NIST Open Media Forensics Challenge

OpenMFC Evaluation Program

Haiying Guan

Yooyoung Lee, Lukas Diduch , and Ilia Ghorbanian

Multimodal Information Group,
Information Access Division , ITL, NIST

OpenMFC2021 Workshop : Day 1, Tuesday, Dec. 7, 2021

Disclaimer

- Certain commercial equipment, instruments, software, or materials are identified in this article in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the equipment, instruments, software or materials are necessarily the best available for the purpose.
- The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.
- All images, graphs, and charts are original works created for DARPA MediFor as well as other NIST Programs.

Acknowledgement

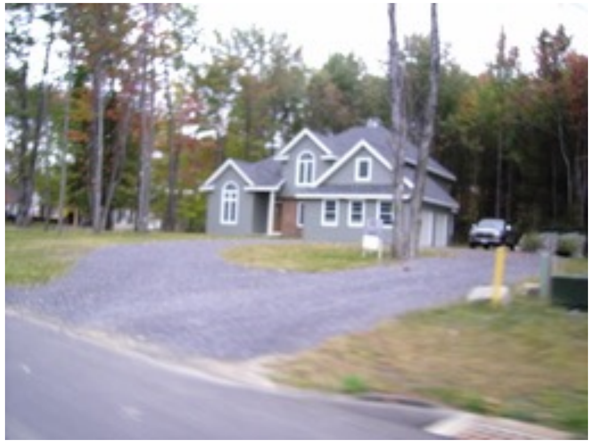
- NIST contributors
  - Jonathan Fiscus
  - Timothee Kheyrkhah
  - Peter Fontana
  - Jesse G. Zhang
- External collaborators:
  - Prof. Siwei Lyu in University at Buffalo
  - Prof. Jennifer Newman, Prof. Yong Guan, Li Lin from Iowa State University, and Prof. Roy A. Maxion from Carnegie Mellon University

OpenMFC Overview Outline

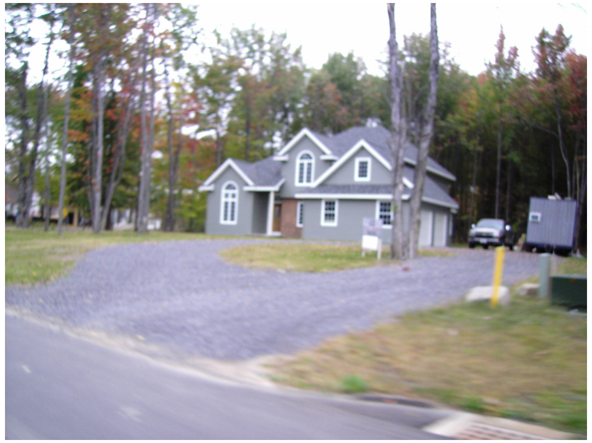
- OpenMFC program overview
  - What, why, who, how
- OpenMFC evaluation design
  - Design challenges, evaluation pipeline, evaluation tasks, evaluation metrics
- OpenMFC 2020-2021 datasets
  - Development dataset, evaluation dataset

Media Forensics: What

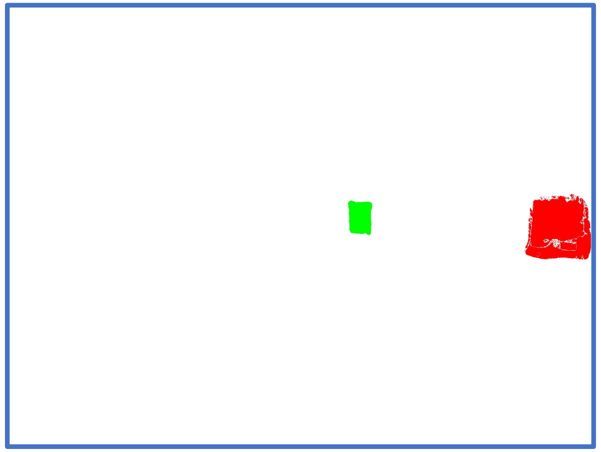
- What is media forensics?
  - “Media Forensic is scientific study into the collection, analysis, interpretation, and presentation of audio, video, and image evidence obtained during the course of investigations and litigious proceedings.” [1]



a. Test Image



b. Original Image

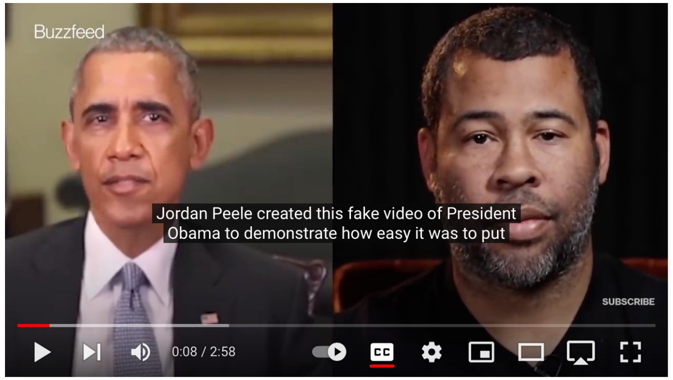


c. Manipulated Mask

[1] https://artsandmedia.ucdenver.edu/areas-of-study/national-center-for-media-forensics/about-the-national-center-for-media-forensics

Media Forensics: Why

- "Seeing is not believing"[4]
  - Excellent image editing software
    - Adobe CC, GIMP, Corel Paintshop Pro, Skylum Luminar, DxO PhotoLab, ON1 Photo RAW, ACDSee Photo Studio Ultimate, Pixlr Editor, Canva, PicMonkey, Snappa, PortraitPro, Fotor, …
  - New advanced technologies
    - Generative Adversarial Network (GAN), Deepfakes[1], CGI, and anti-forensics techniques, …
- Applications
  - Fake news detection in social media platform[2,3]
    - Facebook, Twitter, Instagram, Snapchat, and Google
  - Misinformation or disinformation
  - Academic misconduct[5]
  - Criminal law and private investigation, Security [6]
  - Trustworthiness of media content – authentication


Buzzfeed

Jordan Peele created this fake video of President Obama to demonstrate how easy it was to put

Figure Source: Bloomberg Quicktake, 2018 [1]

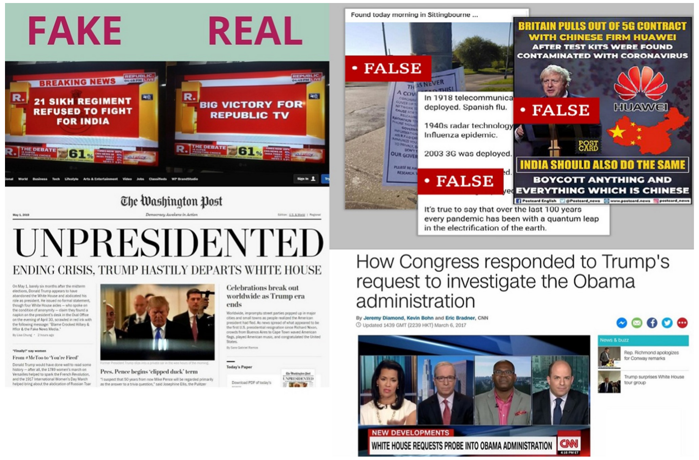

FAKE
REAL
21 SIKH REGIMENT REFUSED TO FIGHT FOR INDIA
BIG VICTORY FOR REPUBLIC TV
The Washington Post
UNPRESIDENTED
ENDING CRISIS, TRUMP HASTILY DEPARTS WHITE HOUSE
• FALSE
• FALSE
• FALSE
BRITAIN PULLS OUT OF 5G CONTRACT WITH CHINESE FIRM HUAWEI AFTER TEST KITS WERE FOUND CONTAMINATED WITH CORONAVIRUS
INDIA SHOULD ALSO DO THE SAME
BOYCOTT ANYTHING AND EVERYTHING WHICH IS CHINESE
How Congress responded to Trump's request to investigate the Obama administration
WHITE HOUSE REQUESTS PROBE INTO OBAMA ADMINISTRATION

Figure Source: EchoFakeD, 2021 [2]

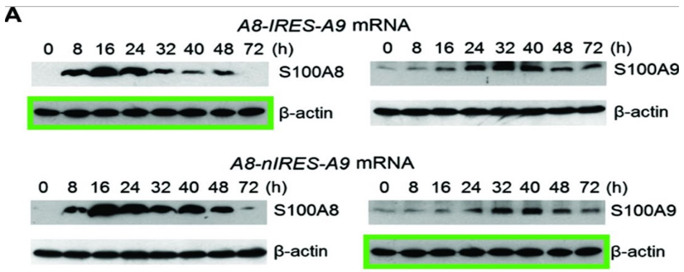

A
A8-IRES-A9 mRNA
0 8 16 24 32 40 48 72 (h)
S100A8
β-actin
0 8 16 24 32 40 48 72 (h)
S100A9
β-actin
A8-nIRES-A9 mRNA
0 8 16 24 32 40 48 72 (h)
S100A8
β-actin
0 8 16 24 32 40 48 72 (h)
S100A9
β-actin

Figure Source: Academic misconduct [5]

[1] https://www.youtube.com/watch?v=gLoI9hAX9dw

[2] https://doi.org/10.1007/s00521-020-05611-1

[3] https://www.computer.org/publications/tech-news/research/social-media-verification-assistant

[4] H. Farid, "Seeing is not believing," in *IEEE Spectrum*, vol. 46, no. 8, pp. 44-51, August 2009, doi: 10.1109/MSPEC.2009.5186556.

[5] https://mbio.asm.org/content/7/3/e00809-16

[6] https://www.cyberscoop.com/social-media-disinformation-represents-security-thr

Media Forensics: Challenges (1)

- What is the difference between media forensics and other research topics?
    - Intrinsic property: No single technology fits all
    - Open topics: manipulations emerging and change all the time
        - Fast evolution
        - Dynamic upgraded
    - Forensics vs. Anti-forensics
        - No traditional steady improvement curve
        - Prediction: who is going to win in the end?
    - Semantics instead of facts

Media Forensics Analytics Technologies: Incomplete Survey

- Copy Move
- Geometric-based Cropping Detector
- Lighting
  - Gradient-Based Illumination Description
  - Light inconsistency on faces
- Face-based tamper detection
  - Facial Expression
- Pixel-based tamper detection
  - JPEG Compression Detection
  - JPEG Dimples
  - Noiseprint
  - Resampling Anomaly
- Color Phenomenology
- Holistic approaches
- Splice detection
- GAN/Deepfakes/AI-Synthesized detection
- Inconsistency detection
  - Audio visual speaker identity inconsistency
  - Audio visual lip out of synchrony
  - Audio visual scene inconsistency detector
  - Light inconsistency
  - Weather / location
  - Codec inconsistency
  - Noise inconsistency
- Video
  - Frame duplication
  - Frame drop
- ENF (Electric Network Frequency)
- Rebroadcast
- Camera verification
- Provenance filtering and graph building

Media Forensics: Challenges (2)

- What is the difference between media forensics and other research topics?
  - Intrinsic property: No single technology fits all
  - Open topic: manipulations emerging and change all the time
    - Fast evolution
    - Dynamic upgraded
  - Forensics vs. Anti-forensics
    - No traditional steady improvement curve
    - Prediction: who is going to win in the end?
  - Semantics instead of facts

OpenMFC Program (1)

- OpenMFC is an evaluation series open to public participants worldwide to support media forensics research and help advance the state-of-the-art imagery (image and video) forensics technologies.
  - Benchmark evaluation is more convincing than self-evaluation
  - Unbiased, neutral position
  - Media forensics analytics is very challenging
  - Keep up with the emerging technologies (GAN, Deepfakes, CGI, anti-forensics, …)
  - A platform for collaboration
- Open evaluation
  - OpenMFC is open to public, seeks to provide the datasets, build a research platform and inspire research worldwide with leaderboard evaluation, and advance the state-of-the-art of media forensics.
- Closed evaluation
  - DARPA MediFor (https://www.darpa.mil/program/media-forensics)
  - DARPA SemaFor (https://www.darpa.mil/program/semantic-forensics)

OpenMFC Program (2)

- The NIST OpenMFC is open worldwide. We invite all organizations including past DARPA MediFor Program participants and current DARPA SemaFor Program participants.
- Participation is free. NIST does not provide funds to participants
- NIST team - design the tasks, provides the data, online evaluation platform (real-time report)
- Participant
  - Registration: to register on the website and complete the data license to download the data.
  - Development: to develop the media forensics algorithm/systems.
  - Test: once your system is functional you will be able to upload your outputs to the challenge website and see your results displayed on the leaderboard.

OpenMFC Program: Objective

- OpenMFC Objective
  - Understand the state-of-the-art performance
  - Provide continuously convincing reports: cross-year comparisons
  - Stimulate/Promote the research in media forensics – system performance analysis
  - Help researchers with system performance analysis to improve their system
  - Bridge the gap between lab-report algorithm performance and in-the-field application performance
- Strategy
  - Collaboration instead of competition: nature of the media forensics
  - Group the media forensics researchers and build a strong connected community

Evaluation Design Challenges

- What to evaluate?
  - Not too easy and not too hard
  - Technical methodology varieties brings big challenges in design a unified evaluation framework
- What resources to use?
  - Hundreds if not thousands of manipulation methods
  - Lack of benchmark datasets: human post-annotation doesn’t work well.
  - Different technologies need different evaluation data
- What we can get from the evaluation?
  - Baseline performance information
  - State-of-the-art, cross-year performance comparison
- How to evaluate?
  - How to handle the dynamic changes of forensic and anti-forensic technologies?
  - How to adapt the changes and provide the evaluation report in time?

Evaluation Task Design Strategy

- MFC task design

Single File Authenticity

Manipulation Detection:

Is the image/video manipulated?

Localization:

Where is the image/video manipulated?

- Spatial
- Temporal
- Temporal-spatial

Authenticity in Context

Image Pair Authenticity

Splice Detection:

Does image1 contain some of image2?

Localization:

- Where in image1 was image2 content spliced?
- Where in image2 is the splice donor?

Image+ Image Collection

Provenance Filtering:

Find related images

Provenance Graph Building:

Construct a phylogeny graph of related images

File+Camera

Camera Verification:

Was an image/video taken by a known camera?

File+Event

Event Verification:

Was an image capture during a known event?

- OpenMFC

single input detection => pair input verification => multi-input analytics integrity

OpenMFC20 Evaluation Tasks

- Image Manipulation Detection and Localization (IMDL)
  - To detect if the image has been manipulated, and then to spatially localize the manipulated region
- Video Manipulation Detection (VMD)
  - To detect if the video has been manipulated
- Image GAN Manipulation Detection (IGMD)
  - To detect GAN-manipulated images (e.g., created by a GAN model, locally/globally modified by a GAN filter/operation, etc.).
- Video GAN/Deepfakes Manipulation Detection (VGMD)
  - To detect GAN/Deepfakes manipulated videos.

OpenMFC20 Evaluation Conditions

- Image Manipulation Detection and Localization (IMDL)
  - Conditions: Image Only (IO) and Image and Metadata (IM)
- Video Manipulation Detection (VMD)
  - Conditions: Video Only (VO) and Video and Metadata (VM)
- Image GAN Manipulation Detection and Localization (IGMDL)
  - Conditions: Image Only (IO)
- Video GAN Manipulation Detection (VGMD)
  - Conditions: Video Only (VO)

Image Manipulation Detection and Localization (IMDL)

System Input

Image(s) + (Metadata)

Probe image

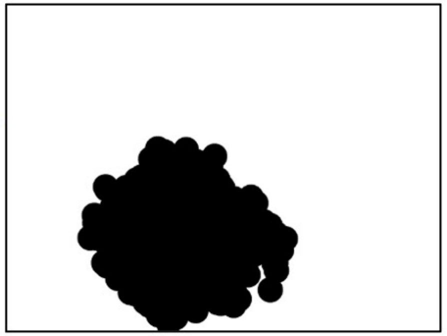
Image Detection and Localization Analytic System

System Output

Confidence score

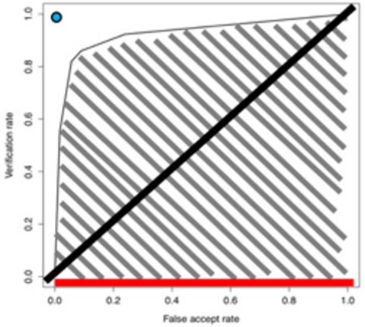
97.86

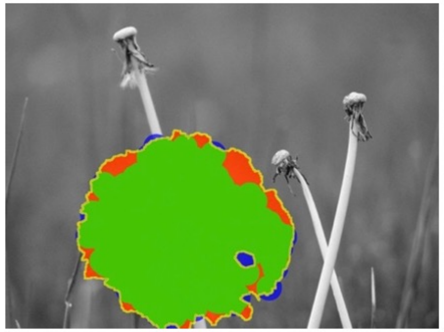
System output mask



Metrics

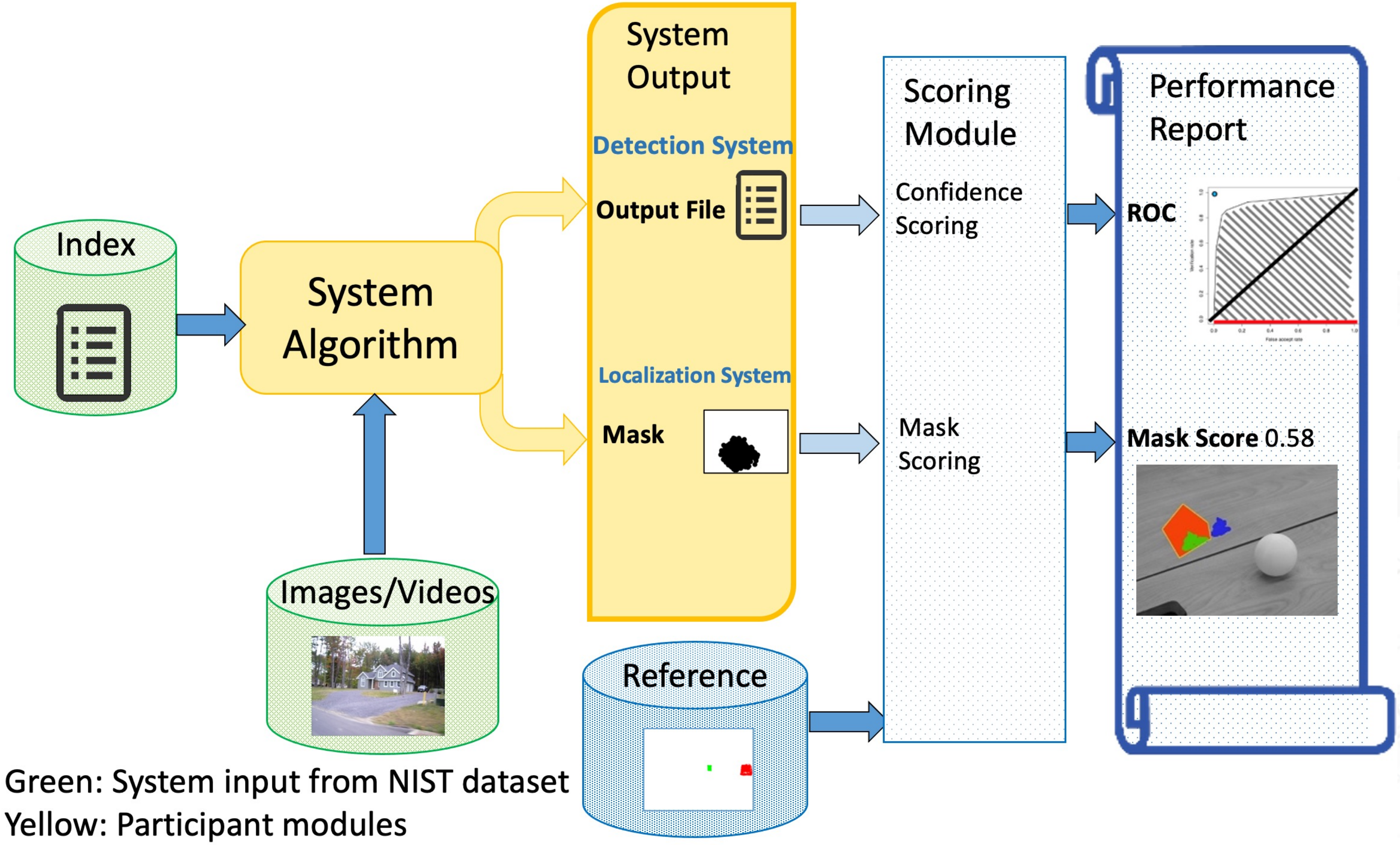
Receiver Operating Characteristic (ROC)
Area Under the Curve (AUC)
Correct Detection (CD) at False Alarm Rate 5%





Manipulated image
Matthews Correlation Coefficient (MCC)

OpenMFC: How to evaluate (1)

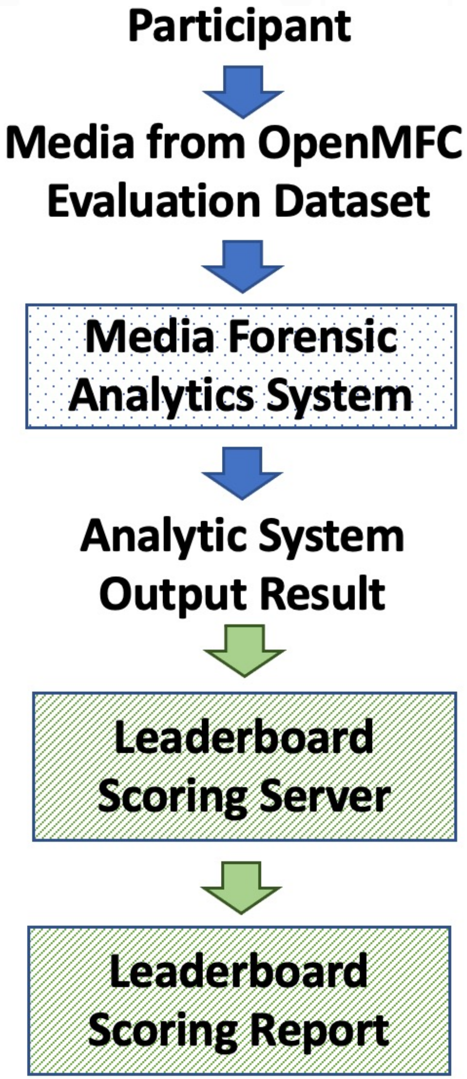


Green: System input from NIST dataset
Yellow: Participant modules
Blue: NIST Evaluation modules

Figure: OpenMFC Evaluation Pipeline

OpenMFC: How to evaluate (2)

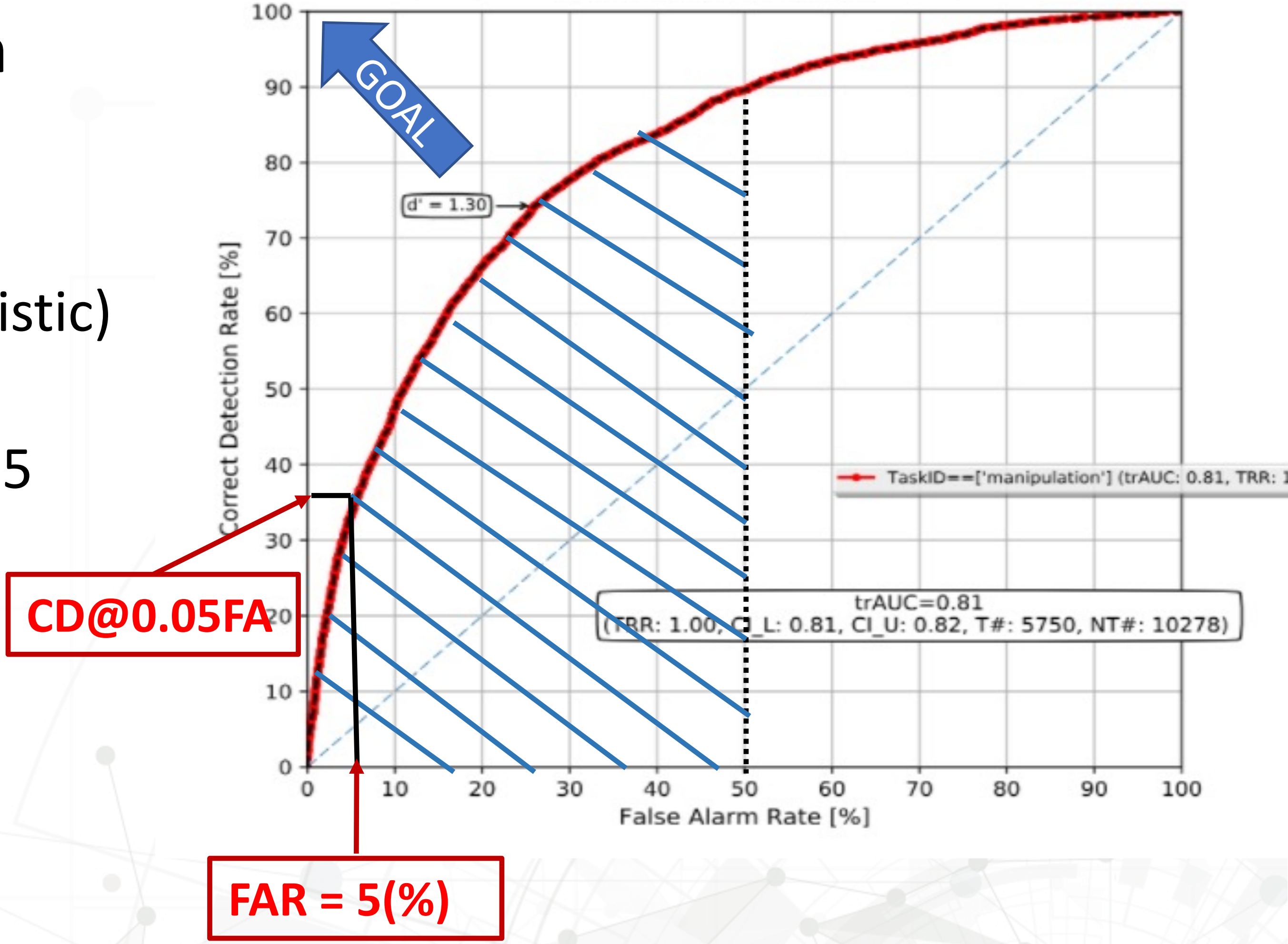
- OpenMFC Take-home evaluation
  - NIST releases test data to participants
  - Participant submits system output
  - Evaluation website provides a leaderboard to report evaluation results



Participant
Media from OpenMFC Evaluation Dataset
Media Forensic Analytics System
Analytic System Output Result
Leaderboard Scoring Server
Leaderboard Scoring Report

Detection System Evaluation Metrics

- Evaluate the accuracy of a system output (e.g., confidence score)
- Evaluation metrics
  - ROC (Receiver Operating Characteristic)
  - AUC (Area Under Curve)
  - CD (Correct Detection) @ FAR = 0.05



CD@0.05FA

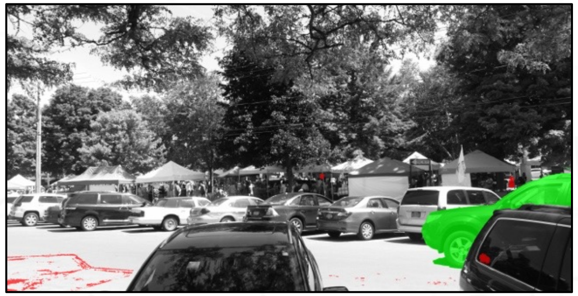
FAR = 5(%)

Localization System Evaluation Metrics

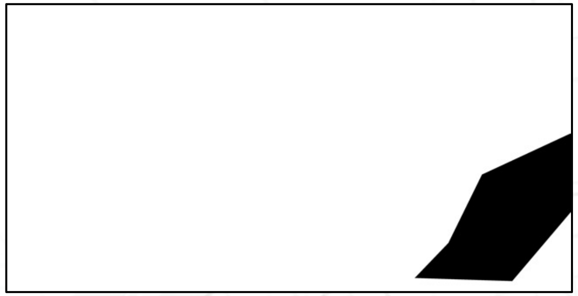
- Metrics
  - Matthews Correlation Coefficient (MCC)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)\cdot(TP+FN)\cdot(TN+FP)\cdot(TN+FN)}} \in [-1,1]$$

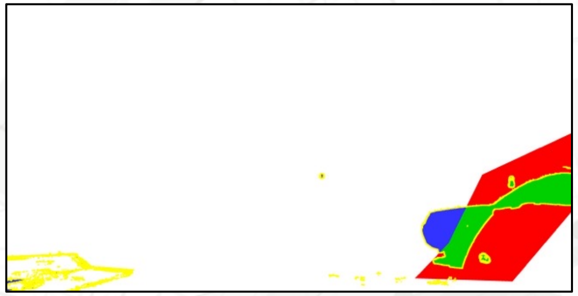
    - 1 denotes perfect accuracy
    - 0 denotes no correlation
    - -1 denotes perfect inaccuracy.
- Optimum MCC
  - The MCC at the optimum grey-scale mask threshold
- Only evaluates on true targets



Probe + ref. mask overlay



System output mask



Color-coded scoring confusion matrix

OpenMFC 2020-2021 Evaluation Data

- Evaluation dataset design challenges
- Manipulation reference collection and annotation
- What information you can obtain from the reference data
- Evaluation data production
- OpenMFC evaluation dataset summary

Evaluation Dataset Design Challenges

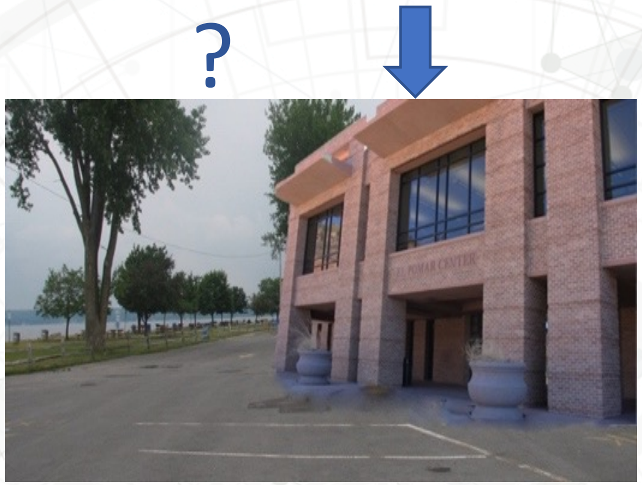
- Evaluation objective
  - Lab algorithm evaluation vs. in-the-field evaluation (real-world applications)
- Media forensics analytics is an open task:
  - emerging software and tools (GAN, Deepfakes, CGI, etc.)
  - Anti-forensics
- Curse of dimensionality
  - Large testing space
    - Media space (image/video/audio, camera/scanner)
    - Manipulation space (manipulator, manipulation operations and software)
    - Anti-forensic technology space
  - The combinatorics of one dimension
    - Suppose a 2-Factorial, single operation experimental design
      - 17,500 images = 70 Operations * 2 levels * 125 examples
      - Not realistic (manipulators routinely stack manipulations)
    - The average graph depth in MFC19 was ~4
      - $6.0*10^9$ images = 70^4 Operations * 2 levels * 125 examples

Manipulation Reference Collection Challenges

- What reference ground-truth data to collect
  - Time, labor, cost, usage, value
- Effective evaluations require knowledge:
  - If the media was manipulated
  - What editing tool was used
  - Who did the manipulation
  - What is the original media
  - What operations were used
  - How the media was manipulated
  - Where the manipulation occurred
  - Semantics of the manipulation: malicious vs. benign
- How
  - Post manipulation interpretation is nearly impossible

Manipulation Reference Collection Approaches

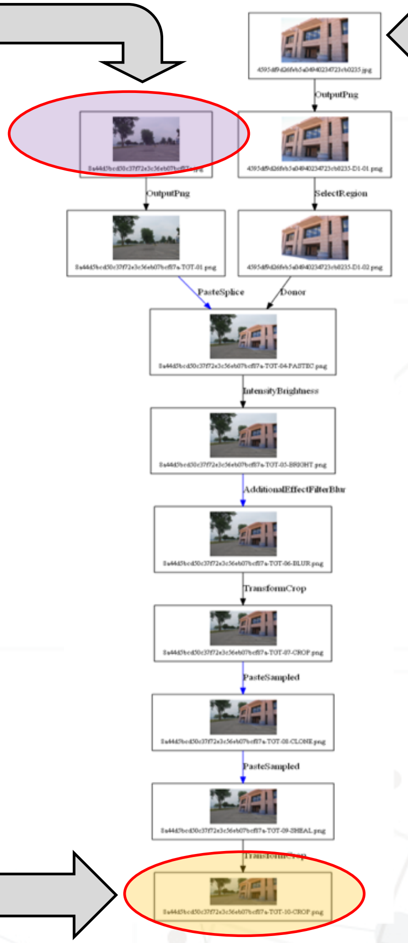
- Separate analytics team from the data generation team
  - Drive the analytics teams from their comfortable zone
- Human + machine for data collection and production
  - Human manipulation (realistic)
  - Automatic manipulation (reduce cost)
  - Extended manipulation (special study)
- Journaling Tools (collaborated with PAR Government Systems)
  - Manipulation journaling tool (JT)
    - Record the manipulation step by step with a graph
    - Automate collection of manipulation region mask
  - Automatic journaling tool (Auto-JT)
  - Extended journaling tool (Extended-JT)
- Real-world simulation: social media laundering (UC. Denver)



Manipulation Journaling Tool

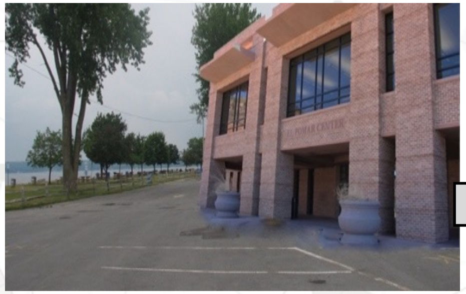
High Provenance

Base Image



Donor Image

Unknown Provenance



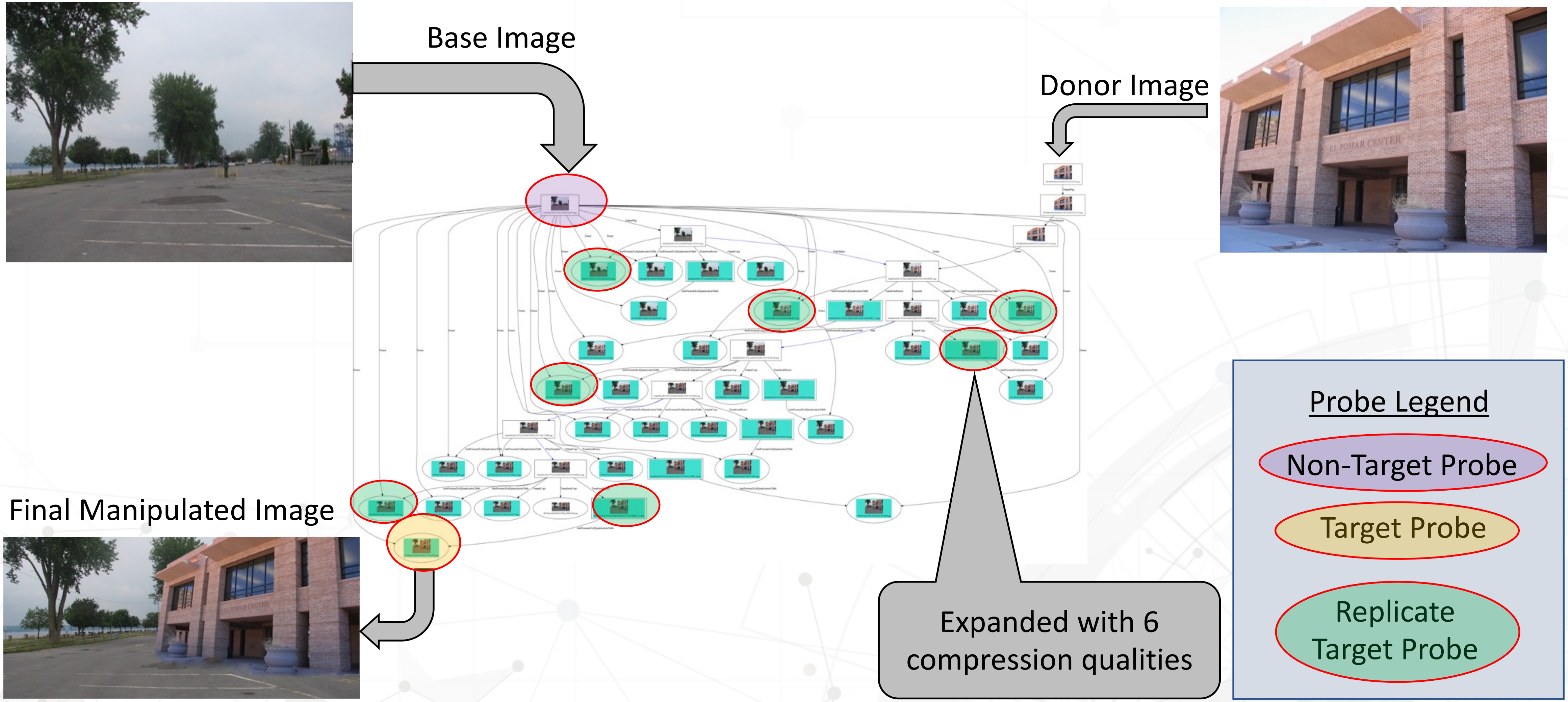
Final Manipulated Image

Probe Legend

Non-Target Probe

Target Probe

Manipulation Journaling Tool (Extended Journal)



Base Image

Donor Image

Final Manipulated Image

Expanded with 6 compression qualities

Probe Legend

Non-Target Probe

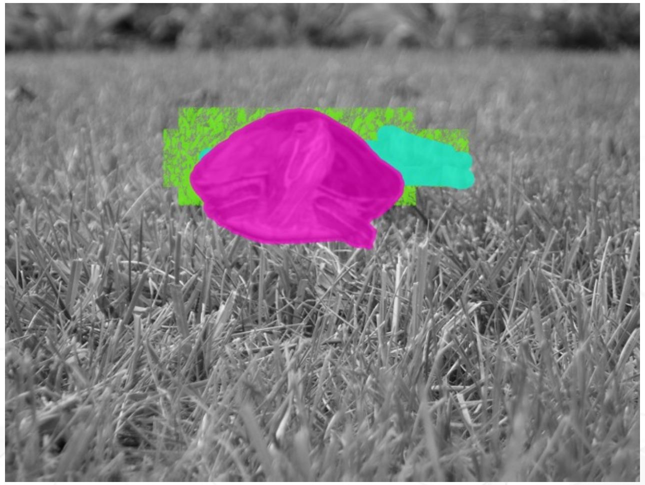
Target Probe

Replicate Target Probe

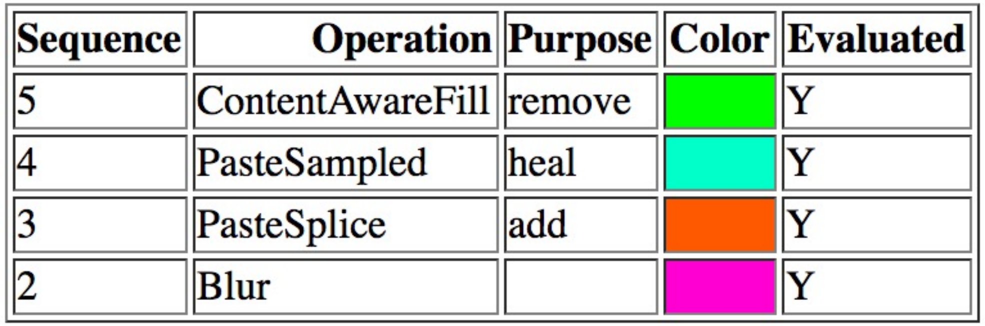
Image Mask for Manipulation Localization

- NC16, NC17 – single layer composite mask
- After MFC18 – multi-layer JPEG 2000 mask
  - Distinct manipulations are recorded in the different layers in JPEG 2000 mask file respectively.
  - Each bit in a byte for a pixel in a single-layer image represents one localizable manipulation.
  - Scoring can thus be extended to specific localizable manipulations in the image.

Manipulated Probe image



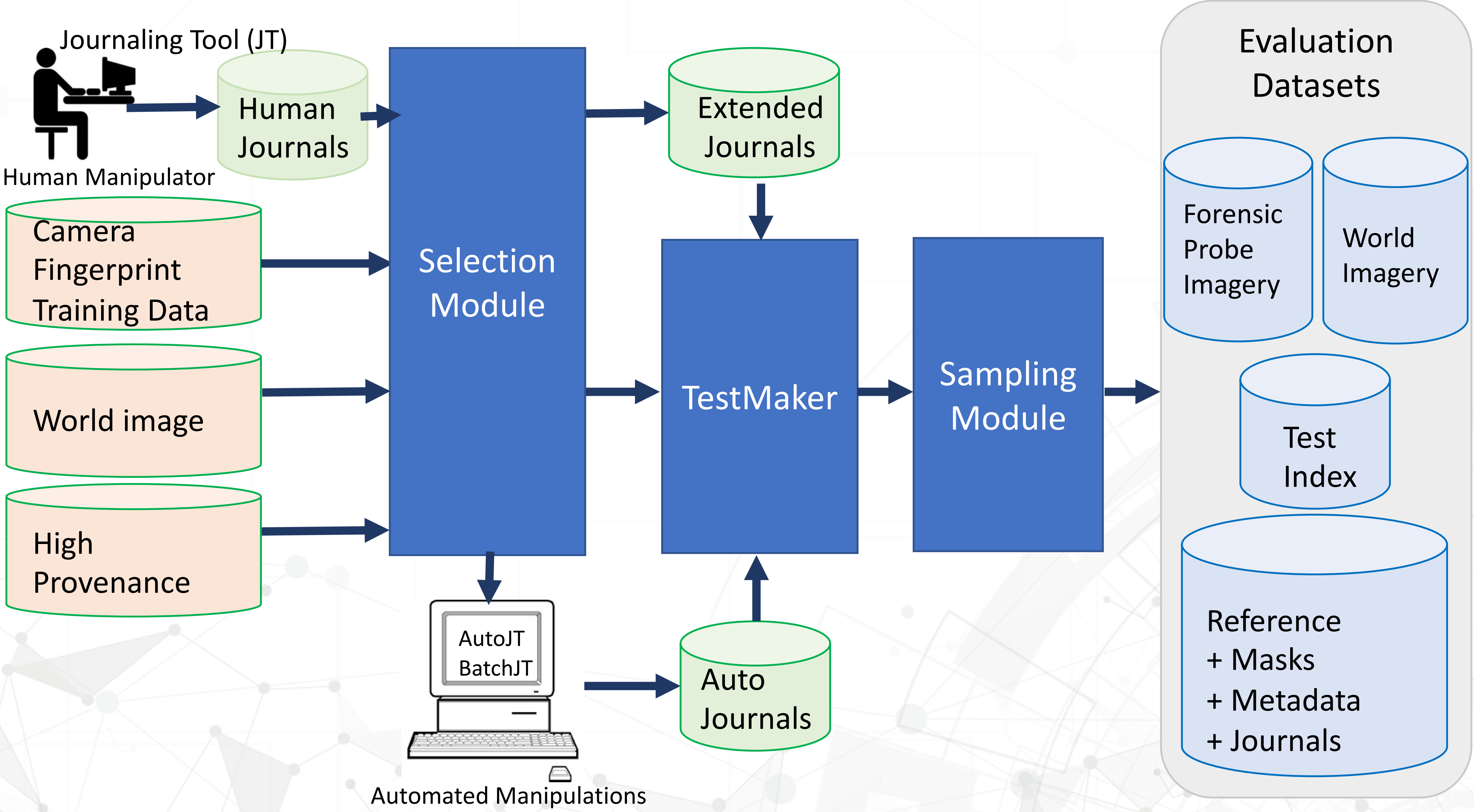
Composite mask



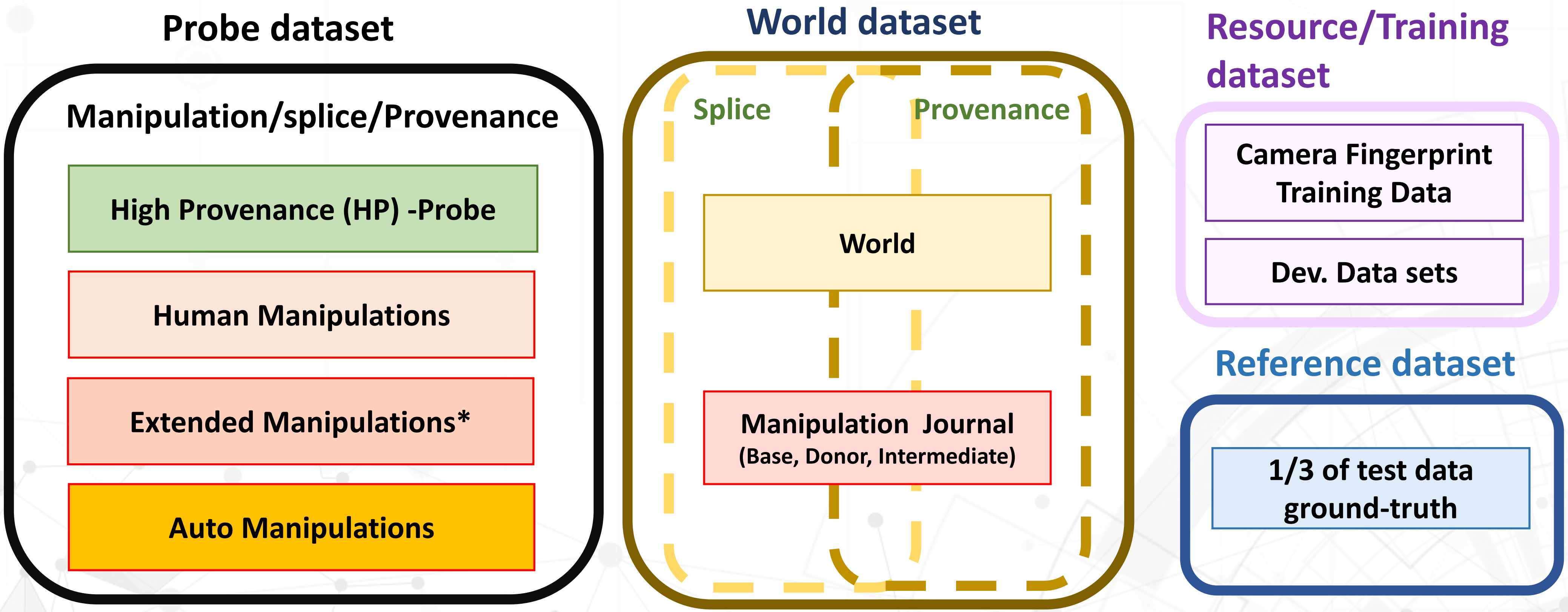
Sequence	Operation	Purpose	Color	Evaluated
5	ContentAwareFill	remove	🟩	Y
4	PasteSampled	heal	🟦	Y
3	PasteSplice	add	🟧	Y
2	Blur		🟪	Y

An animated representation of the information stored by the JPEG2000. Every region is fully represented. The sequence is listed in descending order for node distance from the manipulated probe and may be distinct from the bit placement in the byte.

Evaluation Dataset Production Infrastructure



Data Collection and Evaluation Dataset Overview



Probe dataset

Manipulation/splice/Provenance

High Provenance (HP) -Probe

Human Manipulations

Extended Manipulations*

Auto Manipulations

World dataset

Splice

Provenance

World

Manipulation Journal
(Base, Donor, Intermediate)

Resource/Training dataset

Camera Fingerprint Training Data

Dev. Data sets

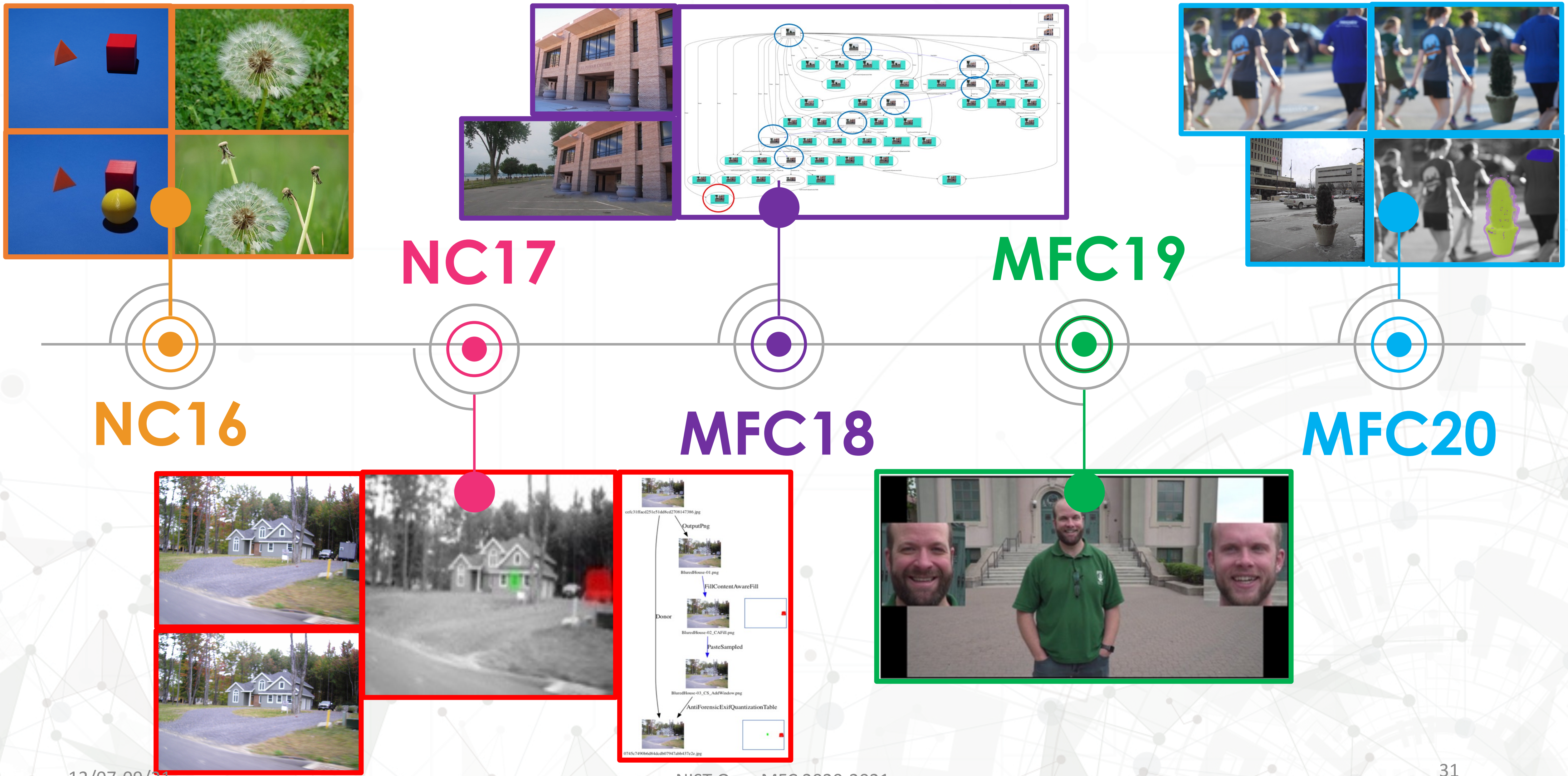
Reference dataset

1/3 of test data ground-truth

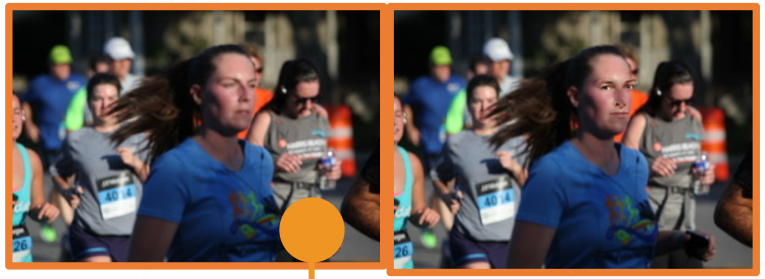
*Extended manipulations: the manipulated images generated by automatic machine manipulation tools in the extended journal described in the previous slide.

l images generated by automatic

OpenMFC Evaluation Dataset (1)

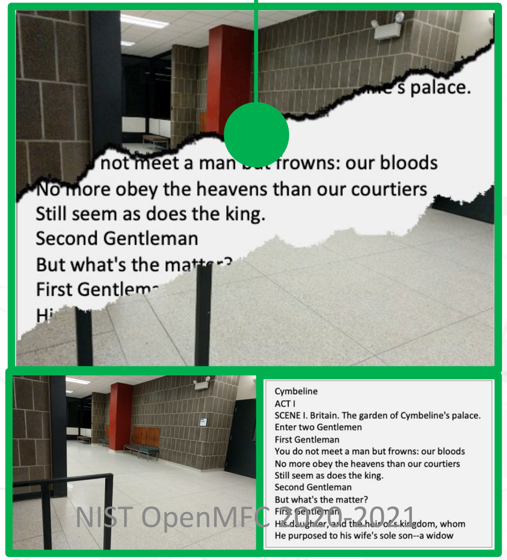

NC16
NC17
MFC18
MFC19
MFC20
OutputPng
BluredHouse-01.png
FillContentAwareFill
Donor
BluredHouse-02_CAFill.png
PasteSampled
BluredHouse-03_CS_AddWindow.png
AntiForensicExifQuantizationTable

OpenMFC Evaluation Dataset (2)

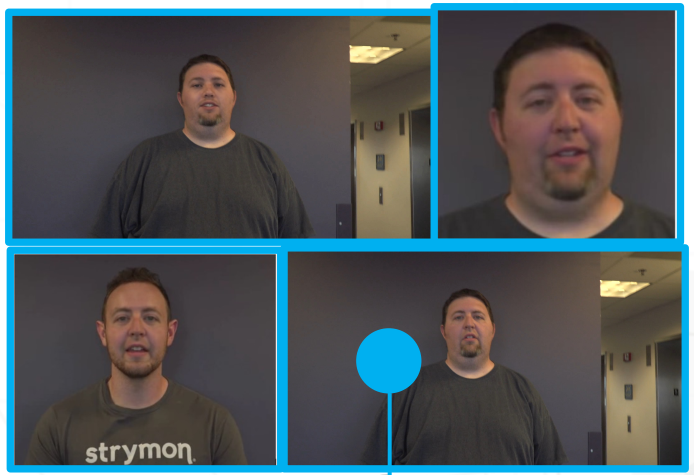


OpenMFC 2022 GAN Image[1]

OpenMFC 2022 StegD[2]


...'s palace.
not meet a man but frowns: our bloods
No more obey the heavens than our courtiers
Still seem as does the king.
Second Gentleman
But what's the matter?
First Gentleman
Hi...

Cymbeline
ACT I
SCENE I. Britain. The garden of Cymbeline's palace.
Enter two Gentlemen
First Gentleman
You do not meet a man but frowns: our bloods
No more obey the heavens than our courtiers
Still seem as does the king.
Second Gentleman
But what's the matter?
First Gentleman
His daughter, and the heir of's kingdom, whom
He purposed to his wife's sole son--a widow



OpenMFC 2022 Deepfake Video[1]

[1] Collaboration with Prof. Siwei Lyu's team

[2] Collaboration with Prof. Jennifer Newman's team

Large-Scale Benchmark Datasets

- Designed and built over 31 evaluation datasets and 13 development datasets to support 7 evaluation tasks and 2 evaluation challenges in MFC.
- Released the public datasets to over 600 individuals from over 200 organizations and 26 countries and regions worldwide (the number is continuously increasing).

Table: NIST released MFC datasets

(Highlighted: the evaluation sets; Grayed: the development sets)

MFC Released Dataset	Media Type	Dev./Eval.	Media Number	Journal #	Create Date
Kick Off (NC16) Image	Image	Dev.	1.1K	400	Jul-16
NC17 Dev Image	Image	Dev.	3.5K	398	Mar-17
MFC18 Dev1 Image	Image	Dev.	5.6K	197	Jan-18
MFC18 Dev2 Image	Image	Dev.	38K	411	Feb-18
NC17 EP1 Image	Image	Eval.	4K	406	Jun-17
MFC18 EP1 Image	Image	Eval.	17K	758	Mar-18
MFC19 EP1 Image	Image	Eval.	16K	1383	Mar-19
MFC18 GAN FULL Image	Image	Eval.	1.3K	267	Apr-18
NC17 Dev Video	Video	Dev.	212	25	Mar-17
MFC18 Dev1 Video	Video	Dev.	116	9	Jan-18
MFC18 Dev2 Video	Video	Dev.	231	21	Feb-18
NC17 EP1 Video	Video	Eval.	360	34	Jun-17
MFC18 EP1 Video	Video	Eval.	1028	114	Mar-18
MFC19 EP1 Video	Video	Eval.	1530	163	Mar-19
MFC18 GAN Video	Video	Eval.	118	19	Jun-18

NIST OpenMFC Resources: https://mfc.nist.gov

- MFC open evaluation datasets
  - Signup: NC16 Kickoff,
  - Signed two agreements:
    - NC17 Evaluation Part 1 (EP1),
    - MFC18 EP1,
  - OpenMFC 2020-2021 Evaluation dataset
    - MFC19 EP1 without ground-truth
- NIST OpenMFC leaderboard scoring server
  - **https://mfc.nist.gov**
- MediScore
  - Github: https://github.com/usnistgov/MediScore/
- Slack – open for public researcher
  - http://openmfc.slack.com
  - Discussion channel: https://app.slack.com/client/T017MTH6RHT/C017MTH7LRK
- Participant Google group – OpenMFC performer only
  - openmfc-performer
  - Mailing list: openmfc-performer@list.nist.gov

Takeaway

- Media Forensics is still in the early stage
- Media Forensics intrinsically is different from other traditional research topics
- OpenMFC tasks
- OpenMFC datasets
- OpenMFC Online website: **https://mfc.nist.gov**
- **Join the OpenMFC program!**

Questions?

OpenMFC team: mfc_poc@nist.gov