsistencies only
  - Cannot identify source region
  - Weak structural supervision

- **Advantage of Three-Class CMFD Training**
  - Role-aware supervision: **source / target / pristine**
  - Learns structured pixel correspondences
  - Models duplication relationships explicitly
  - Provides full forensic traceability

- **Representation-Level Benefit**
  - 2-class → Appearance-based decision boundary
  - 3-class → Correspondence-aware embedding space
  - Stronger inductive bias → Better generalization

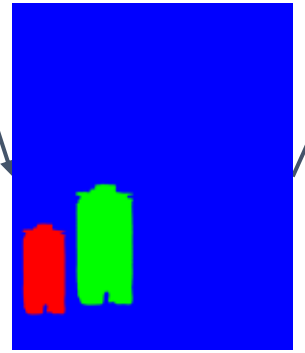


CMFD TRAINING SETUP

PRISTINE IMAGE

SYNTHETIC TRAINING IMAGE GENERATOR

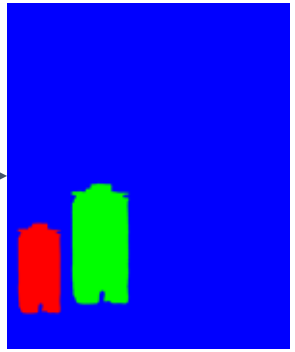
COPYMOVED IMAGE



GENERATED 3-CLASS GT MASK

COPYMOVE FORGERY DETECTOR MODEL

PREDICTED 3-CLASS MASK



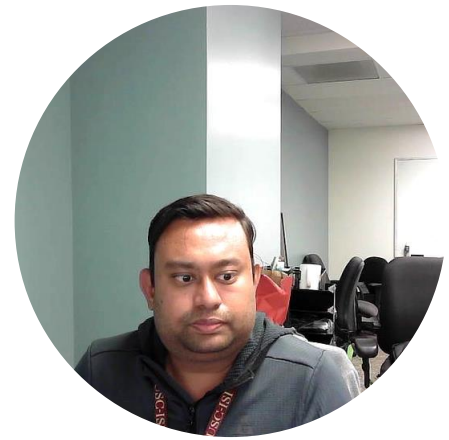


IMAGE SPLICING TRAINING SETUP

PRISTINE DONOR

PRISTINE TARGET

SYNTHETIC TRAINING IMAGE GENERATOR

SPLICED IMAGE

GENERATED BINARY (2-CLASS) GT MASK

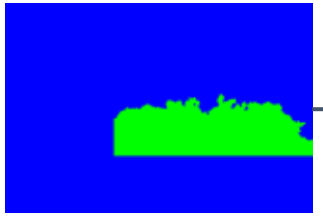
SPLICING FORGERY DETECTOR MODEL

PREDICTED BINARY (2-CLASS) MASK

FORENSIM IMAGE SPLICING TRAINING SETUP

PRISTINE DONOR

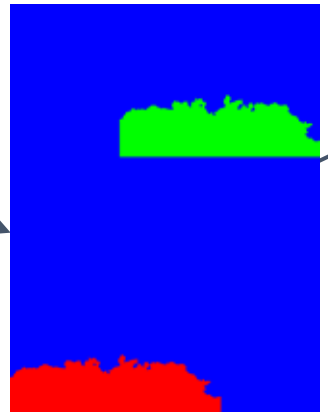
PRISTINE DONOR MASK



PRISTINE TARGET

SYNTHETIC TRAINING IMAGE GENERATOR

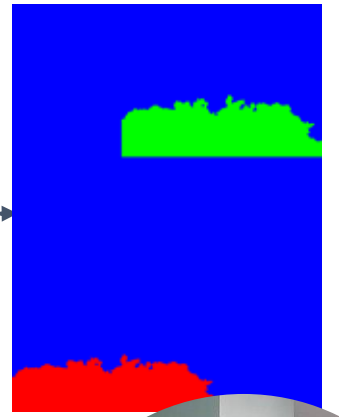
FORGED IMAGE

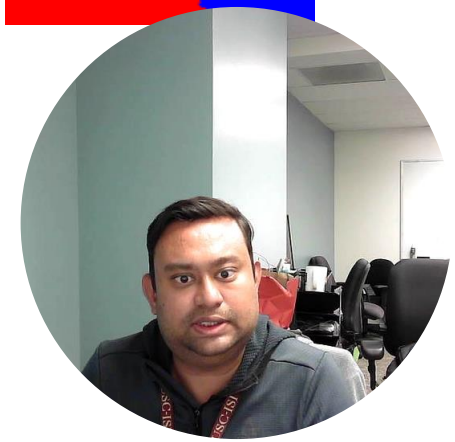


GENERATED 3-CLASS GT MASK

FORENSIM

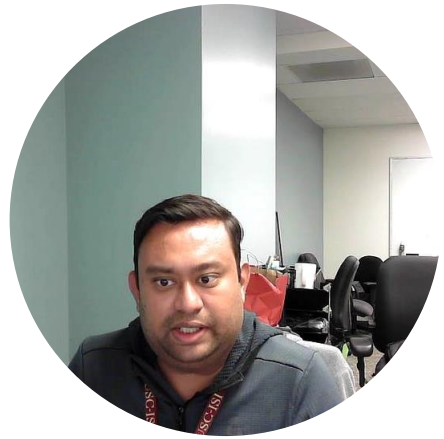
PREDICTED 3-CLASS MASK





ROADMAP

- What is Semantic Forgery Detection?
  - Previous research in Image Forgery Detection
  - What are the Challenges in Image Forgery Detection?
- Solution Proposed – Forensim
  - Forensim Contributions
  - Why Three Class-based Training?
  - **State Space Model Overview**
  - Forensim Modules
  - CMFD-Anything Dataset
  - Evaluation and Results



STATE SPACE MODEL OVERVIEW

What is a State Space Model (SSM)?

- A **linear time-invariant (LTI) system** [1] that models sequential dependencies
- Used in signal processing, similar to the **Kalman Filter**
- Defined by **hidden states** evolving over time with input and projection parameters

SSM Formulation:

Given a continuous input $x(t)$, the system computes the output $y(t)$ using hidden states $h(t)$:

$$h'(t) = \boldsymbol{A}h(t) + \boldsymbol{B}x(t), \quad y(t) = \boldsymbol{C}h(t) \ldots\ldots\ldots\ldots\ldots (1)$$

***A*:** Governs state evolution
***B*:** Projects input into hidden space
***C*:** Maps hidden states to output

[1] Albert Gu, Karan Goel, and Christopher Re. Efficiently modeling long sequences with structured state spaces. ICLR 2022

STATE SPACE MODEL OVERVIEW

Discretization via Zero-Order Hold (ZOH) [1]: Transitions from continuous to discrete SSM

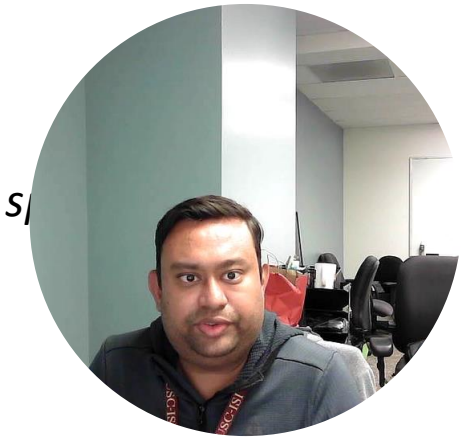
$$\bar{A} = \exp(\Delta A), \quad \bar{B} = (\Delta A)^{-1}(\exp(A) - I)\,\Delta B, \quad \bar{C} = C$$

$$y_k = \bar{C}h_k + \bar{D}x_k \ , \quad h_k = \bar{A}h_{k-1} + \bar{B}x_k \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$\bar{D}$ works as a residual connection and $\bar{B}$ can be approximated using first order Taylor Series as:

$$\bar{B} = (\exp(A) - I)A^{-1}B \approx (\Delta A)(\Delta A)^{-1}\,\Delta B = \Delta B$$

[1] Albert Gu, Karan Goel, and Christopher Re. Efficiently modeling long sequences with structured state s

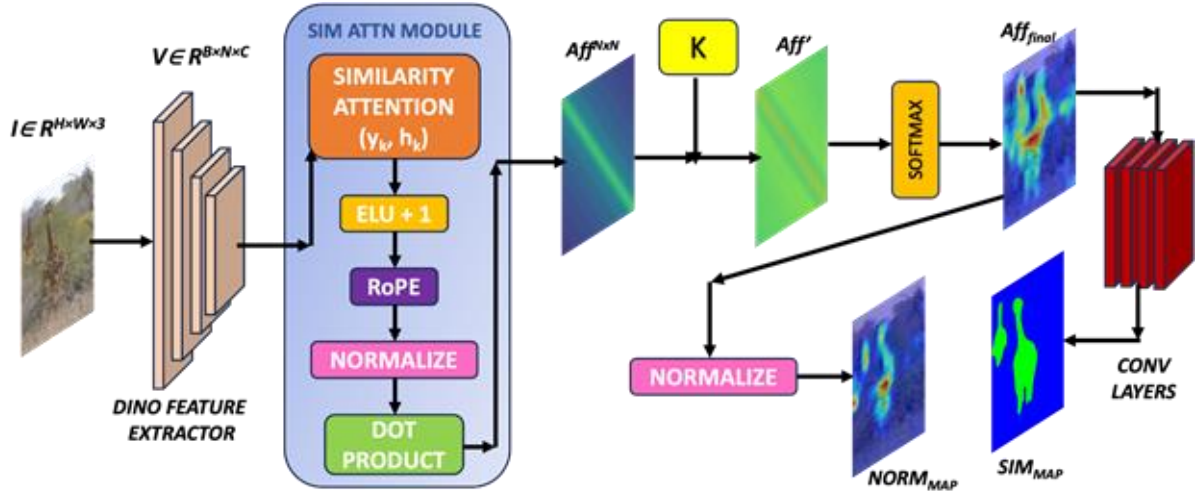


ROADMAP

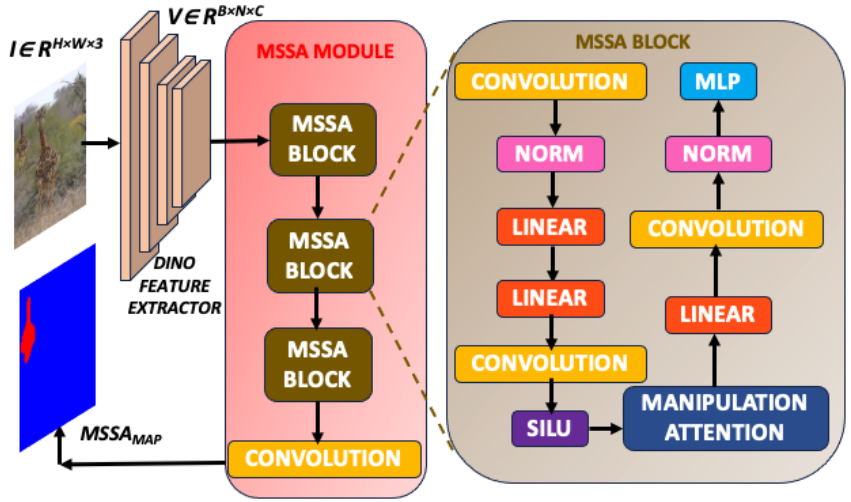
- What is Semantic Forgery Detection?
  - Previous research in Image Forgery Detection
  - What are the Challenges in Image Forgery Detection?
- Solution Proposed – Forensim
  - Forensim Contributions
  - Why Three Class-based Training?
  - State Space Model Overview
  - **Forensim Modules**
  - CMFD-Anything Dataset
  - Evaluation and Results

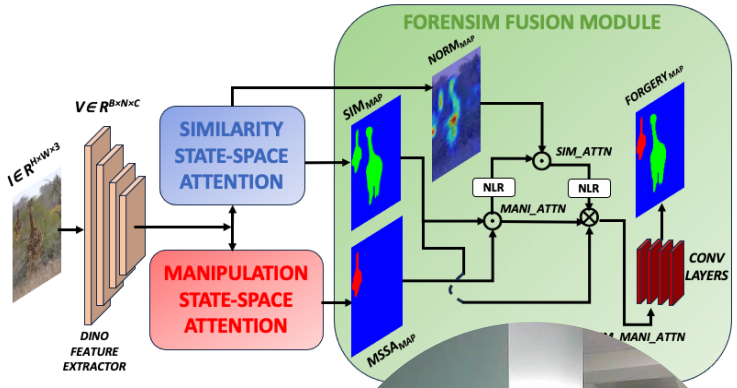
FORENSIM OVERVIEW

Forensim Similarity Attention Module

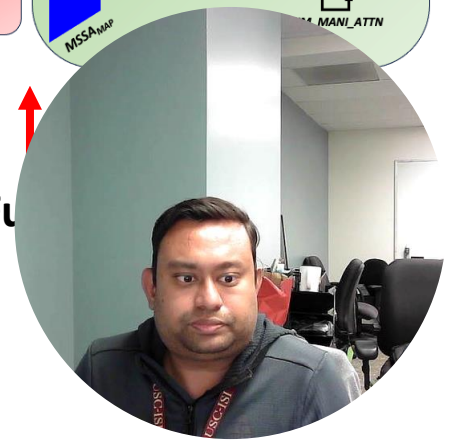


Forensim Manipulation Attention Module





Forensim Fu



FORENSIM ATTENTION SIMILARITY

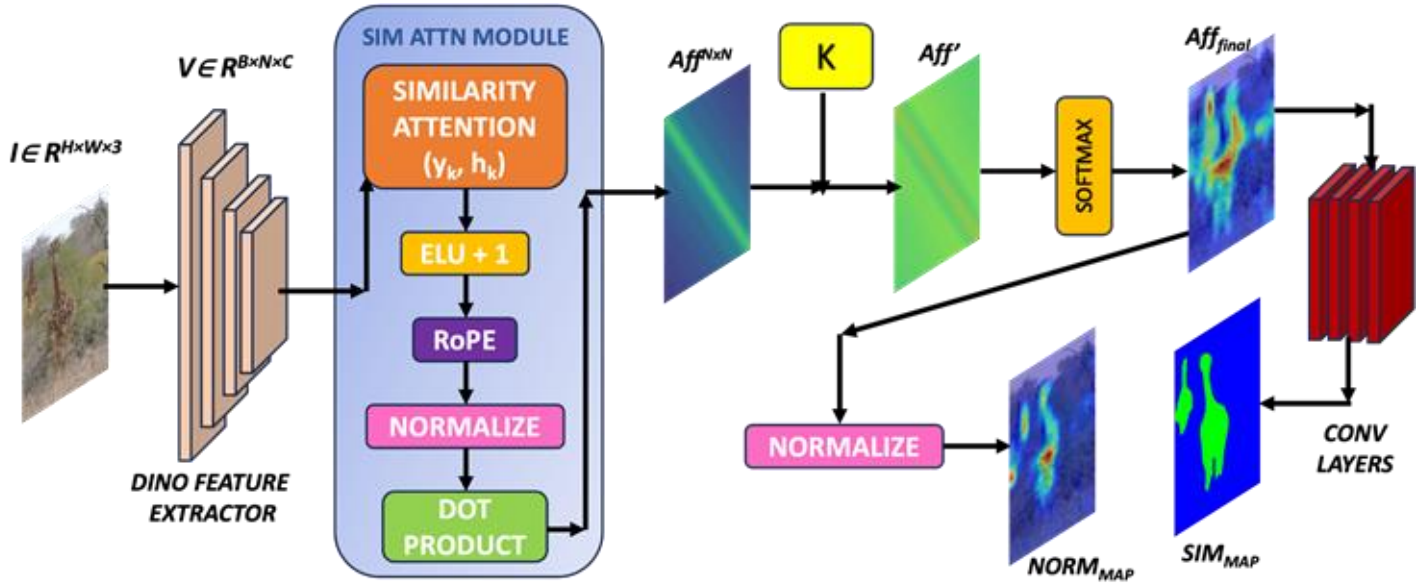


Fig 12: Forensim Similarity Attention Module

What It Does:

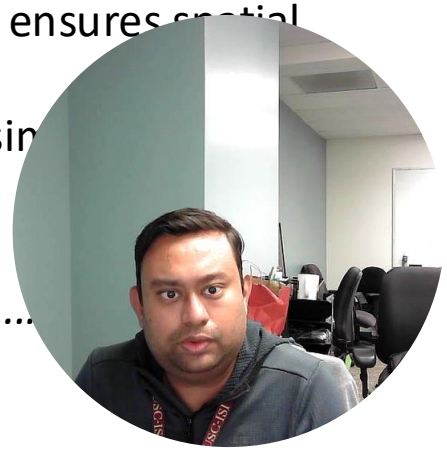
- Computes an **explicit affinity matrix** *Aff* that models region similarity

$$y_k = \bar{C}h_k / \bar{C}n_k + \bar{D}x_k, \qquad h_k = \bar{A}h_{k-1} + \bar{B}x_k \text{.....(3)}$$

$$\text{where, } n_k = \sum_{j=1}^{k} \bar{B}_j$$

- Leverages **State Space Models (SSM) for sequential similarity** computation
- Ensures **global receptive fields** for accurate forgery detection
- **RoPE Positional Encoding** ensures spatial coherence
- **ELU Activation** stabilizes sim computations

$$Aff = \bar{C}\,\bar{B}^{T} \text{.....}$$



FORENSIM ATTENTION SIMILARITY

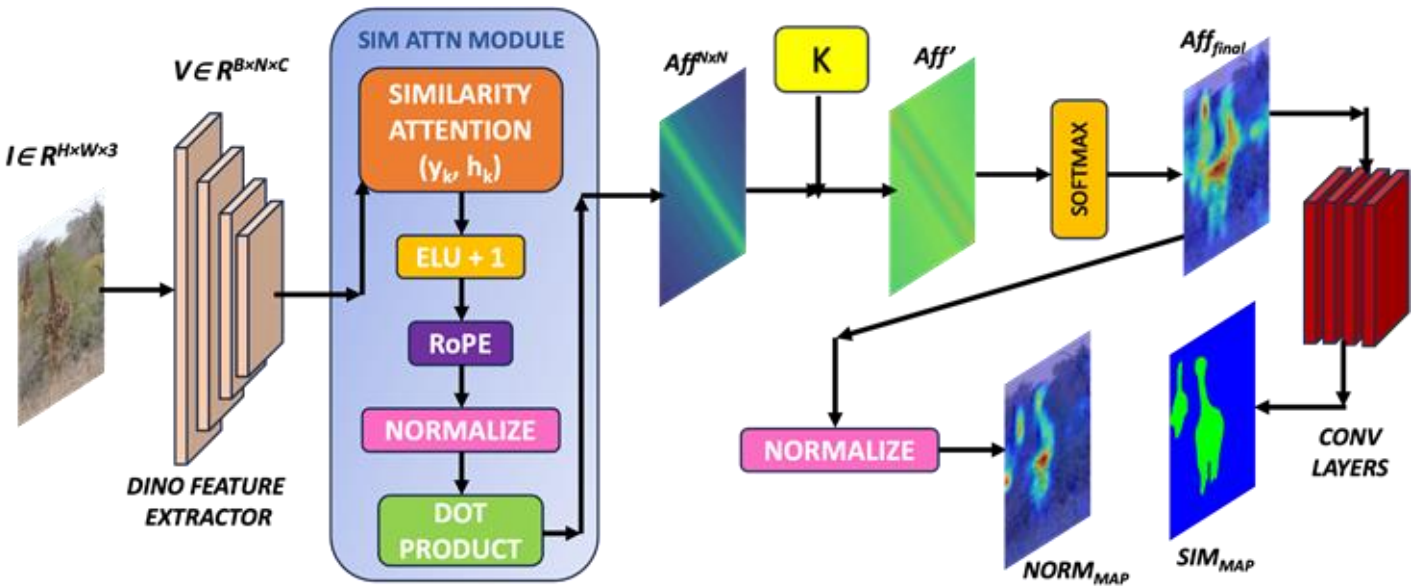


Fig 12: Forensim Similarity Attention Module

Refining the Affinity Matrix:

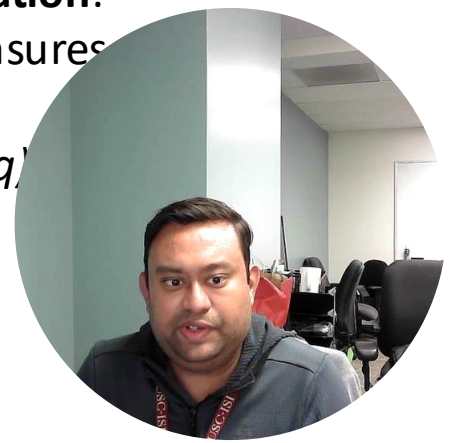
- **Issue**: Higher values along the diagonal (self-correlation)
- **Solution**: Apply spatial smoothing function K(p,q,p',q') to reduce diagonal dominance

$$K(p,q,p',q') = \frac{(p - p')^2 + (q - q')^2}{(p - p')^2 + (q - q')^2 + \sigma^2}$$

Final Affinity Matrix Computation:

- **Bidirectional Softmax** ensures reinforcement:

$$Aff_{final}(p,q) = Aff_{row}(p,q)$$



FORENSIM ATTENTION SIMILARITY

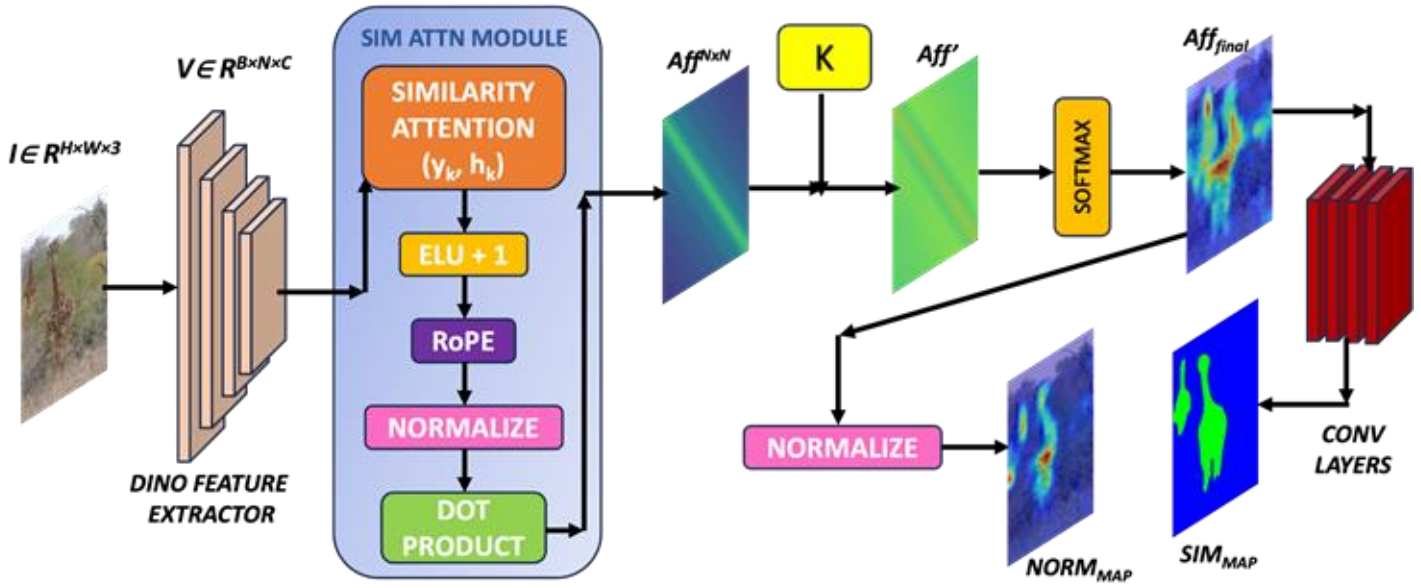


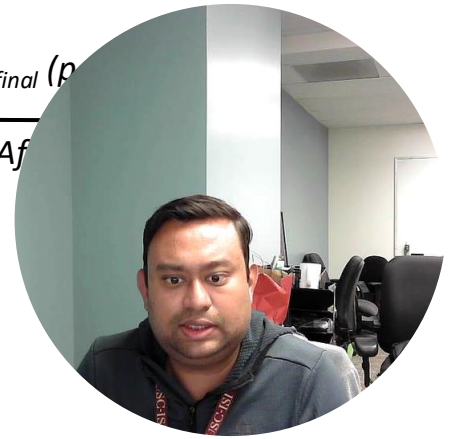
Fig 12: Forensim Similarity Attention Module

Sim_{Map} and $Norm_{Map}$ Generation:

- A **convolutional module** refines the final **Sim_{Map}**
- Outputs **Sim_{Map}**: Highlights **top-k most similar regions per pixel**

- Outputs **$Norm_{Map}$**: Normalized map capturing global dependencies

$$Norm_{Map}(p,q) = \frac{(Aff_{final}(p}{(\Sigma^{N}_{q'=1} Af}$$



FORENSIM MANIPULATION ATTENTION

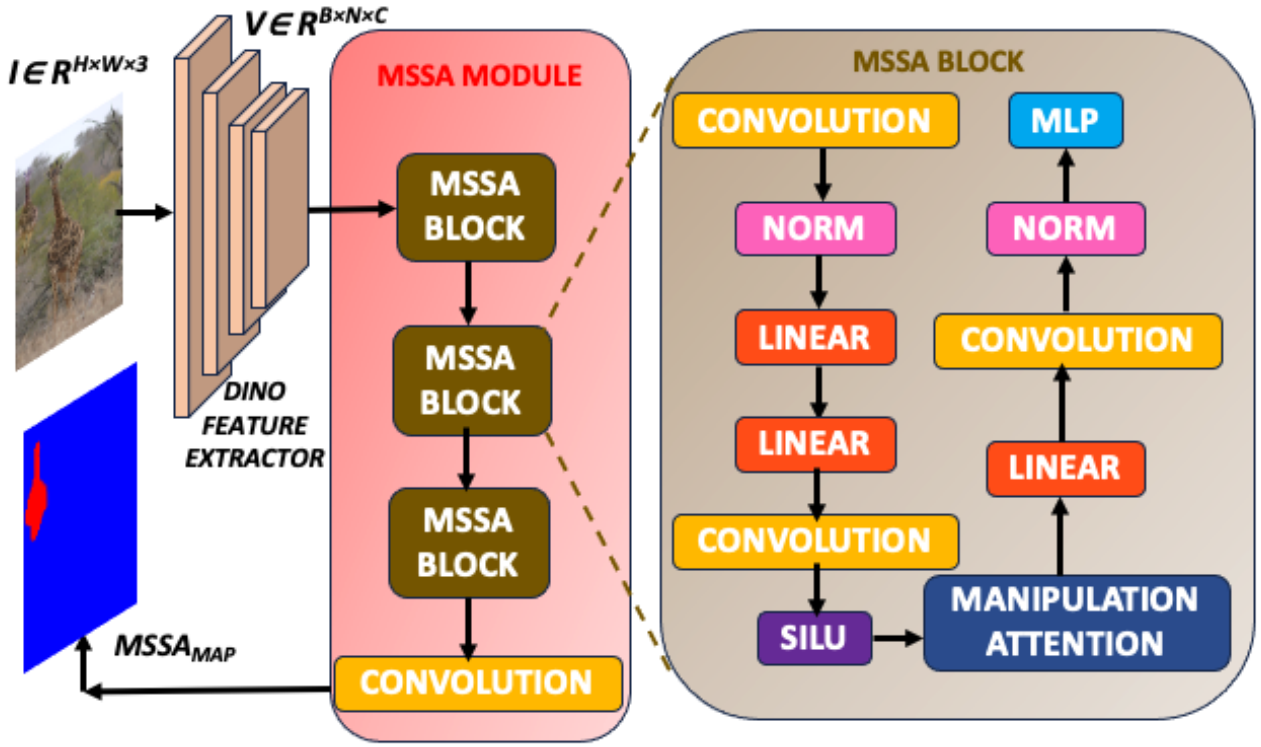


Fig 13: Forensim Manipulation Attention Module

- MSSA enhances **forgery detection** using multi-scale attention
- Composed of **three $MSSA_{Block}$ units** for global & local feature learning
- Outputs **manipulation map** ($MSSA_{Map}$) that highlights forgery regions

$$\bar{C}\bar{B} = ELU(Linear(V_k)) + 1.0 \ldots\ldots (6)$$

FORENSIM MANIPULATION ATTENTION

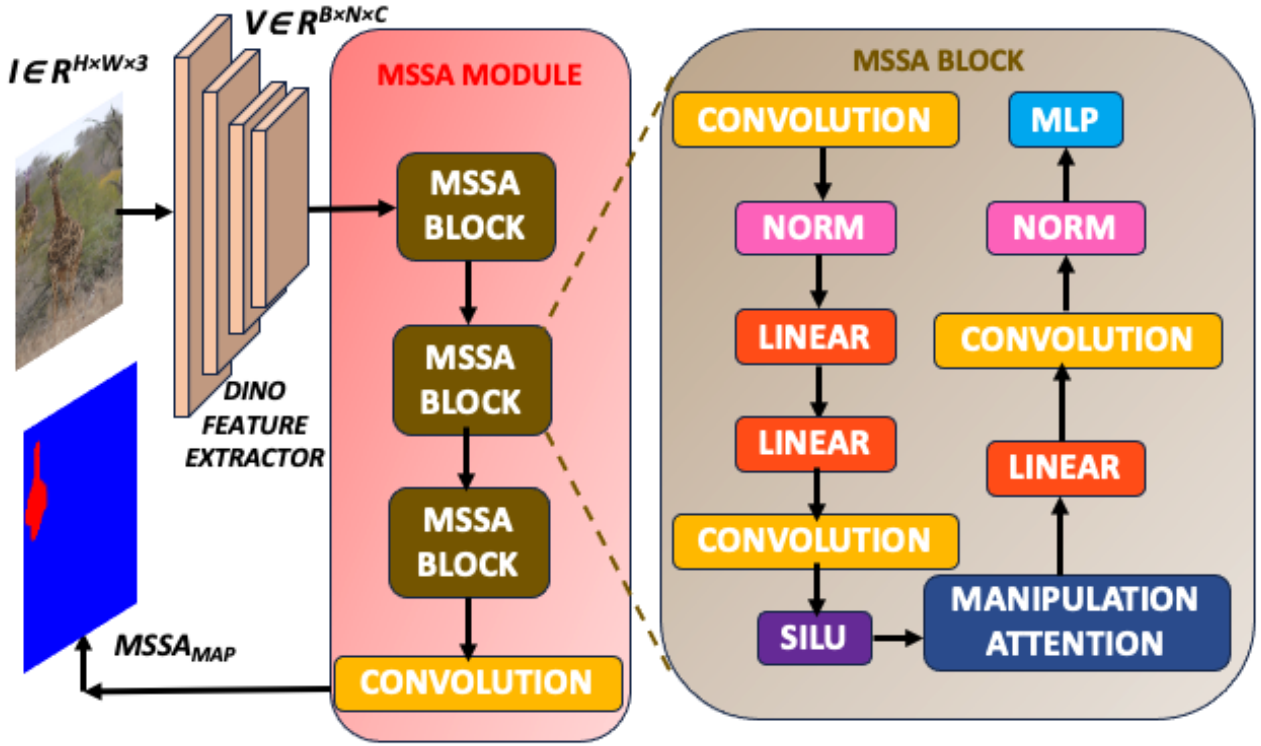


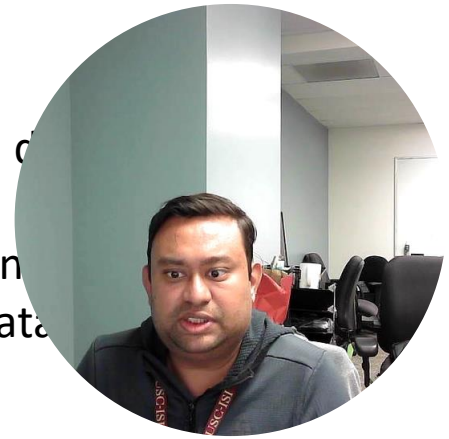
Fig 13: Forensim Manipulation Attention Module

- **Multi-Head Attention** → Captures dependencies across different scales

$$V_k = (\bar{C} \,.\, Rope(\bar{B}))\, \bar{B}^{\top} V_k \,.\, p + LePE(V_k) \ldots\ldots (7)$$

where, $p = 1/(\bar{C}\bar{B}^{T} + \varepsilon)$

- **Efficient Computation** → Reduces memory overhead compared to self-attention
- **Spatial Awareness** → Uses **RoPE** & **LePE** for precise localization
- **Feed-forward MLP** → Refines d
- **Robust Detection** → Consisten manipulated regions across data



FORENSIM FUSION MODULE

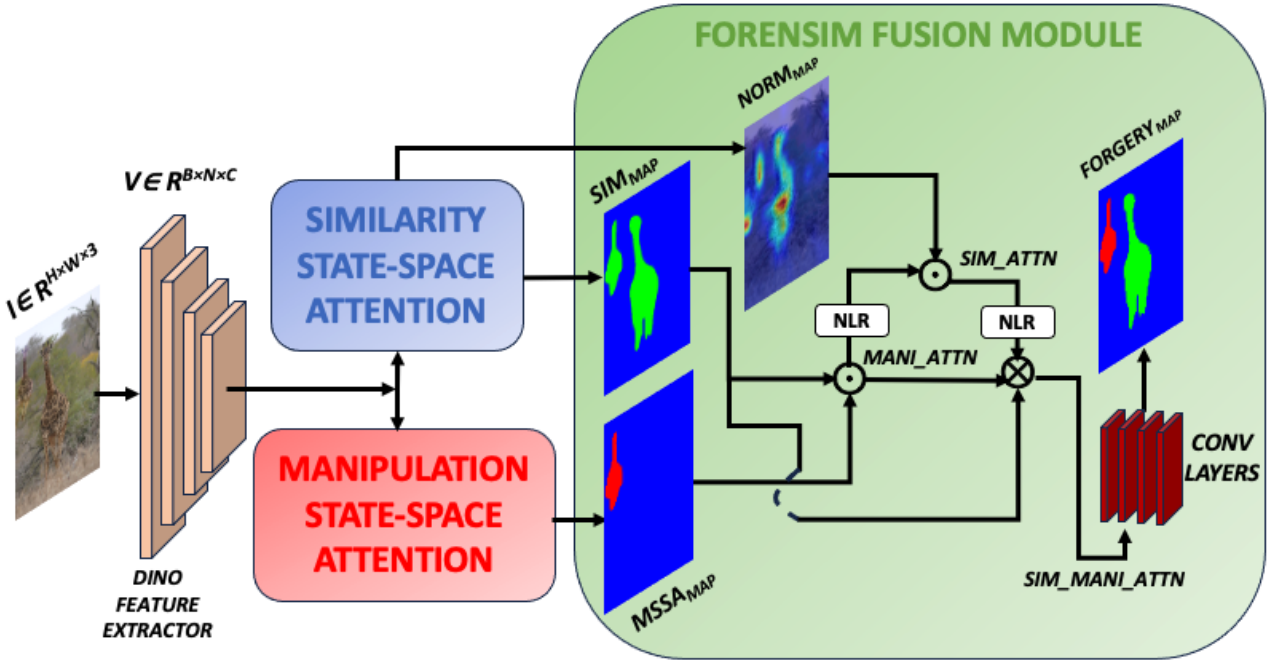


Fig 14: Forensim Fusion Attention Module

- The **Non-Local Refinement (NFR) module** propagates information across spatial locations

$$Sim\_Attn = NFR(MSSA_{MAp} \odot Sim_{Map}) \ldots\ldots\ldots (8)$$

$$Mani\_Attn = NFR(Norm_{Map} \otimes Sim\_Attn) \ldots\ldots (9)$$

$$Sim\_mani\_Attn = concat(Sim\_Attn, Mani\_Attn) \ldots . (10)$$

- Integrates Sim_{Map} , $Norm_{Map}$ and $MSSA_{MAp}$ to enhance contextual coherence

- The fused features are **processed convolutional layers** to refine t **forgery mask** & compute a **det**

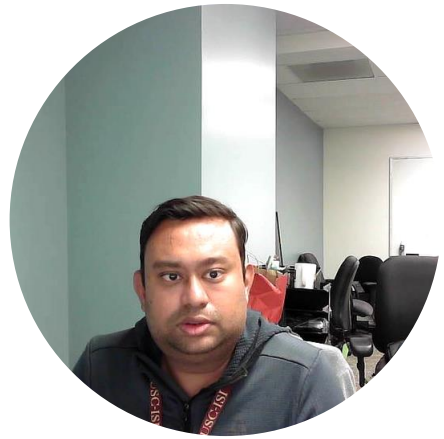
ROADMAP

- What is Semantic Forgery Detection?
  - Previous research in Image Forgery Detection
  - What are the Challenges in Image Forgery Detection?
- Solution Proposed – Forensim
  - Forensim Contributions
  - Why Three Class-based Training?
  - State Space Model Overview
  - Forensim Modules
  - **CMFD-Anything Dataset**
  - Evaluation and Results

NEED FOR A NEW CMFD DATASET

Limitations of Existing Datasets:

- **USC-ISI CMFD**, **CoMoFoD**, and **CASIA CMFD** offer **limited** and **synthetic-looking** images
- Only **941** copy-move forgery images publicly available
- **No source/target region masks**, only binary forgery masks
- Existing forgeries are **easily detectable** by humans, lacking **high-quality**, **realistic**, and **diverse** training data
- **Pristine (negative) samples** are crucial but **missing** in most datasets



NEED FOR A NEW CMFD DATASET

Our Solution: CMFD-Anything Dataset

- Based on **Segment Anything** dataset objects and masks
- **200K** realistic **copy-move forged images**
- **100K** pristine, unaltered images
- Covers **single-object** and **multi-object** copy-move scenarios
- **Refined** using **MGMatting** for seamless duplication
- Split: **80% training**, **10% validation**, **10% testing**

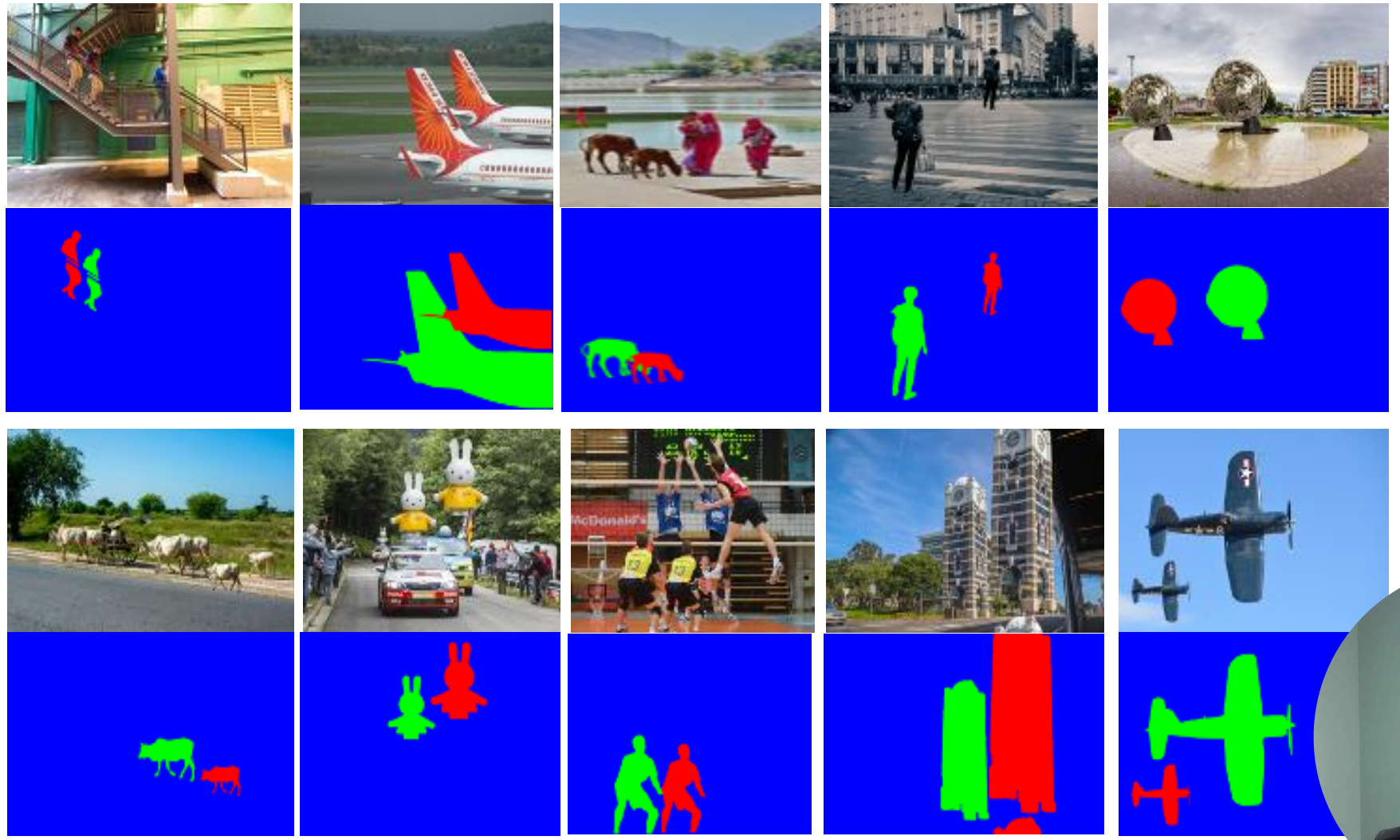
CMFD-ANYTHING DATASET

FORGED IMAGE

CMFD MASK

FORGED IMAGE

CMFD MASK



USC INFORMATION SCIENCES INSTITUTE

USC Viterbi School of Engineering Information Sciences Institute

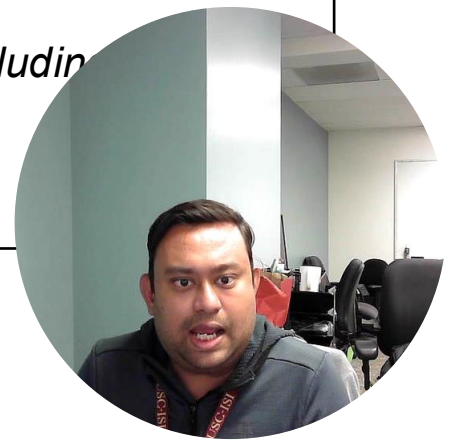
WACV 2026 TUCSON, AZ 3/6 - 3/10

CMFD-ANYTHING IMAGE QUALITY CHECK

Method	Type	Casia_CMFD	CoMoFoD	CMFD_Anything
PaQ-2-PiQ (↑)	No-Reference	3.64	3.42	**4.11**
CLIP-IQA (↑)	No-Reference	6.58	6.42	**7.23**
SSIM (↑)	Full-Reference	0.832	0.779	**0.881**
LPIPS (↓)	Full-Reference	0.179	0.243	**0.124**

Table 1: Image Quality Assessment (IQA) on CMFD datasets ↑ indicates higher is better; ↓ indicates lower is better.

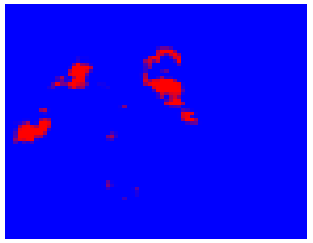
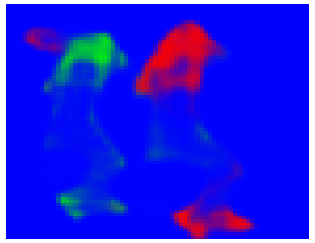
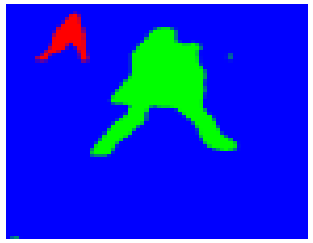
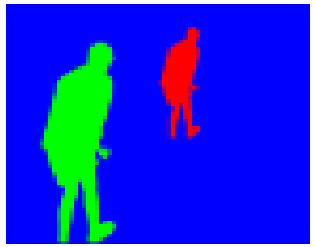
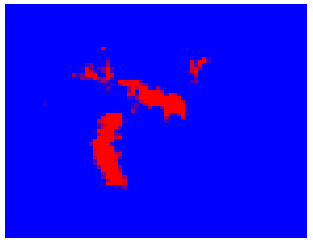
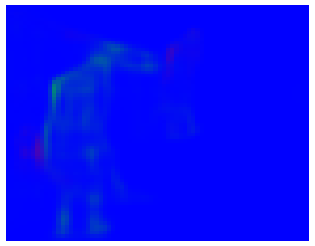
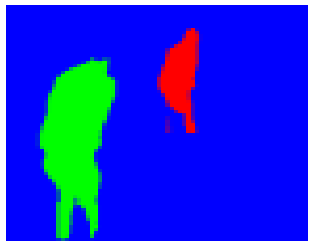
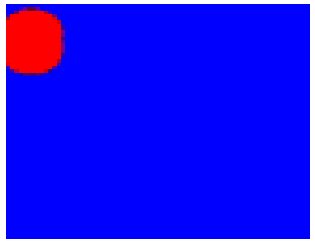
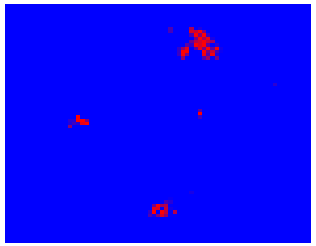
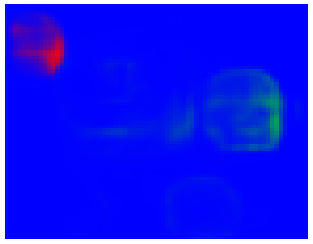
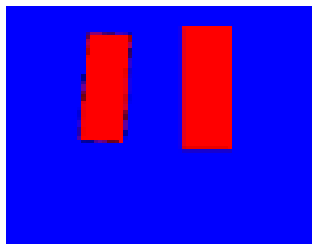
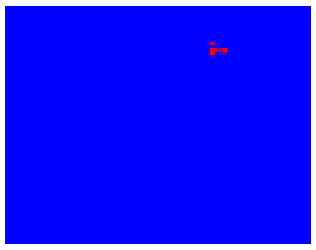
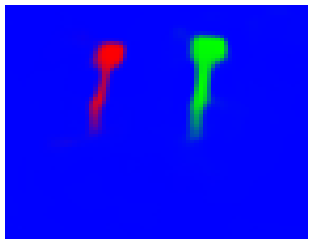
- *For forged images alone, we employ no-reference metrics such as PaQ-2-PiQ and CLIP-IQA to evaluate perceptual naturalness in the absence of ground truth.*
- *For comparing forged images against their pristine counterparts, we use full-reference metrics includin*
  LPIPS to measure structural similarity and perceptual distance.
- **(↑)** indicates higher is better; **(↓)** indicates lower is better. ***Bold - Best***, *Underline - Second Best*



ROADMAP

- What is Semantic Forgery Detection?
  - Previous research in Image Forgery Detection
  - What are the Challenges in Image Forgery Detection?
- Solution Proposed – Forensim
  - Forensim Contributions
  - Why Three Class-based Training?
  - State Space Model Overview
  - Forensim Modules
  - CMFD-Anything Dataset
  - **Evaluation and Results**

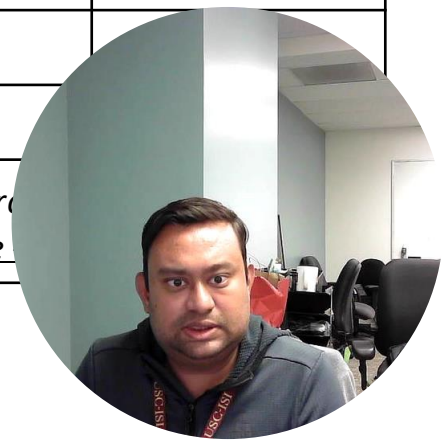
FORENSIM QUALITATIVE CMFD

	FORGED IMAGE	GT MASK	BUSTERNET MASK	MANTRANET MASK	DOA_GAN MASK	FORENSIM MASK
USC-ISI CMFD						
CMFD ANYTHING						
CASIA						
COMOFOD						

FORENSIM CMFD PERFORMANCE

METHOD	PRECISION [LOCALIZATION]			RECALL [LOCALIZATION]			F1 [LOCALIZATION]			PRECISION [DETECTION]	RECALL [DETECTION]	F1 [DETECTION]
	P	S	T	P	S	T	P	S	T			
USC-ISI CMFD TEST SET												
BusterNet	93.71	55.85	53.84	99.01	38.26	48.73	96.15	40.84	48.33	89.26	80.14	84.45
Mantranet	93.50	8.66	48.53	99.22	2.28	28.43	96.08	2.97	30.58	68.72	85.82	76.32
DOA_GAN	96.99	76.30	85.60	98.87	63.57	80.45	97.69	66.58	81.72	96.83	96.14	96.48
HiFi-Net	92.80	7.10	46.00	98.80	1.90	26.00	95.30	2.50	29.00	66.00	84.00	74.00
TruFor	96.88	77.10	86.42	99.01	65.92	81.93	97.99	67.82	82.89	96.95	96.88	96.59
SparseViT	97.01	77.85	87.23	99.10	66.91	82.47	98.06	68.29	83.44	97.10	97.08	
Forensim	**97.43**	**79.61**	**89.29**	**99.34**	**70.49**	**99.67**	**99.92**	**74.83**	**84.48**	**98.47**	**97.64**	

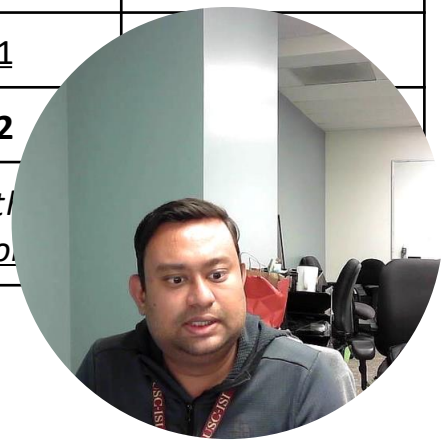
*Table 2: CMFD Results on USC-ISI CMFD test set. Localization valuated using pixel-level precision, recall, and F1 score metrics acr (Pristine), S (Source), and T (Target). Detection evaluated using image-level Precision, Recall, and F1. **Bold - Best**, Underline*



FORENSIM CMFD PERFORMANCE

METHOD	PRECISION [LOCALIZATION]			RECALL [LOCALIZATION]			F1 [LOCALIZATION]			PRECISION [DETECTION]	RECALL [DETECTION]	F1 [DETECTION]
	P	S	T	P	S	T	P	S	T			
CMFD-ANYTHING TEST SET												
BusterNet	47.34	36.88	35.16	53.42	26.78	34.96	47.87	28.34	31.26	44.56	42.61	43.12
ManTraNet	48.65	7.16	37.48	64.98	2.21	24.85	48.16	2.41	27.89	52.62	64.93	57.14
DOA_GAN	53.48	48.72	52.67	61.83	32.12	43.74	73.36	37.62	44.83	70.06	73.44	71.42
HiFi-Net	47.50	6.50	36.00	**73.00**	1.90	23.00	47.20	2.10	26.00	51.00	63.00	55.50
TruFor	56.91	50.83	57.63	65.42	36.21	47.32	77.83	41.29	49.51	72.45	74.62	73.01
SparseViT	57.74	52.14	59.95	67.88	38.04	49.89	79.36	43.16	51.73	73.12	75.81	
Forensim	**59.41**	**54.82**	**64.77**	69.74	**41.91**	**53.43**	**82.67**	**50.61**	**54.16**	**75.12**	**77.42**	

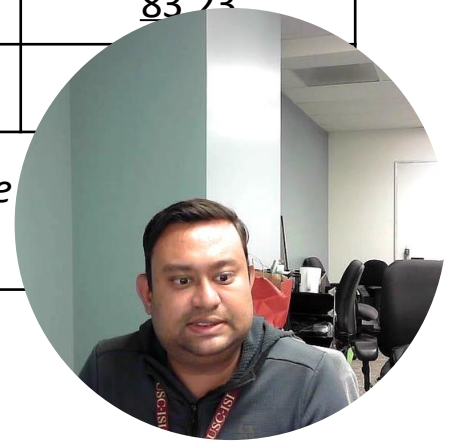
Table 3: CMFD Results on CMFD_Anything test set. Localization valuated using pixel-level precision, recall, and F1 score metrics across t
*(Source), and T (Target). Detection evaluated using image-level Precision, Recall, and F1. **Bold - Best**, Underline - Seco*



FORENSIM CMFD PERFORMANCE

METHOD	PRECISION [LOCALIZATION]	RECALL [LOCALIZATION]	F1 [LOCALIZATION]	PRECISION [DETECTION]	RECALL [DETECTION]	F1 [DETECTION]
BusterNet	42.15	30.54	33.72	48.34	75.12	58.82
ManTraNet	38.12	27.42	30.85	52.13	67.14	58.76
DOA_GAN	54.70	39.67	41.44	63.39	77.00	69.53
TruFor	<u>55.67</u>	41.83	43.62	<u>78.92</u>	86.32	82.43
SparseViT	54.32	<u>42.91</u>	<u>45.87</u>	77.81	<u>87.43</u>	<u>83.23</u>
Forensim	**61.87**	**47.24**	**58.12**	**84.24**	**91.71**	

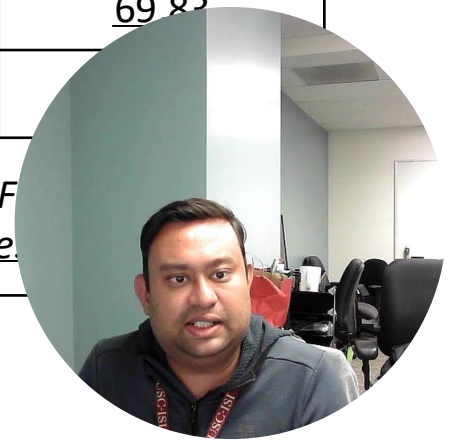
Table 4: CMFD Results on Casia CMFD test set. Localization valuated using pixel-level precision, recall, and F1 score evaluated using image-level Precision, Recall, and F1. ***Bold - Best****, <u>Underline - Second Best</u>*



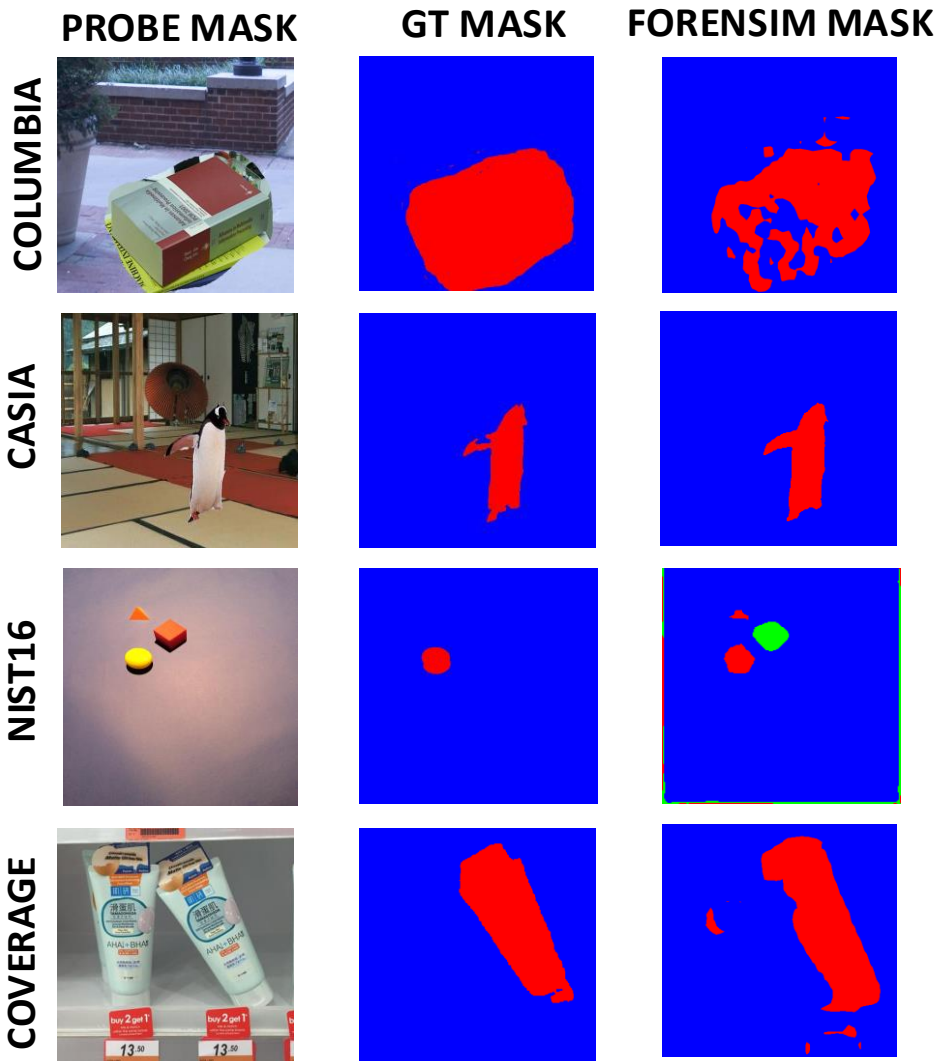
FORENSIM CMFD PERFORMANCE

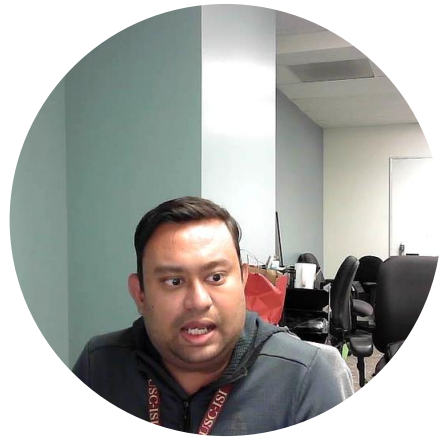
METHOD	PRECISION [LOCALIZATION]	RECALL [LOCALIZATION]	F1 [LOCALIZATION]	PRECISION [DETECTION]	RECALL [DETECTION]	F1 [DETECTION]
BusterNet	51.25	28.20	35.34	53.20	57.41	55.22
ManTraNet	36.11	25.48	28.34	50.34	54.92	52.48
DOA_GAN	48.42	37.84	36.92	60.38	65.98	63.05
TruFor	49.83	<u>38.74</u>	36.54	66.91	<u>71.25</u>	68.94
SparseViT	<u>51.15</u>	37.20	<u>37.92</u>	<u>67.70</u>	70.41	<u>69.82</u>
Forensim	**56.41**	**43.57**	**47.82**	**76.58**	**77.94**	

Table 5: CMFD Results on CoMoFoD CMFD test set. Localization valuated using pixel-level precision, recall, and F
Detection evaluated using image-level Precision, Recall, and F1. ***Bold - Best****, <u>Underline - Second Be</u>*



FORENSIM QUALITATIVE IMDL

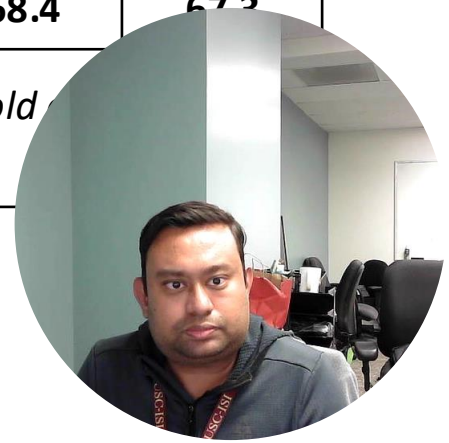




FORENSIM IMDL PERFORMANCE

METHOD	NIST16		Columbia		Coverage		Casia		Average	
	Best	*Fixed*	*Best*	*Fixed*	*Best*	*Fixed*	*Best*	*Fixed*	*Best*	*Fixed*
ManTraNet	21.9	19.3	47.5	46.2	21.1	19.6	38.2	32.7	32.2	29.5
TruFor	35.6	34.8	89.4	88.5	47.3	45.7	83.5	<u>81.8</u>	63.9	62.7
SparseViT	<u>39.4</u>	<u>38.4</u>	**95.9**	**95.9**	<u>52.5</u>	<u>51.3</u>	**84.2**	**82.7**	<u>68.0</u>	<u>67.1</u>
Forensim	**40.2**	**39.1**	<u>94.2</u>	<u>93.8</u>	**55.7**	**54.6**	<u>83.4</u>	<u>81.8</u>	**68.4**	**67.3**

Table 6: Pixel-level F1 localization on Benchmark IMDL datasets using best threshold and a fixed threshold
Bold - Best, *<u>Underline - Second Best</u>*



FORENSIM MODEL ABLATION STUDIES

Variant	Normalization	Casia	CoMoFoD	CMFD_Anything
None	Raw	31.6	29.2	33.8
Raw-Only	Partial	<u>44.8</u>	<u>36.7</u>	<u>54.3</u>
BiSoftmax	Full	**58.1**	**47.8**	**62.4**

Table 7: Pixel-level F1 across Normalization types in Similarity Attention.
Bold - Best, *<u>Underline - Second Best</u>*

DATASET ABLATION STUDIES

Ablation Condition	224 X 224	256 X 256	320 X 320	512 X 512
No MGMAtting	<u>56.6</u>	<u>57.7</u>	<u>58.5</u>	<u>61.2</u>
Binary Mask	53.9	55.2	56.3	59.4
Noisy Groudtruth	53.4	54.6	55.5	58.4
CMFD_Anything	**59.6**	**60.8**	**62.1**	**62.4**

Table 8: Pixel-level F1 for Ablation Studies on CMFD_Anything using Forensim.
Bold - Best, *<u>Underline - Second Best</u>*

DATASET ABLATION STUDIES

Training Data	MCC	F1 (Target)	AUC	BAcc
Casia only (Splicing)	0.418	0.316	0.598	0.683
Casia only (Copy-Move)	0.401	0.302	0.587	0.671
Casia only (Removal)	0.406	0.308	0.592	0.677
CoMoFoD only (Splicing)	0.389	0.292	0.579	0.664
Casia + CoMoFoD	0.510	0.374	0.623	0.712
CMFD_Anything only (Copy-Move)	<u>0.622</u>	<u>0.590</u>	<u>0.682</u>	<u>0.766</u>
All Combined for Forensim Training	**0.681**	**0.624**	**0.700**	**0.812**

Table 9: All models are evaluated on the CMFD Anything test set with Forensim. We vary the training composition by dataset and manipulation type. Performance improves with diverse forgery exposure. ***Bold - Best****, <u>Underline - Second Best</u>*

FORENSIM ROBUSTNESS ANALYSIS

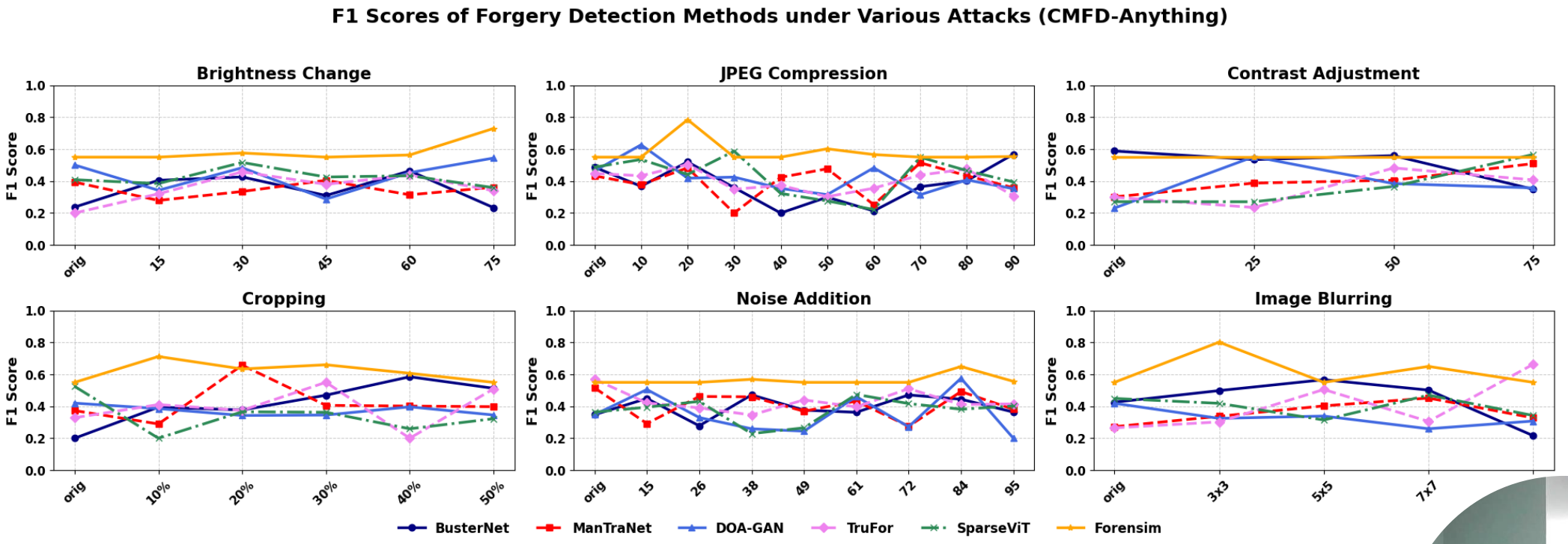


Fig 15: Robustness analysis of Forensim compared to baselines on the CMFD-Anything dataset across six c
perturbations—following the protocol in ManTraNet [CVPR 2019]

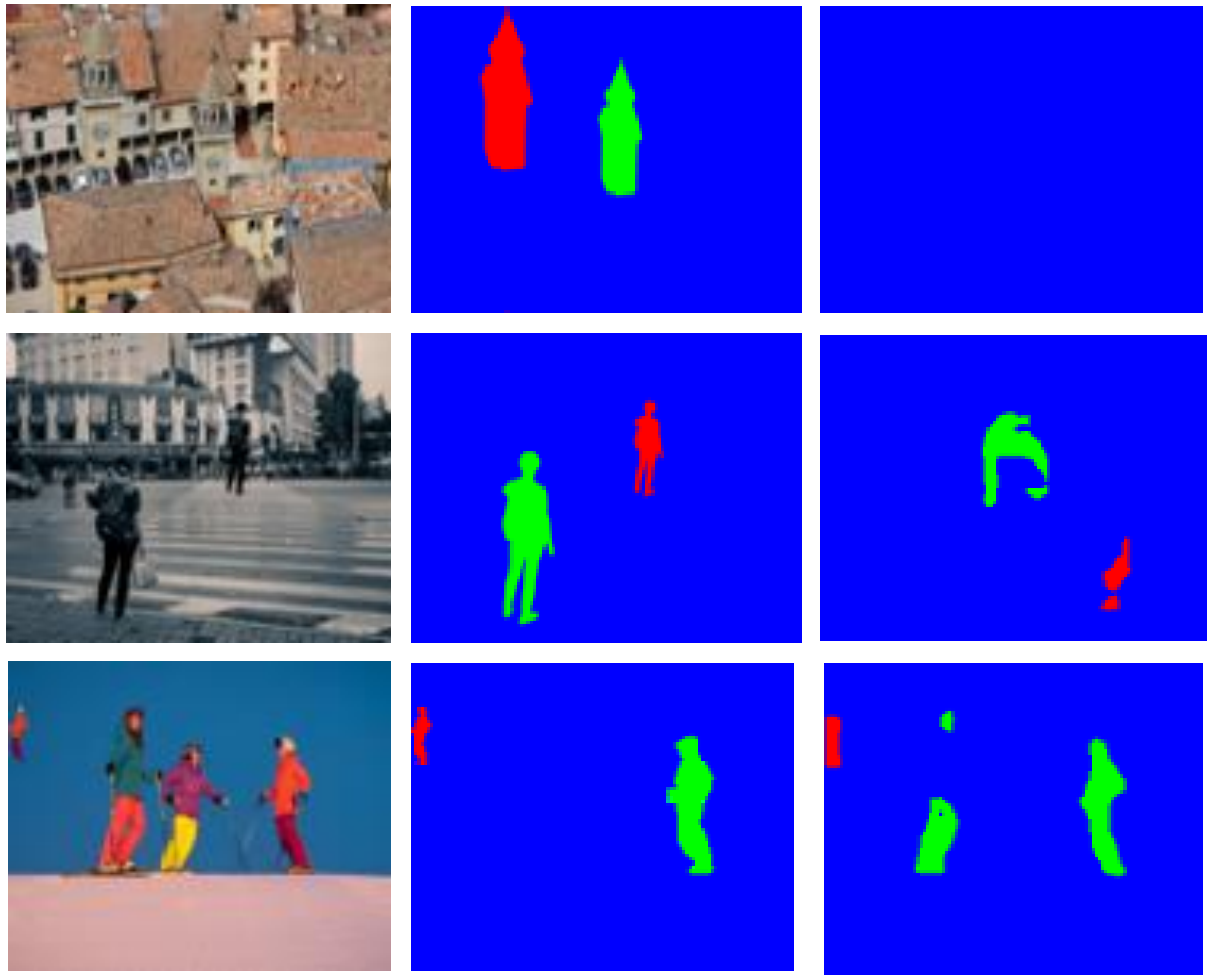
FORENSIM MODEL COMPLEXITY

Method	Conference	Backbone	Input Size	Parameters	FLOPs
BusterNet	ECCV 2018	VGG	256 X 256	<u>15.5 M</u>	45.7 G
ManTraNet	CVPR 2019	VGG	256 X 256	**3.9 M**	274.0 G
DOA_GAN	CVPR 2020	VGG	256 X 256	26.9 M	46.7 G
HiFi-Net	CVPR 2023	HRNet	512 X 512	17.2 M	<u>32.9 G</u>
TruFor	CVPR 2023	ViT	512 X 512	68.7 M	236.5 G
SparseViT	AAAI 2025	ViT	512 X 512	50.3 M	46.2 G
Forensim	WACV 2026	SSM	512 X 512	36.7 M	**28.7 G**

Table 10: Model Complexity with respect to Resolution, Parameters, FLOPS.
Bold - Best, <u>*Underline - Second Best*</u>

FORENSIM LIMITATIONS

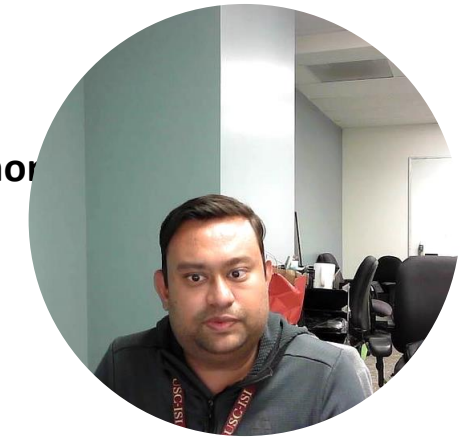
PROBE MASK | GT MASK | PREDICTED MASK



← Background-matching splices where low contrast suppresses $MSSA_{map}$

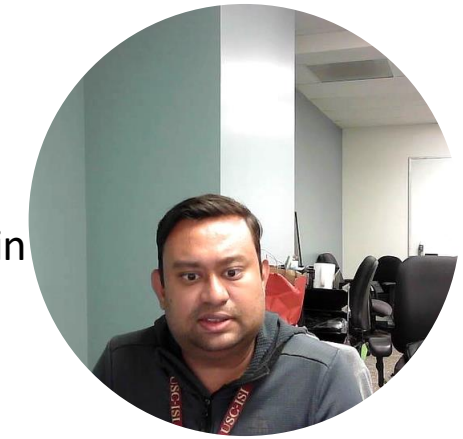
← Repetitive structures causing many-to-many matches in Aff_{final}

← Rare errors on low-color/mor



FORENSIM TAKE AWAY

- **Beyond Copy-Move**
  - Designed Image Forgery Detection **Training** from the perspective of **three class-based images**
  - Forensim demonstrated **robust performance across various types of image forgeries**, including splicing and inpainting
- **End-to-End CMFD Solution**
  - Forensim is a vision state-space attention-based model that detects both **source** and **target** regions in forged images
- **Novel Attention Mechanisms**
  - Introduced **Similarity Attention** and **Manipulation Attention** modules, for efficient and precise forgery detection
- **CMFD-Anything Dataset**
  - Curated a **high-resolution**, **realistic** dataset to address the scarcity of quality training data in CMFD
- **State-of-the-Art Performance**
  - Extensive experiments show Forensim **outperforms existing CMFD methods**, advancing the SOTA in



Questions?

Thank You