

ARO Sponsored Workshop on Assured Autonomy

Media Forensics Challenge Evaluation Overview


Jonathan Fiscus and **Haiying Guan**

Multimodal Information Group

Information Access Division

Information Technology Laboratory

National Institute of Standards and Technology (NIST)

June 24, 2020 

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 Program.



Thanks to colleagues and collaborators!

- Colleagues in NIST Team

Multimodal Information Group, ⁺ Image Group

Jonathan Fiscus (co-PI),

Haiying Guan (co-PI),

Dr. Yooyoung Lee,

Dr. Amy Yates⁺,

Andrew Delgado,

Daniel Zhou,

Timothee Kheyrkhah,

Dr. Xiongnan Jin

- DARPA Program collaborators

- DARPA Media Forensic (MediFor) Team

- PAR Government

- National Center for Media Forensics,
University of Colorado Denver

- RankOne

- Data Machines Incorporated

- Next Century

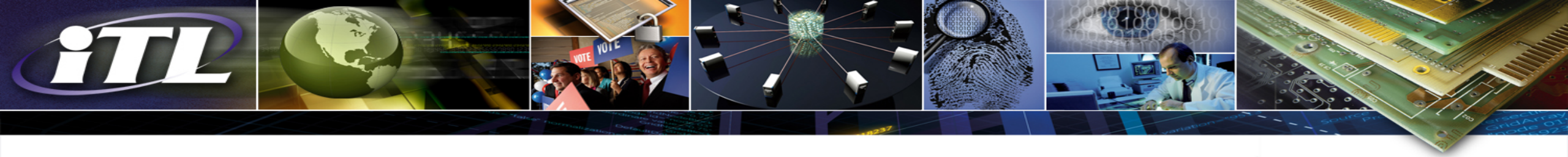
- Air Force Research Lab



Outline

- Media Forensics Challenge (MFC) evaluation design
- MFC evaluation tasks
- MFC evaluation datasets
- MFC evaluation metrics
- MFC evaluation platform
- MFC performance report



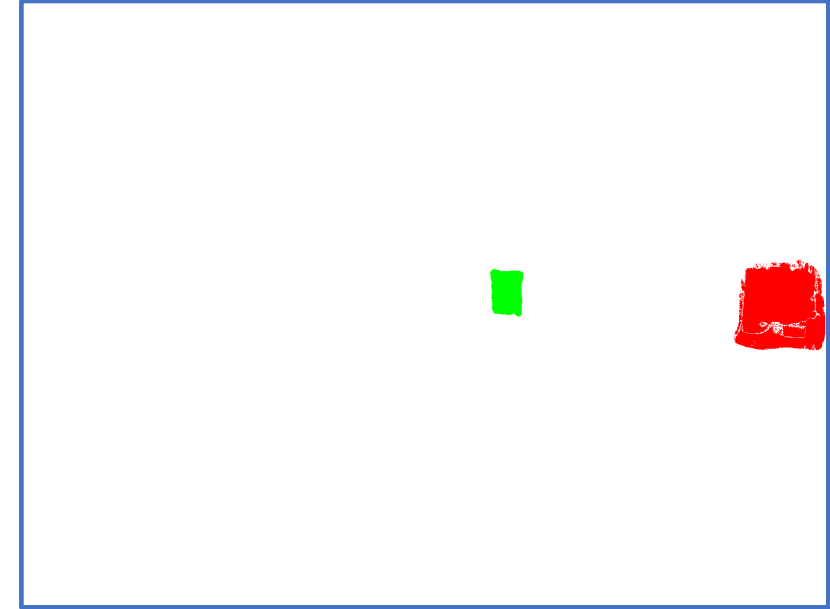


Media Forensics Challenge Evaluation Design

- What is Media Forensic?
- DARPA MediFor program
- Evaluation design challenges
- NIST MFC evaluation design



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

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

DARPA MediFor Evaluation Challenges

- DARPA MediFor Program 2017-2020
 - Objective in BAA: “develop technologies for the automated assessment of the integrity of an image or video.”
- Evaluation design challenges
 - What to evaluate?
 - Technical methodology varieties brings big challenges in design a unified evaluation framework
 - What resources to use?
 - Lack of benchmark datasets
 - Different technologies needs different evaluation data
 - What we can get from the evaluation?
 - Lack of baseline performance information
 - Lack of state-of-the-art performance information

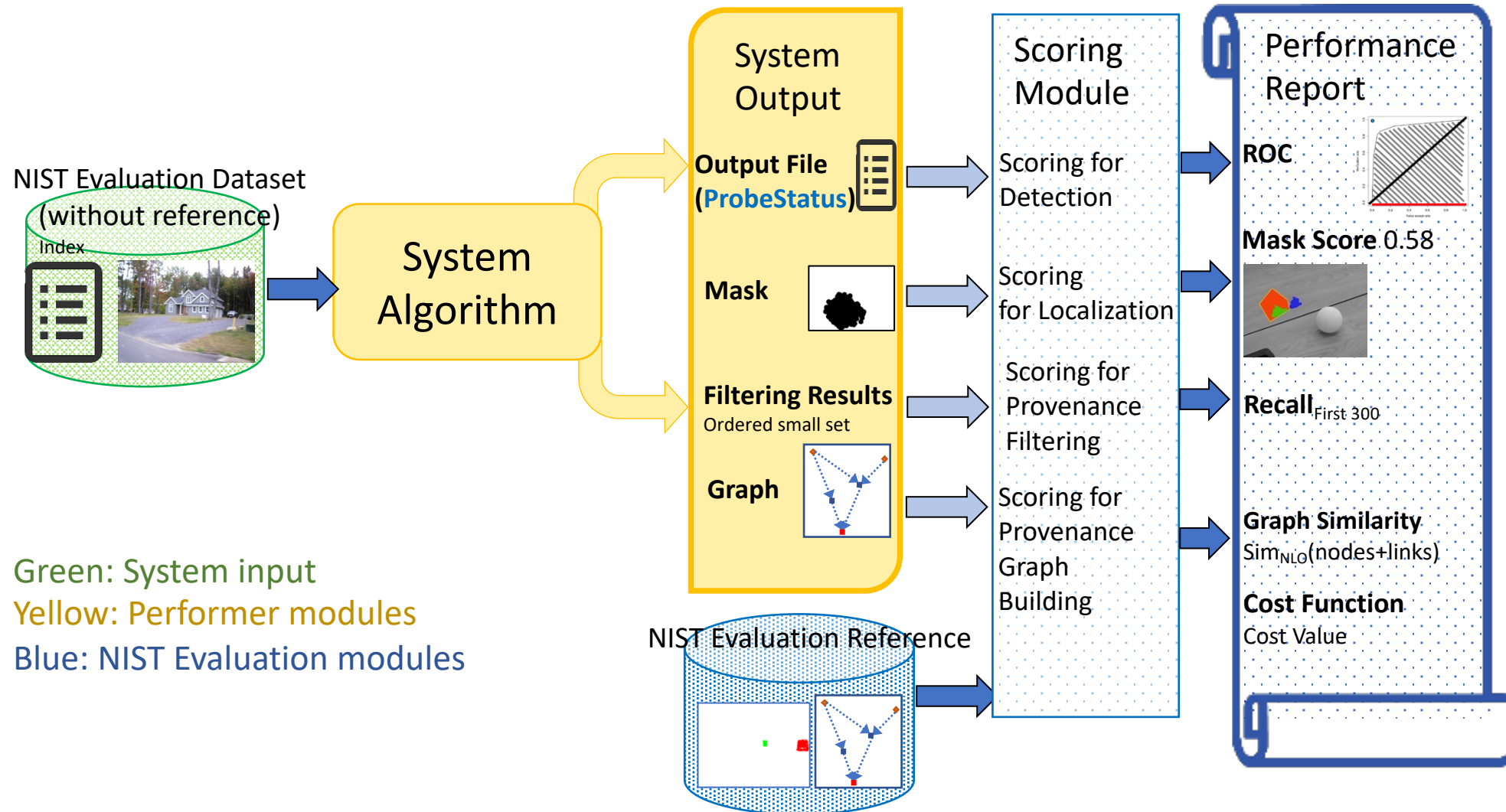


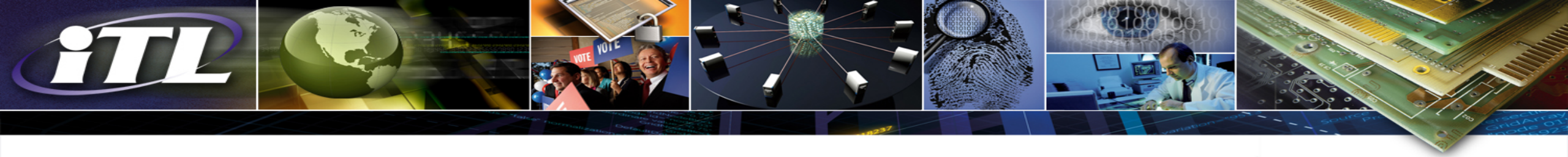
NIST MFC Evaluation Design and Approaches

- Evaluation tasks
 - Task design philosophy
 - 6 task definitions
- Evaluation data with reference ground-truth
 - Training, development, special study, and evaluation datasets
- Evaluation scoring metrics and software
 - Metrology for holistic vs. “Opt-In” evaluation
 - Factor Analysis: selective scoring, special studies
- Evaluation approaches and platforms
 - Take Home vs. Container evaluation
 - Open evaluation vs. Sequestered evaluation



Media Forensic Challenge Evaluation Infrastructure





Media Forensic Challenge Evaluation Tasks

- Task design philosophy
- Task definition

Media Forensics Challenge Evaluation Task Overview

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?

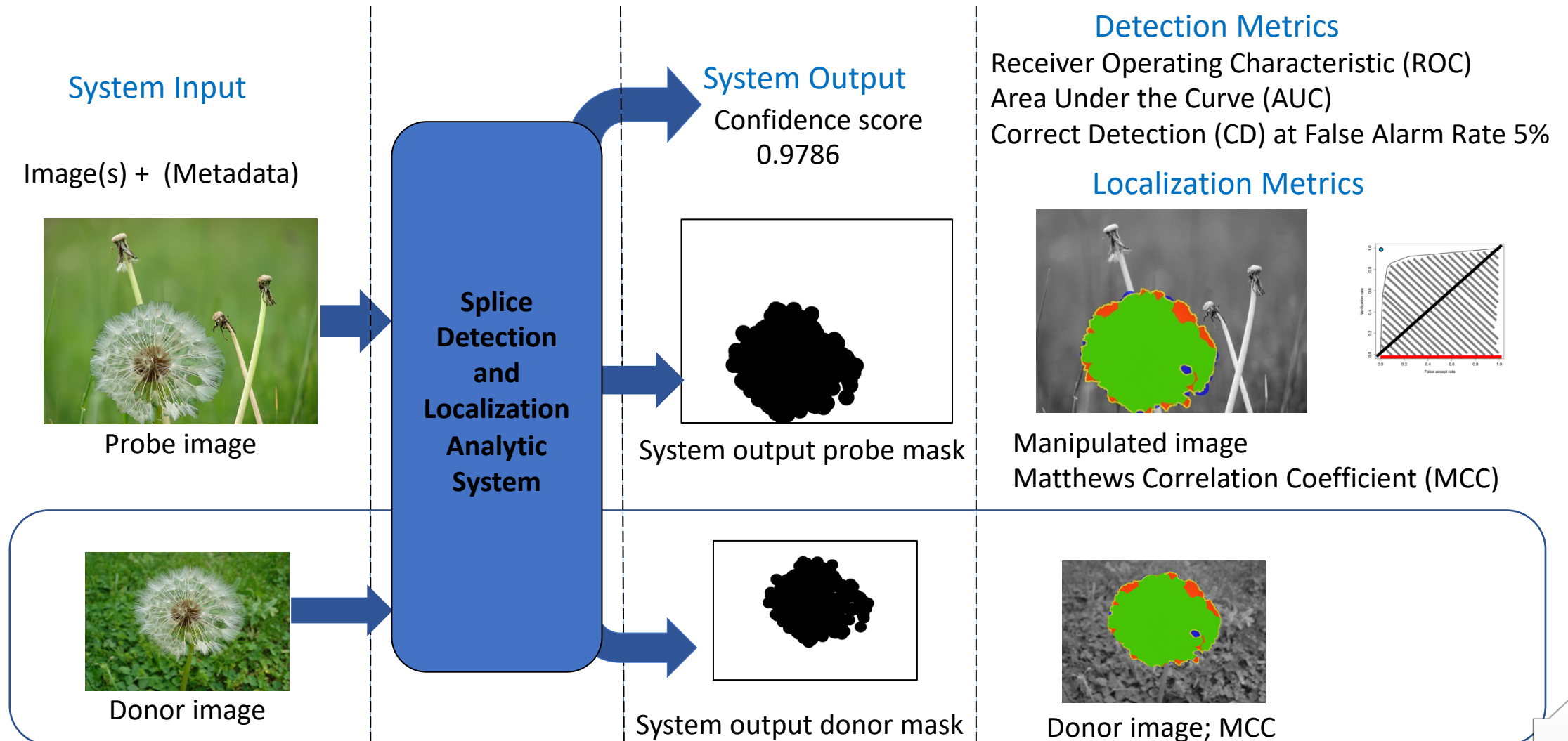
File+Event

Event Verification:

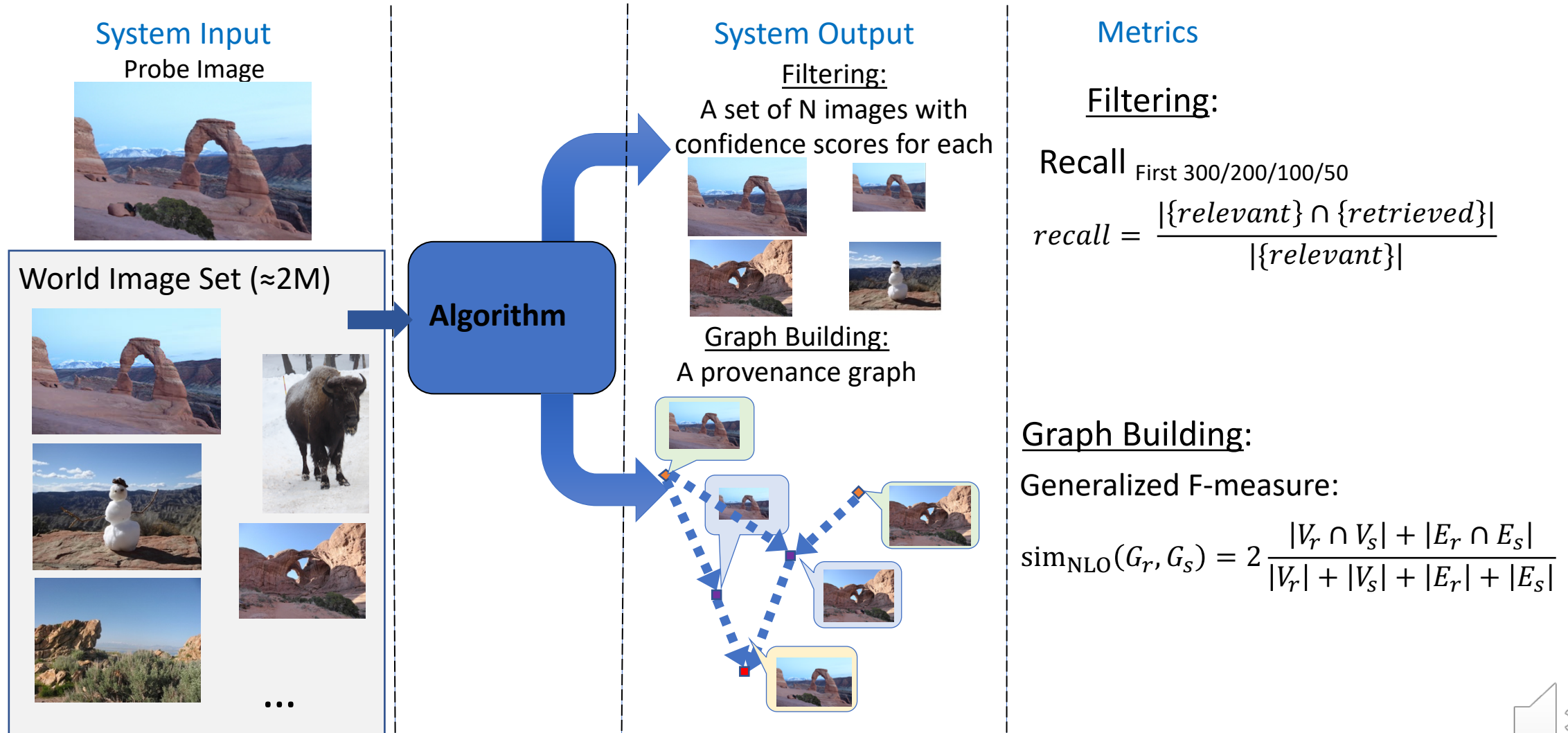
Was an image capture during a known event?



Image/Splice Manipulation Detection and Localization



Provenance Filtering and Graph Building



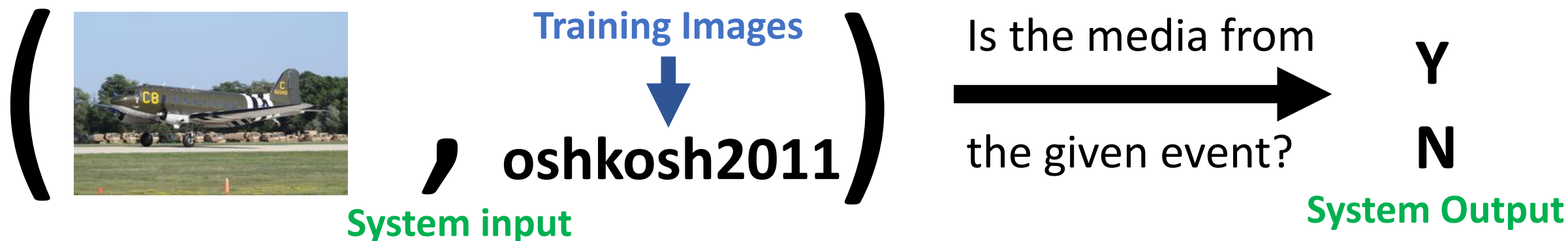
Camera ID Verification Task

- Task: Determine if a probe is from a claimed camera; If manipulated, localize the changes.



Event Verification Task

- Task Definition: Given a collection of images and videos from the event, determine if a probe is from the claimed event.



- 12 events (air show, hurricane, marathon, blizzard, etc.)



oshkosh2011



oshkosh2010



hurricane_katrina



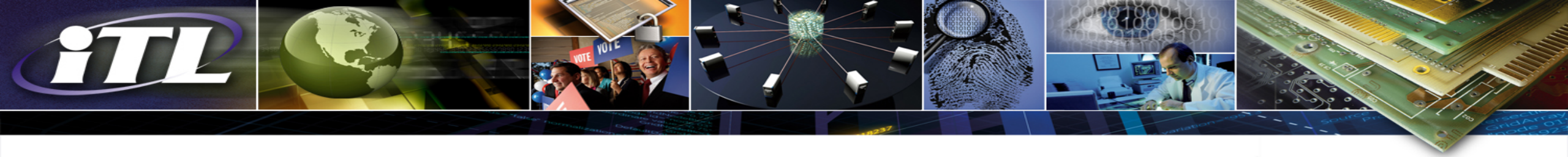
hurricane_ike



berlin_marathon



chicago_blizzard



Media Forensic Challenge Evaluation Data

- Manipulation reference collection design
- Data collection
- Evaluation data production
- MFC evaluation dataset

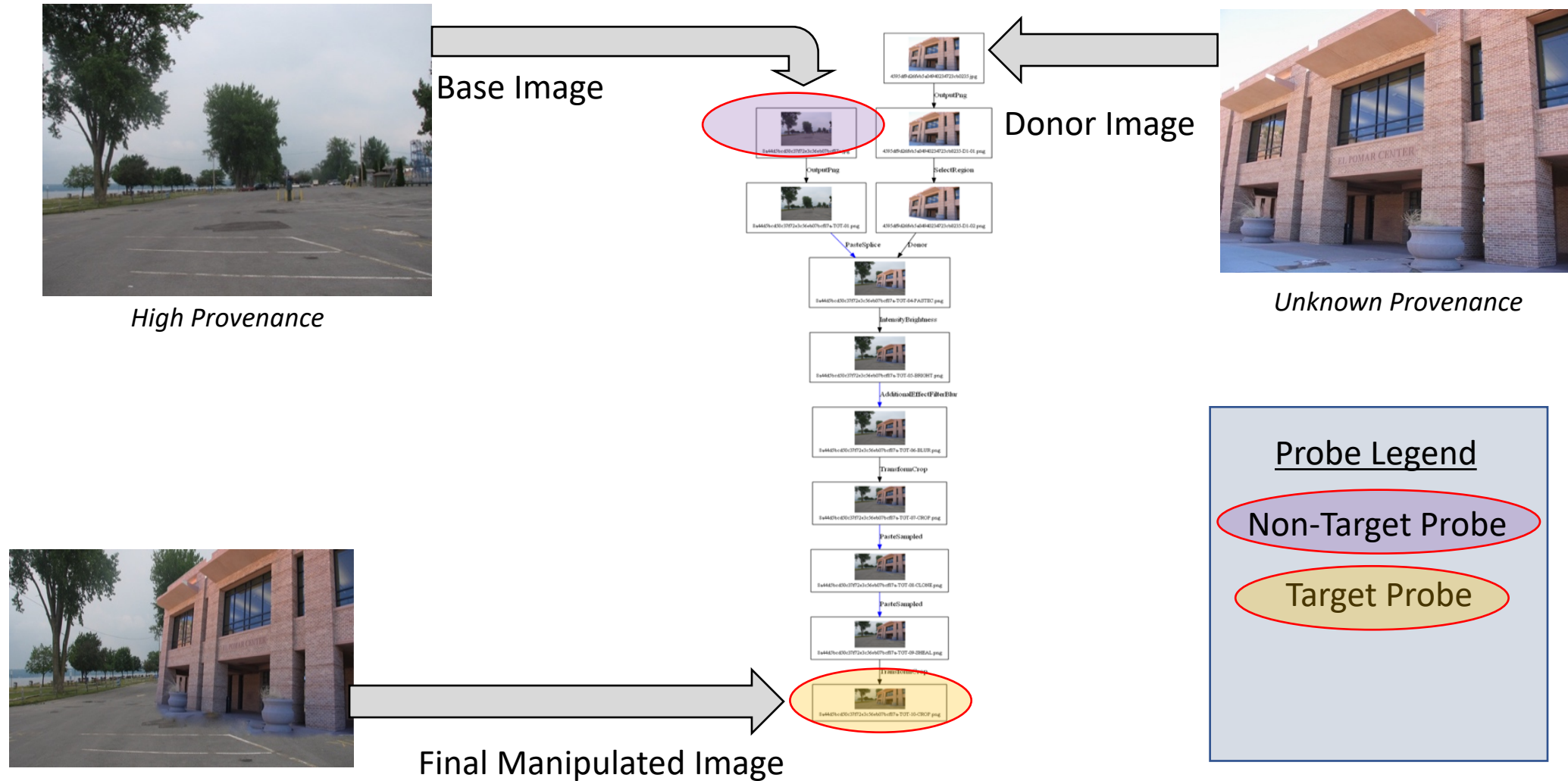


Manipulation Reference Collection Challenge

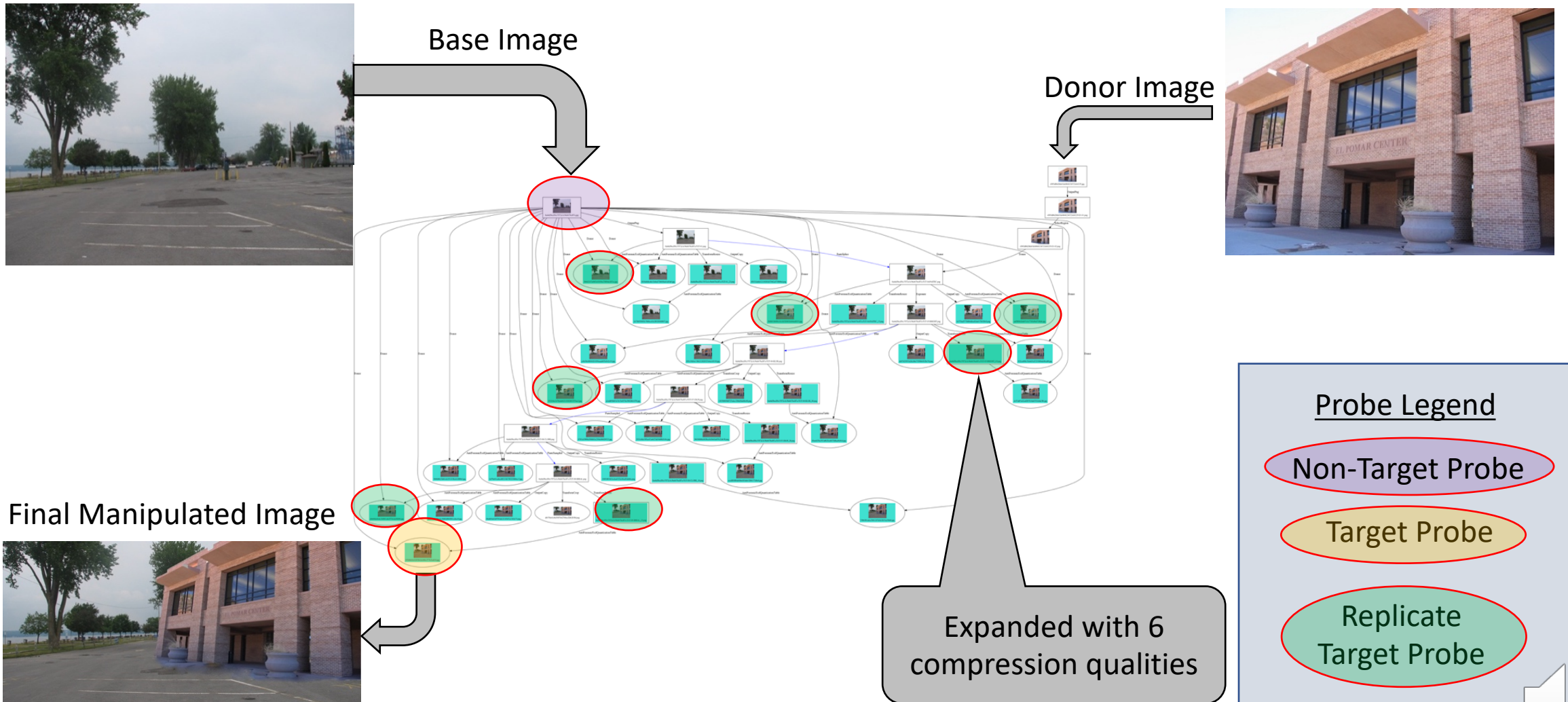
- Post manipulation interpretation is nearly impossible
- Effective evaluations require knowledge:
 - Where the manipulation occurred
 - What tool was used
 - What operation was used
 - Semantics of the manipulation: remove vs. add
- MFC Approach (human and machine annotation):
 - Record steps with PAR's Journaling Tool
 - Automate collection of manipulation region mask



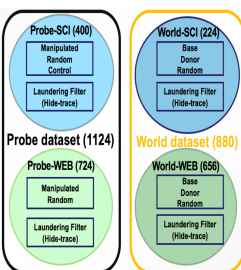
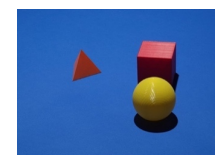
Manipulation Journaling Tool



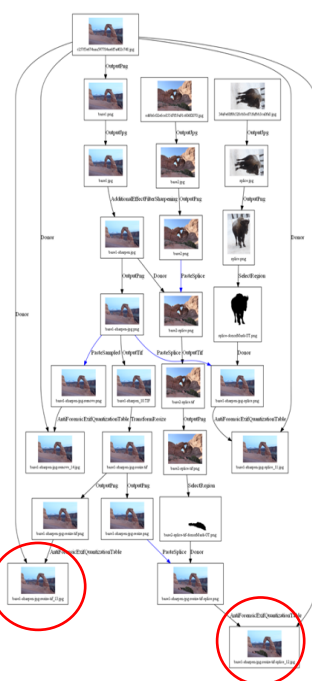
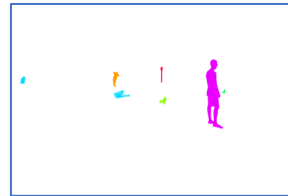
Manipulation Journaling Tool (Extended Journal)



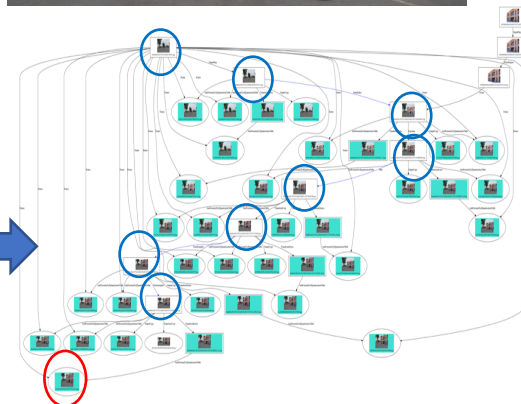
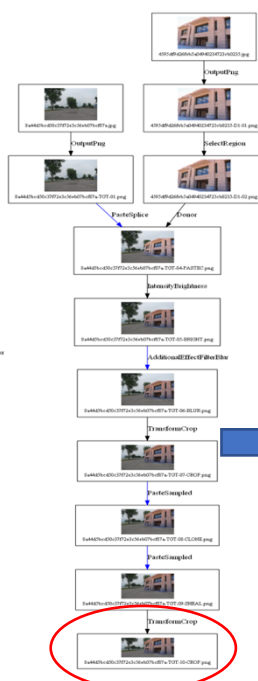
MFC Evaluation Dataset History



**Kick-off
2016 Dataset**

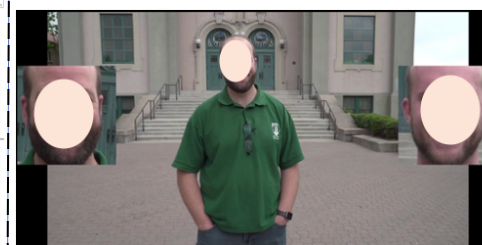
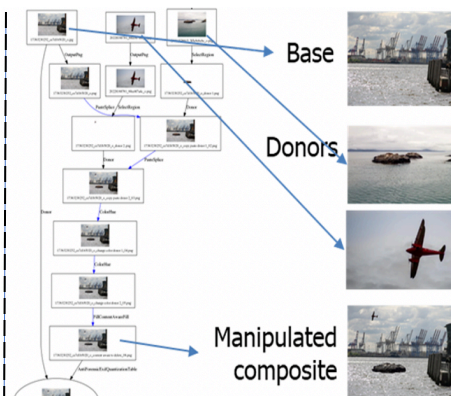


**Provenance
Auto Journaling Tool (JT)
Nimble Challenge 2017**



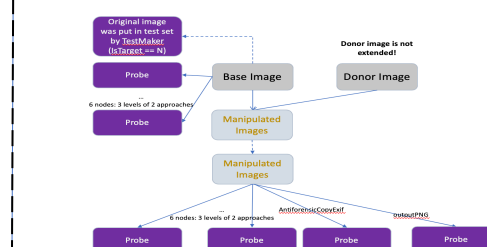
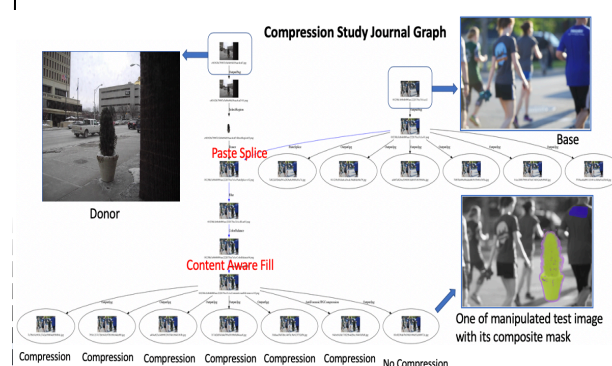
**New Manipulations
(CGI, Recapture, ...)
Extended JT, AutoJT**

MFC 2018



- Camera ID Eval. datasets
- Video Temporal Spatial
- Additional Manipulation Operations (GAN etc.)
- Extended JT, AutoJT

MFC 2019

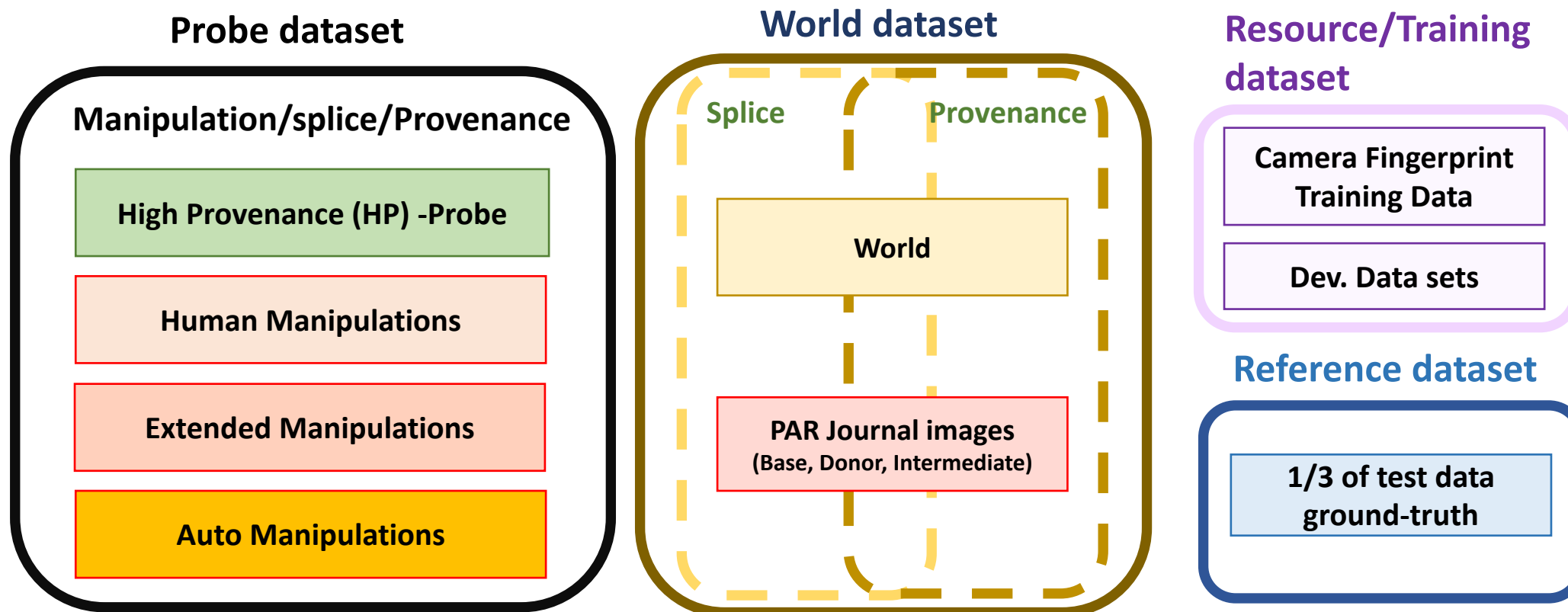


- Special study data**
- Compression
 - Global Blur
 - Single Operation
 - Social Media Laundering
 - Frame Drop/Dup.

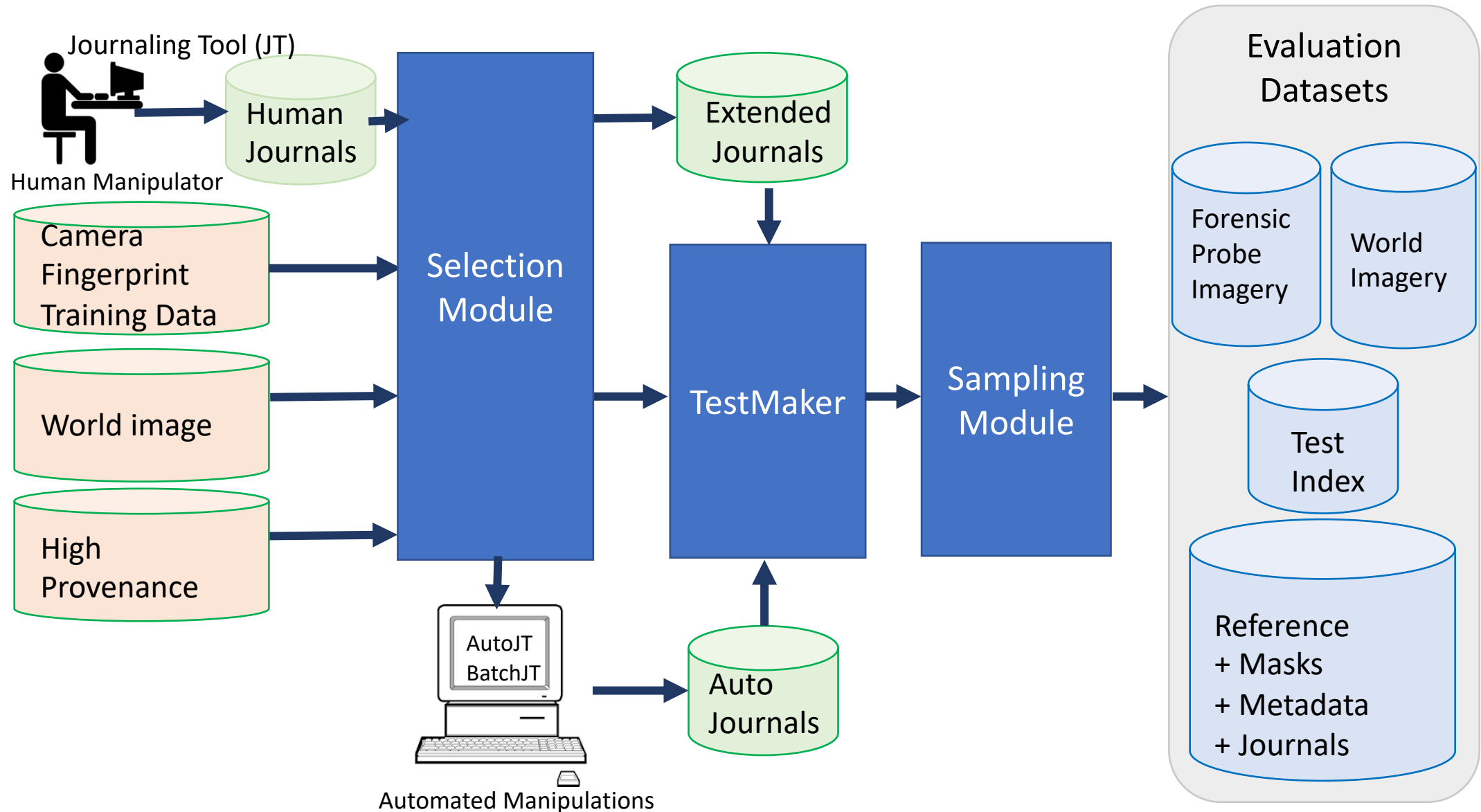
MFC 2020



MFC General Data Collection Overview

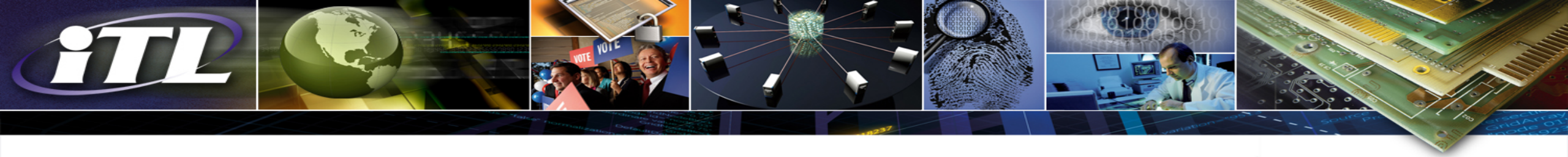


Evaluation Dataset Production Infrastructure



MFC Evaluation Dataset Summary

Task(s)	NC17 EP1	MFC18 EP1	MFC19 EP1	MFC20 EP1
Image	4K	17K	16K	20K
Video	0.36K	1K	1.5K	2.5K
Provenance	1K Probe 1M World	10K Probe 1M World	9.4K Probe 2M World	5.9K Probe 2M World



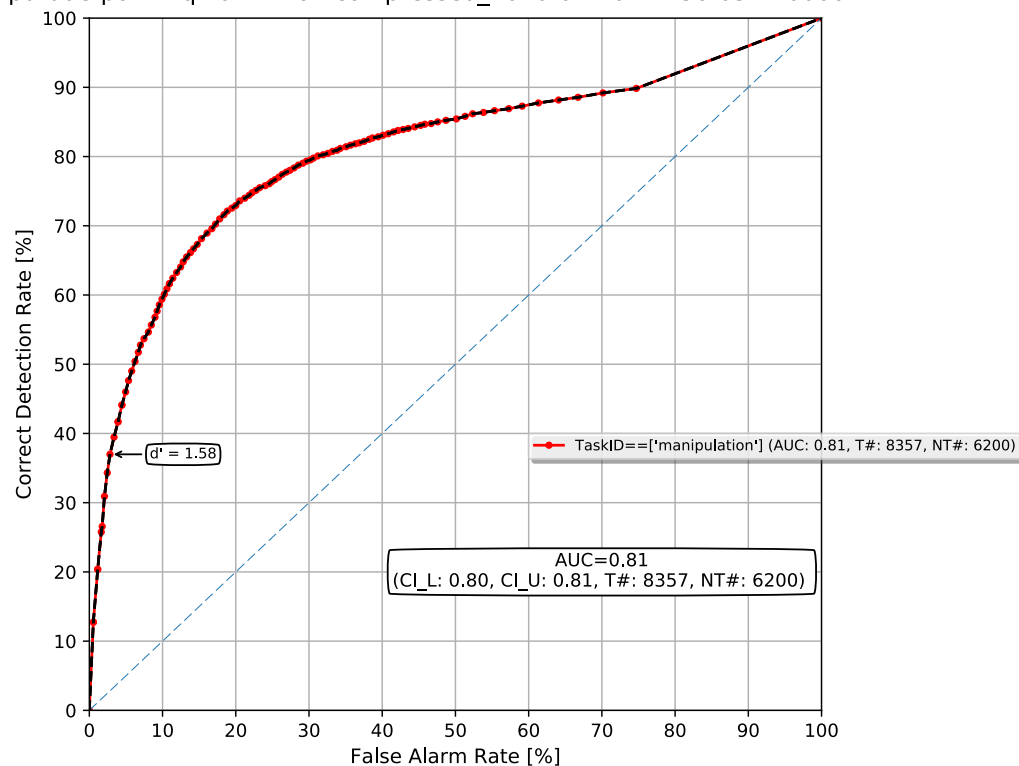
Media Forensic Challenge Scoring Methodology

- Metrics
- Holistic vs. Opt-In Technologies
- Factor Analysis with Selective Scoring
- Special Studies

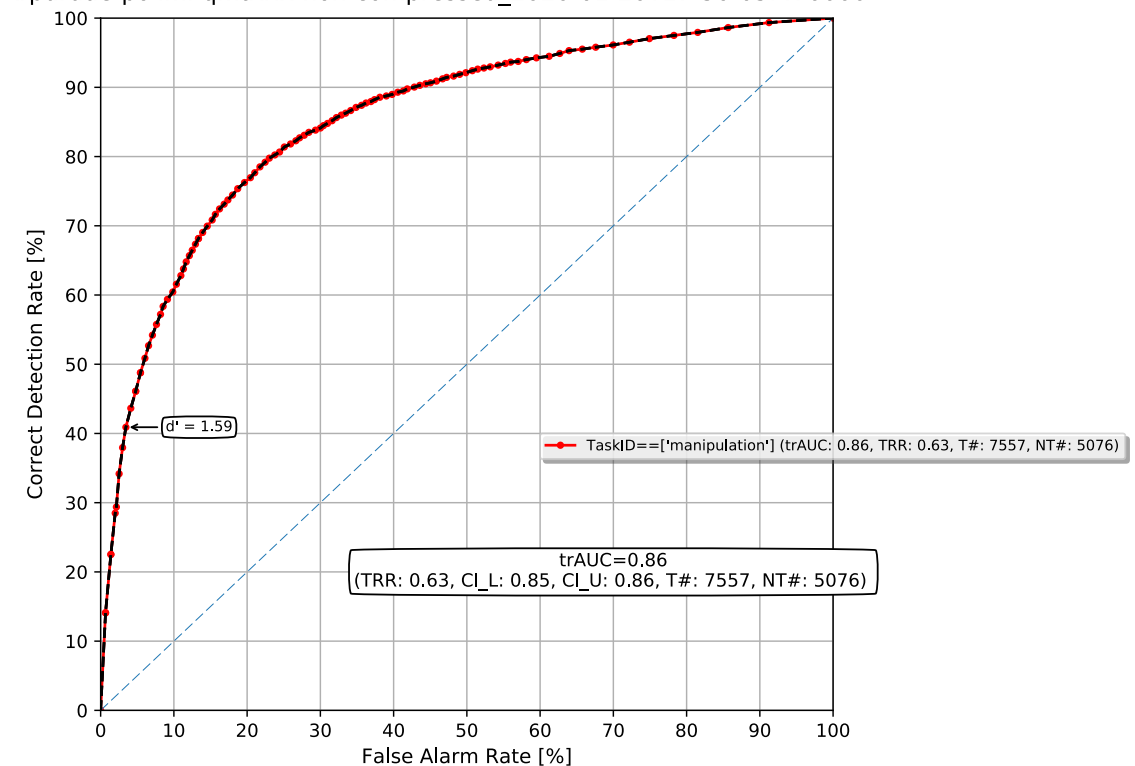


Holistic vs. Opt In Technologies

- Some media forensic systems only response to a certain media
 - e.g., jpeg compression systems should not respond if input is not in jpeg format



(a) Holistic



(b) Opt In



Challenges and Approaches

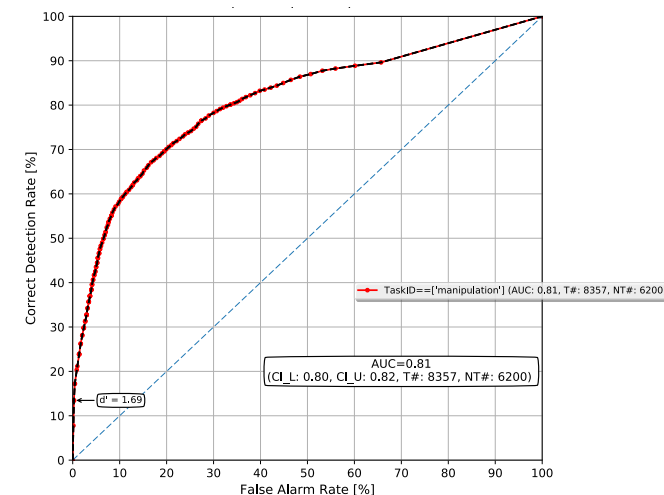
- Core challenge: curse of dimensionality
 - Media space (image/video/audio, camera/scanner)
 - Manipulation space (manipulator, manipulation operations and software)
 - Anti-forensic technology space
- MFC data production approaches:
 - Human manipulation journals (realistic)
 - Automatic manipulation journals (reduce cost)
 - Extended manipulation journal (special study)
- MFC performance analysis approaches:
 - Overall manipulation performance
 - Selective Scoring Analysis
 - Special Study Analysis



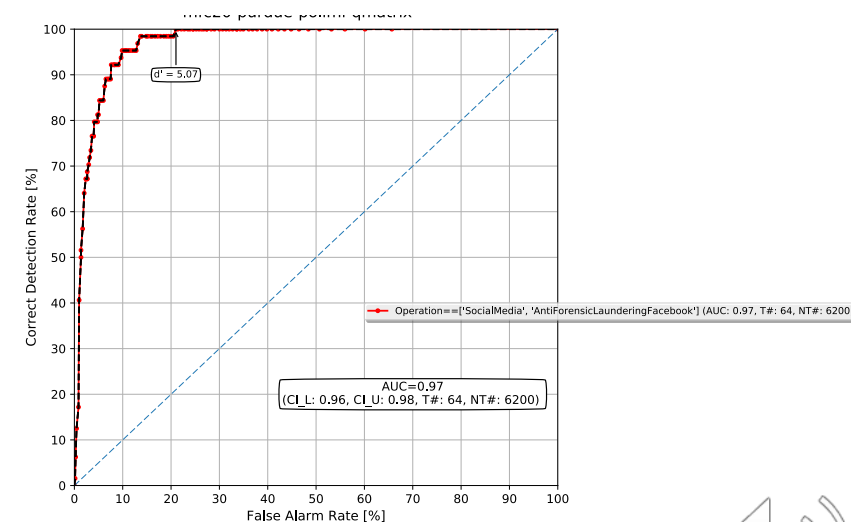
MFC20 Image Selective Scoring

Name	Definition
Splice	Any operation that takes a region from a donor media and pastes it into a probe
Clone	Pixels are sampled from the image and pasted back in different area of the image
Splice/Clone	Pixels are pasted within or between the images
Crop	Outer pixel regions from a probe image are removed
Resize	Image dimensions from a probe image are changed
Intensity	A range of intensity pixel values is changed
Antiforensic	Any techniques that erase processing history of image manipulations
Antiforensic-PRNU	Any techniques that use PRNU
Antiforensic-CFA	Any techniques that use CFA
Social Media	Any techniques that use social media related operations
Global Blur/Smooth	Any techniques that use a low-pass filter (globally) to remove outlier pixels (e.g., noise)
Local Blur/Smooth	Any techniques that use a low-pass filter (locally) to remove outlier pixels (e.g., noise)
GAN	Any operations that use GAN-based techniques locally/globally
NonGAN-CGI	Any operations that use non-GAN CGI
Distortion	Deformation of images
Remove	Remove a set of pixels.
Face Manipulation	Any manipulation done to a face.
All	All data without selective scoring

Figure: The same system performances on two evaluation conditions



(1) All Manipulations (AUC = 0.81)



(2) Social Media Laundering (AUC = 0.97)



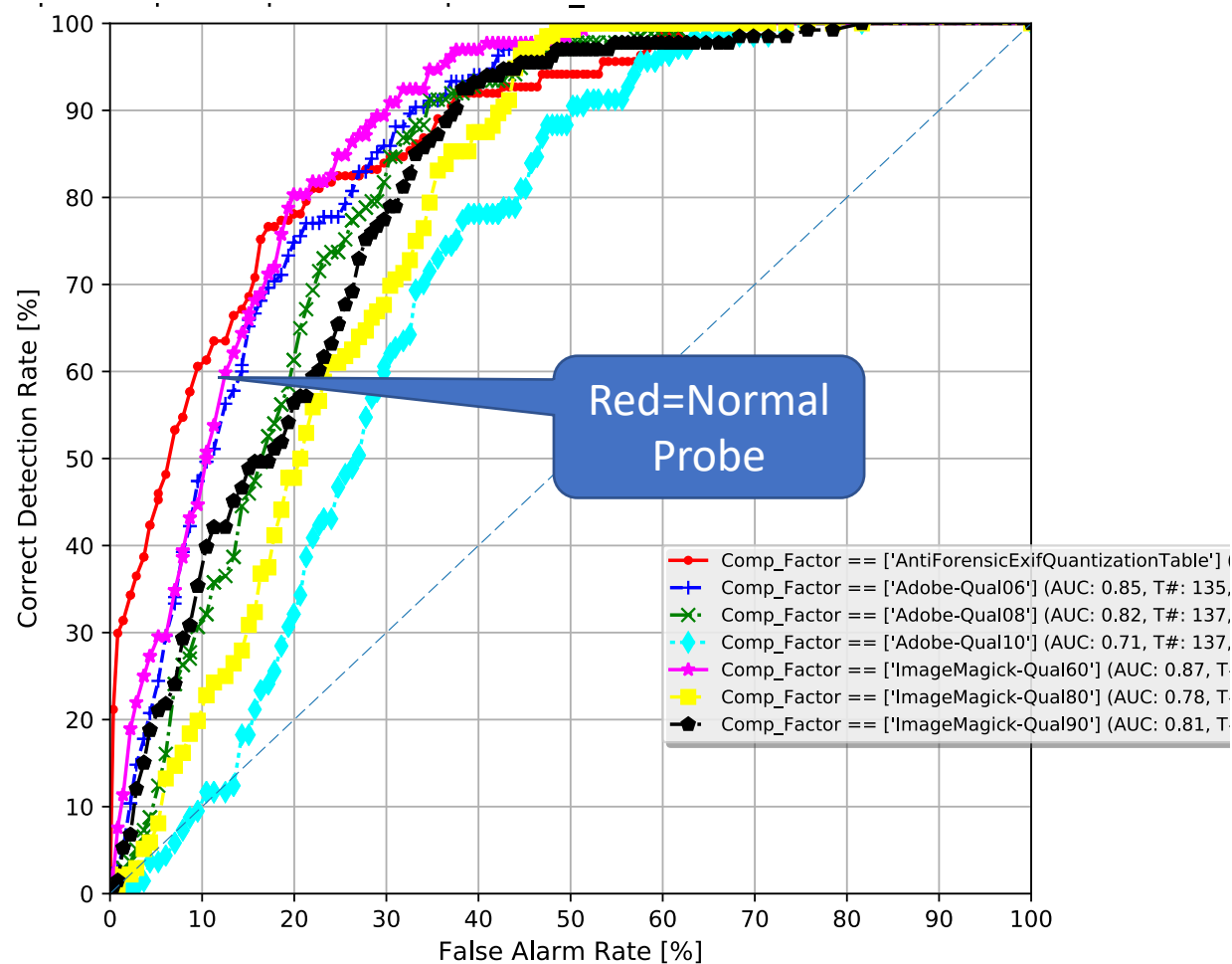
Factor Analysis: Special Studies

- Special Study approach
 - Build specific data sets to answer specific performance assessment questions.
 - Enables two new views of performance assessment: Operation Only Detection and Facet Detection
- MFC20 Special Studies
 - (Image) Compression
 - (Image) Global Blur
 - (Image) Social Media Laundering – Image
 - (Image) Single Operation (Paste-Splice)
 - (Video) Frame Drop/Duplication
 - (Video) Social Media Laundering - Video

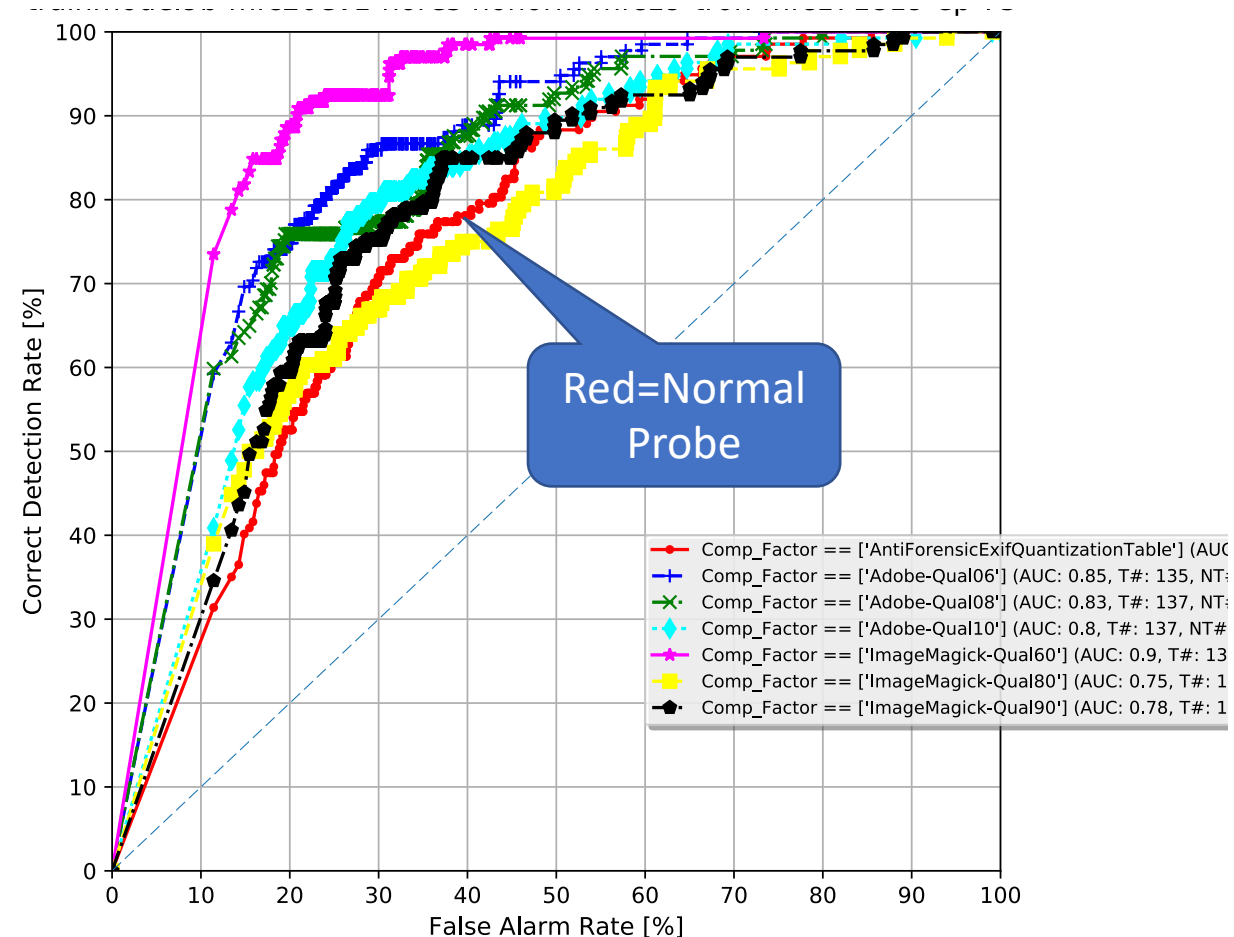


Compression Study Example:

7 Conditions:
 1: EXIF Copy
 3: Adobe Levels (6,8,10)
 3: ImageMagick Levels (60,80,90)



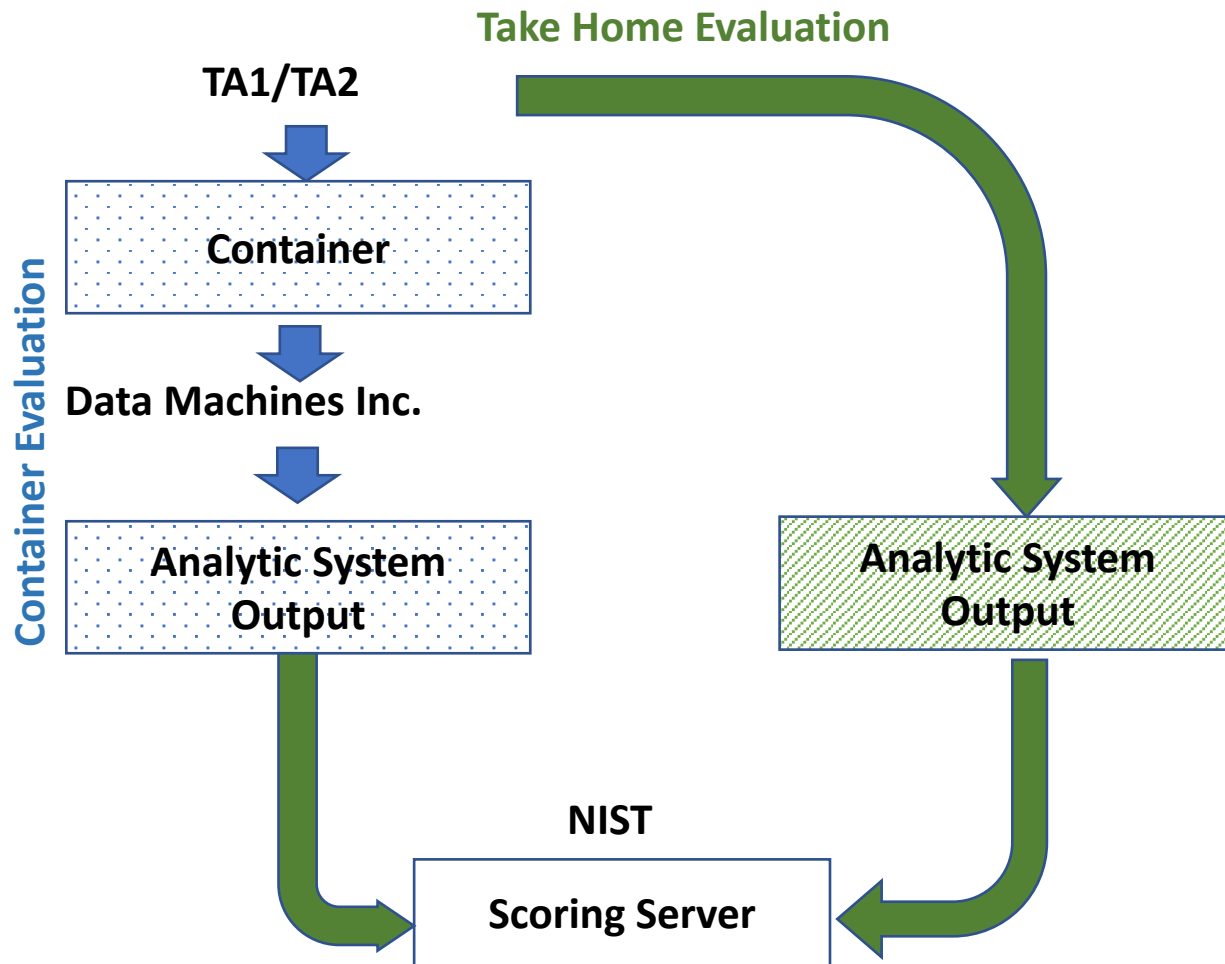
#2458



#2516



Take Home vs. Container Evaluations

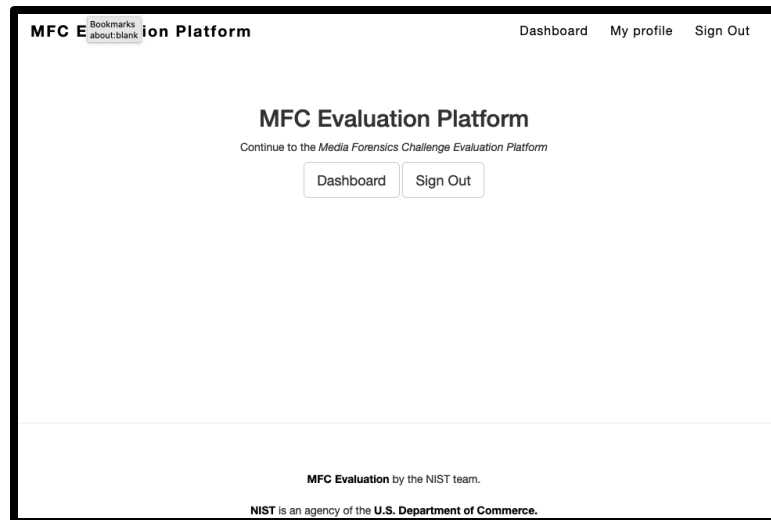


- Take home
 - NIST releases test data to performer
 - Performer submits system output
- Container
 - Performer submits system containers
 - Evaluation team run system on sequester evaluation data

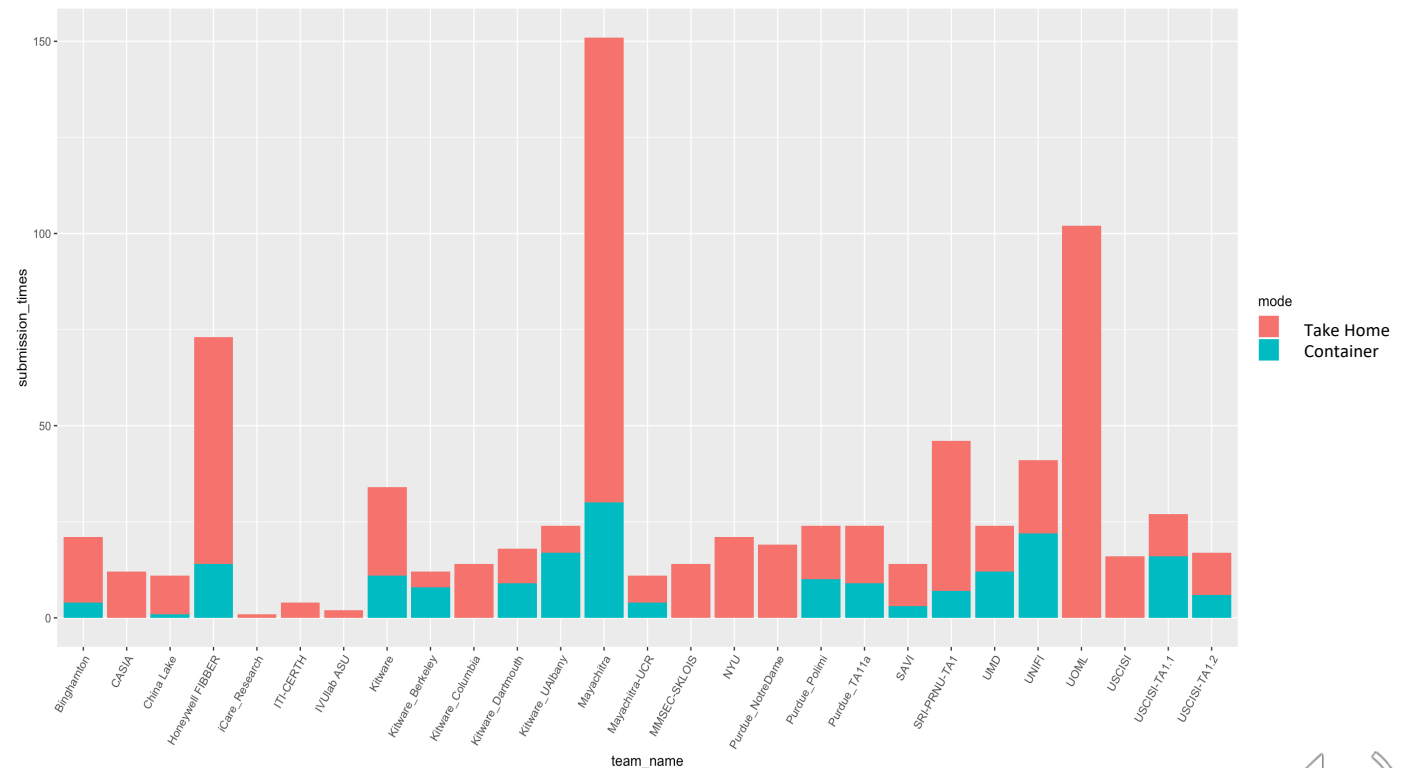


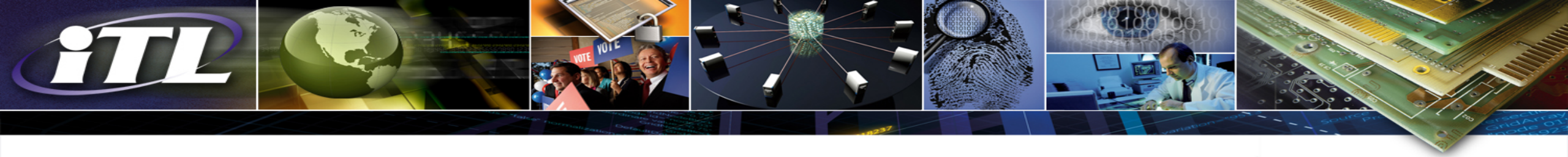
NIST MFC Scoring Server

- Performers had access to an automated scoring server
 - 65 MFC Data sets
 - Supports 6 evaluation tasks
 - Thousands submissions
 - 12K scoring runs



Distribution of Submissions per Team
TakeHome (Orange) and Container (Blue)





Media Forensic Challenge Result Reports

- MFC20 Results
- Cross-Year comparison



MFC20 Results on Evaluation Part 1 Dataset

Image Manipulation Detection

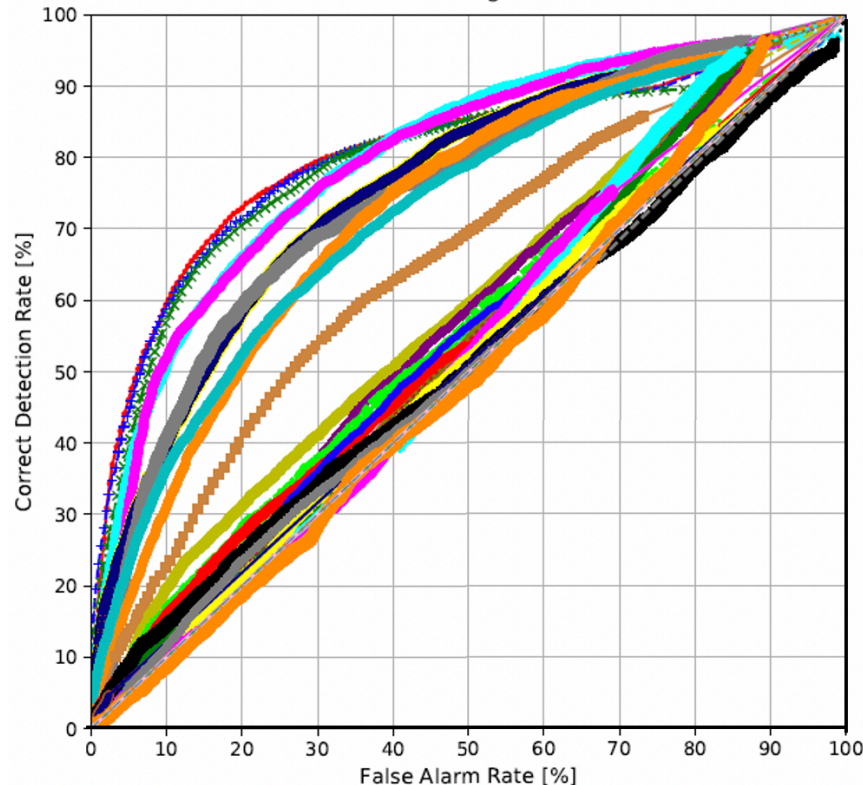


Figure: ROC on MFC20 EP1 Image
Highest AUC = 0.8, CD@0.05FA = 0.44

Video Manipulation Detection

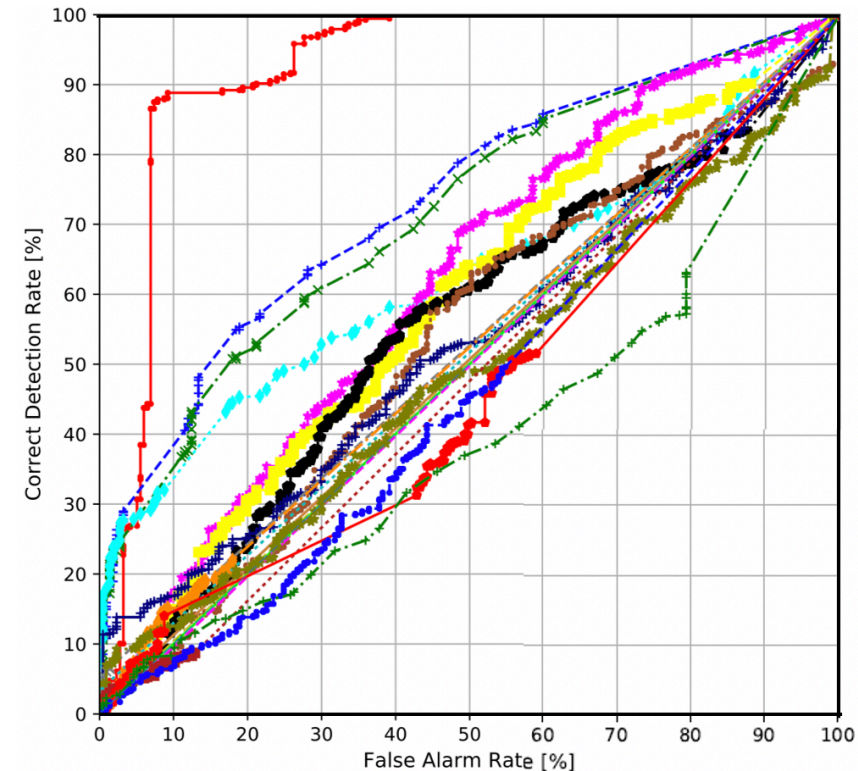
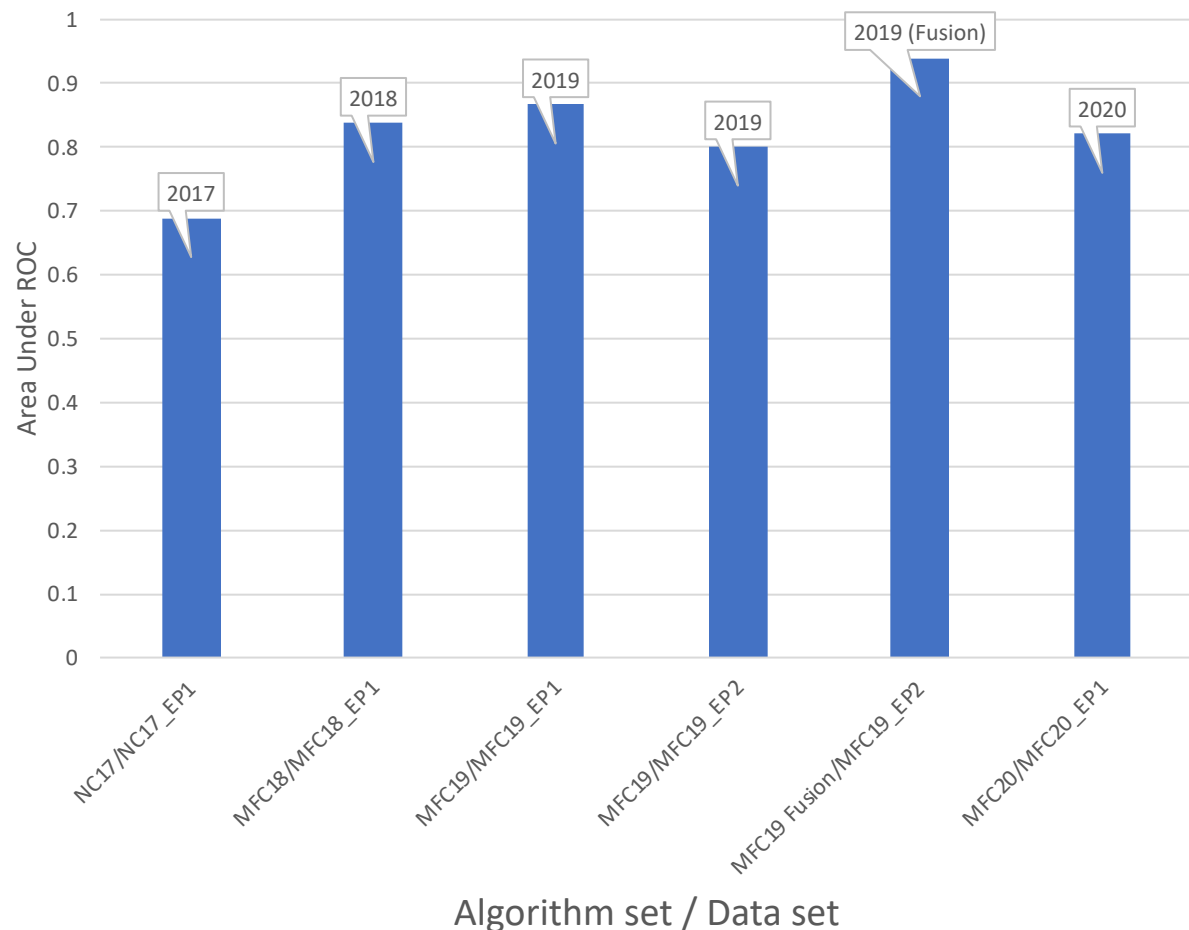


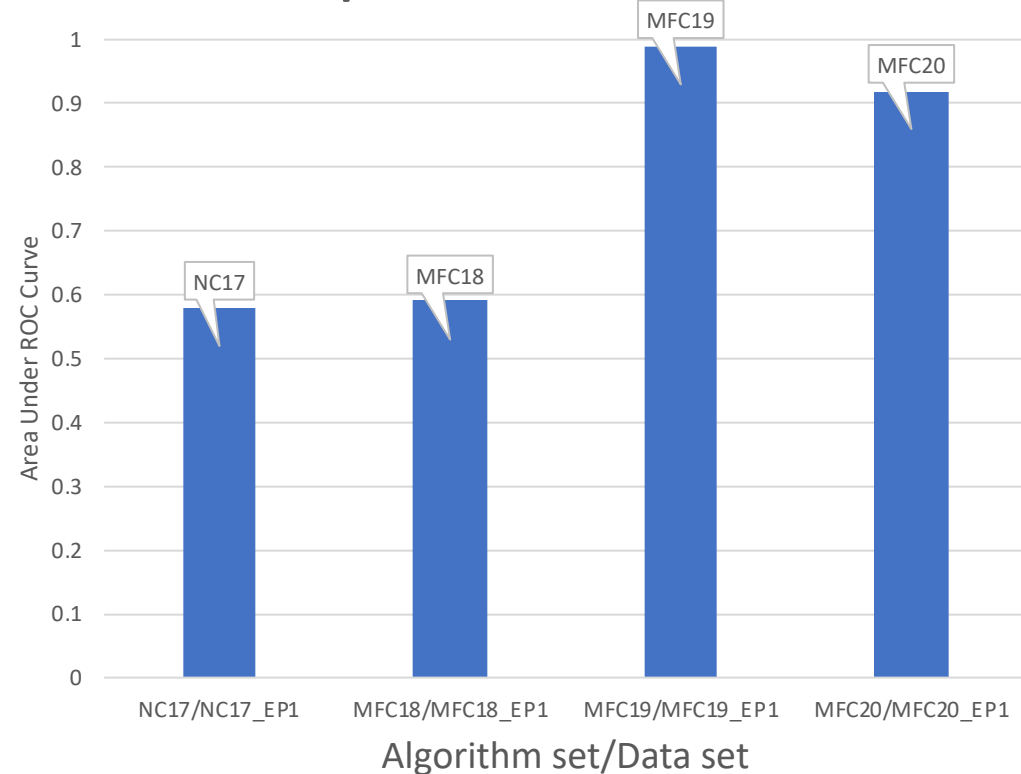
Figure: ROC on MFC20 EP1 Video
Highest AUC = 0.92, CD@0.05FA = 0.27

Year-to-year improvements in detection performance

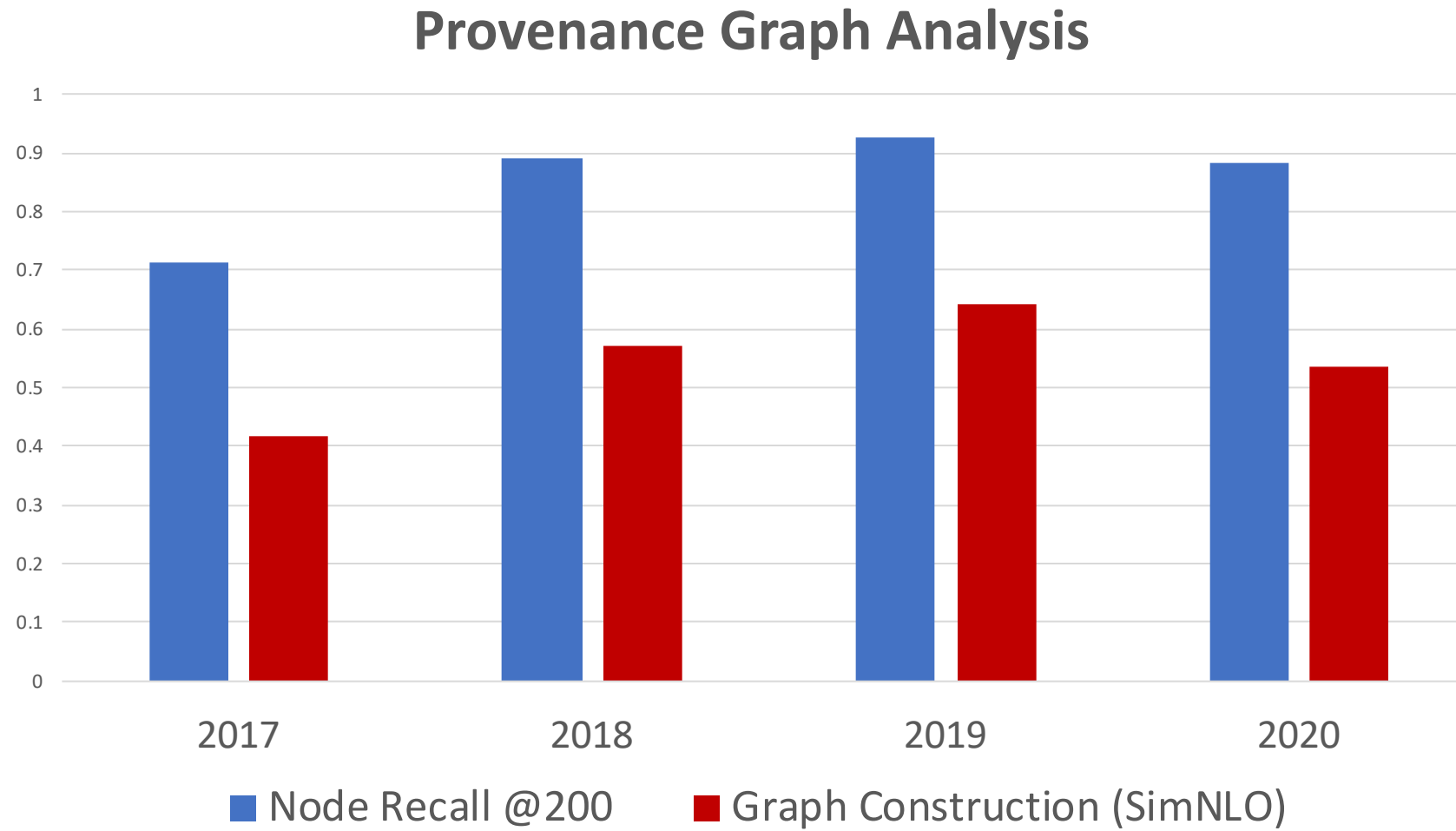
Image Manipulation Detection Performance



Video Manipulation Detection Performance



Year-to-year in provenance task performance



NIST MFC resources

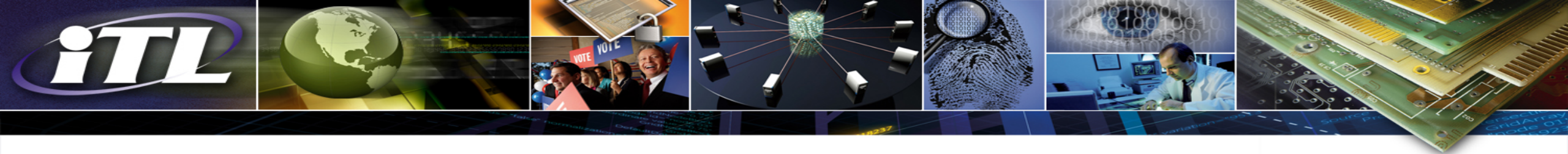
- MFC open evaluation datasets
 - NC16 Kickoff, NC17 Evaluation Part 1 (EP1), MFC18 EP1, MFC19 EP1, MFC20 EP1
 - Till now, we have released our datasets to about **230** individuals, **140** organizations, and **26** countries and regions worldwide
- MediScore
 - Git: <https://gitlab.mediforprogram.com/jfiscus/MediScore>
- NIST MFC scoring server
 - MFC20: 2.6K submissions and 12K scoring runs
 - Leaderboard version (coming soon!)



Assured Autonomy Evaluation: Initial Thoughts

- What is the state-of-the-art of this domain?
 - Baseline performance
 - Different stages focus on different types of evaluation measurements
- What are the key evaluation metrics in current stage?
 - Initial stage (idea/algorithm/preliminary results in research lab)
 - Potential capability, algorithm performance, ROC etc.
 - Prototype stage (target on real-world applications)
 - System hardware/software requirements, processing speed (real-time?)
 - Product stage (interface, human factors)
 - safety, reliability, responsibility, usability
- How to build the evaluation to drive the research directions?
 - Task design
 - Benchmark evaluation dataset design
 - Evaluation infrastructure design





Thank You for Your Attention!

- NIST Media Forensic Challenge (MFC) website: <https://www.nist.gov/itl/iad/mig/media-forensics-challenge>
- NIST Media Forensic team contact email: mfc_poc@nist.gov
- Presenter email: haiying.guan@nist.gov

