# NIST Open Media Forensics Challenge
## (OpenMFC)

## Presented by Haiying Guan

NIST Team: Haiying Guan, Ilia Ghorbanian*, Lukas Diduch^, Yooyoung Lee, Baptiste Chocot+, and Christopher Tu#

Multimodal Information Group (MIG), Information Access Division(IAD),
Information Technology Laboratory (ITL), National Institute of Standards and Technology (NIST)

*PREP Student, ^ Software Engineer at Dakota Consulting, Inc., + Foreign Guest Researcher, # Volunteer Student

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

# Disclaimer

Certain equipment, instruments, software, or materials are identified in this presentation in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement of any product or service by NIST, nor is it intended to imply that the materials or equipment identified 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.

If not specified in the figure title, all images, graphs, and charts are original works created for NIST OpenMFC program or DARPA MediFor Program under NIST IRB #ITL- 0018.

# Outline

What

Our Mission

How

Insights

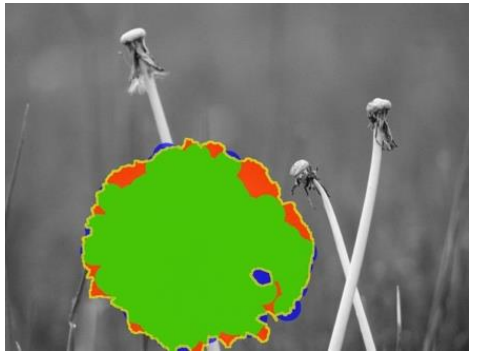
Ongoing Effort

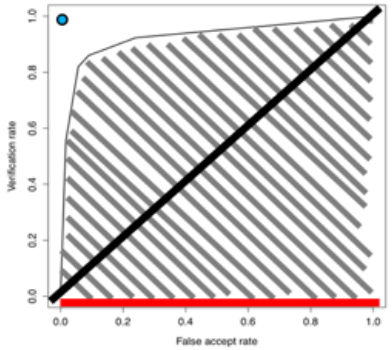
Vision/Plan

# What is OpenMFC?

**Media Forensic Systems**

**System Performance Reports**






| RANK | SUBMISSION ID | SUBMISSION DATE | TEAM NAME | SYSTEM NAME | AUC | CDR@0.05FAR |
|---|---|---|---|---|---|---|
| 1 | 90 | 2021-07-10 09:56:14 | TEAM_ID_1 | test | 0.689716 | 0.207018 |
| 2 | 93 | 2021-07-30 18:13:11 | TEAM_ID_1 | test | 0.683956 | 0.187135 |
| 3 | 75 | 2021-06-14 04:12:30 | TEAM_ID_2 | dry-run | 0.554261 | 0.012865 |
| 4 | 52 | 2021-06-01 15:41:20 | TEAM_ID_2 | dry-run | 0.547125 | 0.009357 |
| 5 | 82 | 2021-06-23 09:21:26 | TEAM_ID_1 | System_ID_1 | 0.500033 | 0.051077 |

OpenMFC is a public evaluation program that invites participants worldwide to contribute to media forensics research and advance state-of-the-art forensic technologies for image and video analysis.

# Why Are We Building OpenMFC?

- Generative and manipulated media are ubiquitous
- “Seeing is not believing”
- Impact Areas:
  - Academic misconduct
  - Journalism integrity
  - Criminal and terrorism investigation
  - Geospatial intelligence
  - Cybersecurity
- Benchmark evaluations support forensic researchers in developing advanced technologies

# Our Mission

**Self-report**

**vs.**

**External report**

- System performance is highly sensitive to data and task variations.
- Self-evaluation reports may not accurately reflect real-world performance.
- Neutral and unbiased benchmark evaluations are essential for reliable assessment.

**One-time eval.**

**vs.**

**Recurring eval.**

- Media manipulation and generation technologies are continuously evolving.
- Recurring evaluations are more effective than one-time assessments for tracking technological advancement.

# OpenMFC Goals and Challenges

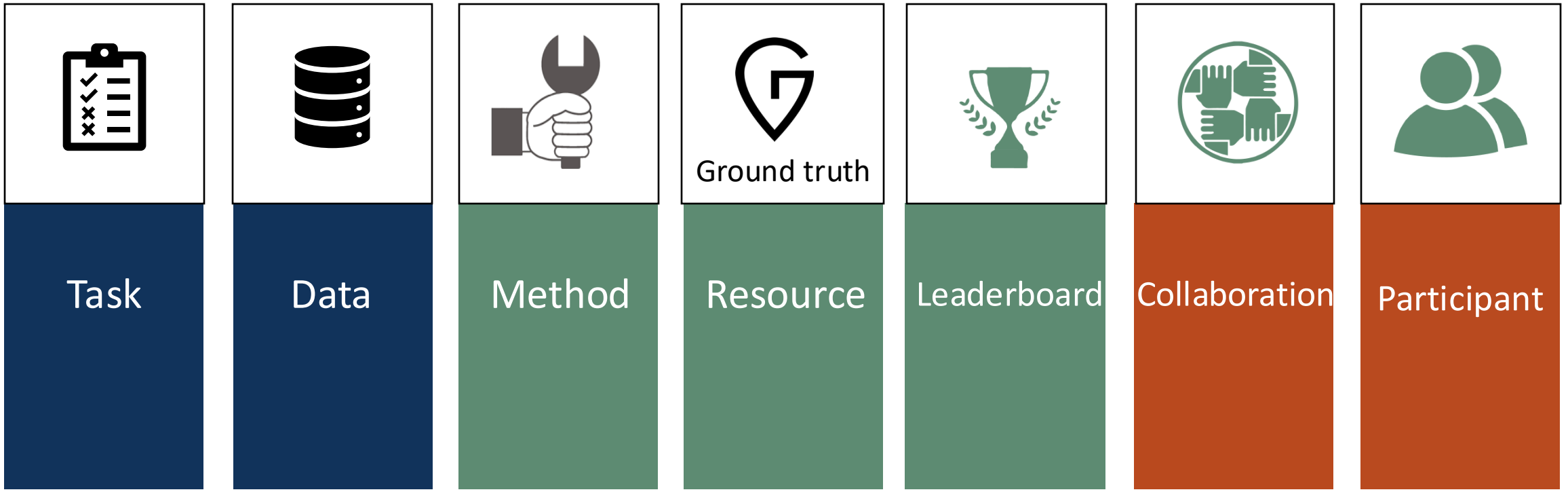
**Goals**

- Foster research in media forensics.
- Gain insights into state-of-the-art techniques through recurring evaluations.
- Provide continuous, year-over-year comparison reports.
- Assist researchers in improving their systems through detailed performance analysis.
- Support the transition from lab-based research to real-world field evaluations.

**Challenges**

- **Media forensics vs. Anti-forensics:** A non-linear trajectory in technological advancements.
- **Target drift:** Rapid evolution of technologies (e.g., GANs, deepfakes, stable diffusion, DALL·E 2, etc.)
- **Data Challenges:** Issues with availability, scalability, privacy, and high resource demands.

# OpenMFC Evaluation Design



# Task Design Strategy

## Detection

Identifying whether a given piece of media has been manipulated or synthesized

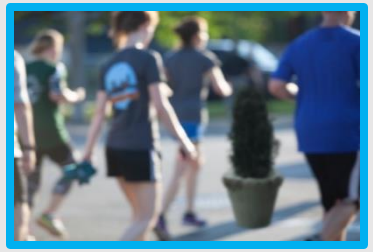
Y

?

N

## Verification

Determining whether a given claim is supported by available evidence

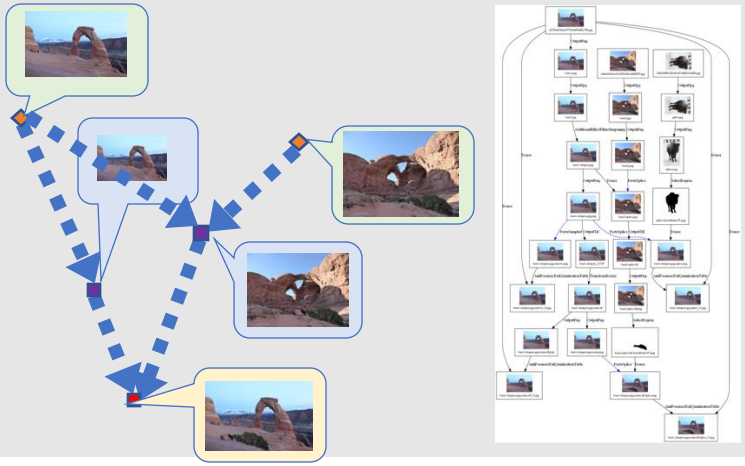


?

Boston Marathon

## Association

Establishing relationships or connections between different elements



## Deepfake Detection

## Event Verification

## Provenance Graph Building

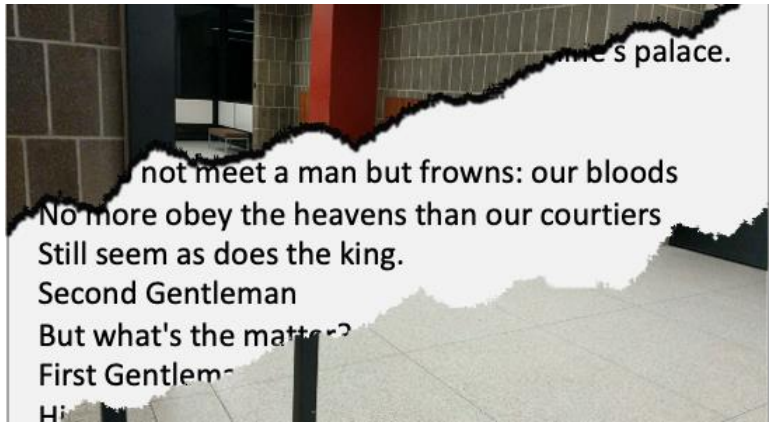
# OpenMFC Tasks

## Manipulation Detection (MD)

- Image_MD (IMD)
- ImageSplice_MD (ISMD)
- Video_MD (VMD)

## Deepfake Detection (DD)

- Image Deepfake/GAN Detection (IDD)
- Video_DD (VDD)




## Steganography Detection (StegD)

- StegD

# Example Task: Deepfake Detection

**System Input** | **System** | **System Output**

**Probe media**

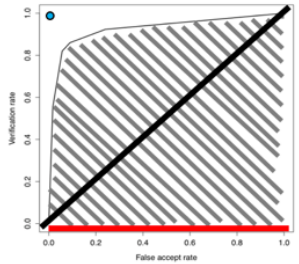
**Detection System**

**Confidence score**

Figure: An OpenMFC Evaluation Task Example

## Leaderboard

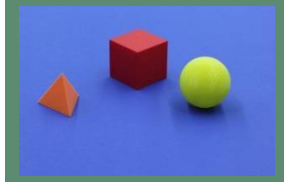
**Receiver Operating Characteristic (ROC)**



**Area Under the Curve (AUC)**
**Correct Detection (CD) at False Alarm Rate 5%**

| RANK | SUBMISSION ID | SUBMISSION DATE | TEAM NAME | SYSTEM NAME | AUC | CDR@0.05FAR |
|---|---|---|---|---|---|---|
| 1 | 90 | 2021-07-10 09:56:14 | TEAM_ID_1 | test | 0.689716 | 0.207018 |
| 2 | 93 | 2021-07-30 18:13:11 | TEAM_ID_1 | test | 0.683956 | 0.187135 |
| 3 | 75 | 2021-06-14 04:12:30 | TEAM_ID_2 | dry-run | 0.554261 | 0.012865 |
| 4 | 52 | 2021-06-01 15:41:20 | TEAM_ID_2 | dry-run | 0.547125 | 0.009357 |
| 5 | 82 | 2021-06-23 09:21:26 | TEAM_ID_1 | System_ID_1 | 0.500033 | 0.051077 |

# Evaluation Data Challenges



vs.

Lab Algorithm Evaluation

vs.

Real-World Field Application

**Curse of Dimensionality & Dynamic Shift**

Media complexity

Manipulation complexity

- Hundreds of manipulation tools and operations
- Emerging technologies: Stable Diffusion, Deepfake, GAN, Generative AI, etc.
- Anti-forensic technologies

**Ground truth labeling**

**Post interpretation is nearly impossible**

- Has it been manipulated? Was it malicious?
- Who performed the manipulation?
- What was the original media source?
- Which editing tool was used?
- What specific operations were applied?
- How was the media manipulated?
- Where did the manipulation occur?

**Concerns & Constraints**

Availability

Scalability

Privacy, Ethics, Consent

Resource intensity (time, labor, cost)

Distribution

# Media Manipulation Journaling

a. Manipulated image

b. Original image (Base image)

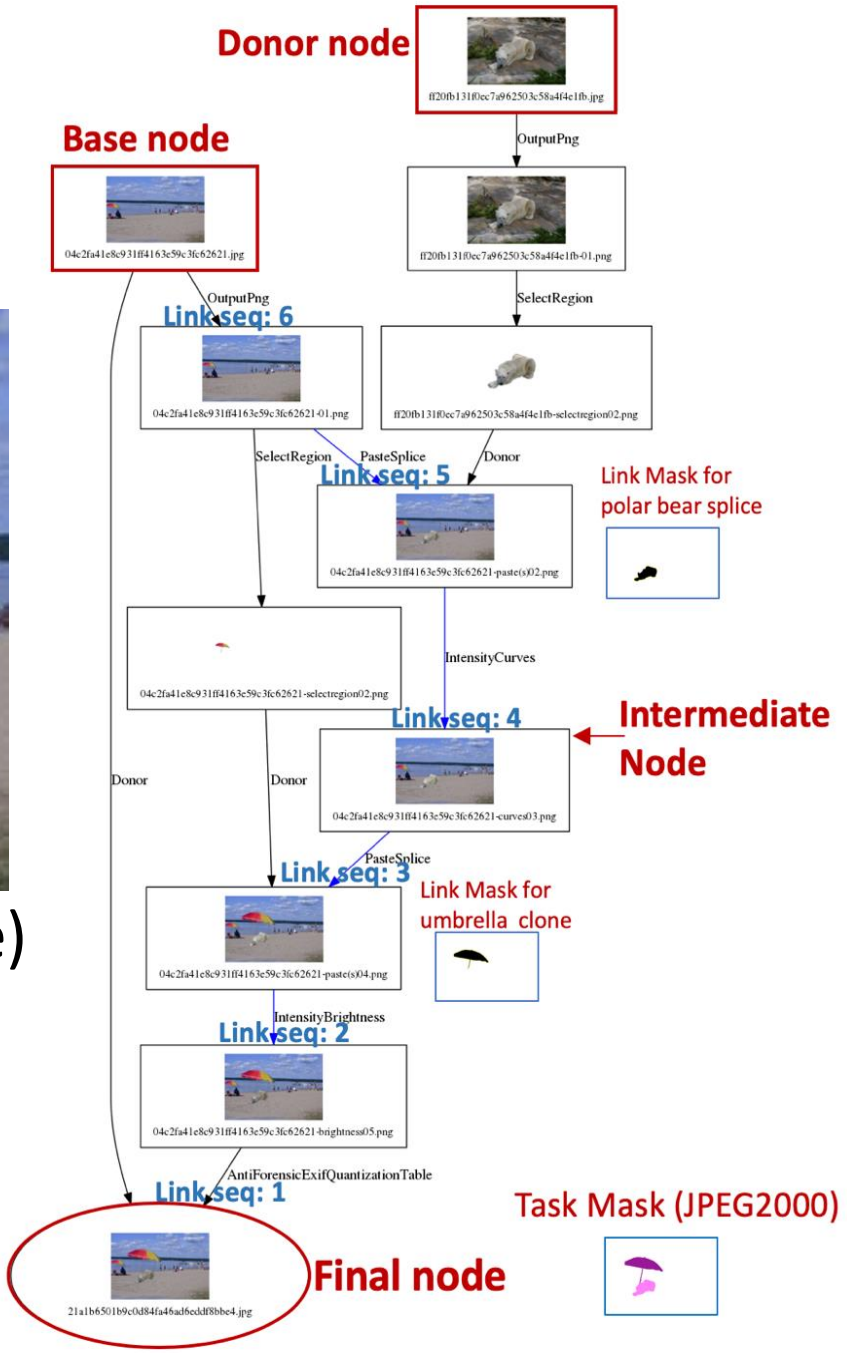



Figure: Example of a Manipulation Journal Log

# Evaluation Reference Collection

## 1 Data Generation

- Establish a dedicated data generation team that operates independently from the analytic teams
- Prevent evaluation and data bias
- Encourage analytics teams to focus on developing and evaluating more realistic test scenarios.

## 2 Annotation: Human + Machine

- Human Manipulation (realistic)
- Automatic Manipulation (reduce cost)
- Extended Manipulation (special study)

## 3 Annotation: Journaling Tools

- Manipulation Journaling Tool (JT)
- Automatic Journaling Tool (Auto-JT)
- Extended Journaling Tool (Extended-JT)

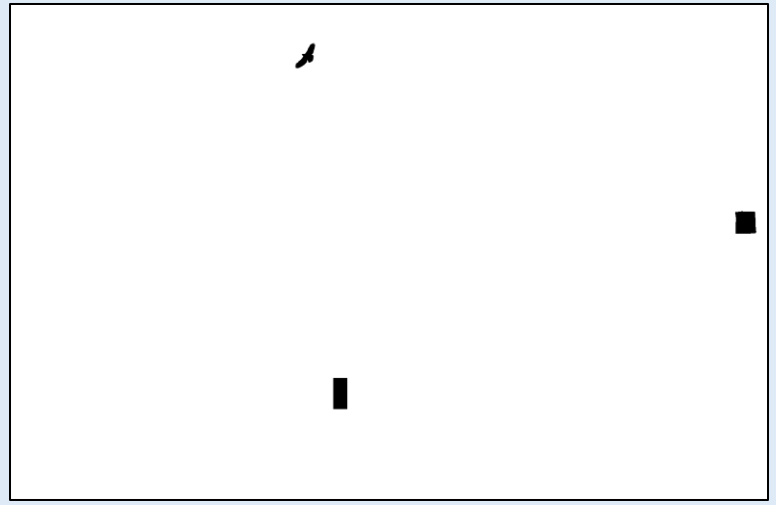
## 4 Real-World Simulation

- Post-Processing
- Anti-Forensics
- Recapture

# Image Manipulation Examples

1.1 Manipulated image

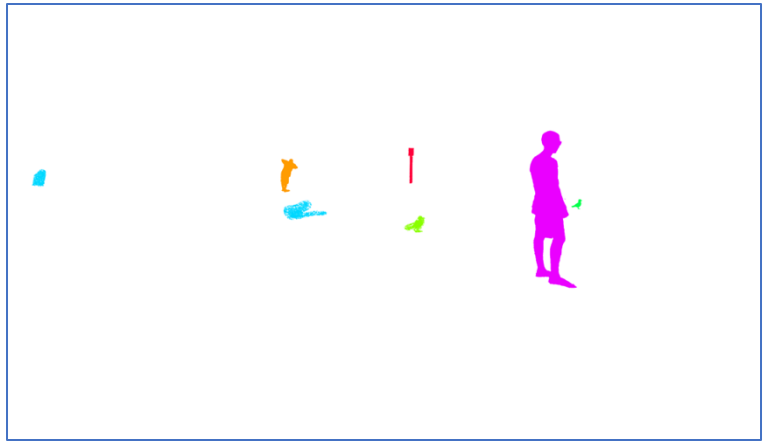
1.2 Original image



1.3 Manipulation Mask

2.1 Manipulated Image

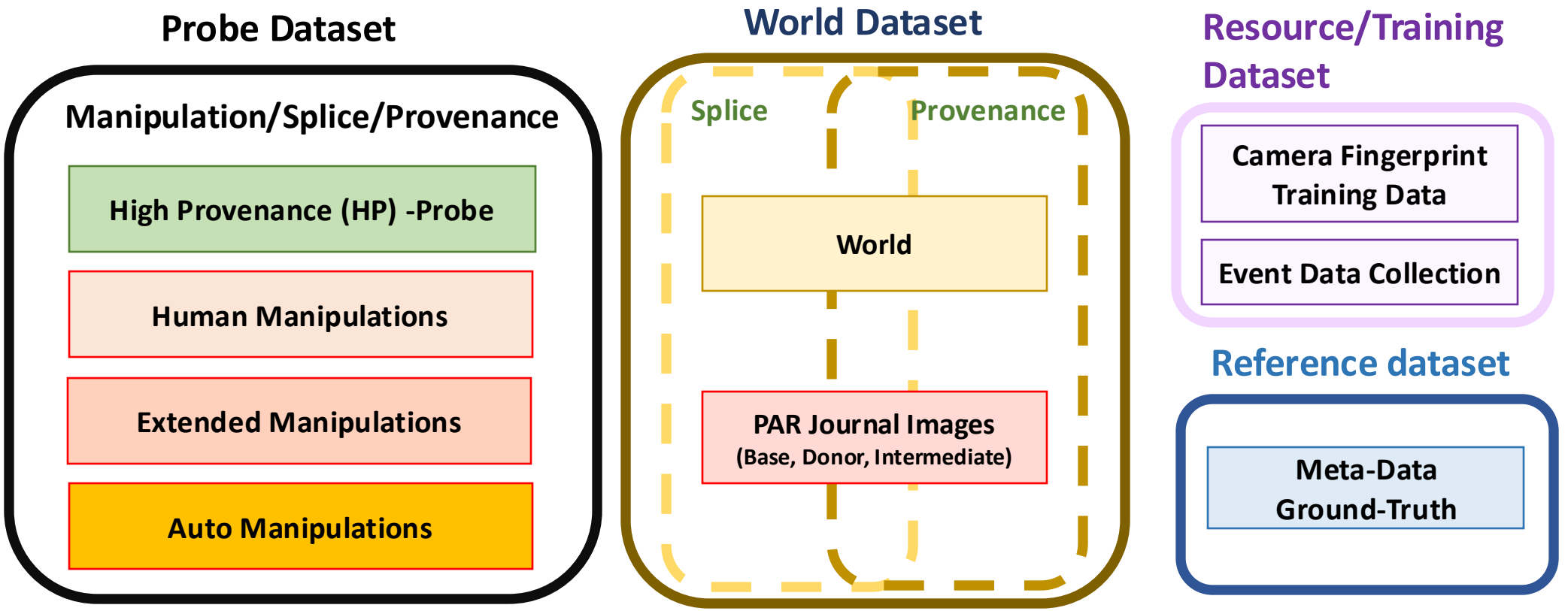
2.2 Original Image



2.3 Manipulation Mask (JPEG 2000)

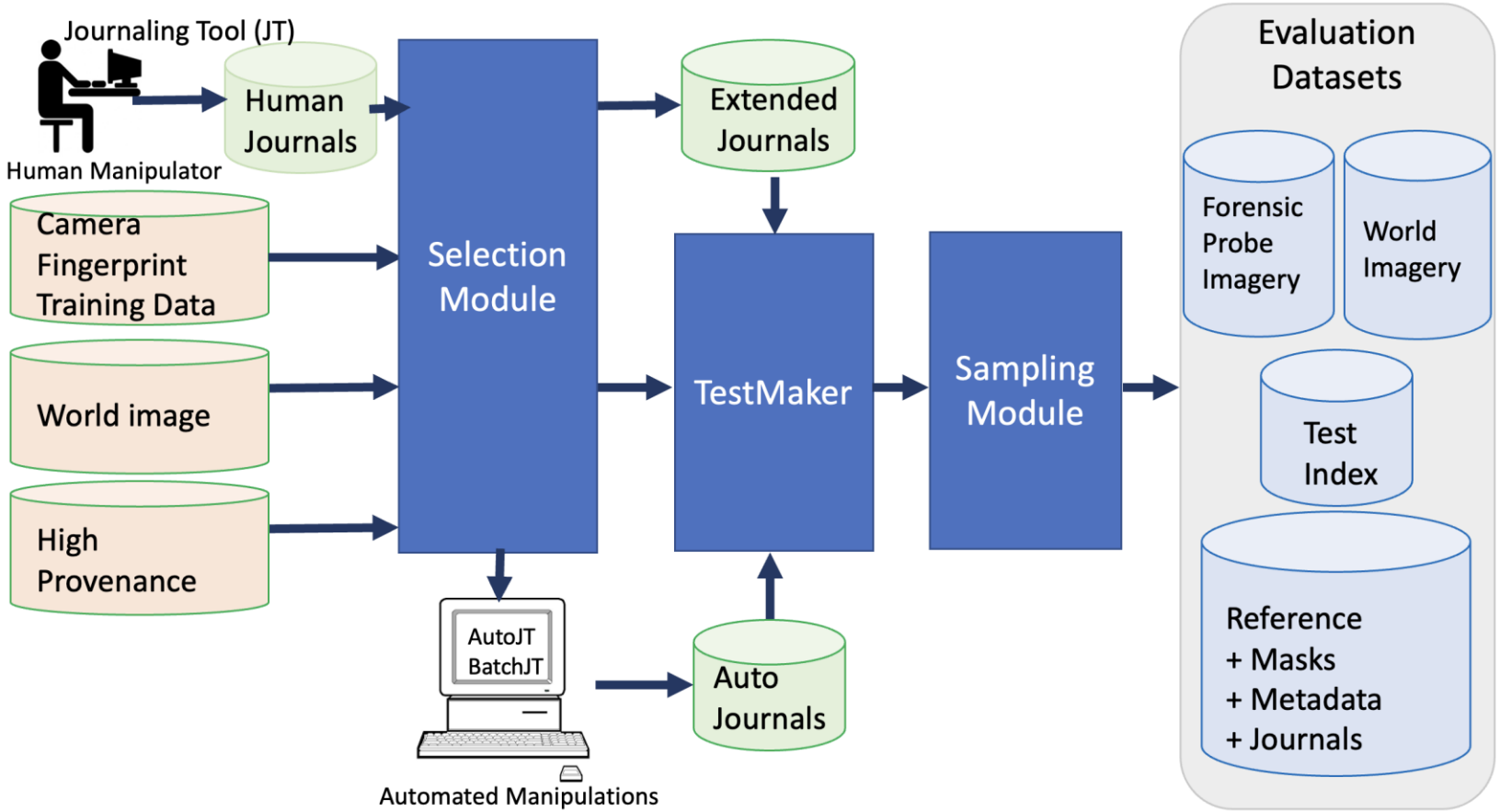
# Evaluation Data Overview

**We optimize data utilization across all tasks while minimizing overall costs by leveraging the diverse evaluation data collected by various teams. This includes original media, manipulation journal graphs, globally sourced media, and camera training data.**



# Evaluation Dataset Production Infrastructure

**We have developed an infrastructure for generating evaluation datasets that incorporates various types of data, metadata, and ground-truth references, enabling the continuous updating of evaluation data.**



# Evaluation Data Diversity

**How effectively do the algorithms detect different types of manipulations, given the wide range of existing image and video editing methods, techniques, and software?**

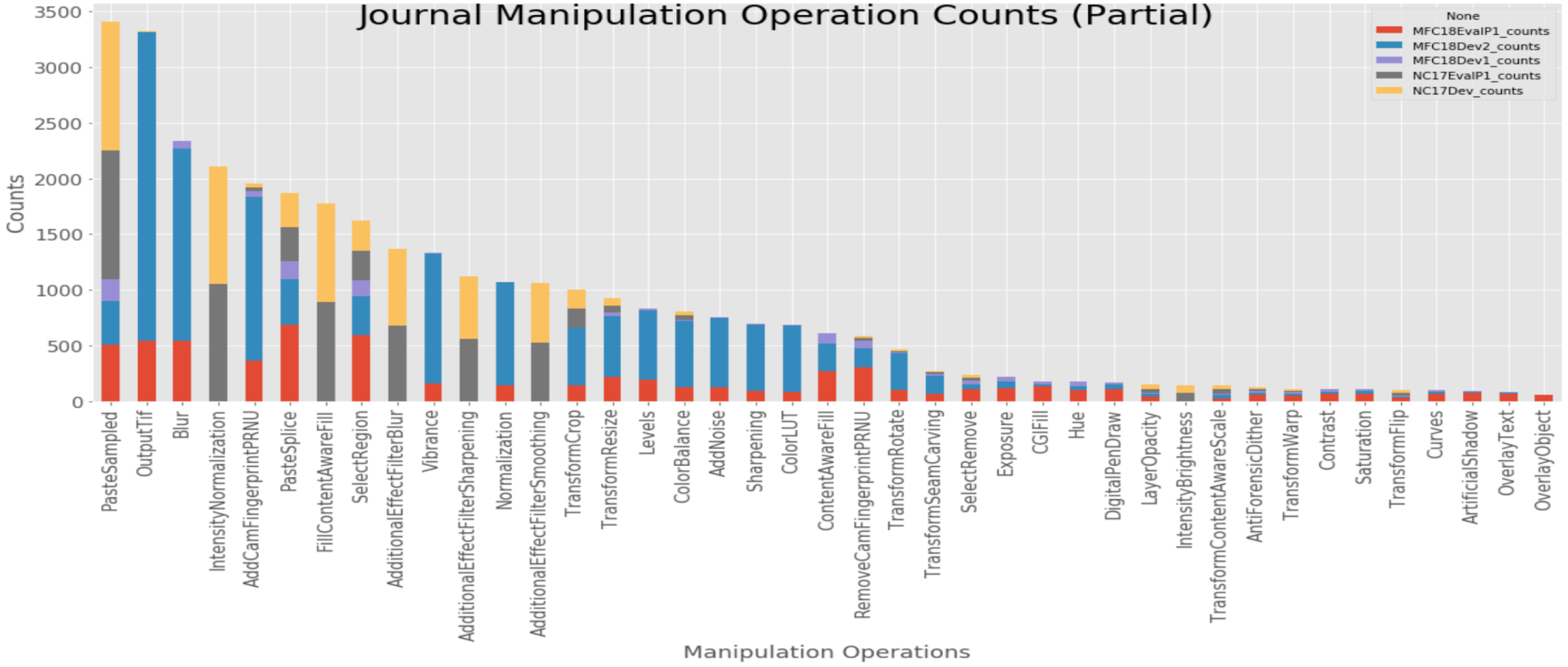
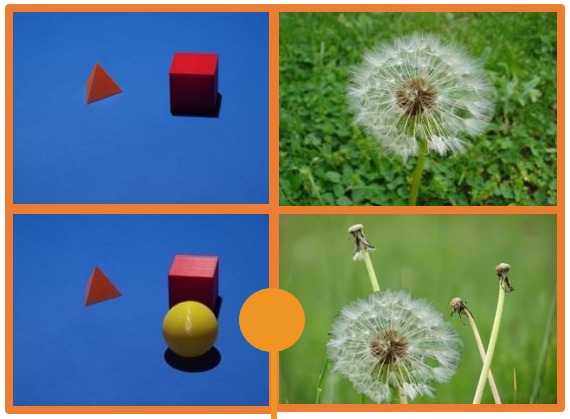


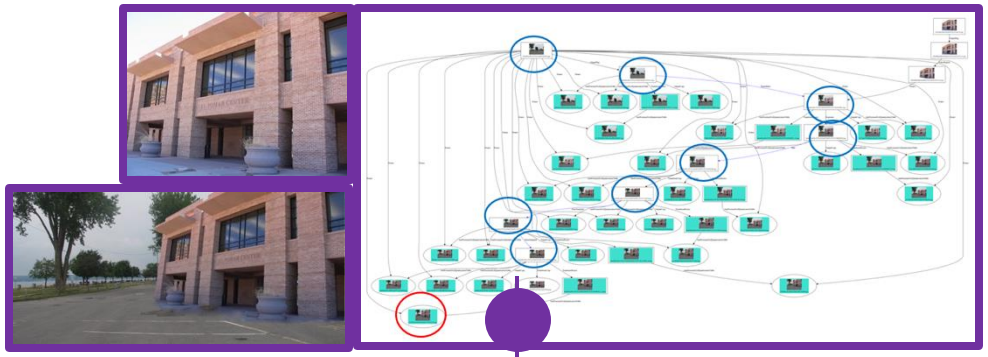

Figure: Histogram of Manipulation Operations Across Datasets

# OpenMFC and MediFor Evaluation Datasets (1)

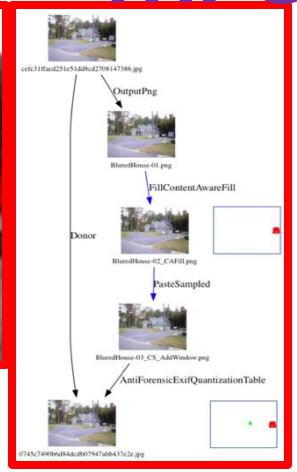


NC16

NC17

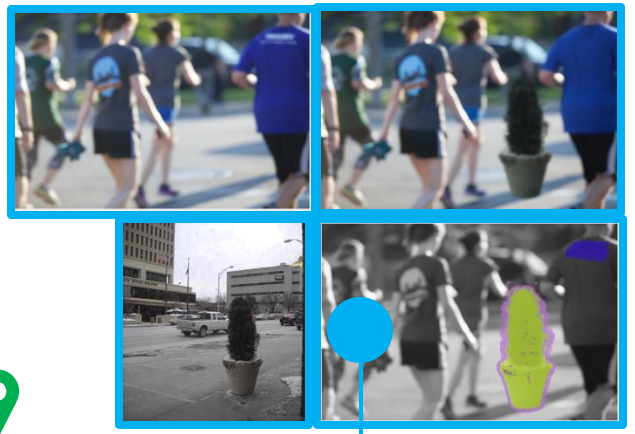


MFC18

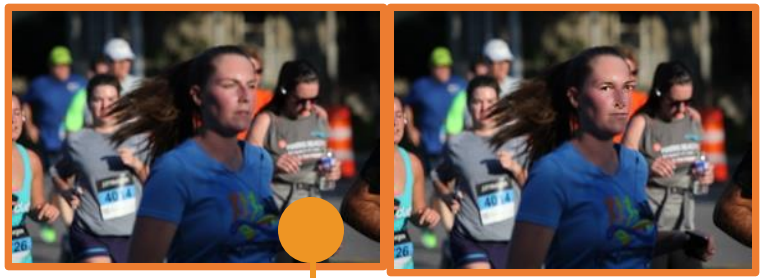



MFC19

MFC20

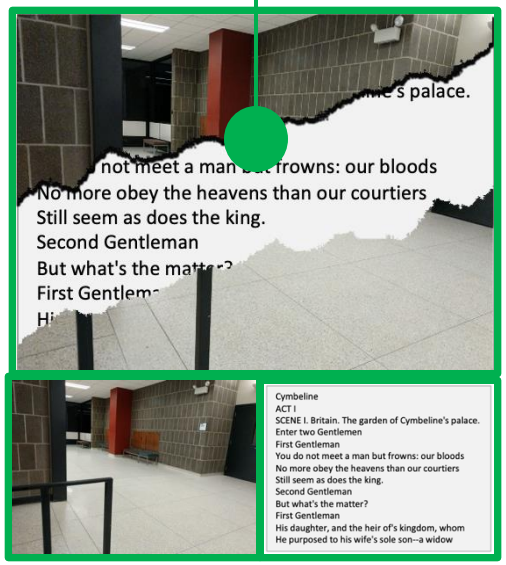


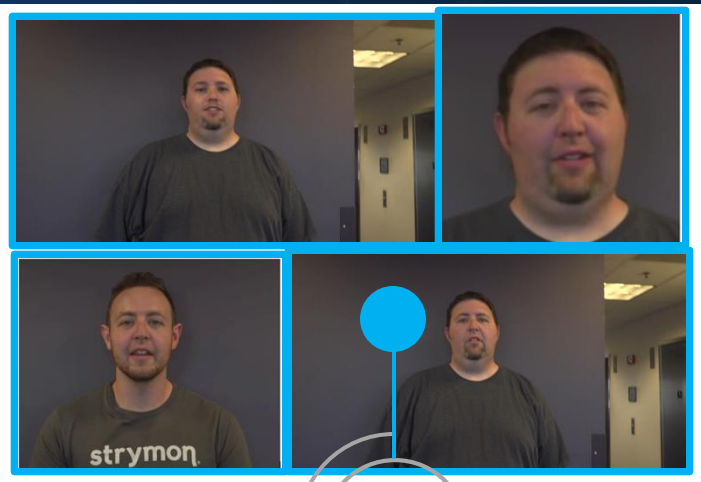
# OpenMFC and MediFor Evaluation Datasets(2)



## OpenMFC GAN Image

## OpenMFC StegD



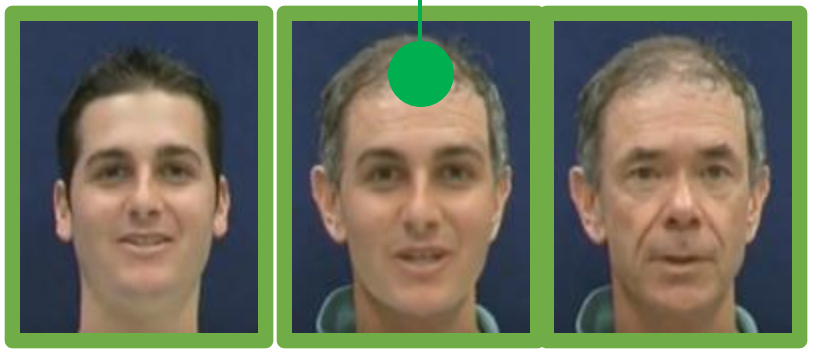

## OpenMFC Deepfake Video

## OpenMFC Image Splice

# OpenMFC Evaluation Dataset (3) - Coming

Deepfake Image

Deepfake Video



Deepfake Audio

# OpenMFC Evaluation Datasets

Table: OpenMFC2020 Evaluation datasets

| OpenMFC 2022 Evaluation Dataset | OpenMFC 2022 Task | Corresponding MFC Dataset Name | Media Type | Media Number | Journal # | Create Date |
|---|---|---|---|---|---|---|
| **OpenMFC20_Image_MD** | IMD | MFC19 EP1 Image | Image | 16K | 1383 | 2022 |
| **OpenMFC20_Video_MD** | VMD | MFC19 EP1 Video | Video | 1.5K | 163 | 2022 |
| **OpenMFC20_Image_DD** | IDD | MFC18 GAN FULL Image | Image | 1.3K | 267 | 2022 |
| **OpenMFC20_Video_DD** | VDD | MFC18 GAN Video | Video | 118 | 19 | 2022 |


Table: OpenMFC2022-2023 Evaluation datasets

| OpenMFC 2022 Evaluation Dataset | OpenMFC 2022 Task | Media Type | Media Number | Create Date |
|---|---|---|---|---|
| **OpenMFC22_SpliceImage_MD** | ISMD | Image | 2K | 2022 |
| **OpenMFC22_Image_StegD** | StegD | Image | 480 | 2022 |
| **OpenMFC23_Image_DD** | IDD | Image | (TBD) | 2023 |
| **OpenMFC23_Video_DD** | VDD | Video | (1.5K) | 2023 |

# OpenMFC Evaluation Pipeline

**NIST has developed the flexibility to customize the evaluation strategy to meet various assessment requirements, including both take-home and sequestered evaluations, each with its own specific needs and prerequisites.**

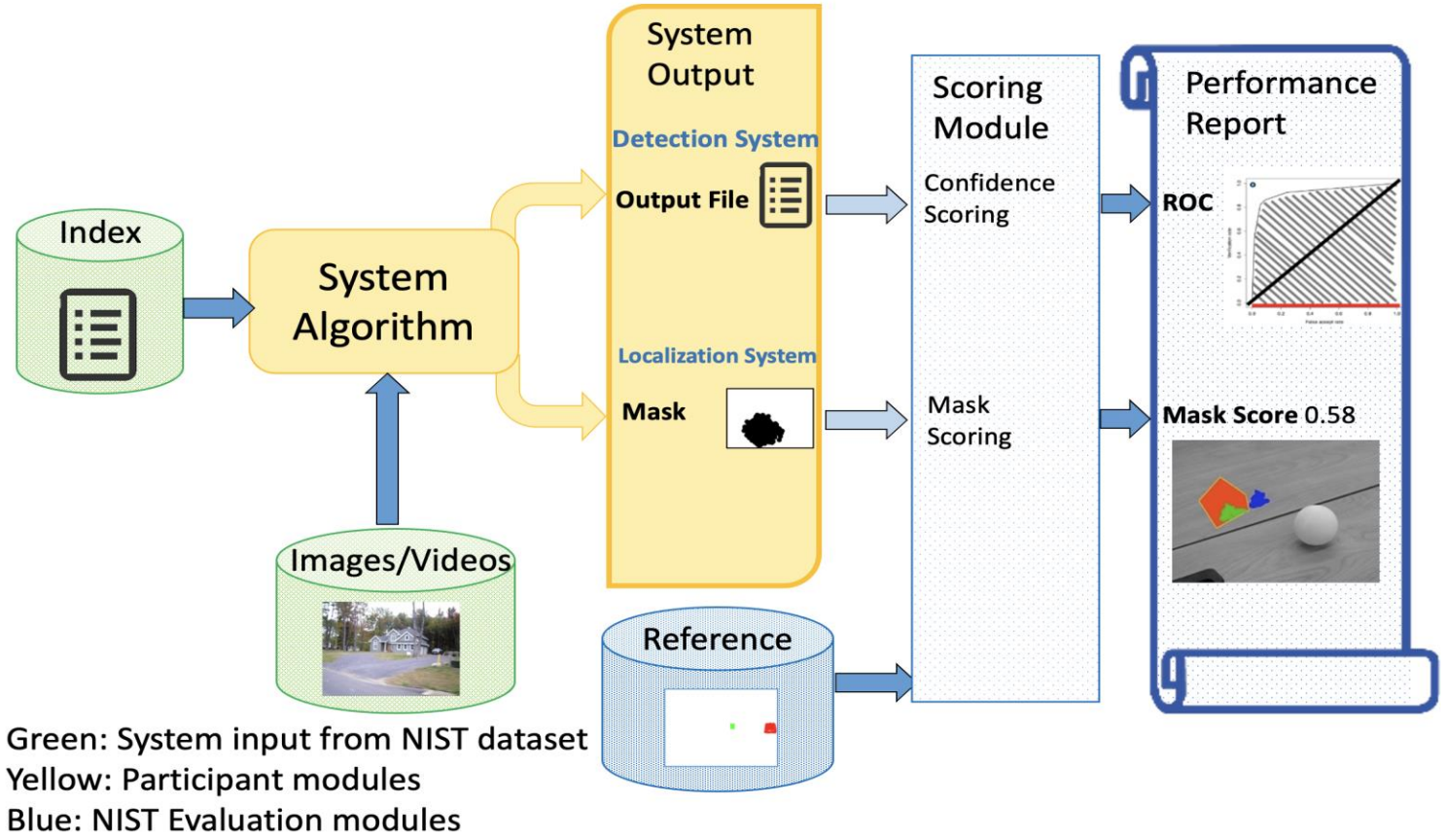



Figure: Evaluation Pipeline

# Evaluation Scoring Software: MediScore

https://github.com/usnistgov/MediScore/

**Unified evaluation framework to support:**

- Detection tasks
  - Image/Video Manipulation Detection (Single Input)
  - Splice Detection task (Pair Input)
- Localization tasks
  - Image/Video Manipulation Localization task (Single Input)
  - Splice Localization task (Pair Input)
- Verification tasks
  - Camera ID Verification
  - Event Verification
- Provenance tasks
  - Provenance filter
  - Graph builder

OpenMFC

Major Functions

1. Validator
2. DetectionScorer
3. MaskScorer
4. Provenance
   - ProvenanceValidator
   - ProvenanceGraphBuildingScorer
   - ProvenanceFilteringScorers

Flexible Scoring Options

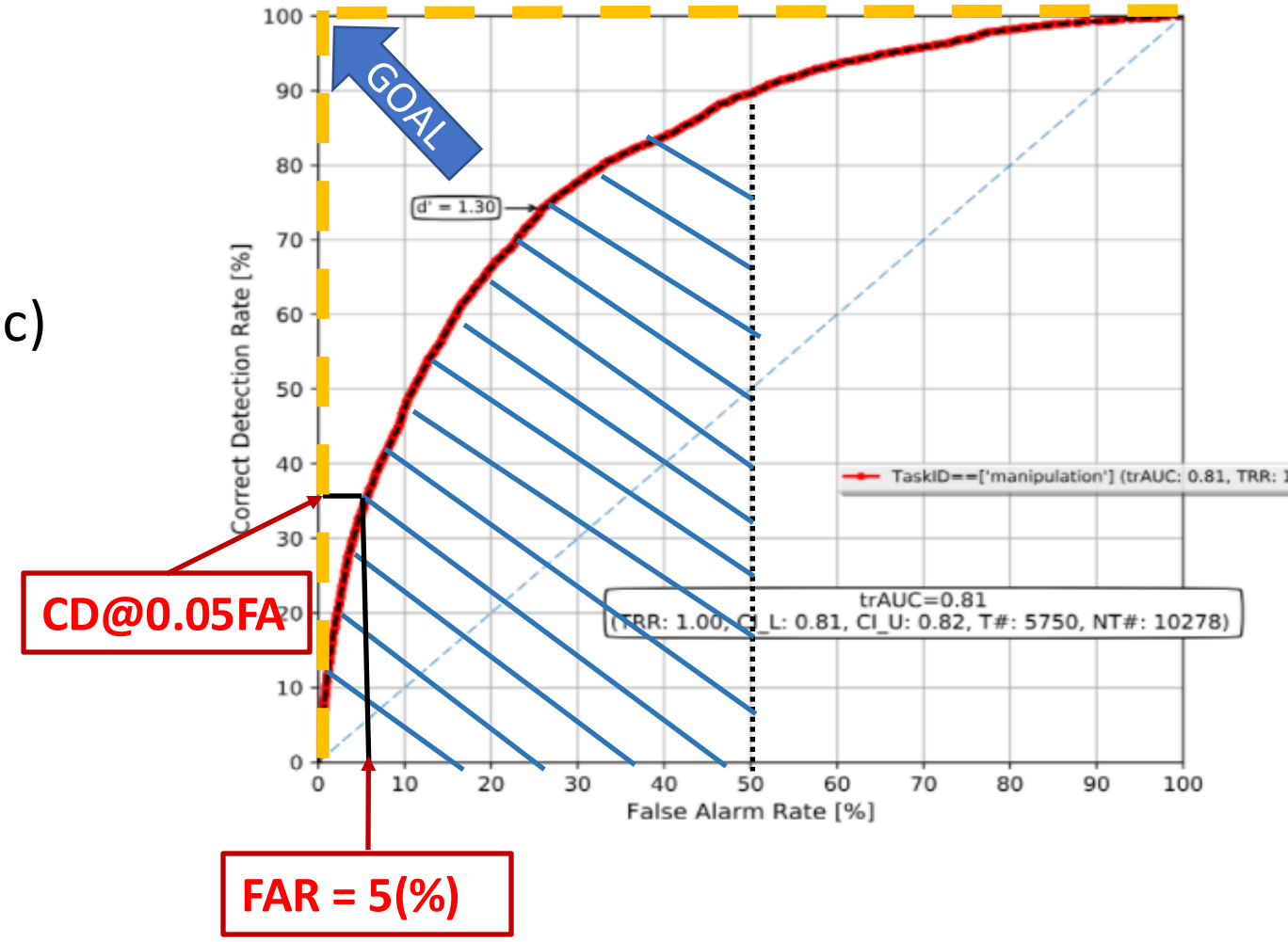
1. OptOut

   The testing data is defined by user
2. SelectiveScoring

   The testing data is defined by selected manipulation operation

# 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



# Media Forensic System Performance

- When reviewing the system performance report, it is important to clearly understand the characteristics of the evaluation data, including its diversity and realism.
- The NIST cross-year evaluation report does not offer a straightforward apples-to-apples comparison due to the increasing complexity of datasets over time.
- The performance of  media forensic systems may decline due to the growing diversity and complexity of datasets, including factors such as manipulation complexity, the introduction of new manipulations, and the emergence of anti-forensic technologies.
- The effectiveness of media forensic systems can be influenced by factors such as resizing, compression, post-processing, and the use of anti-forensic tools.

# Insights and Takeaways

## Dynamic Update

The landscape of media forensic research is continuously evolving with the advancement of media editing and generation technologies.

## Media Forensics vs. Anti-Forensics

The performance of forensic systems may decline as anti-forensic technologies adapt and improve.

## Data Diversity

A wide range of diversity in real applications.

## Forensic System Performance

There is still considerable room for improvement in media forensic technologies.

**Challenges**

## Continuous Evaluation

The data generation infrastructure supports regularly update.

## Red team vs. Blue team evaluation

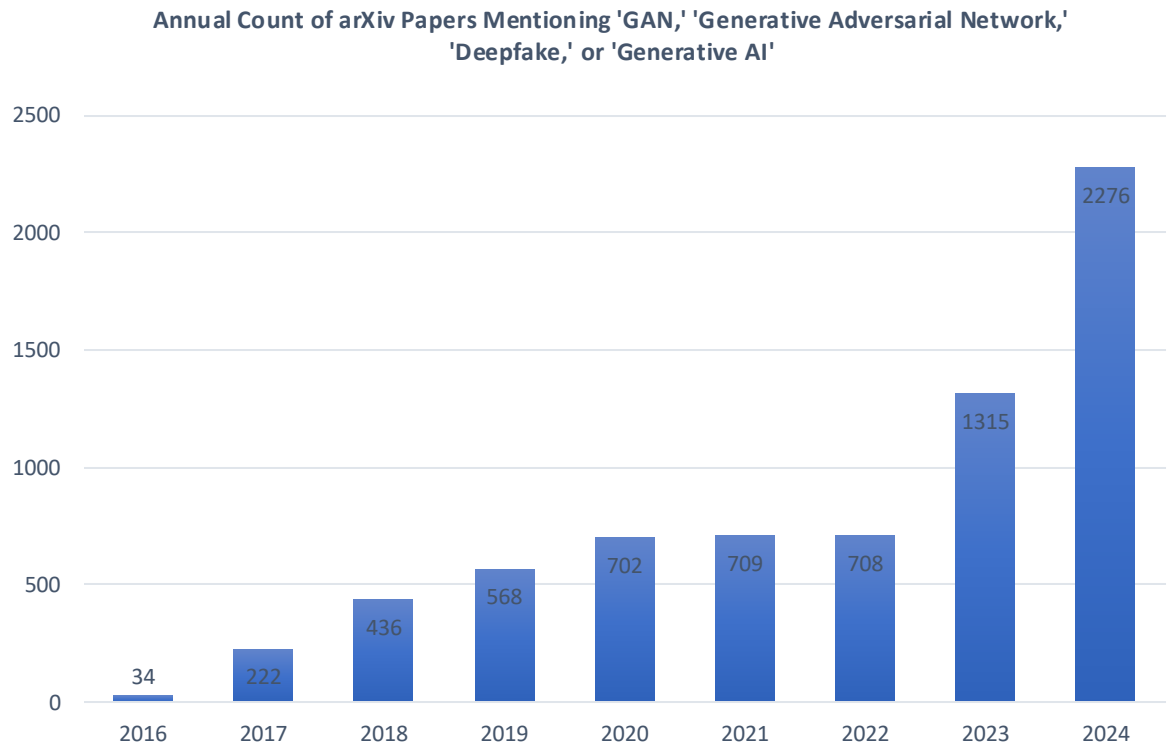
## Collaboration Platform

Extensive experience in dataset design, collection, annotation, and generation

## Flexible and Customized Evaluation Strategies

**NIST Solutions**

# Ongoing Efforts: Deepfake Challenge




Annual Count of arXiv Papers Mentioning 'GAN,' 'Generative Adversarial Network,' 'Deepfake,' or 'Generative AI' as 12/20/2024

## Generative AI models:

- GAN (Generative Adversarial Networks) (2014)
- Text-to-Image Synthesis (2015)
- Image-to-Image Generation (2017)
- Deepfakes (2018)
- Diffusion Models (2021)
- Text-to-Audio and Voice Cloning Models (2023)
- Multimodal Models (2024)

## Focus:

- Media Types: Image, Video, and Audio
- Content Types: Face, nonface,

# NIST In-House Evaluation Data

## AI Media Generation Tools

- DeepFaceLab, StyleGAN 1/2/3, Celeb-DF
- Post-processing or Anti-forensics

## Data Release IRB approval

Collaborators: Prof. Conrad Sanderson (VidTIMIT); DARPA MediFor data teams from U.C. Denver and PAR Government

## Capable for Special Studies

- Enhanced with additional meta-data collection

## Support for Forensic System Development

Collaborators: Prof. Siwei Lyu, Dr. Shan Jia, and Yan Jun at University at Buffalo


Figure: Example of a Deepfake and Its Two Source Videos

# Ongoing Efforts: Generative AI Evaluation

## Text-to-Image

DALL·E 2 (OpenAI), Midjourney, Runway, Adobe Firefly, Picsart, Craiyon, Jasper, Canva, PicSo, Bing Image Creator

## Image-to-Image

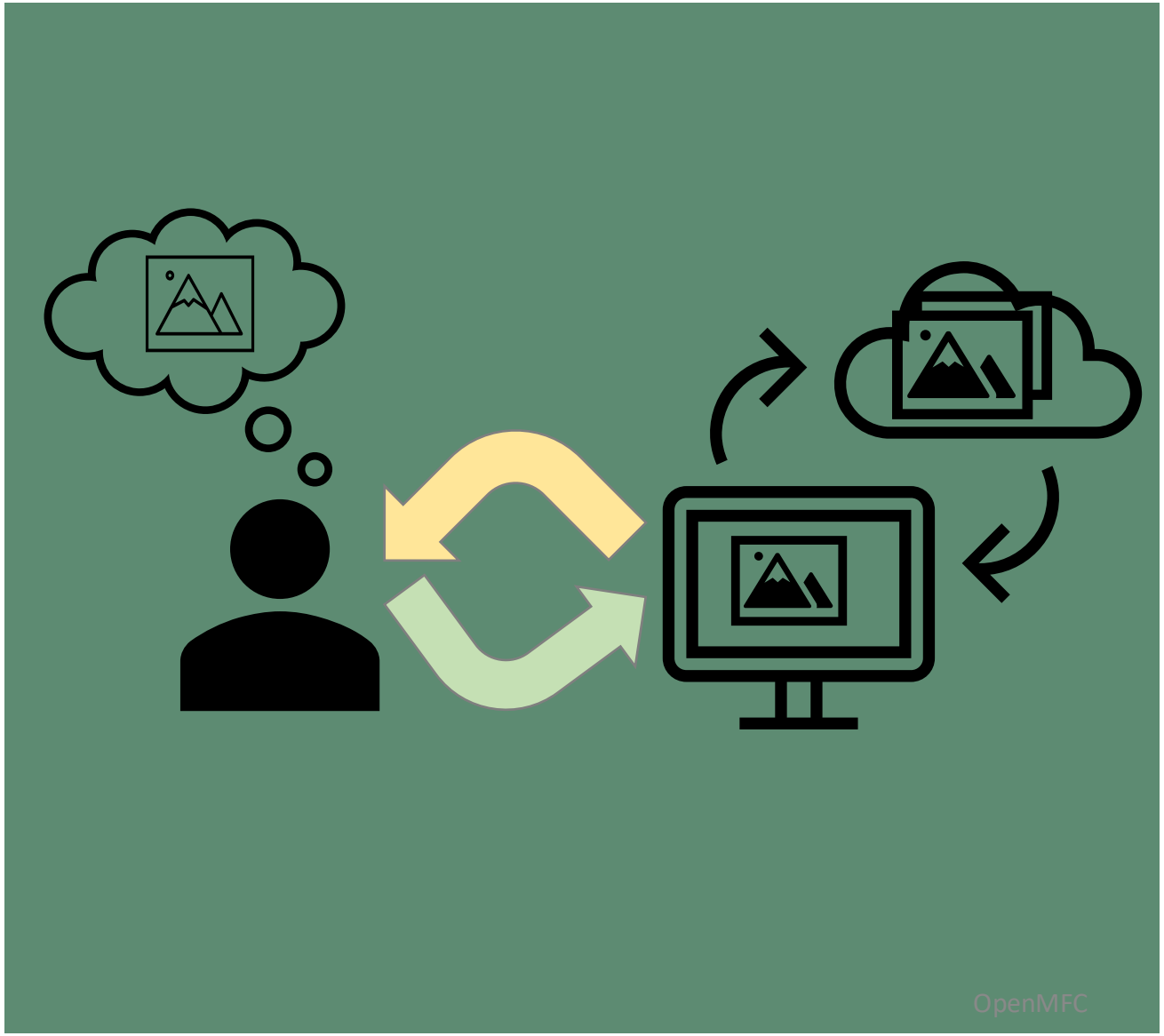
RunwayML, Media.io, AI Avatar, Midjourney (Discord), Stable diffusion

## Beyond Forensics

Art and content creation, image inpainting, video synthesis, character animation, super-resolution, etc.

# Future Direction: New Evaluation Methodology

## Bridging the Gap Between Traditional AI and HMI Assessment Paradigms




- **Objective Evaluation**

  Traditional AI evaluation methods for detection, verification, retrieval systems
  - Is it real or generated?
  - What is the provenance?
- **Subjective Evaluation**

  Human as an evaluator
  - Creativity
  - User satisfaction
  - User interaction experiences
- **Human-Machine Interaction (HMI)**

  Humans as a users

  HMI evaluation methods
  - Usability and design
  - Overall performance and effectiveness

# How to organize an evaluation program?

## 1 Artists - Media Creation and Manipulation

Data collection

Data manipulation

Data annotation

Data creation

## 2 Forensic Experts

Media forensics specialists

Forensic practitioners

## 3 Anti-Forensic Experts

Post-processing algorithms

Recapture

Anti-Forensic systems

## 4 AI and HMI Experts

Generator model

Detector model

Forensic model

Human-Computer-Interaction

# Vision & Plan: Collaboration Platform

- Challenges
  - Data
    - expert, time, cost, labor, IRB agreements
  - Systems
    - forensic baseline systems
    - anti-forensic systems
  - Participants
    - high demands on time and efforts
- Collaboration Platform
  - Cross-team collaboration: share knowledge, skill, tools, and resources
  - Intra-team collaboration: foster cooperation within teams rather than competition

OpenMFC

NIST evaluations

- Maintain a neutral and unbiased stance
- Provide ongoing assessments of emerging technologies
- Support a diverse range of stakeholders, including government policymakers, industry product developers, and academic researchers

# NIST OpenMFC at a Glance

800+[1] people

200+ organizations

26+ countries

4 + 2 evaluations

15+ publications

5+ collaborative organizations

[1] Over the years of MFC and OpenMFC, more than 800 individuals have been involved as organizers, program developers, participants, stakeholders, government agency representatives, registered users of the evaluation website, or recipients of our development and evaluation datasets, among other roles.

# Publications (1)

Yan Ju, Shan Jia, Jialing Cai, Haiying Guan, Siwei Lyu, "GLFF: Global and Local Feature Fusion for AI-synthesized Image Detection," IEEE Transactions on Multimedia, 2023, https://ieeexplore.ieee.org/document/10246417, Digital Object Identifier: 10.1109/TMM.2023.3313503.

Haiying Guan, Yooyoung Lee, Lukas Diduch, Jesse Zhang, Ilia Ghorbanian, Timothee Kheyrkhah, Peter Fontana, Jonathan Fiscus and James J Filliben, "Open Media Forensics Challenge (OpenMFC) 2020-2021: Past, Present, and Future," NIST Interagency/Internal Report (NISTIR) Number 8396, https://doi.org/10.6028/NIST.IR.8396.

Haiying Guan, Andrew Delgado, Yooyoung Lee, Amy N. Yates, Daniel Zhou, Timothee Kheyrkhah, and Jonathan Fiscus, "User Guide for NIST Media Forensic Challenge (MFC) Datasets," NIST Interagency/Internal Report (NISTIR) Number 8377, July 2021, https://doi.org/10.6028/NIST.IR.8377.

Haiying Guan, Mark Kozak, Eric Robertson, Yooyoung Lee, Amy N. Yates, Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah, Jeff Smith and Jonathan Fiscus, "MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation," 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 63-72, 2019, https://ieeexplore.ieee.org/document/8638296.

Eric Robertson, Haiying Guan, Mark Kozak, Yooyoung Lee, Amy N. Yates, Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah, Jeff Smith and Jonathan Fiscus, "Manipulation Data Collection and Annotation Tool for Media Forensics," IEEE Computer Vision and Pattern Recognition conference 2019, Workshop on Applications of Computer Vision and Pattern Recognition to Media Forensics, 2019. https://openaccess.thecvf.com/content_CVPRW_2019/html/Media_Forensics/Robertson_Manipulation_Data_Collection_and_Annotation_Tool_for_Media_Forensics_CVPRW_2019_paper.html.

# Publications (2)

Haiying Guan, Baptiste Chocot, Ilia G. Bajgiran, Lukas Diduch, Yooyoung Lee, Siwei Lyu, Yan Ju, Shan Jia, Conrad Sanderson "Deepfake Challenge 2023 Video Dataset ReadMe," NIST Publication ID 936550, ERB Control#: N2023-0762.

Haiying Guan, Yooyoung Lee, and Lukas Diduch, "Open Media Forensics Challenge 2022 Evaluation Plan," https://www.nist.gov/publications/open-media-forensics-challenge-2022-evaluation-plan.

Haiying Guan, Yooyoung Lee, Lukas Diduch, Ilia Ghorbanian, and Jim Horan, "Open Media Forensics Challenge (OpenMFC) 2021 Workshop Presentation Slides," https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=933793.

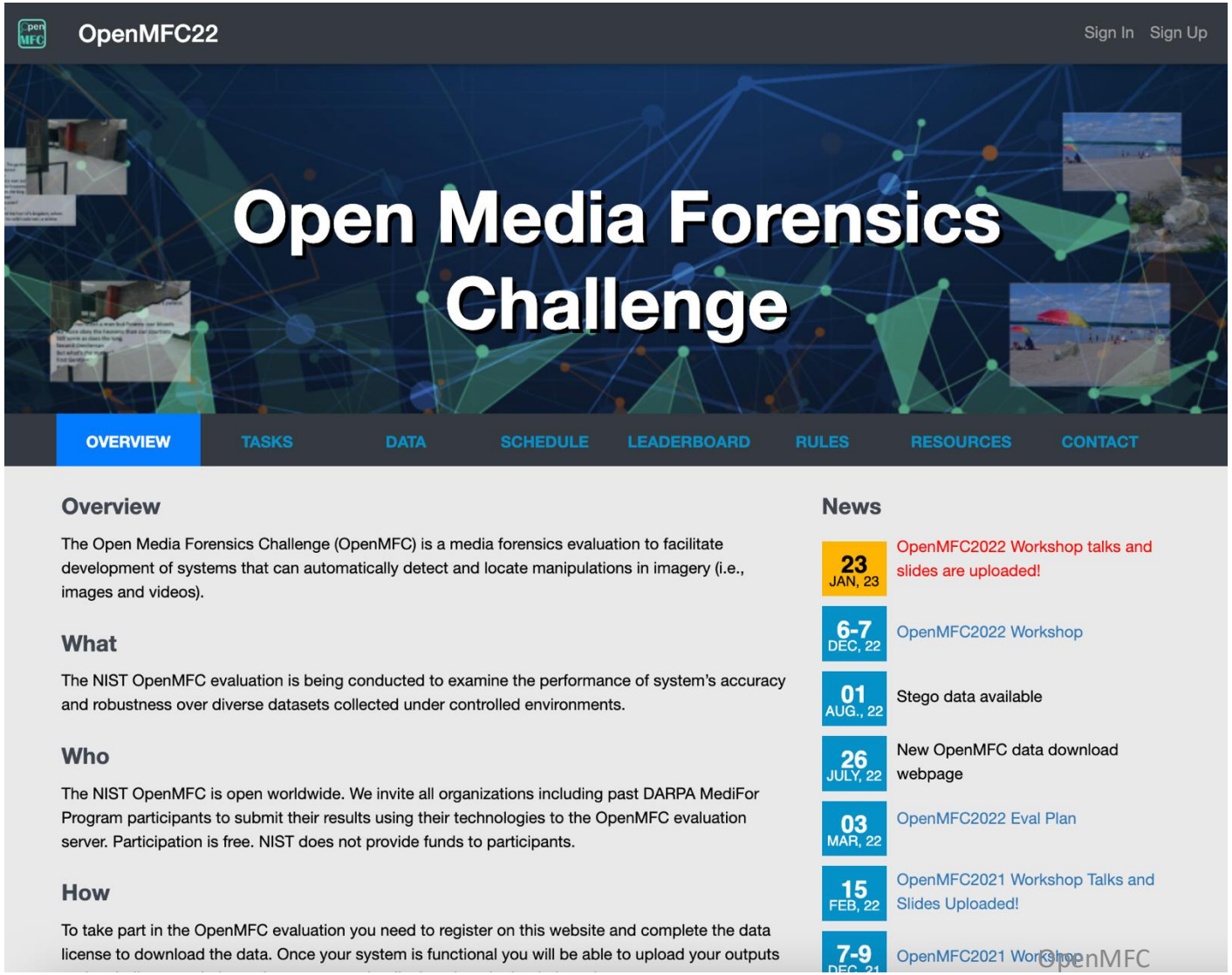
Amy N. Yates, Haiying Guan, Yooyoung Lee, Andrew Delgado, Timothee Kheyrkhah and Jonathan Fiscus, "Open Media Forensics Challenge 2020 Evaluation Plan," September 2020, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=930801.

Jonathan Fiscus, Haiying Guan, Yooyoung Lee, Amy N. Yates, Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah, "Open Media Forensics Challenge 2020 - DARPA MediFor Demo Day Presentation," August 2020, https://www.nist.gov/publications/open-media-forensics-challenge-2020-darpa-medifor-demo-day.

Jonathan Fiscus, Haiying Guan, Yooyoung Lee, Amy N. Yates, Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah, and Xiongnan Jin, "NIST Media Forensic Challenge (MFC) Evaluation 2020 - 4th Year DARPA MediFor PI meeting," July 2020, https://www.nist.gov/publications/nist-media-forensic-challenge-mfc-evaluation-2020-4th-year-darpa-medifor-pi-meeting.

Jonathan Fiscus, Haiying Guan, "Media Forensics Challenge Evaluation Overview Talk on ARO Sponsored Workshop on Assured Autonomy," June 2020, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=930628.

# NIST OpenMFC Resource




## Website

https://mfc.nist.gov

## Slack

http://openmfc.slack.com

## Scoring Software

https://github.com/usnistgov/MediScore/

## Contact

**mfc_poc@nist.gov**

# Acknowledgement


- External collaborators:
  - Prof. Siwei Lyu, Dr. Shan Jia, and Yan Ju at University at Buffalo
  - Prof. Conrad Sanderson
  - DARPA MediFor Program
  - Data team from National Center for Media Forensics, U.C. Denver
  - Data team from PAR Government
  - Prof. Jennifer Newman, Li Lin, Prof. Yong Guan from Iowa State University, and Prof. Roy A. Maxion from Carnegie Mellon University
- NIST contributors
  - James Horan
  - Jonathan Fiscus
  - Timothee Kheyrkhah
  - Daniel Zhou
  - Edmond Golden
  - Noah Dove
  - Peter Fontana
  - Jesse G. Zhang

# Questions?

# Stay In Touch!

haiying.guan@nist.gov