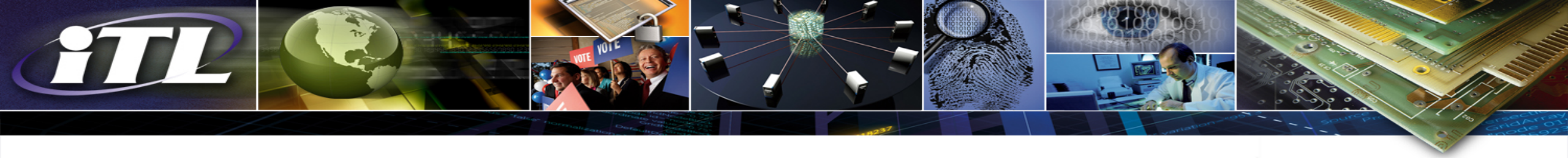


2020 MediFor Challenge Evaluation Overview

Jonathan Fiscus (Co-PI), Dr. Haiying Guan (Co-PI), Dr. Yooyoung Lee,
Dr. Amy Yates[+], Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah,
Dr. Xiongnan Jin

Multimodal Information Group, [+] Image Group
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology (NIST)

April 21-25, 2020

Disclaimer

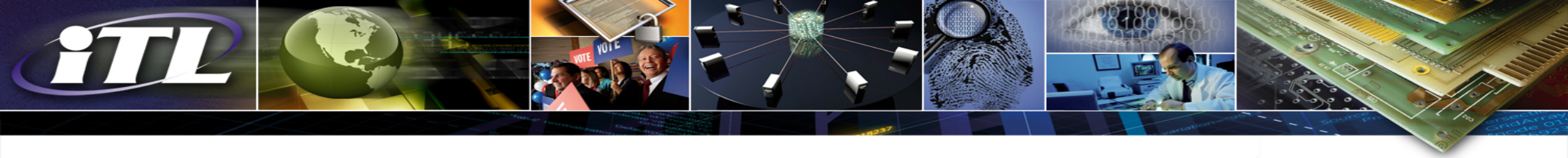
- Certain commercial equipment, instruments, software, or materials are identified in this article in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the equipment, instruments, 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 the Test and Evaluation Team!

- Program Administration
  - DARPA Media Forensic (MediFor) Team
- TA3 Data Production and Curation
  - PAR Government
  - National Center for Media Forensics, University of Colorado Denver
  - RankOne
  - Rochester Institute of Technology
  - Drexel University
  - University of Michigan
- Container Execution
  - Data Machines Incorporated
- MediFor Demo System
  - Next Century
- Contracting
  - Air Force Research Lab
- Evaluation Design and Implementation
  - NIST MediFor Team

Media Forensic Challenge (MFC) Overview Outline

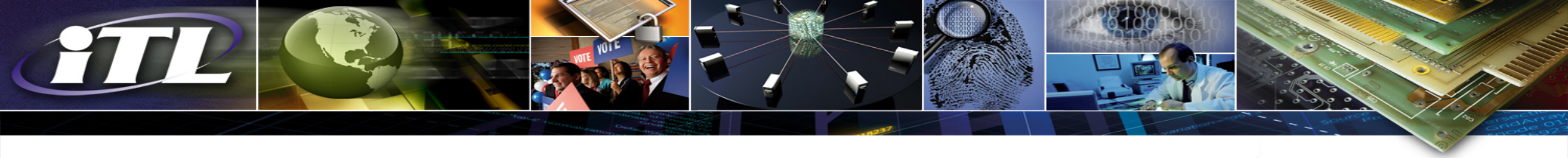
- MediFor evaluation requirements and challenges
- MediFor common evaluation tasks
- Holistic vs. ”Opt-In” technologies
- Manipulation journaling for data production
- Understanding System Performance with Factor Analysis
- MediFor Data Set Summary
- MFC20 results preview



MediFor Evaluation Requirements and Challenges

Media Forensic Evaluation Requirements and Challenges

- 4th MediFor Program Evaluations 2017-2020
- Program requirements from the initial kickoff
  - Support as many common evaluations as possible
  - Support evaluation of integrity indicators (TA1) via a selectable menu
  - Support evaluation of integrity reasoning over indicators (TA2)
  - Understand system performance
- Fundamental metrology challenges from the modest starting point
  - Metrology for holistic vs. “Opt-In” media forensic systems
  - Manipulation journaling for data production
  - Factor Analysis: selective scoring vs. Special collections
  - Take Home vs. Container evaluations



Six Common MediFor Evaluation Tasks

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 Manipulation Detection and Localization

System Input

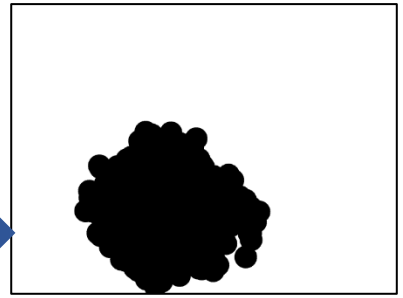
Image(s) + (Metadata)

Probe image

Image Detection and Localization Analytic System

System Output

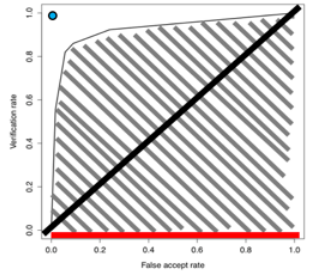
Confidence score
97.86



System output probe mask

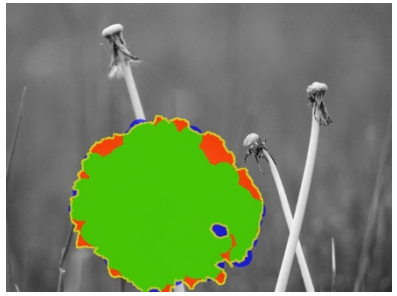
Detection Metrics

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



Localization Metrics

Matthews Correlation Coefficient (MCC)
Symmetric Range: [-1:1]
- 1 denotes perfect accuracy
- 0 denotes no correlation
- -1 denotes perfect inaccuracy.



$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

Splice Manipulation Detection and Localization

System Input

Image(s) + (Metadata)

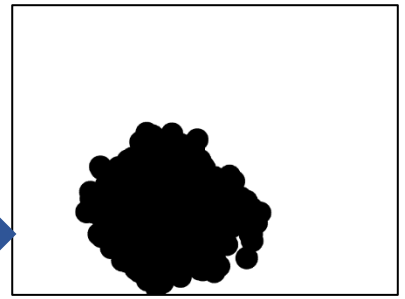
Probe image

Donor image

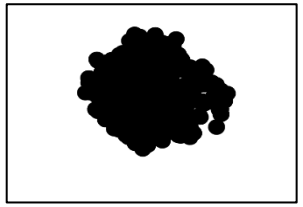
Splice Detection and Localization Analytic System

System Output

Confidence score
97.86



System output probe mask

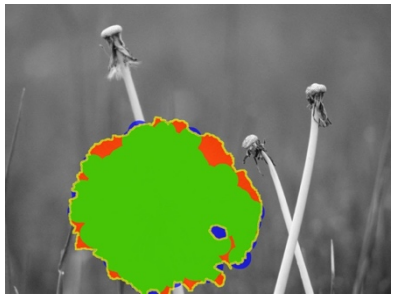


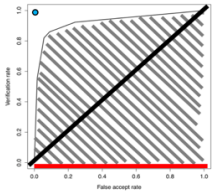
System output donor mask

Detection Metrics

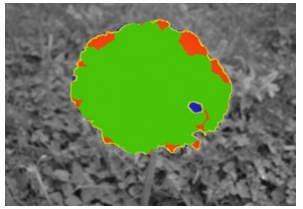
Receiver Operating Characteristic (ROC)
Area Under the Curve (AUC)
Correct Detection (CD) at False Alarm Rate 5%

Localization Metrics



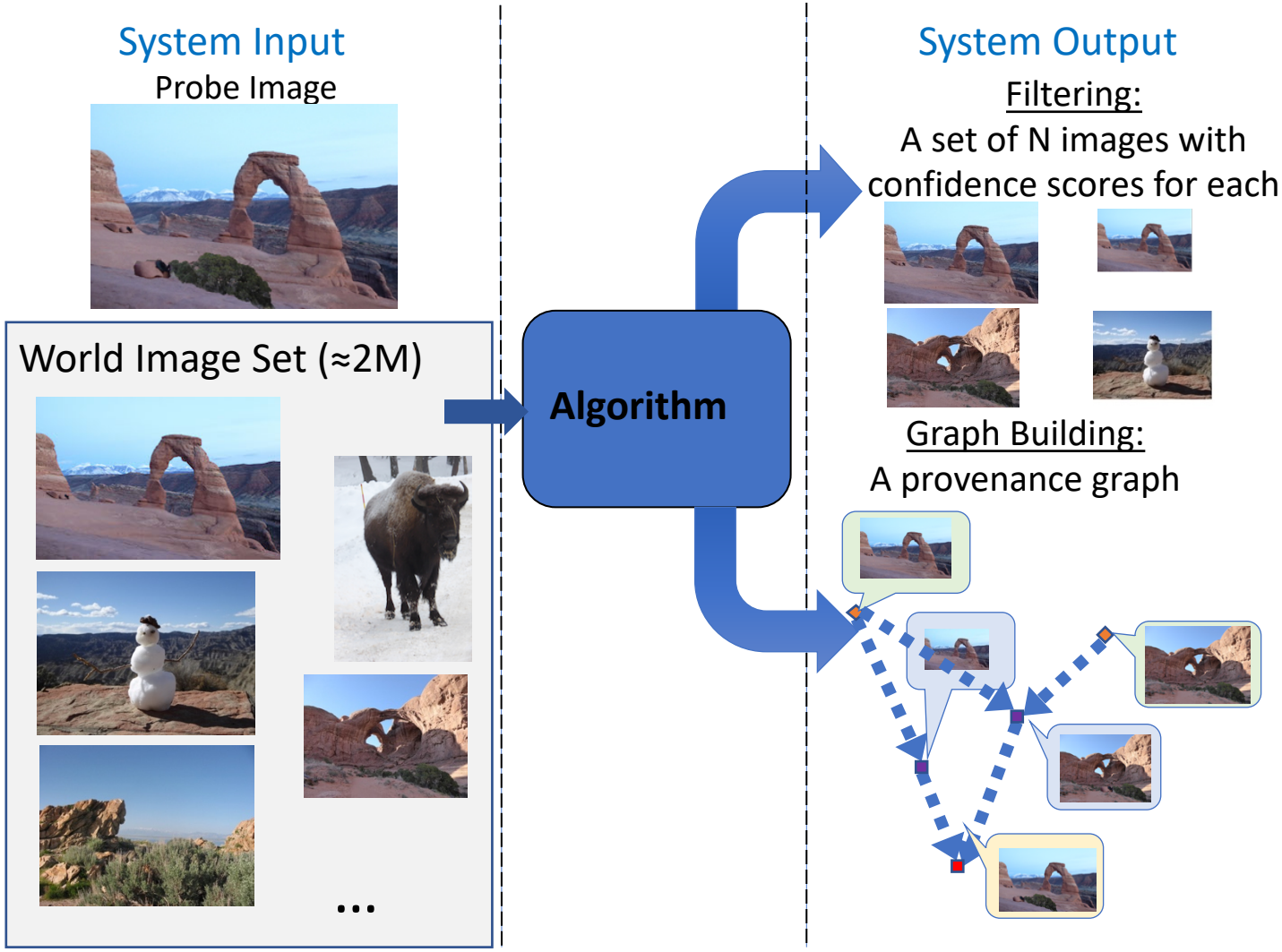


Manipulated image
Matthews Correlation Coefficient (MCC)



Donor image; MCC

Provenance Filtering and Graph Building



Metrics

Filtering:

Recall First 300/200/100/50

$$recall = \frac{|\{relevant\} \cap \{retrieved\}|}{|\{relevant\}|}$$

Graph Building:

Generalized F-measure:

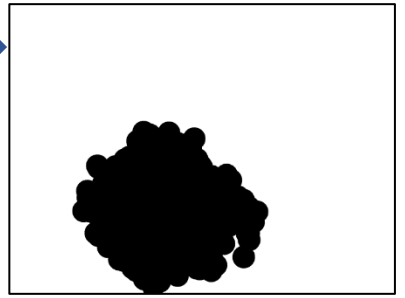
$$\mathrm{sim}_{\mathrm{NLO}}(G_r, G_s) = 2\frac{|V_r \cap V_s| + |E_r \cap E_s|}{|V_r| + |V_s| + |E_r| + |E_s|}$$

Camera ID Verification Task

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

System Output

Confidence score
97.86



System output probe mask

Event Verification Task

- Task Definition: Given a collection of images and videos from the event, determine if a probe is from the claimed event.
- MFC20 Events
  - 12 events: hurricane_matthew, hurricane_sandy, hurricane_harvey, hurricane_katrina, hurricane_Irma, hurricane_Ike, oshkosh2011, oshkosh2010, berlin_air_show, berlin_marathon, chinese_new_year_london_2014, chicago_blizzard_2011.


oshkosh2011


oshkosh2010


hurricane_katrina


hurricane_ike


berlin_marathon


chicago_blizzard_2011

Video Manipulation Detection and Temporal Localization

- Video Detection metrics
  - Receiver Operating Characteristic (ROC)
  - Area Under the Curve (AUC)
  - Correct Detection (CD) at False Alarm Rate (FAR) of 5%
- Video Temporal Localization
  - Metrics: Matthew Correlation Coefficient (MCC)

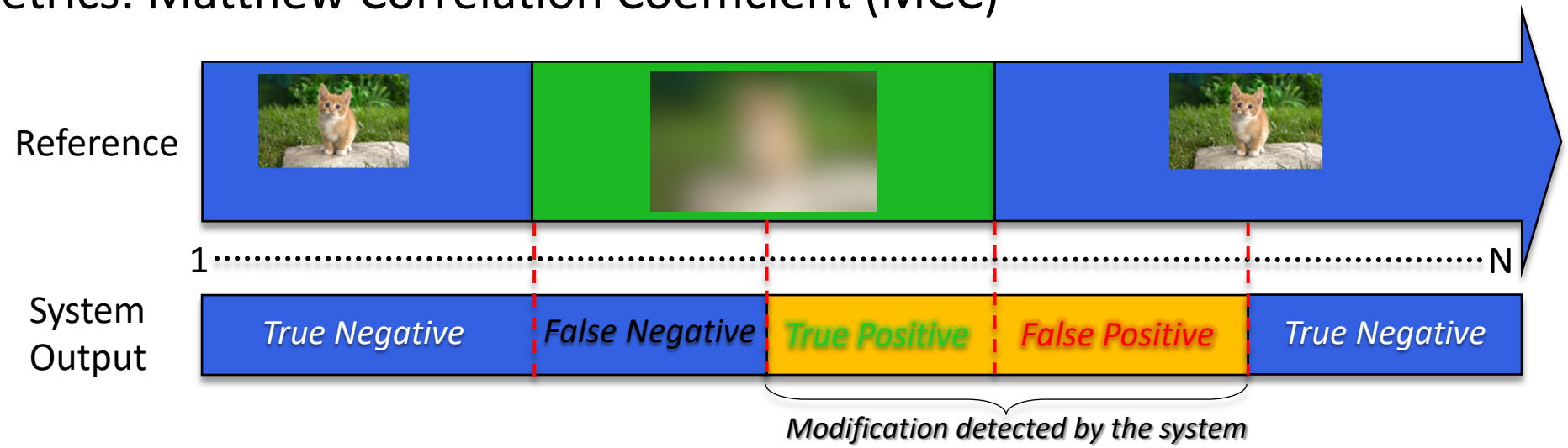
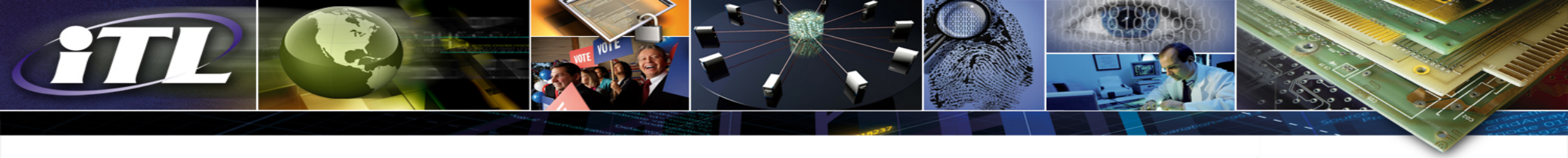


Figure: Video Temporal Detection and Localization



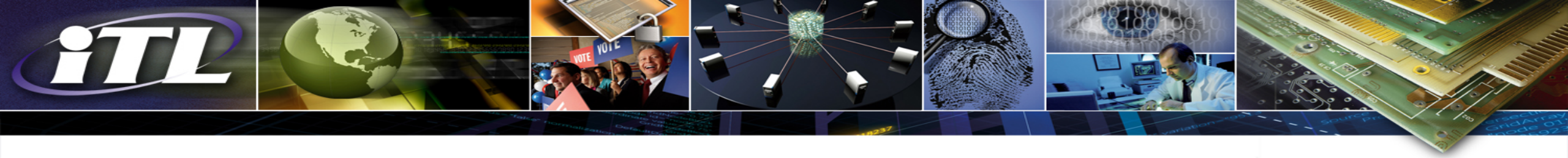
Holistic vs. Opt-In Technologies

Holistic vs. Opt In Technologies

- Evaluation challenge:
  - Some media forensic systems determine a response should not be returned
    - E.g., face illumination consistency systems should not respond if no face was found the image

Probe Status	Description
Processed	probe was fully processed
OptOut	the system **determined** a response should not be returned
OptOutLocalization	the system, **determined** a detection response but not a localization response should be returned
NonProcessed	A system failure of some kind occurred and will be scored with low probability

- NIST reports:
  - Holistic performance measures: score all trials
  - Opt In performance measures:
    - Trial Response Rate (TRR) – Percent of processed, NonProcessed, and FailedValidation images
    - Performance measures excluding opt'd out probes



Manipulation Journaling for Data Production

Manipulation Journaling: Describing Manipulations

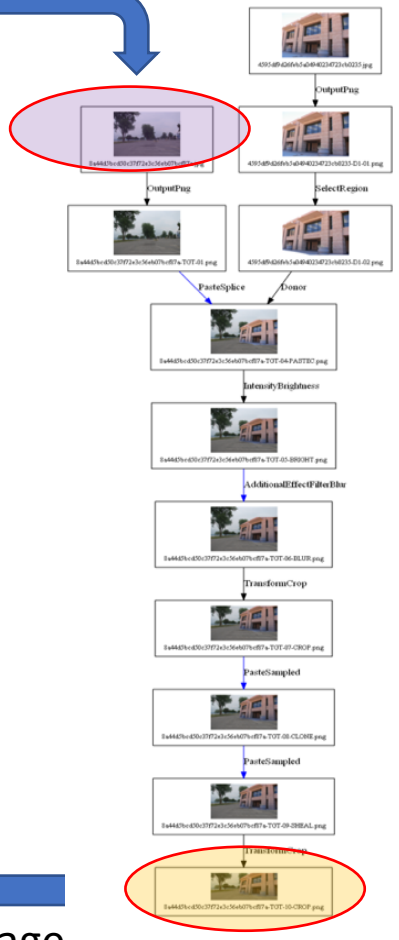
- Challenge:
  - Post manipulation interpretation of imagery changes is nearly impossible
  - Effective evaluations require knowledge:
    - Where the manipulation occurred
    - What tool was used
    - What operation was used
    - Semantics of the manipulation: remove vs. add

- MediFor Approach:
  - Record steps with PAR's Journaling Tool
  - Automate collection of localization

Manipulation Journaling - Operation Logging

High Provenance

Base Image



Donor Image

Unknown Provenance

Final Manipulated Image

Probe Legend

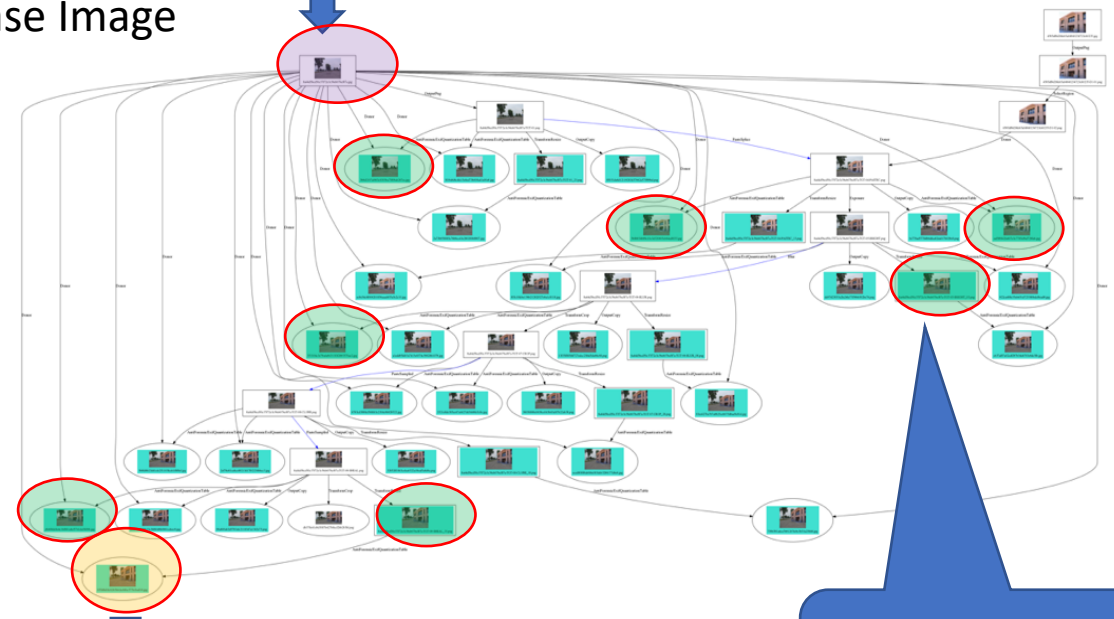
Non-Target Probe

Target Probe

Manipulation Journaling – Journal Expansion

Base Image

Donor Image



Final Manipulated Image

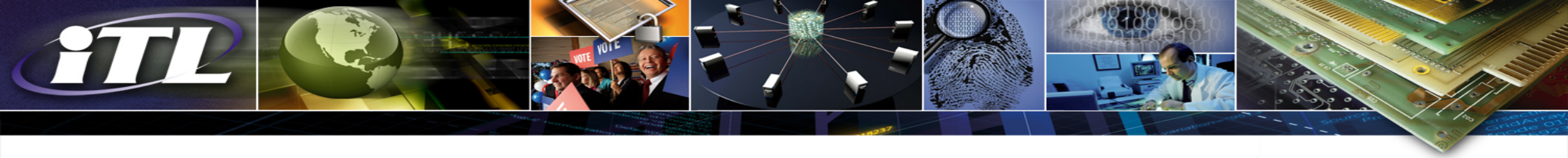
e.g., Expanded with 6 compression qualities

Probe Legend

Non-Target Probe

Target Probe

Replicate Target Probe



Understanding System Performance with Factor Analysis

Factor Analysis: Selective Scoring vs. Replicate Trials

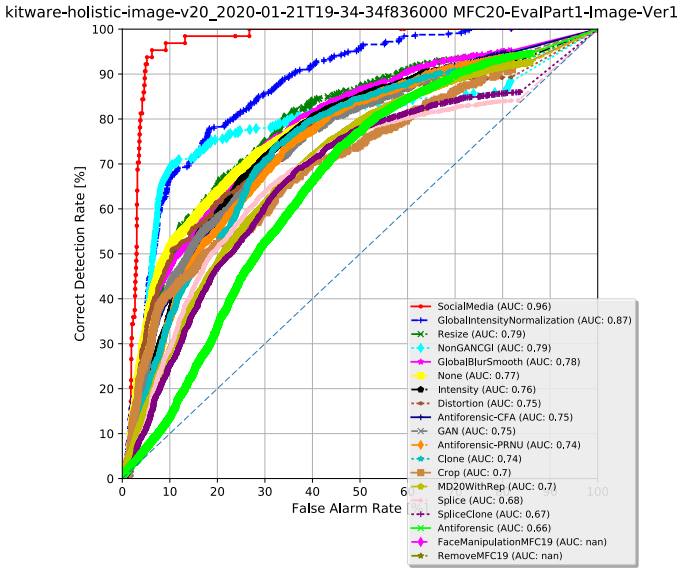
- Core challenge: the combinatorics of manipulation
  - Suppose a 2-Factorial, single operation experimental design
    - 17,500 images = 70 Operations * 2 levels * 125 examples
    - Not realistic (manipulators routinely stack manipulations)
  - The average graph depth in MFC ‘19 was ~4
    - $6.0*10^9$ images = 70^4 Operations * 2 levels * 125 examples
    - Laughably over execution budget
- MediFor data production approaches:
  - Human’s build realistic manipulations
  - Automatically extend journals with final node variations
  - Semi- and fully- automatic journal creation
- MediFor performance analysis approaches:
  - Overall manipulation performance
  - Selective Scoring Analysis
  - Special Study Analysis

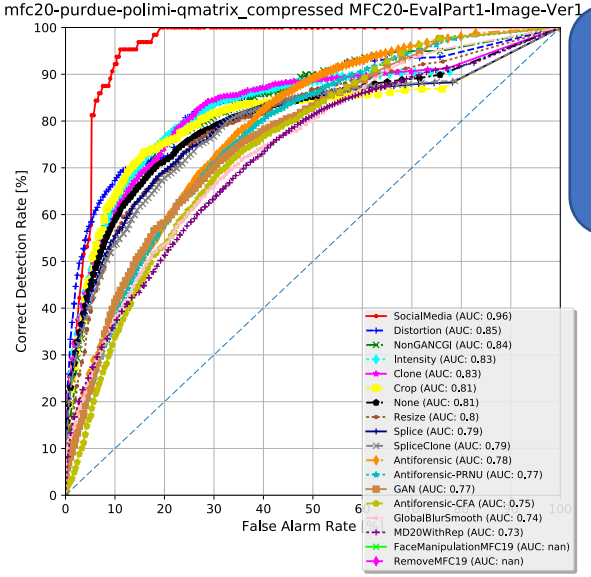
Factor Analysis: Selective Scoring

- Selective Scoring approach:
  - Non-Targets: Unmanipulated probes of known provenance
  - Targets: Manipulated probes containing the selected manipulations
- Strength of approach:
  - Insight into the effect of manipulation type on performance
- Weakness of the approach:
  - Confounding factors not controlled.

MFC20 Image Selective Scoring Queries

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 techinques 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





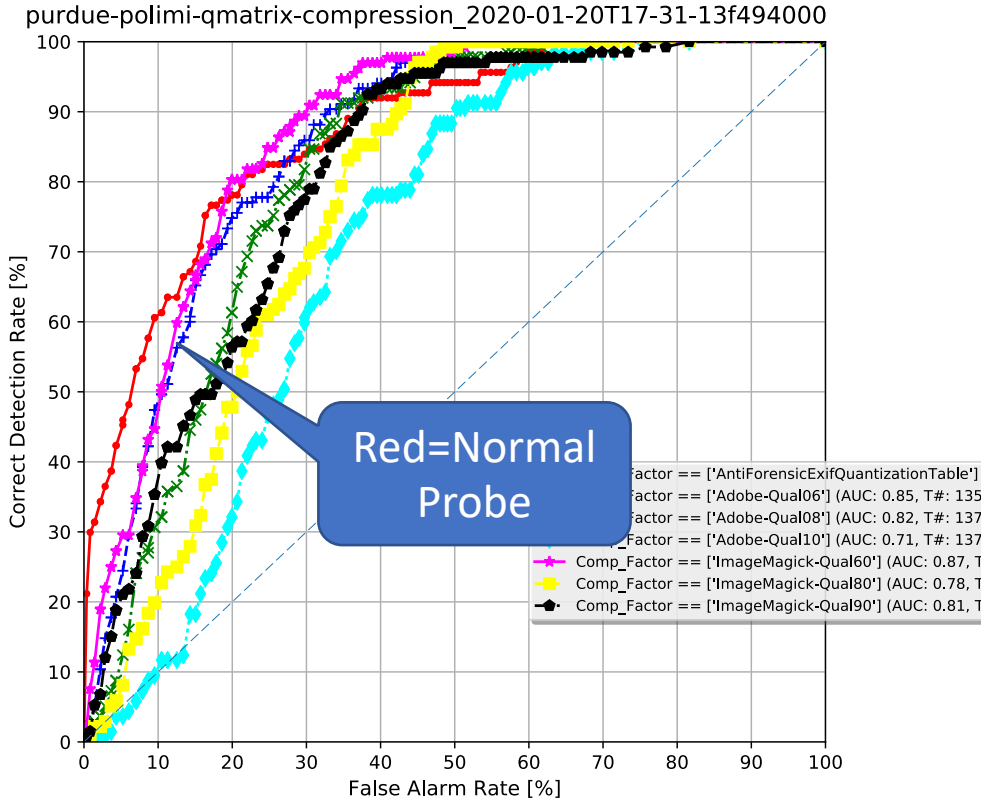
Similar Performance, Different Variance

Factor Analysis: Special Studies

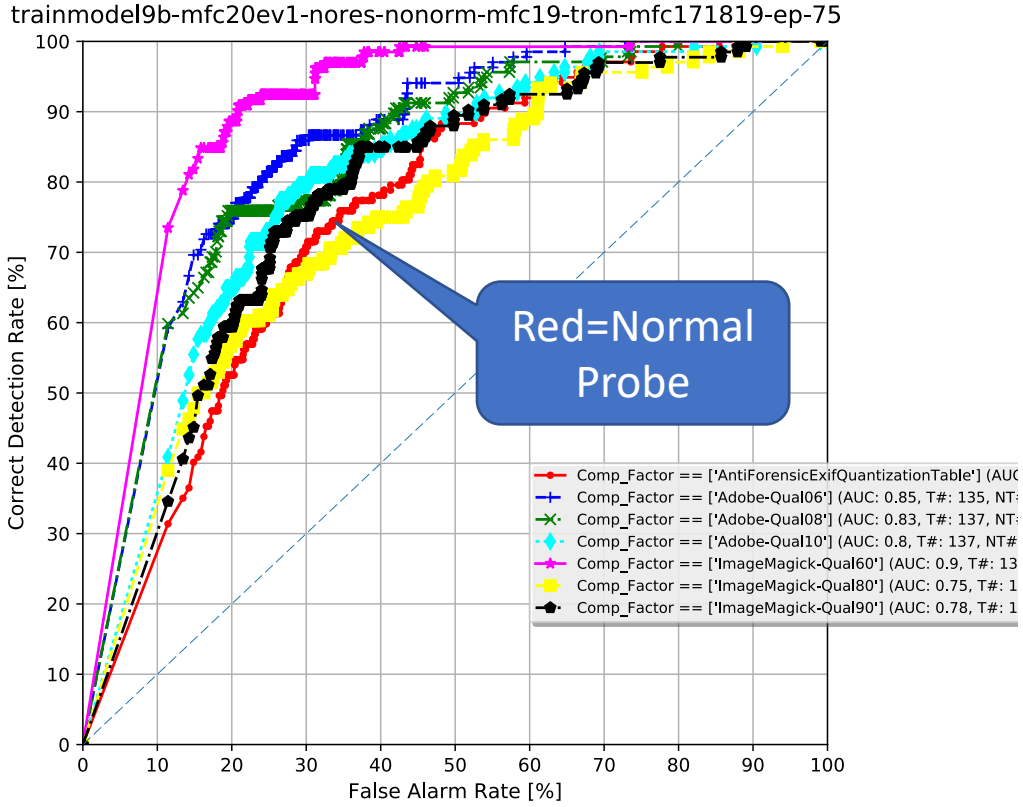
- Special Study approach:
  - Build specific data sets to answer specific performance assessment questions.
  - Enables two new views of performance assessment
    - Operation Only Detection
    - Facet Detection
- MFC20 Special Studies
  - Image
    - Compression
    - Global Blur
    - Social Media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

Compression Study Example:

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

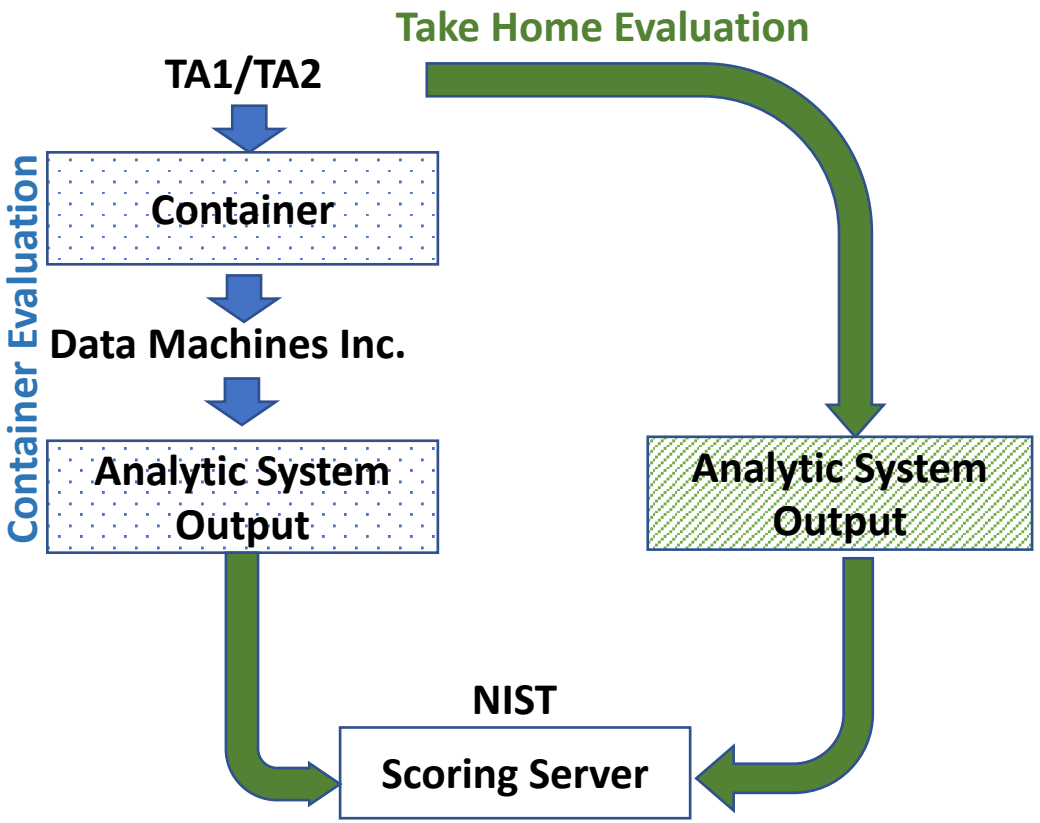


Purdue_Polimi #2458



Mayachitra #2516

Take Home vs. Container Evaluations

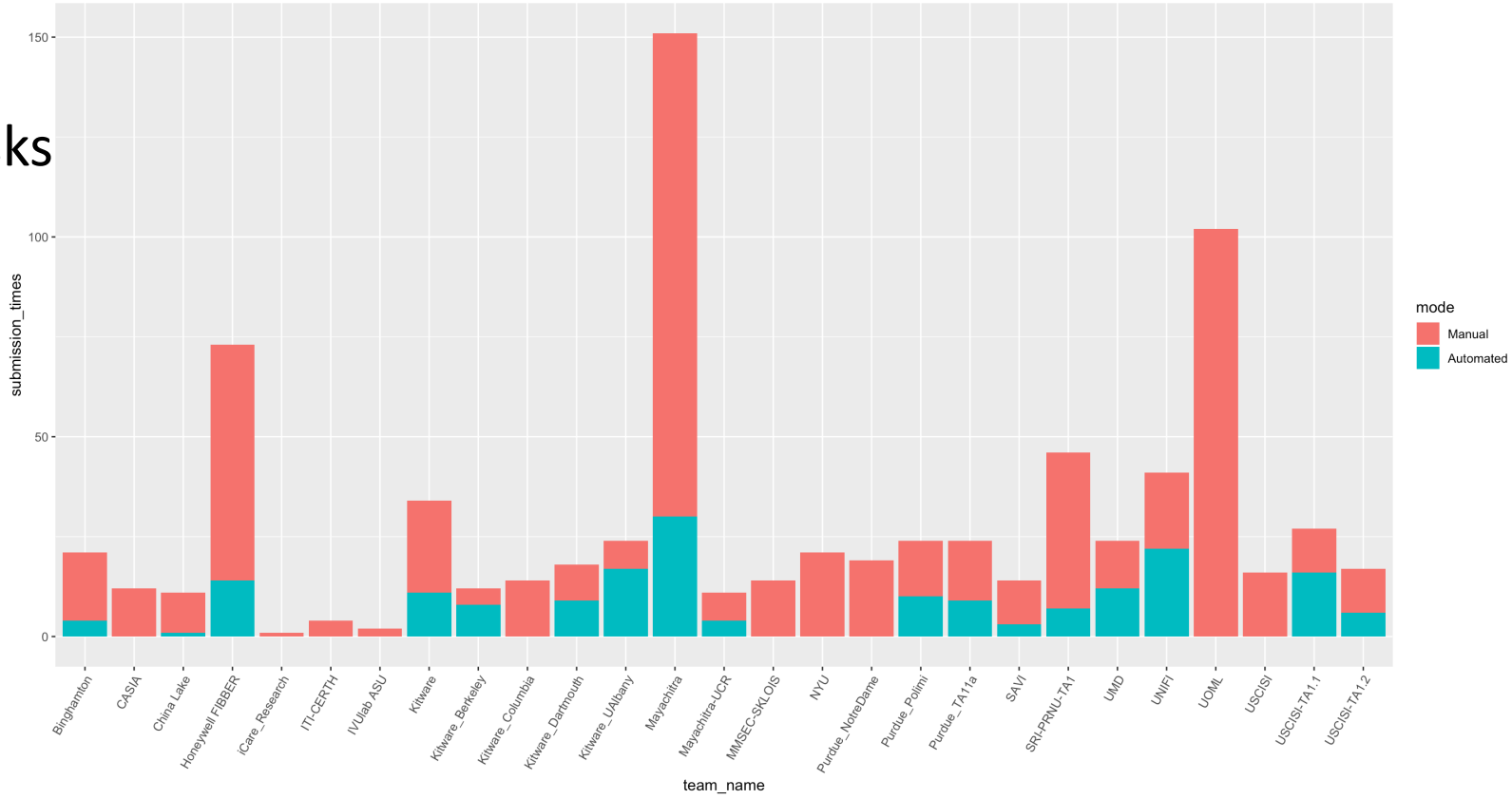


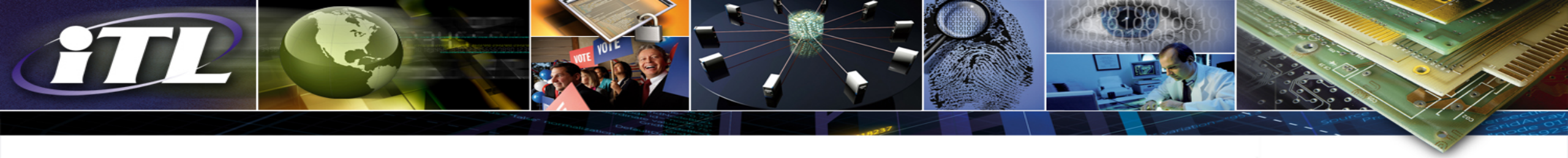
- Integration by TA2 requires access to algorithms for training
- History:
  - MFC ‘18, Data Machines Inc. completed a proof of concept to accept containers and processing data
  - MFC ‘19, Image Manip. Detection and Loc. and Video Manip. Detection and Temporal Loc., TA2 tested with delivered containers
  - MFC ‘20: Added Video Spatial Localization, Provenance Tasks, and Camera Verification

NIST MFC Scoring Server

- Performers had access to an automated scoring server
  - 65 MFC Data sets
  - Supports 6 evaluation tasks
  - Over 2596 submissions (1170 Active)
  - 12142 scoring runs

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





Media Forensic Challenge (MFC) Evaluation Datasets

MFC Image and Video Data Sets for Detection

NIST Data Sets	Image		Video	
	Probe	Journal	Probe	Journal
NC17 EvalPart1	4,000	406	360	47
MFC18 EvalPart1	17,000	758	1,000	114
MFC19 EvalPart1	16,000	1383	1,500	163
MFC20 EvalPart1	20,000	2536	2,500	217

MFC Image Data Sets for Provenance Tasks

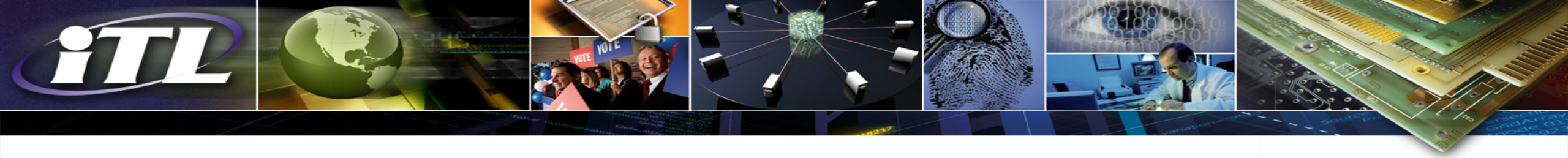
NIST Data Sets	Image Probe	Image Journal	World
NC17 EvalPart1	1K	406	1M
MFC18 EvalPart1	10K	641	1M
MFC19 EvalPart1	9420	1025	2M
MFC20 EvalPart1 2M	5926	1571	2M

MFC Image Date Sets for Splice Detection.

NIST Splice Data Sets	Image Probe	Image Journal
NC17 EvalPart1	329K	156
MFC18 EvalPart1	18K	381
MFC19 EvalPart1	18K	621
MFC20 EvalPart1	18K	1266

MFC Data Sets for Camera Verification

		MFC18			MFC19			MFC20		
Test	Train	Probe Pair	Cam.	Jour.	Probe Pair	Cam.	Jour.	Probe Pair	Cam.	Jour.
Image	Image	5275	39	452	8804	73	844	11288	106	1454
	Video	3383	25	410	6845	57	802	9346	88	1411
	Multimedia	3383	25	410	6845	57	802	9346	88	1411
Video	Image	289	11	67	351	23	81	788	35	87
	Video	289	11	67	337	22	81	767	34	87
	Multimedia	289	11	67	337	22	81	767	34	87



Select MFC20 Results Preview

MFC20 Team and Task Participation Summary

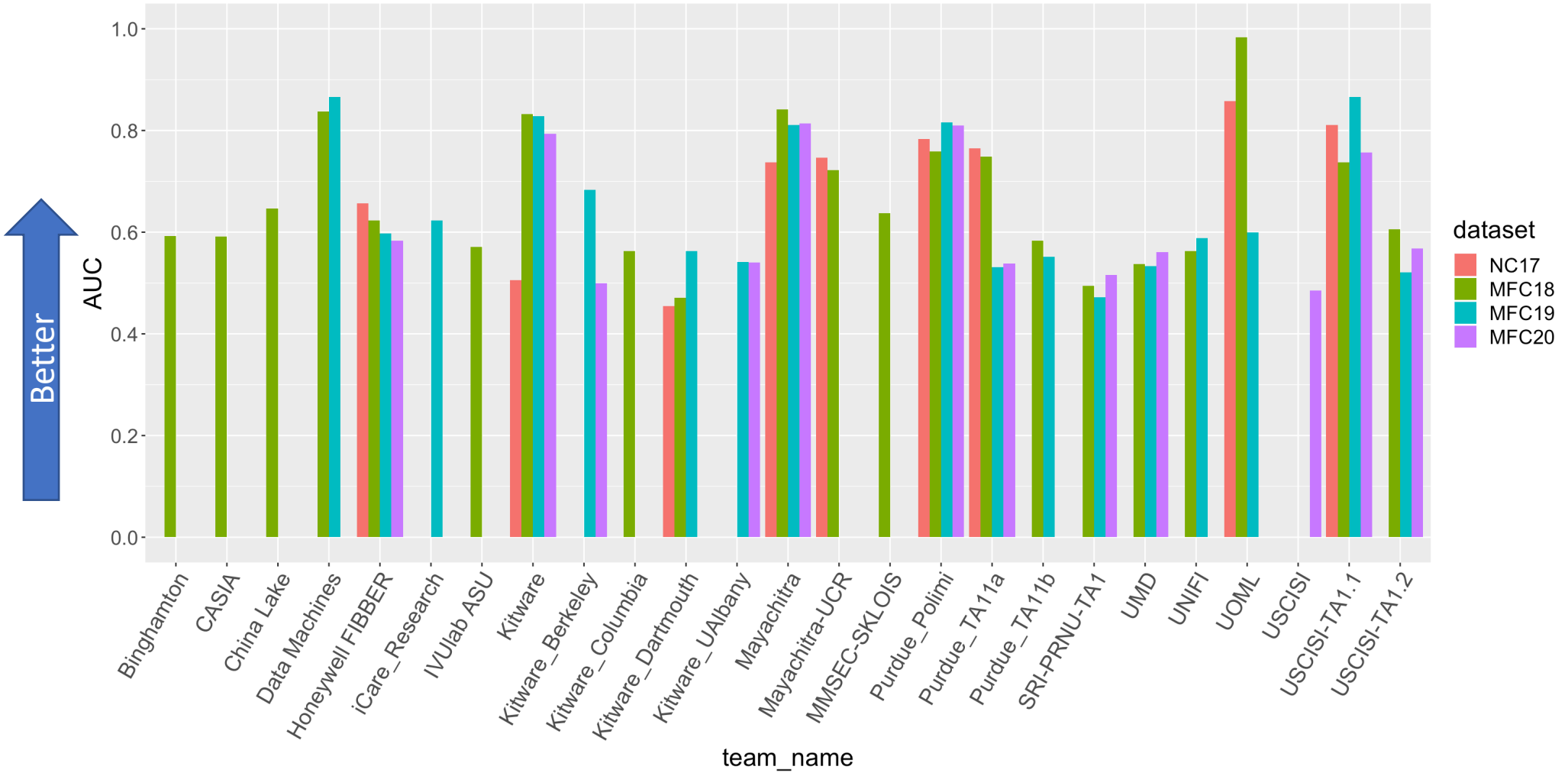
Image Systems

	Camera	Event	Manipulation	Provenance Filtering	Provenance Graph Building	Splice
Binghamton	1					
Honeywell FIBBER			15			
Kitware			11			1
Kitware_Berkeley			1			
Kitware_Columbia				1	1	
Kitware_UAlbany			2			
Mayachitra		6	34			
Purdue_NotreDame				1	1	
Purdue_Polimi			9			
Purdue_TA11a			4			
SRI-PRNU-TA1	1		2			
UMD			4			
USCISI			2	1	1	
USCISI-TA1.1			16			
USCISI-TA1.2			2			

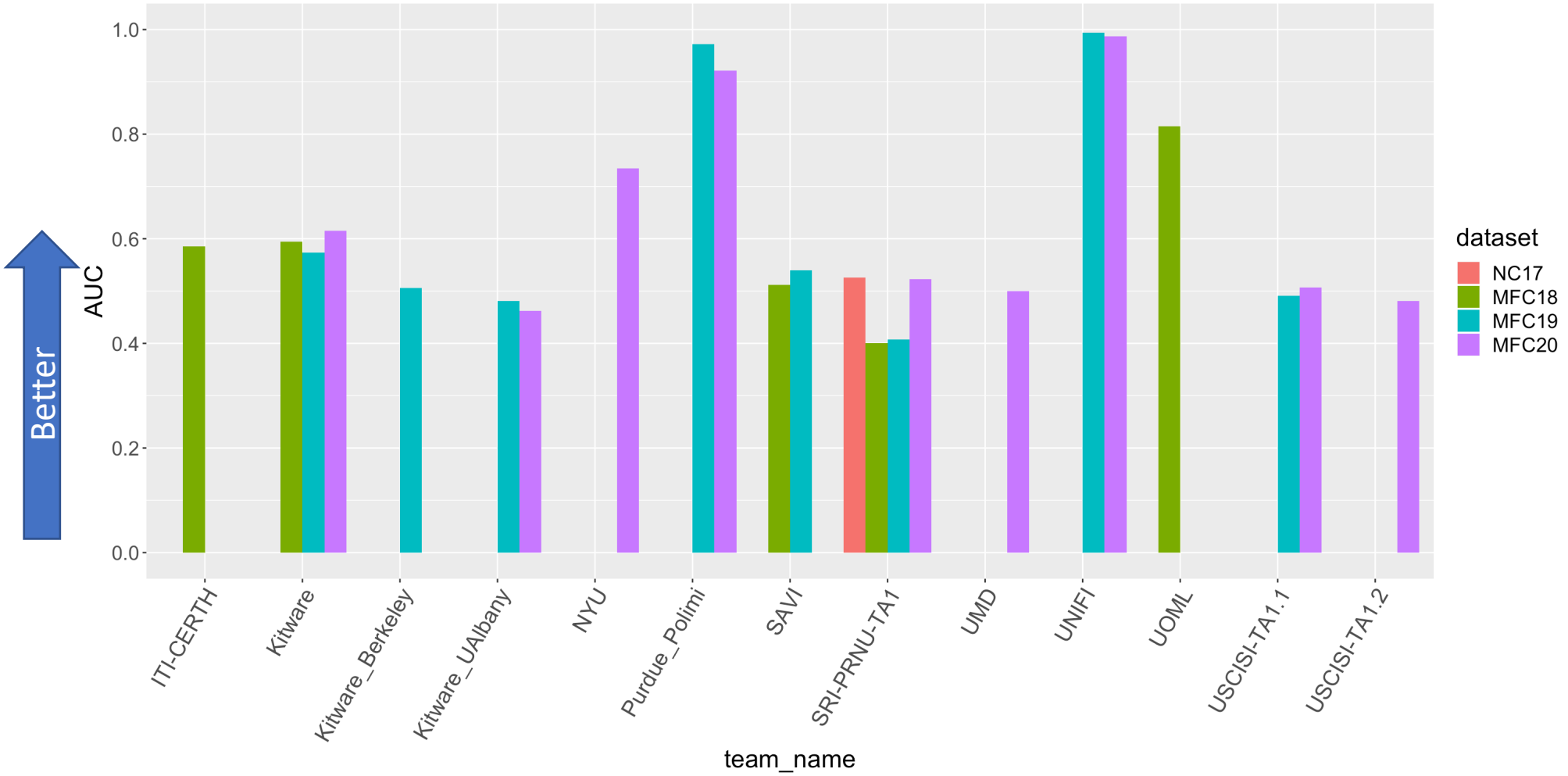
Video Systems

	Camera Verification	Manipulation
Kitware		8
Kitware_UAlbany		1
NYU		8
Purdue_Polimi		2
SRI-PRNU-TA1	2	1
UMD		1
UNIFI		2
USCISI-TA1.1		2
USCISI-TA1.2		1

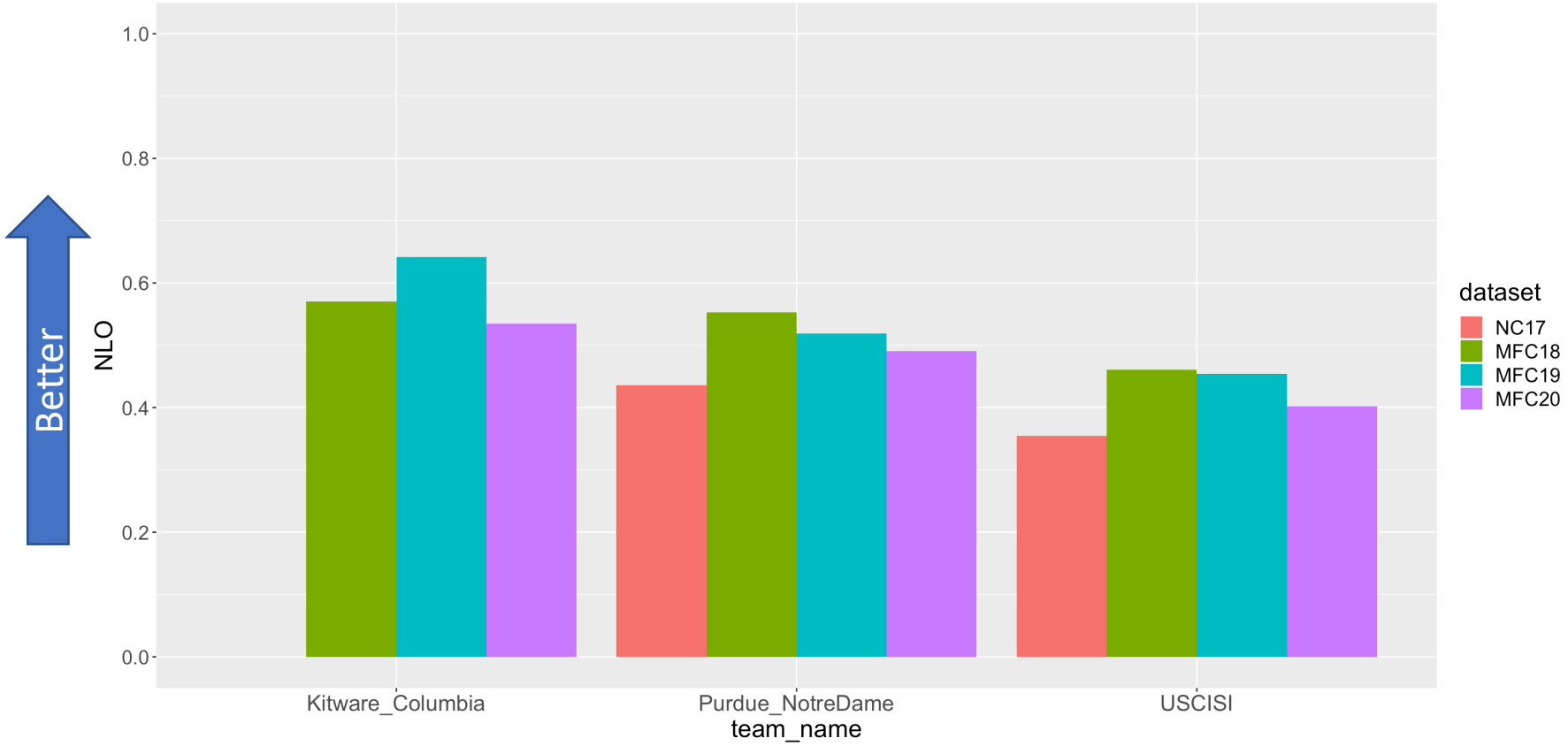
Image Manipulation Detection: NC17-MFC20



Video Manipulation Detection: NC17-MFC20



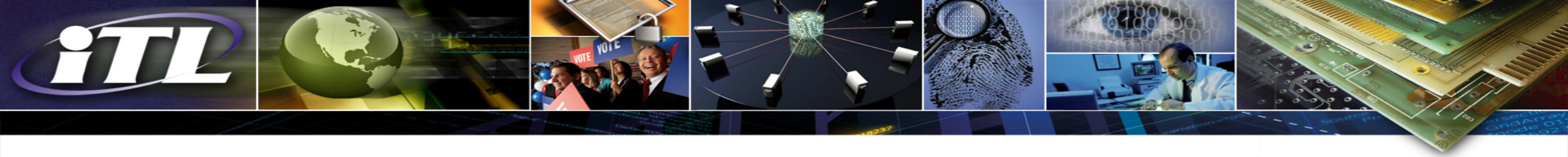
Provenance Graph Building: NC17-MFC20



NIST Data Sets	Probe	World
NC17 EvalPart1	1K	1M
MFC18 EvalPart1	10K	1M
MFC19 EvalPart1	9420	2M
MFC20 EvalPart1	5926	2M

MFC Overview Summary and What's Next

- Introduced the 6 common evaluation task for the MediFor Program
- Introduced the data creation approach for the MediFor Program
- Subsequent NIST talks during this meeting will be deep dives
- This is the final MediFor evaluation; NIST has plans to continue open evaluations of media forensic systems. Details to follow.



MFC20 Image Evaluation Results Deep Dive

Jonathan Fiscus (Co-PI), **Dr. Haiying Guan** (Co-PI), Dr. Yooyoung Lee,
Dr. Amy Yates[+], Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah,
Dr. Xiongnan Jin

Multimodal Information Group, [+] Image Group
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology (NIST)

April 21-25, 2020

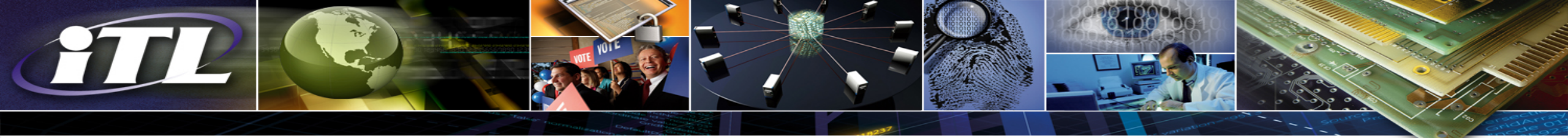


Image Manipulation Detection and Localization

Image Manipulation Detection and Localization

System Input

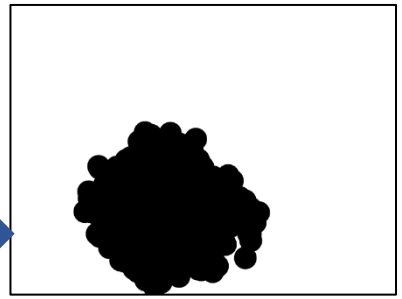
Image(s) + (Metadata)

Probe image

Image Detection and Localization Analytic System

System Output

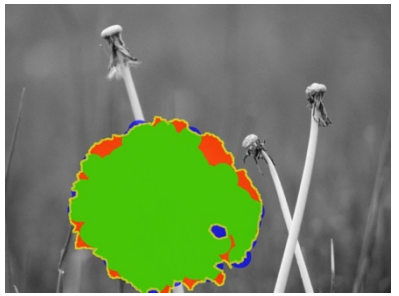
Confidence score
97.86

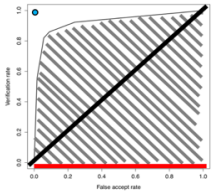


System output probe mask

Metrics

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

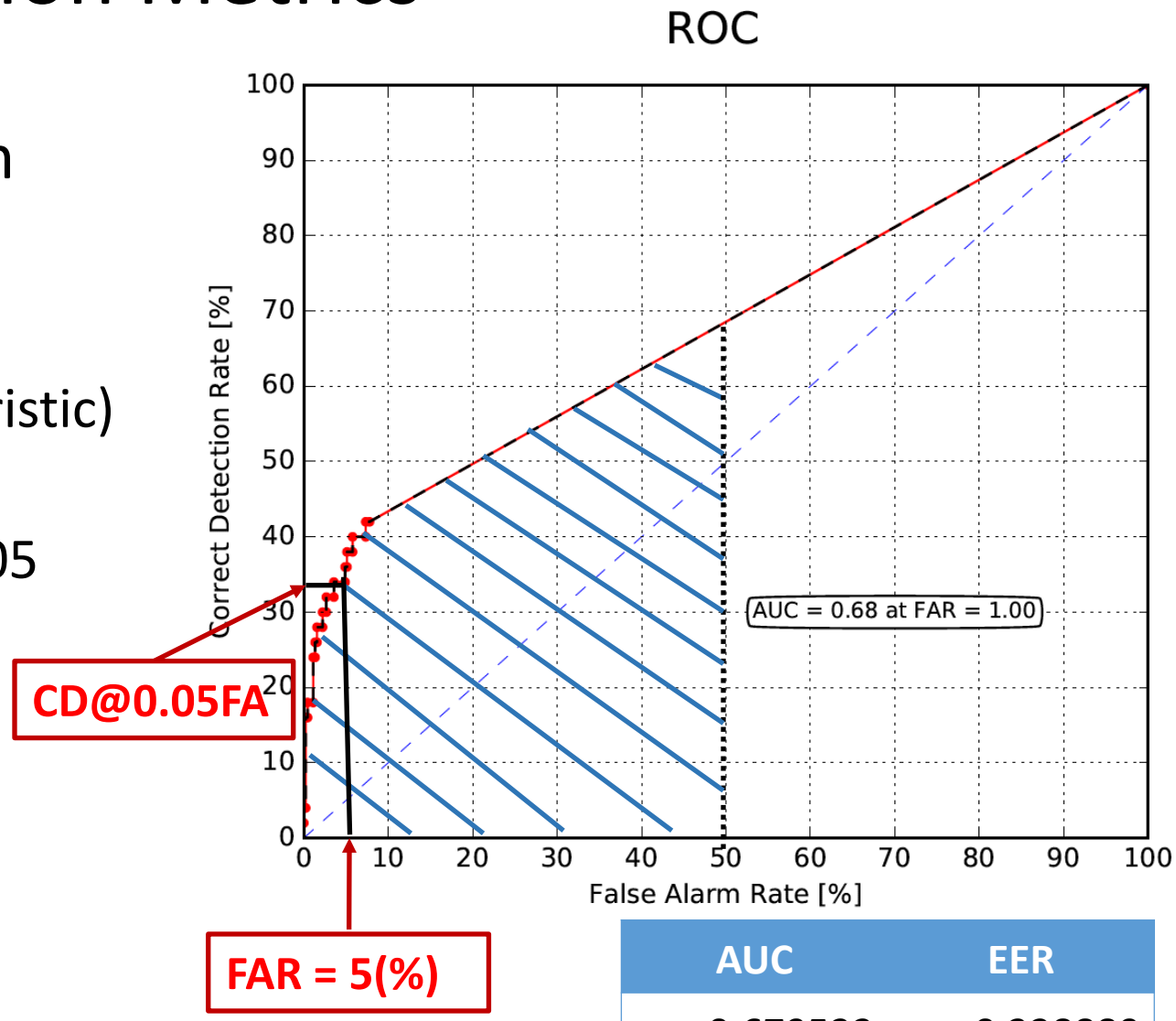




Manipulated image
Matthews Correlation Coefficient (MCC)

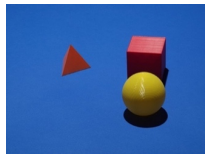
Detection System Evaluation Metrics

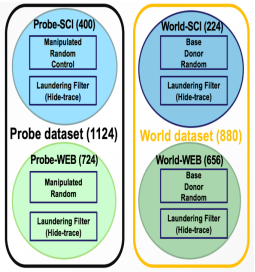
- Evaluate the accuracy of a system output (e.g., confidence score)
- Evaluation metrics
  - ROC (Receiver Operating Characteristic)
  - AUC (Area Under Curve)
  - CD (Correct Detection) @ FAR = 0.05
  - EER (Equal Error Rate)

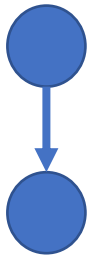


AUC	EER
0.679533	0.328889

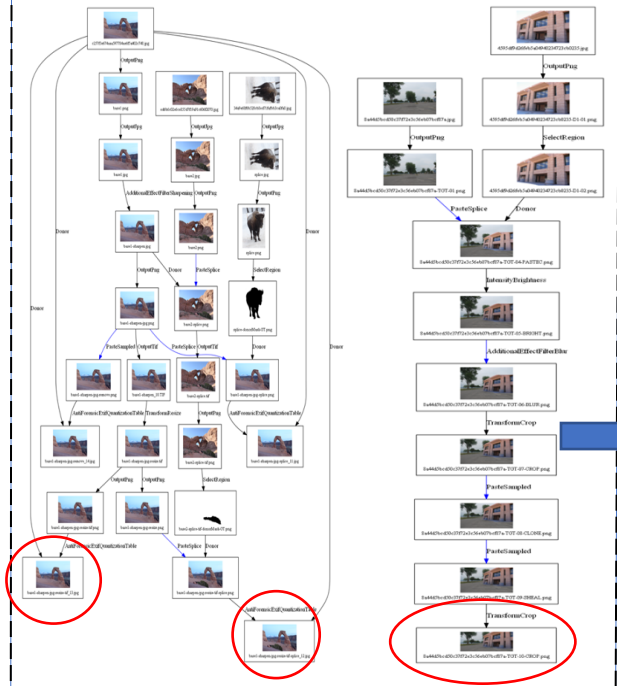
MFC Evaluation Dataset History







Kick-off
2016 Dataset

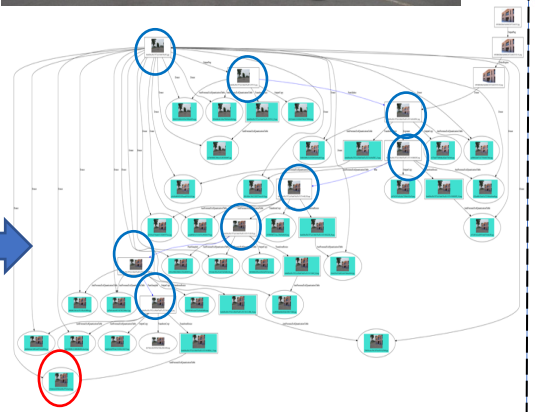


Provenance

Image

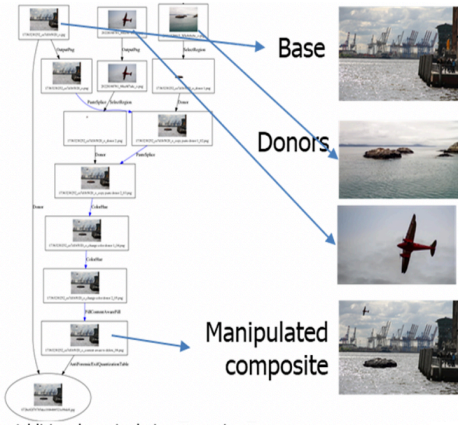
Auto Journaling Tool (JT)

Nimble Challenge 2017



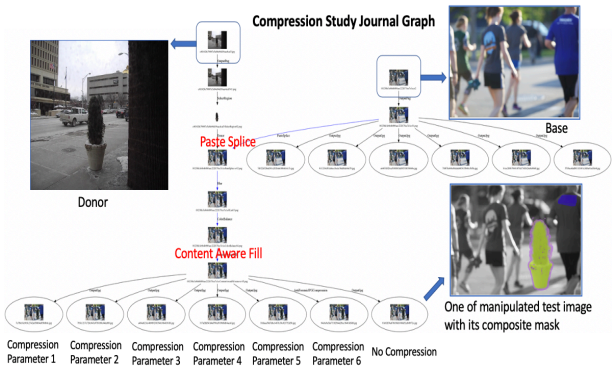
New Manipulations
(CGI, Recapture, …)
Extended JT, AutoJT

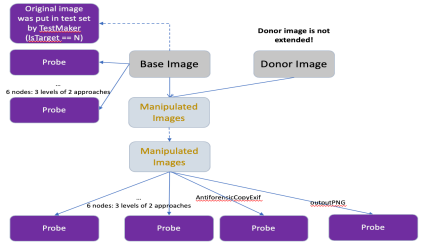
MFC 2018



- 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 Image Evaluation Open Dataset Summary

NIST Data Sets	Image Probe	Image Journal	Date
NC17 EvalPart1	4K	406	06/2017
MFC18 EvalPart1	17K	758	03/2018
MFC19 EvalPart1	16K	1383	03/2019
MFC20 EvalPart1	**20K**	**2536**	**03/2020**

Holistic vs. Opt In Technologies
- Allowing Systems to Respond When/If Appropriate

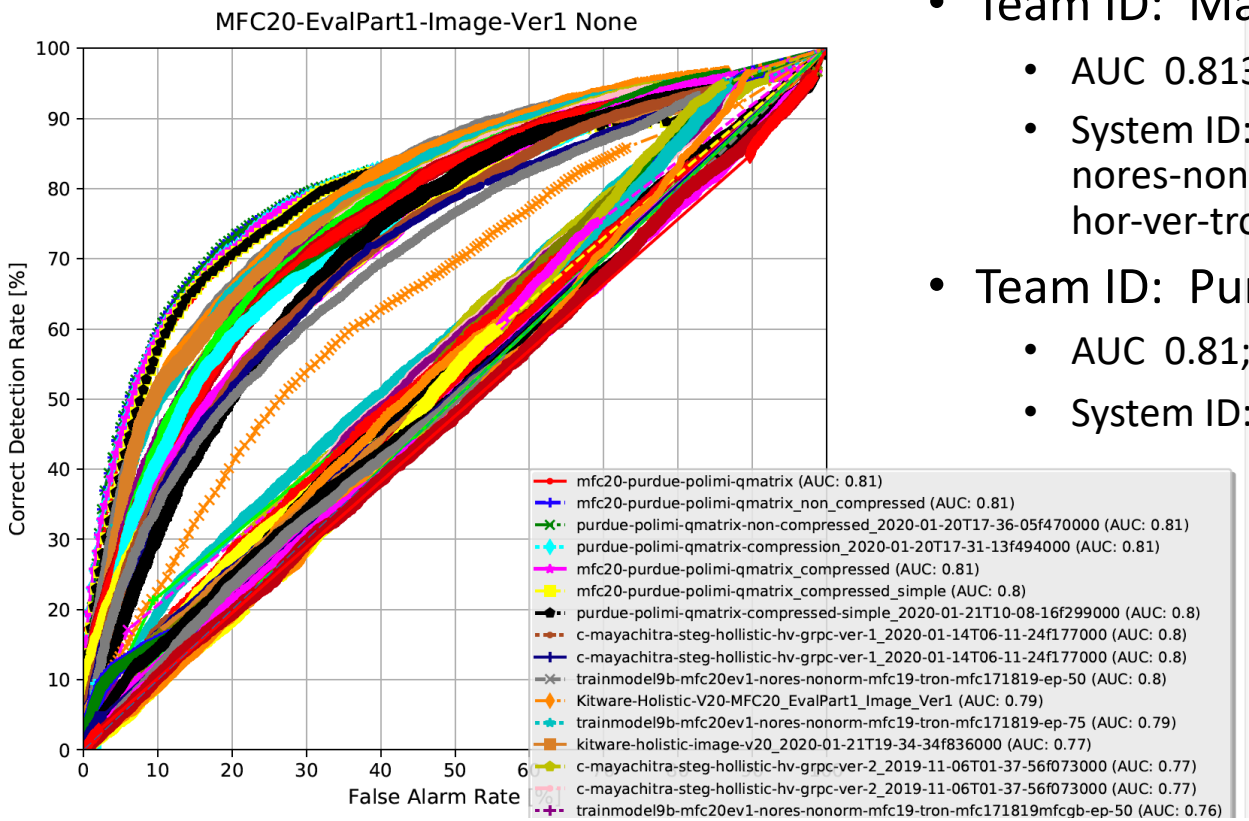
- Evaluation challenge:
  - Some media forensic systems **determine** a response should not be returned
    - E.g., face illumination consistency systems should not respond if no face was found the image

Probe Status	Description
Processed	probe was fully processed
OptOut	the system **determined** a response should not be returned
OptOutLocalization	the system, **determined** a detection response but not a localization response should be returned
NonProcessed	A system failure of some kind occurred and will be scored with low probability

- NIST reports:
  - Holistic performance measures: score all trials
  - Opt In performance measures:
    - Trial Response Rate (TRR) – Percent of processed, NonProcessed, and FailedValidation images
    - Performance measures excluding opt'd out probes

Image Manipulation Detection Results: Full Data

- 20K probe images
- 12 teams:
  - Honeywell FIBBER
  - Kitware_Berkeley
  - Kitware_UAlbany
  - Kitware
  - Mayachitra
  - Purdue_Polimi
  - Purdue_TA11a
  - SRI-PRNU-TA1
  - UMD
  - USCISI-TA1.1
  - USCISI-TA1.2
  - USCISI
- 82 image detection systems as 04/09/2020.



- Highest AUC on full MFC20 EP1:
  - Team ID: Mayachitra
    - AUC 0.81384;
    - System ID: trainmodel9b-mfc20ev1-nores-nonorm-adam-def-apr2020-mfc19-hor-ver-tron-mfc171819-ep-40
  - Team ID: Purdue_Polimi
    - AUC 0.81; (CD@0.05FA = 0.436)
    - System ID: mfc20-purdue-polimi-qmatrix

Figure: TA1 system MFC20 EP1, All probes (regardless of OptIn)

Image Manipulation Detection Results: Opt In (1)

- OptIn Systems on MFC20 EP1:
  - 77 systems as 04/09/2020
  - Highest AUC:
    - AUC 0.855; (CD@0.05FA = 0.47, TRR = 0.63)
    - Team ID: Purdue_Polimi
    - System ID: purdue-polimi-qmatrix-non-compressed_2020-01-20T17-36-05f470000

Figure: TA1 system MFC20 EP1, Opt In probes

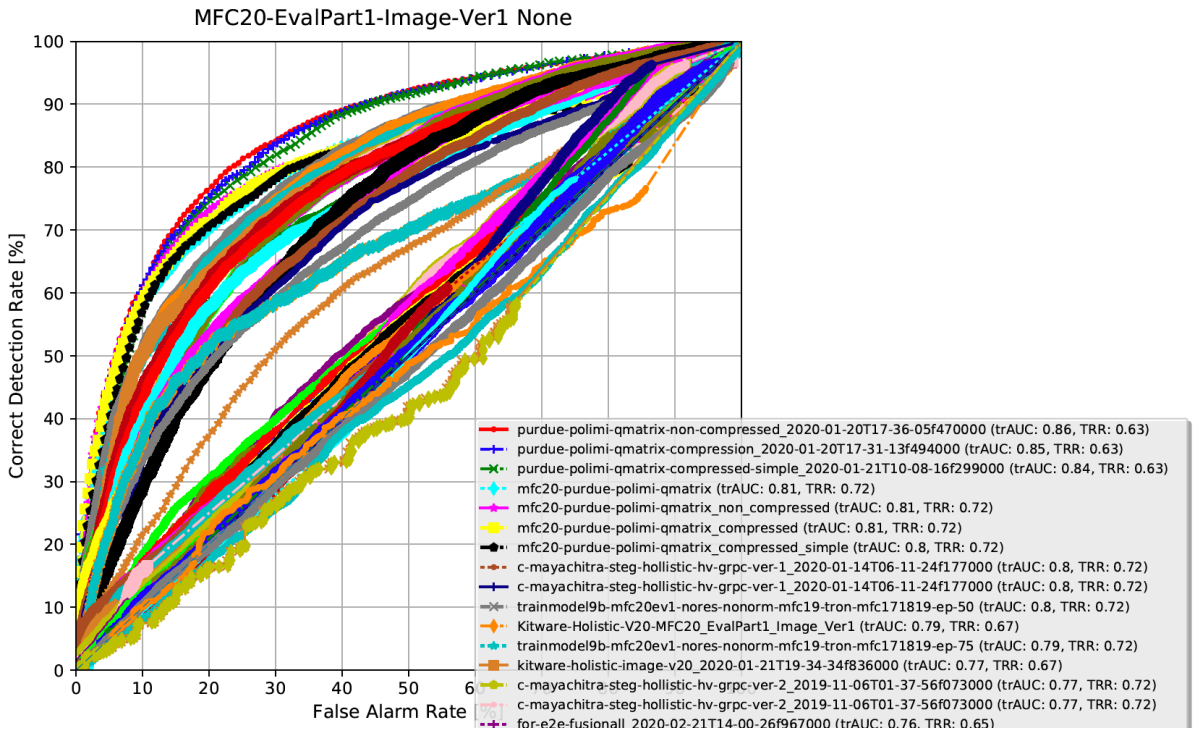
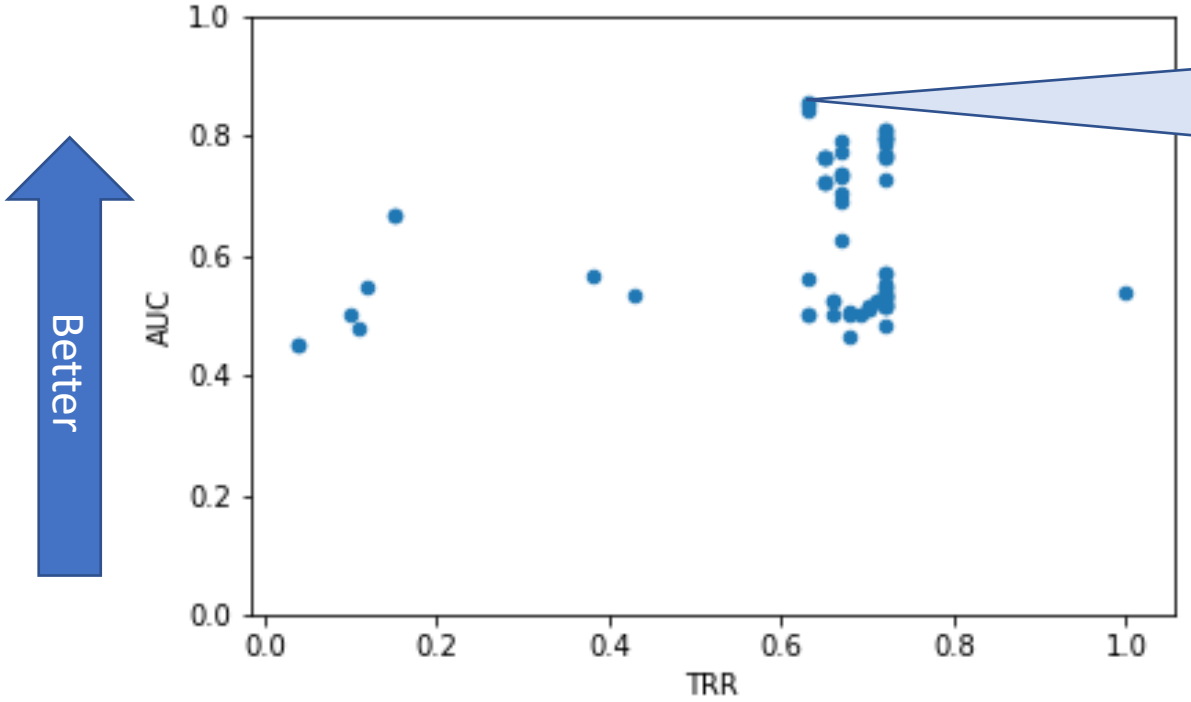


Image Manipulation Detection Results: Opt In (2)



Trial Respond Rate (TRR)

Highest AUC = 0.855:

Team ID: Purdue_Polimi

System ID: purdue-polimi-qmatrix-non-compressed_2020-01-20T17-36-05f470000

CD@0.05FA: 0.47

TRR: 0.63

Figure: Image detection Opt In system Area Under the Curve (AUC) vs. Trial Response Rate (TRR) on MFC20 EP1 Image dataset (each point is an analytic system)

Factor Analysis: Selective Scoring

- Evaluate the system performance on a certain type of data
- Manipulation Detection (MD)
  - Target = Any manipulated media (image or video)
  - NonTarget = HP media
- Selective Scoring on Manipulation Detection (S-MD)
  - Target = Media contains defined manipulations; other operations may also be present in the media
  - NonTarget = HP media

Image Manipulation Detection Results
- Selective Scoring on GlobalBlurSmooth (Full Data)

- MFC20 EP1
- Selective Scoring on GlobalBlurSmooth
- 14 systems as 04/09/2020
- Highest AUC:
  - AUC 0.802; (CD@0.05FA = 0.191)
  - Team ID: Honeywell FIBBER
  - System ID: p-inhnoi_2020
- Highest CD@0.05FA:
  - AUC 0.771; (CD@0.05FA = 0.293)
  - Team ID: Mayachitra
  - System ID: trainmodel9b-mfc20ev1-nores-nonorm-mfc19-hvhv-fusn-tron-mfc171819mfcgb-ep-50

Figure: TA1 system MFC20 EP1, All probes (regardless of OptIn) Selective Scoring on GlobalBlurSmooth

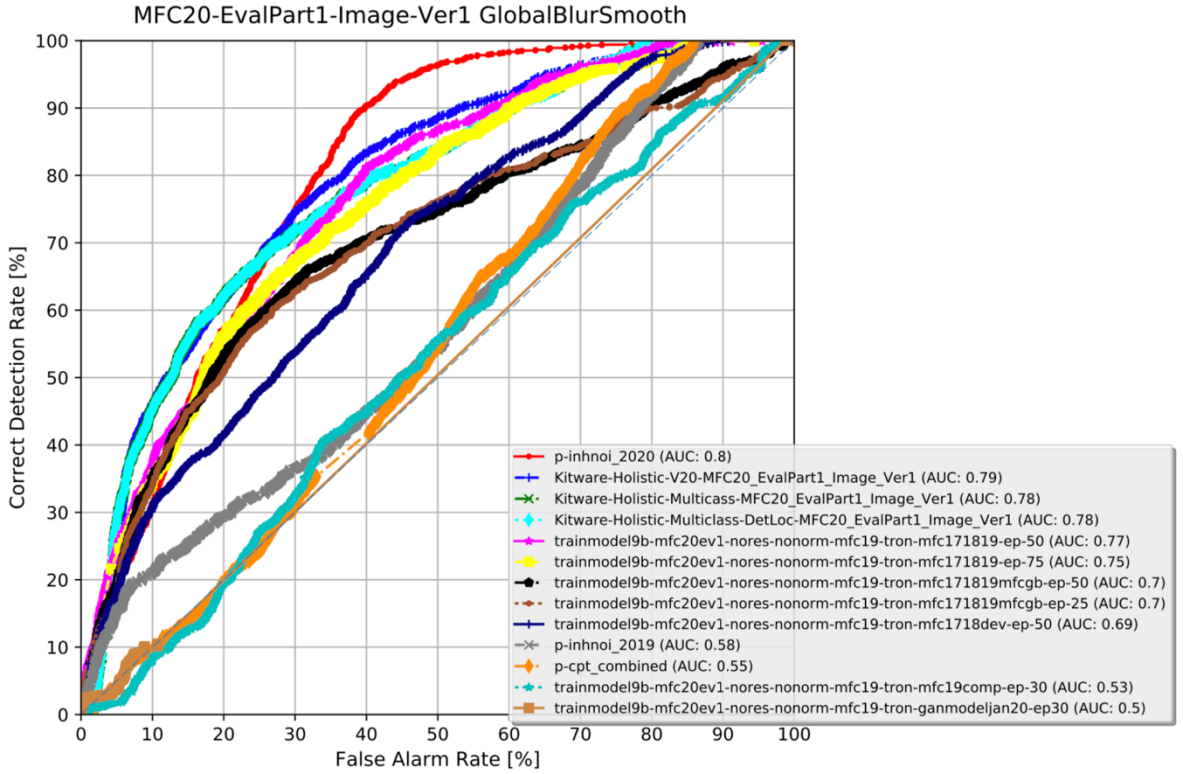


Image Manipulation Detection Results
- Selective Scoring on SocialMedia (Opt In)

- MFC20 EP1
- Selective Scoring on SocialMedia (Opt In)
- 10 systems as 04/09/2020
- Highest AUC:
  - AUC 0.954; (CD@0.05FA = 0.766, TRR = 0.27)
  - Team ID: Kitware
  - System ID: Kitware-Holistic-V20-MFC20_EvalPart1_Image_Ver1

Figure: TA1 system MFC20 EP1, Selective Scoring on SocialMedia OptIn probes

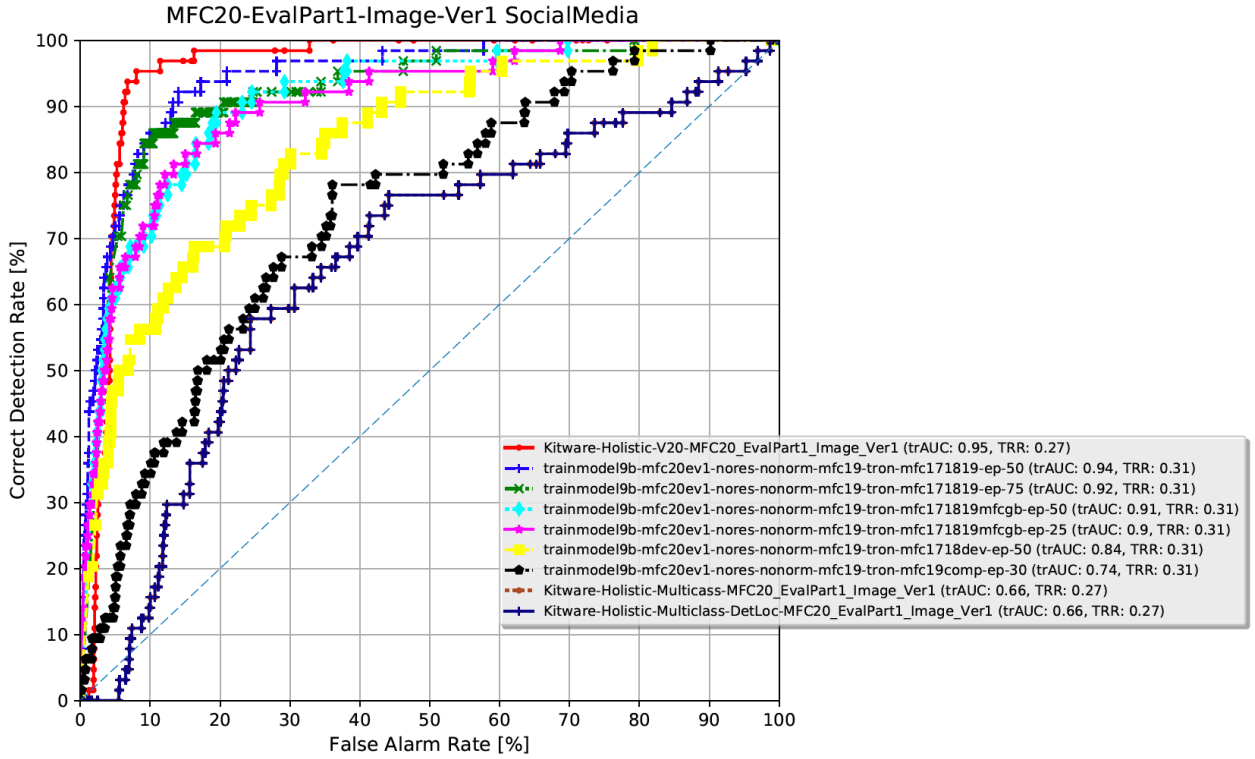


Image Manipulation Detection Results
- Selective Scoring: GlobalIntensityNormalization (Opt In)

- MFC20 EP1
- Selective Scoring on GlobalIntensityNormalization (Opt In)
  - 23 systems as 04/09/2020
- Highest AUC:
  - AUC 0.928; (CD@0.05FA = 0.674, TRR = 0.32)
  - Team ID: Mayachitra
  - System ID: trainmodel9b-mfc20ev1-nores-nonorm-mfc19-hvhv-fusn-tron-mfc171819mfcgb-ep-50

Figure: TA1 system MFC20 EP1,
Selective Scoring: GlobalIntensityNormalization
OptIn probes

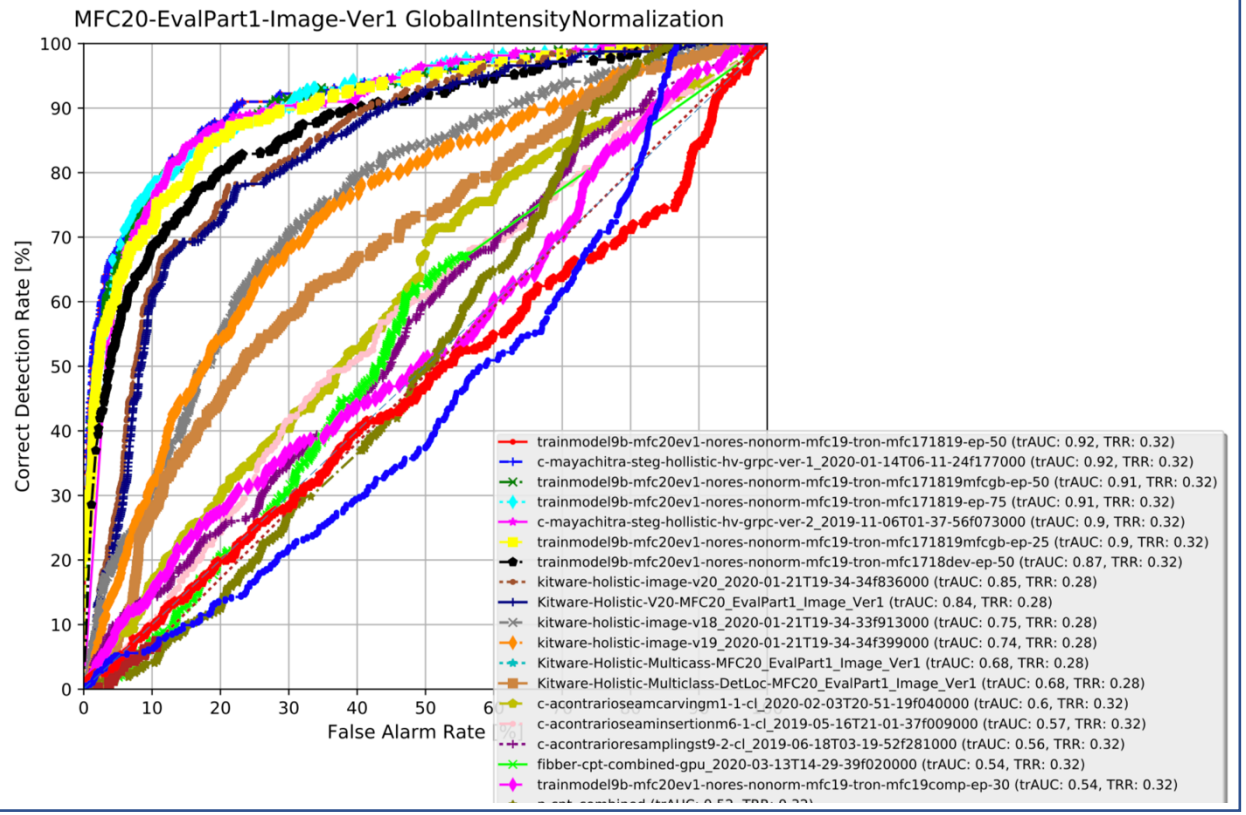


Image Manipulation Detection System - Team Performance Comparison Across Years

(Full Data)

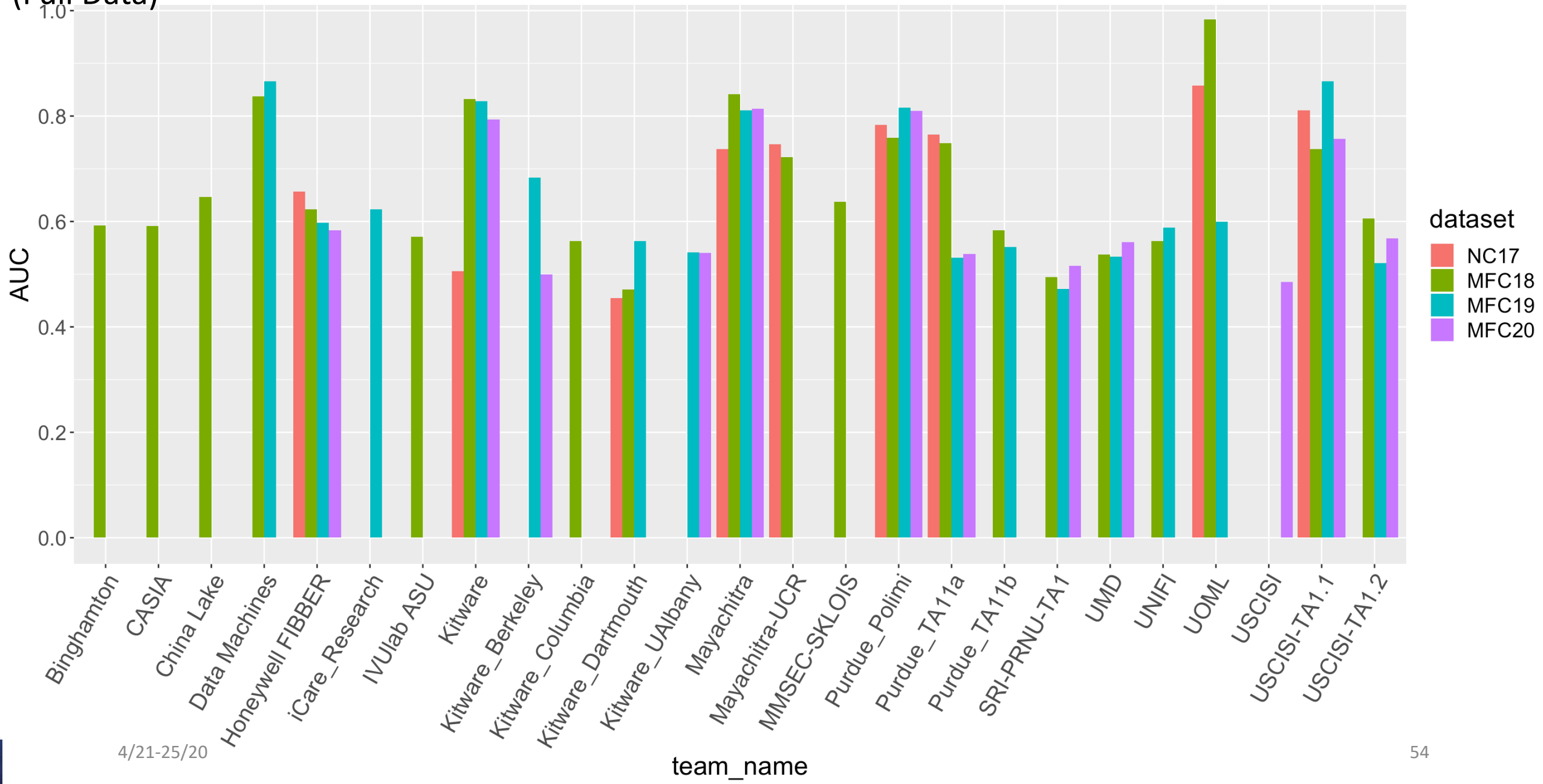


Image Manipulation Detection System - Team Performance Comparison Across Years (Opt In)

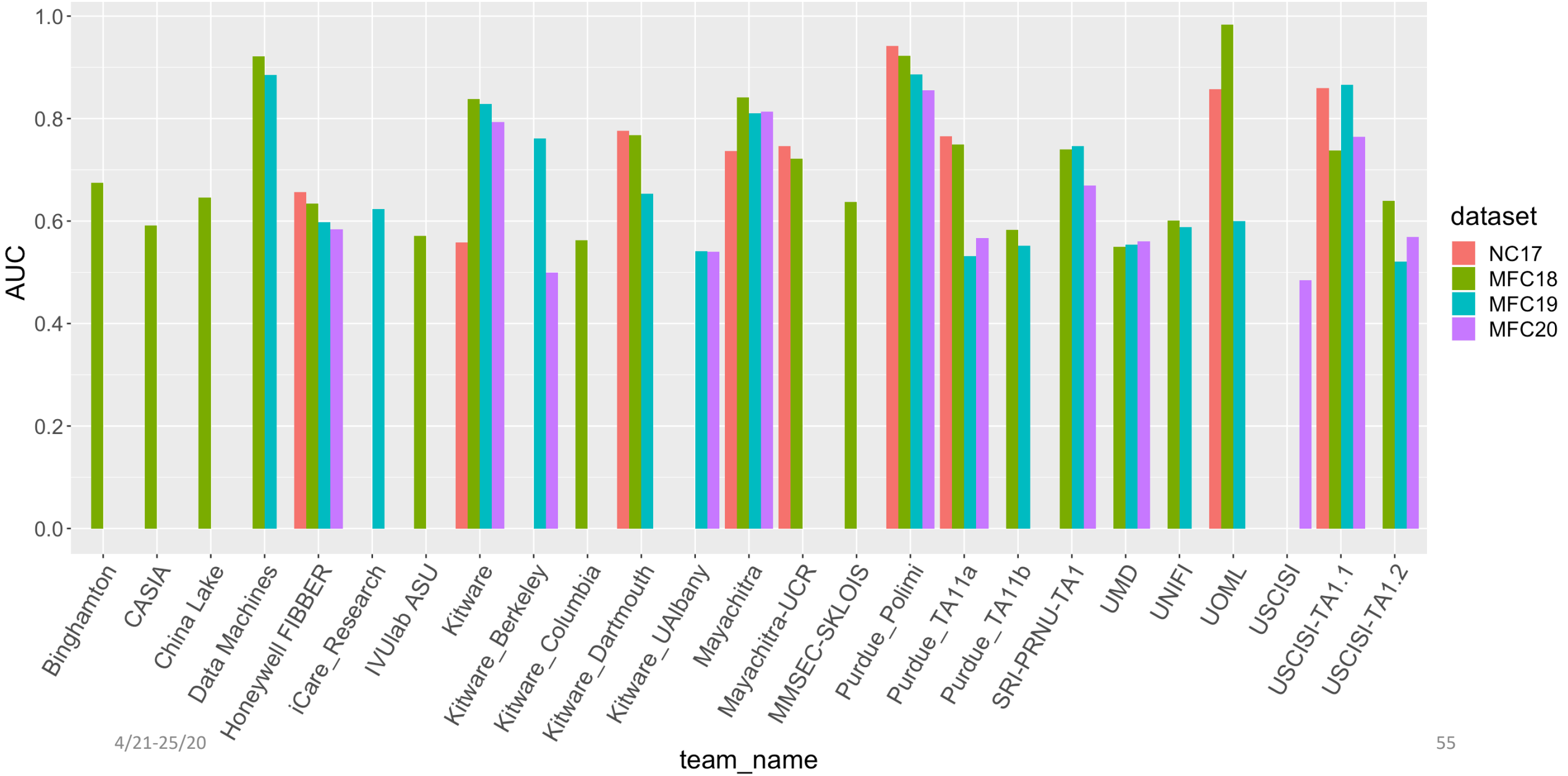


Image Manipulation Detection System

- Best Team Performance Comparison Across Years (Full Data, ImageOnly)

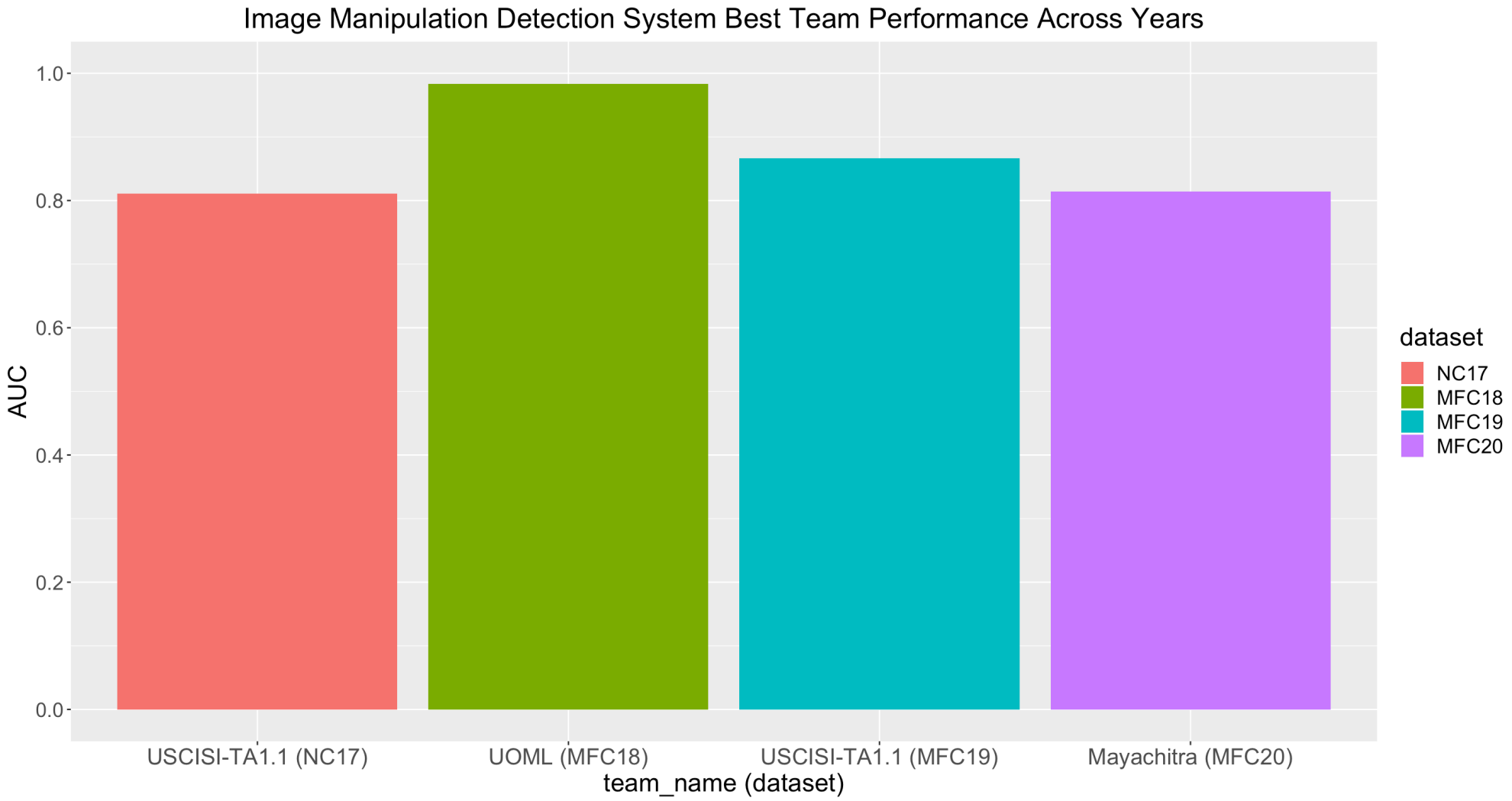
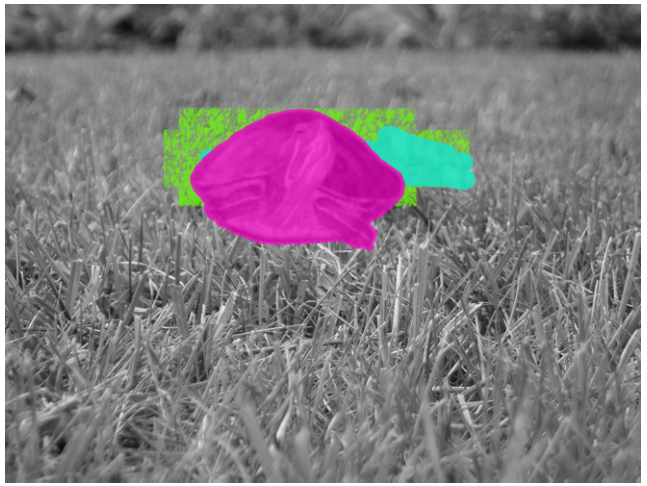


Image Localization Task

- Jpeg 2000 composite mask
  - Distinct manipulations are recorded in the different layers in Jpeg2000 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

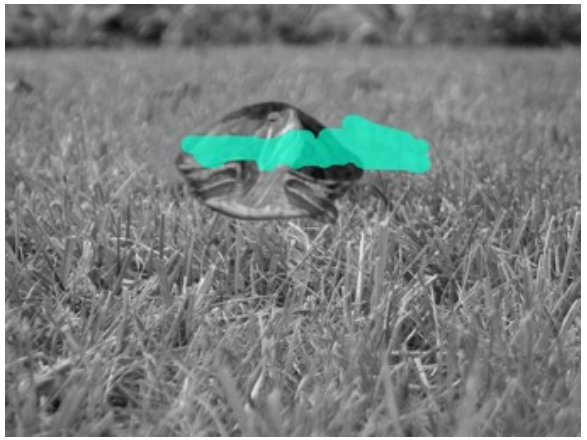
Sequence	Operation	Purpose	Color	Evaluated
5	ContentAwareFill	remove		Y
4	PasteSampled	heal		Y
3	PasteSplice	add		Y
2	Blur			Y

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.

Image Localization Selective Scoring

- The JPEG2000 masks encode bits that can be used to store information from multiple overlapping manipulations.
- Scoring can now be done on manipulations from any recorded layer.
- Example of selective scoring query:
  - Operation == ['PasteSplice'] or (Operation == ['PasteSampled'] and Purpose == ['Clone'])

Content Aware Fill



Paste Sampled

Paste Splice

Blur

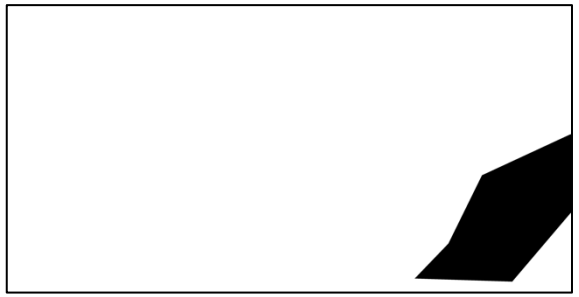
Image Localization Metrics

- Metrics
  - Matthews Correlation Coefficient (MCC)

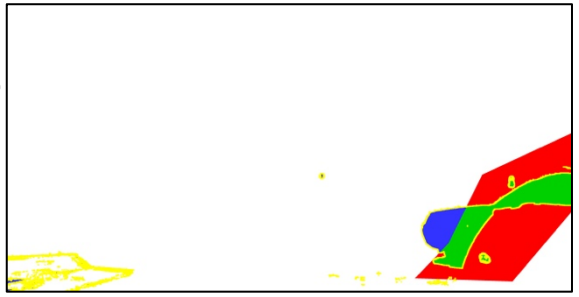
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \in [-1,1]$$

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

Probe + ref. mask overlay



System output mask



Color-coded scoring confusion matrix

Image Manipulation Localization Container Results

- 8 teams:
  - Honeywell FIBBER
  - Kitware
  - Purdue_TA11a
  - SRI-PRNU-TA1
  - UMD
  - USCISI-TA1.1
  - USCISI-TA1.2
  - USCISI
- 21 image localization systems as 04/09/2020:
  - Highest MCC = 0.247;
  - TRR = 0.939191
  - Team ID: USCISI-TA1.1
  - System ID: noiseprint-loc-3_2019-03-08T15-18-11f736000

Image Manipulation Localization Container Results: Opt In

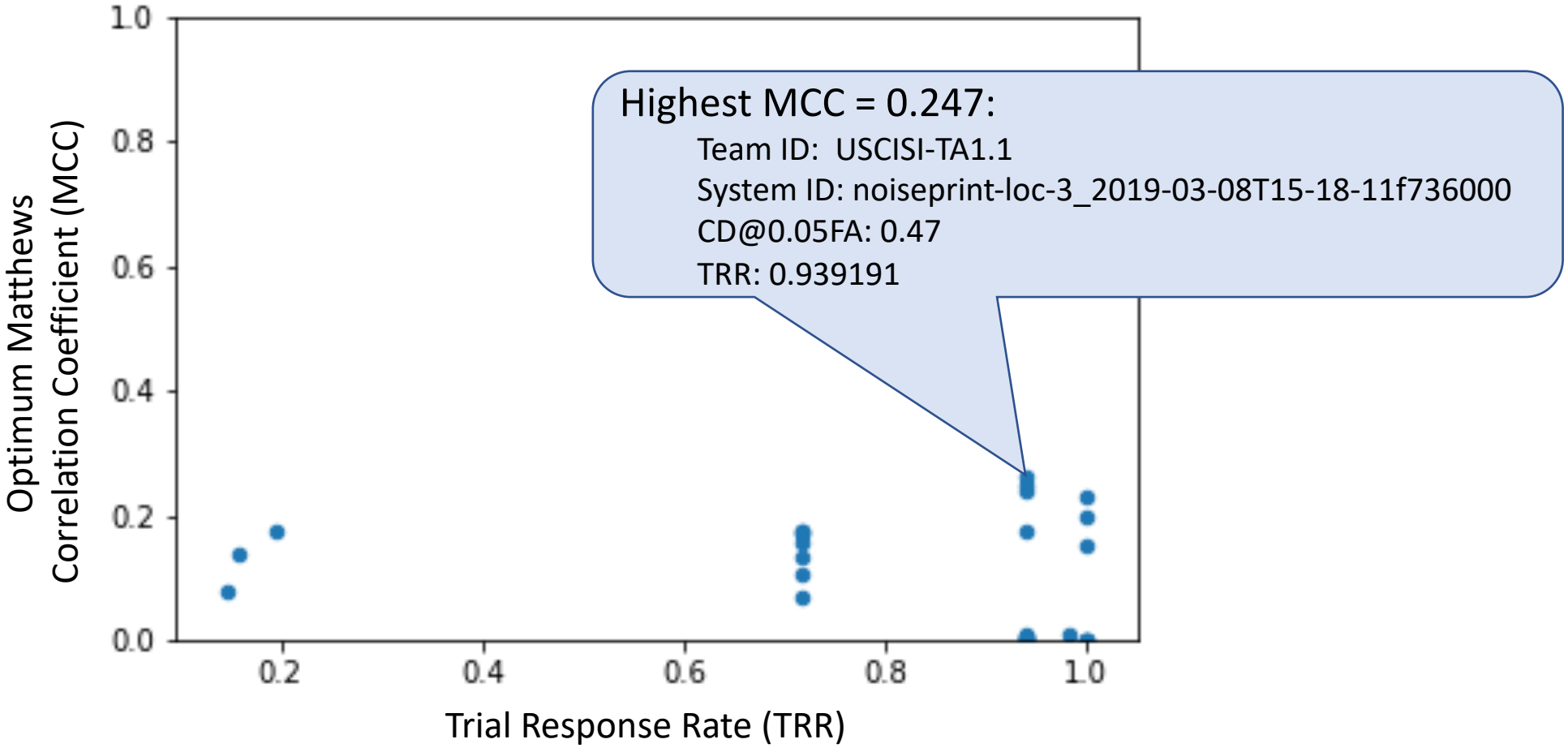
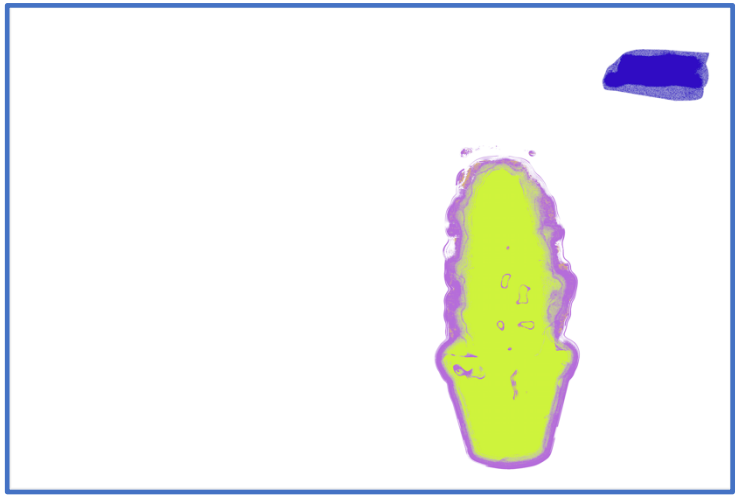


Figure: Image localization OptIn system Optimum MCC vs. TRR performance on MFC20 EP1 Image dataset

Image Localization Evaluation Example: Manipulation Mask

Base

Manipulated



Manipulated Region Mask

Donor

Sequence	BitPlane	Operation	Purpose	Color	Evaluated
6		OutputPng			Y
5	1	PasteSplice	add		Y
4	2	Blur			Y
3	3	ColorBalance			Y
2	4	ContentAwareFill	remove		Y
1		AntiForensicJPGCompression			Y

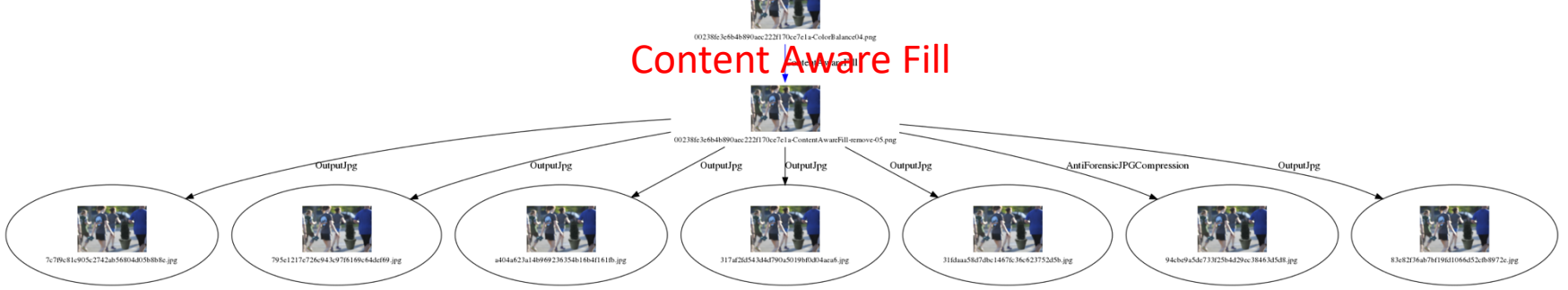
Manipulation Operation

Overlay of manipulated region mask

Image Localization Evaluation Example: Journal Graph

Donor

Base

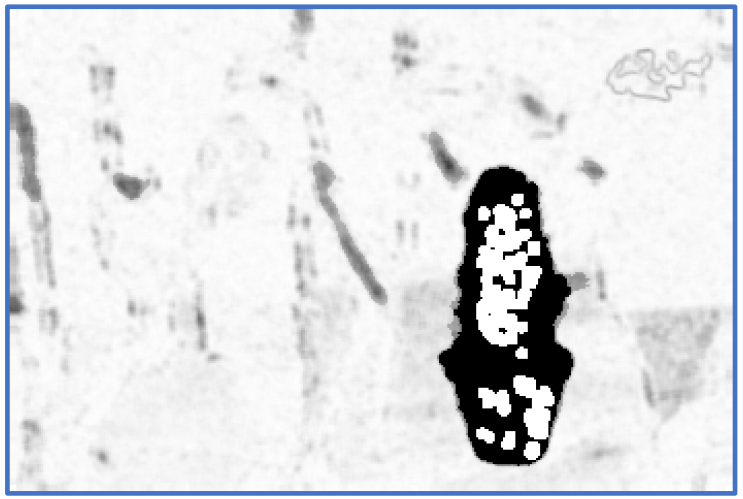


Journal Graph

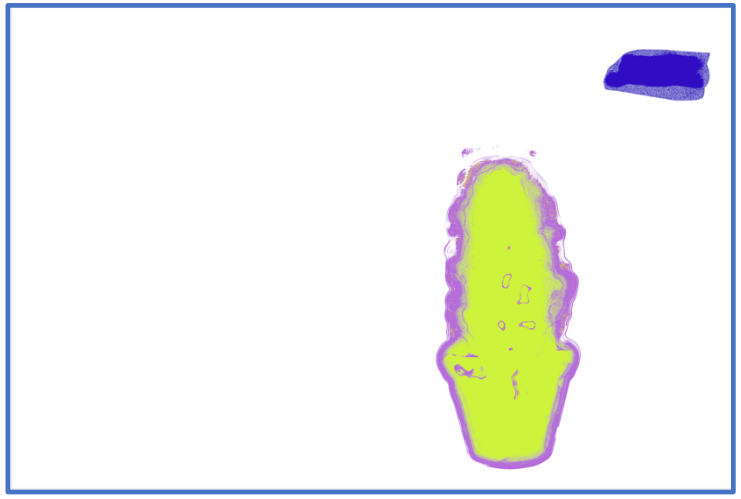
Manipulated

Image Localization Evaluation Example: Localization System

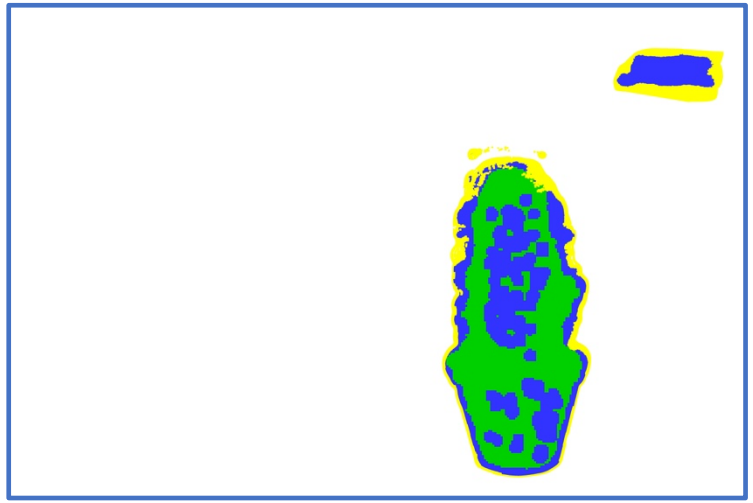
Overlay of manipulated region mask



System localization output



Manipulated Region Mask



Evaluation Results

Optimum Threshold: 29

Localization Metrics	Optimum
Nimble Mask Metric (NMM)	0.029
Matthews Correlation Coefficient (MCC)	0.699
Binary Weighted L1 Loss (BWL1)	0.048
Grayscale Weighted L1 Loss (GWL1)	0.095

Total Pixels (N): 20621162

Confuson Measures	OptimumPixelCount	OptimumProportion
True Postives (TP: green)	1052791	0.051
False Postives (FP: red)	0	0.000
True Negatives (TN: white)	18575635	0.901
False Negatives (FN: blue)	992736	0.048

No-Score Measures	Pixels	Proportion
Boundary No-Score Zone (BNS: yellow):	405142	0.019
Selective No-Score Zone (SNS: pink):	0	0.000
System Opt Out No-Score Zone (PNS: purple):	0	0.000
Total No-Score Zone:	405142	0.019

Image Manipulation Localization System - Team Performance Comparison Across Years (FullData)

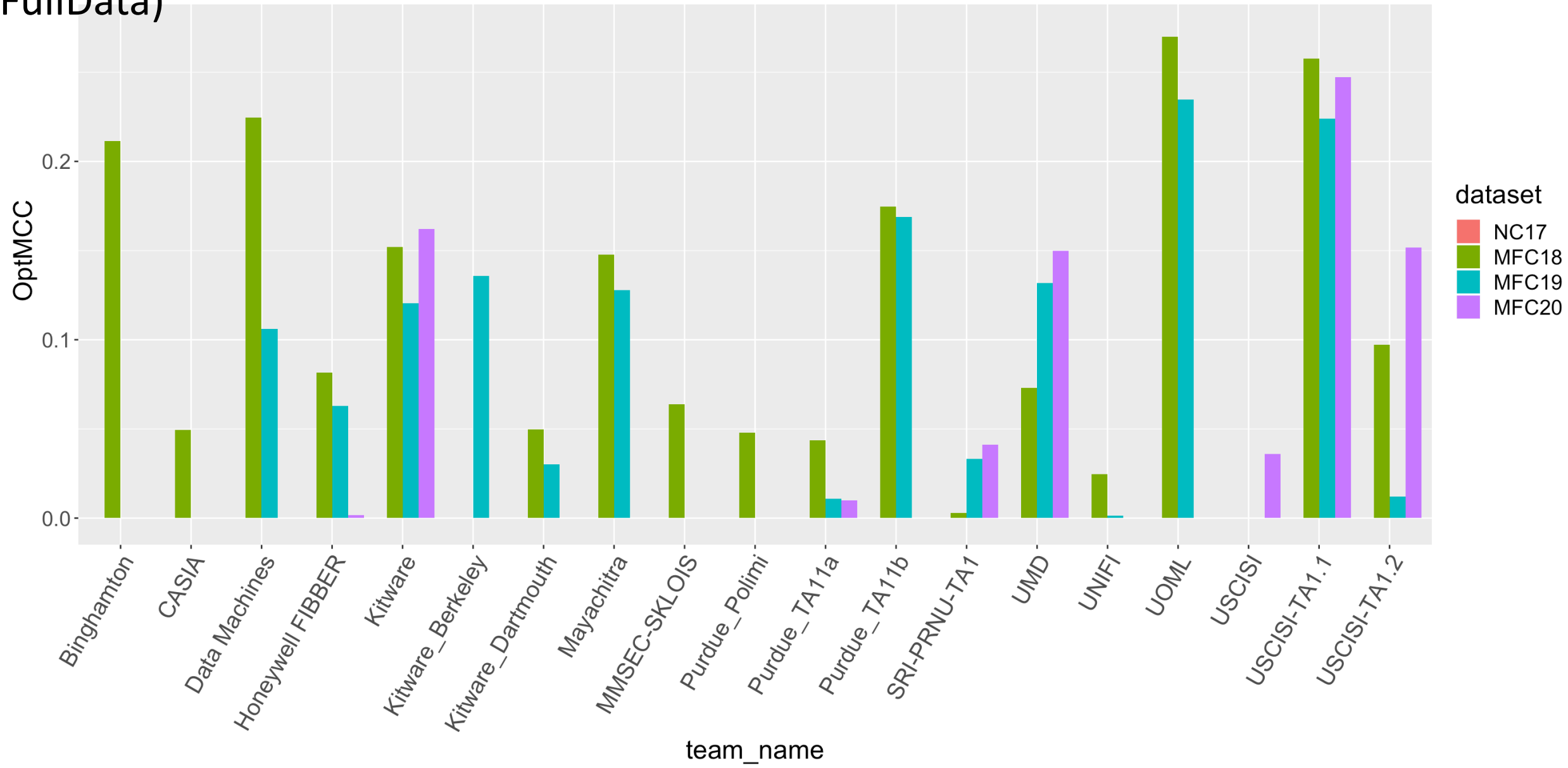
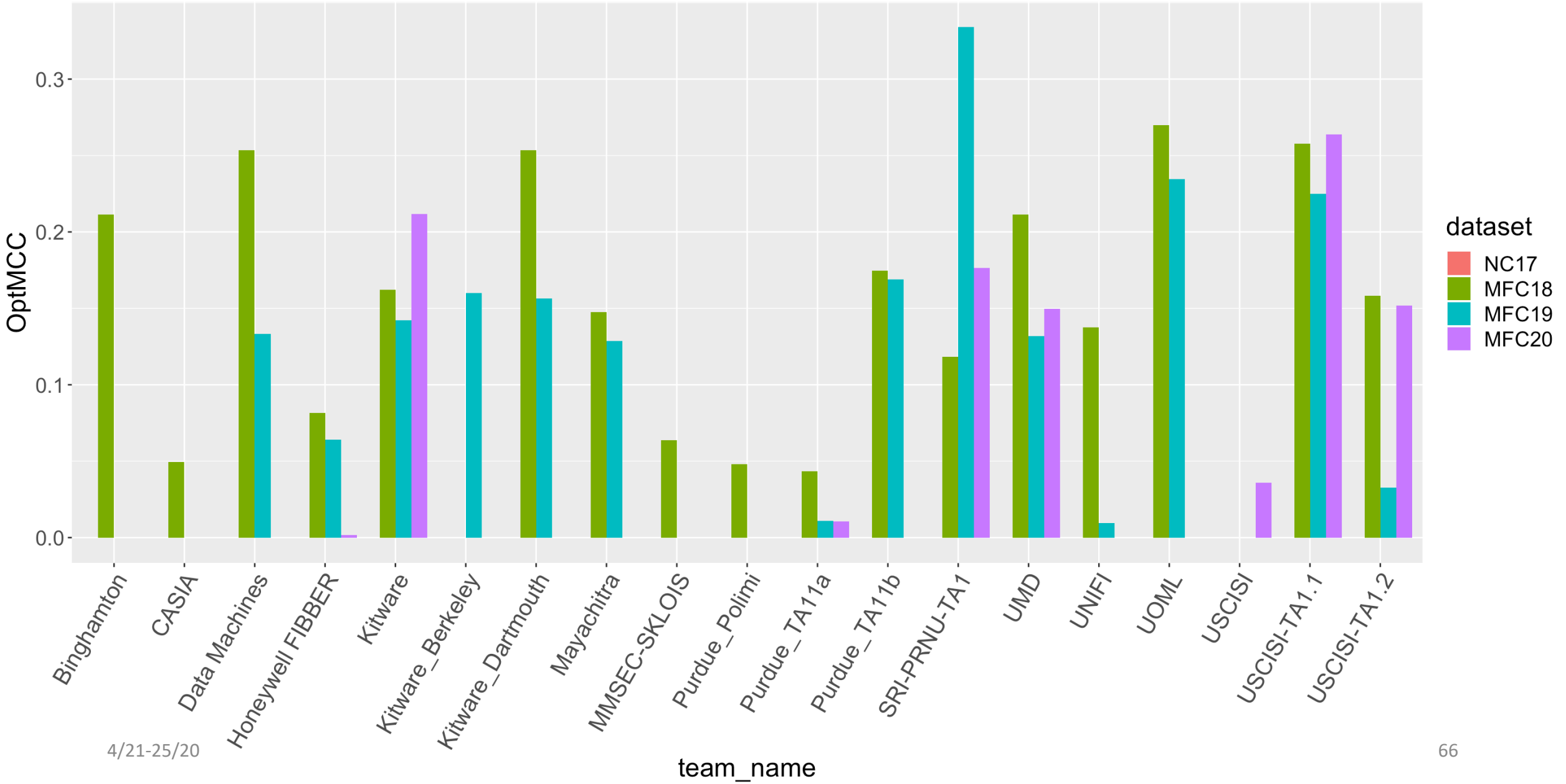
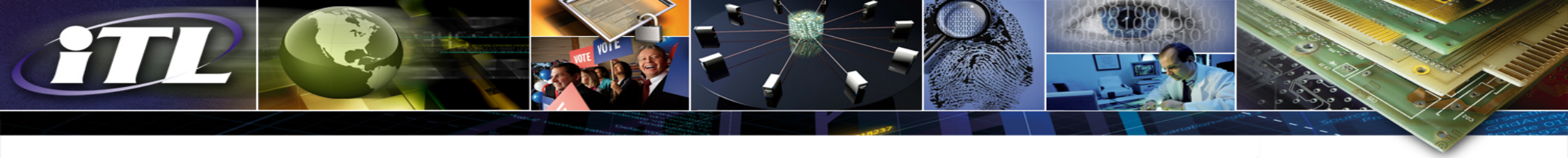


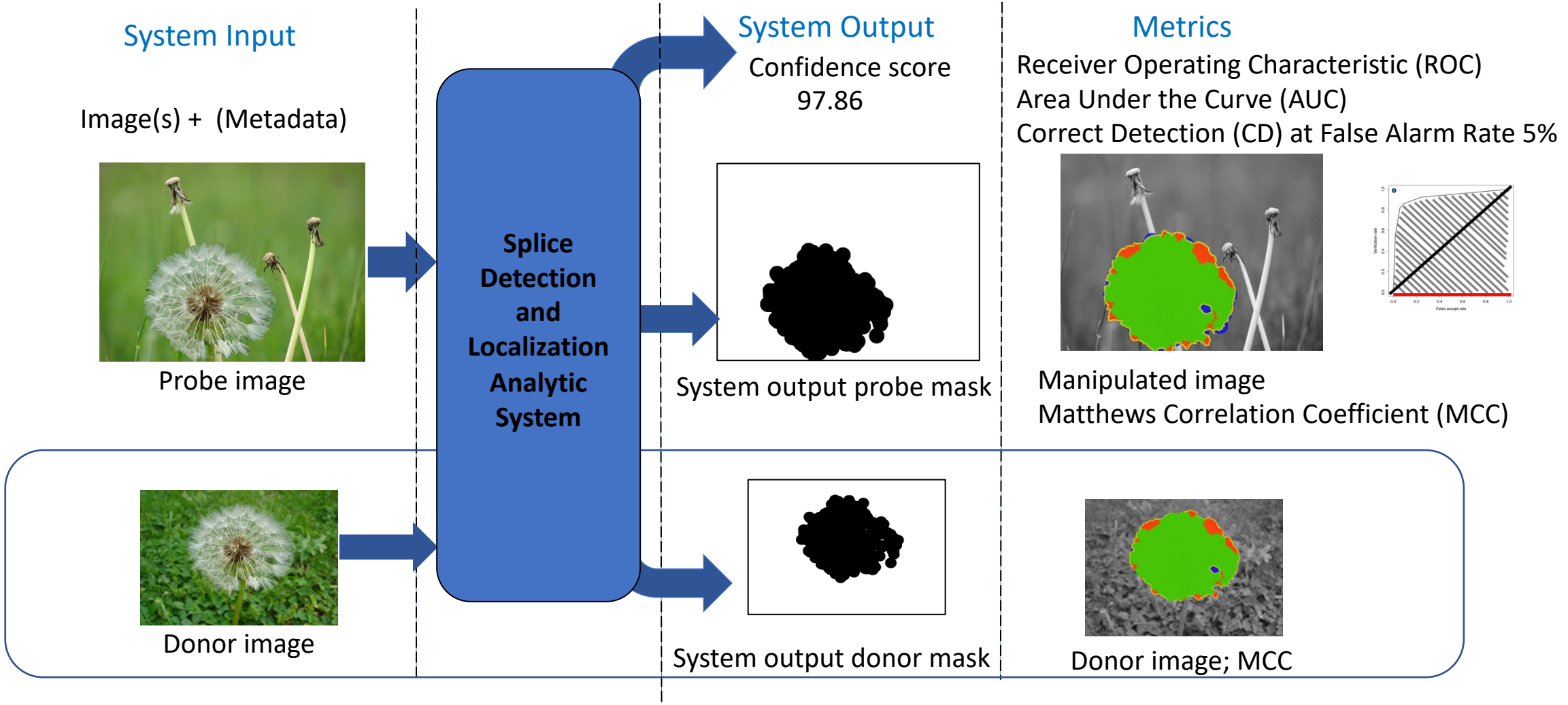
Image Manipulation Localization System - Team Performance Comparison Across Years (Opt In)





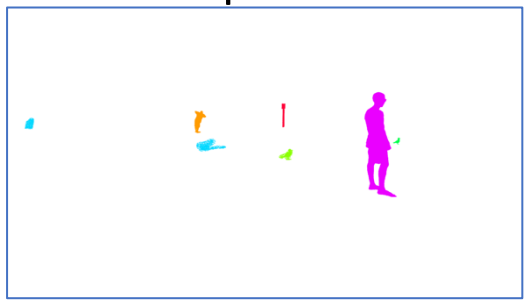
Splice Manipulation Detection and Localization Outline

Splice Manipulation Detection and Localization



Splice Localization Mask Example

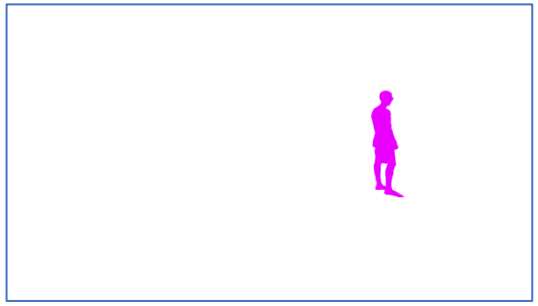
Color Composite Mask



Base Image

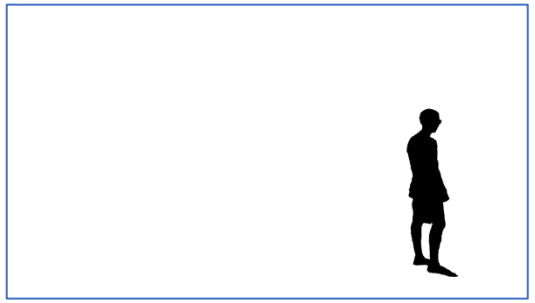
Probe Image

Reference Probe Mask Given the Donor



Donor Image

Donor Mask



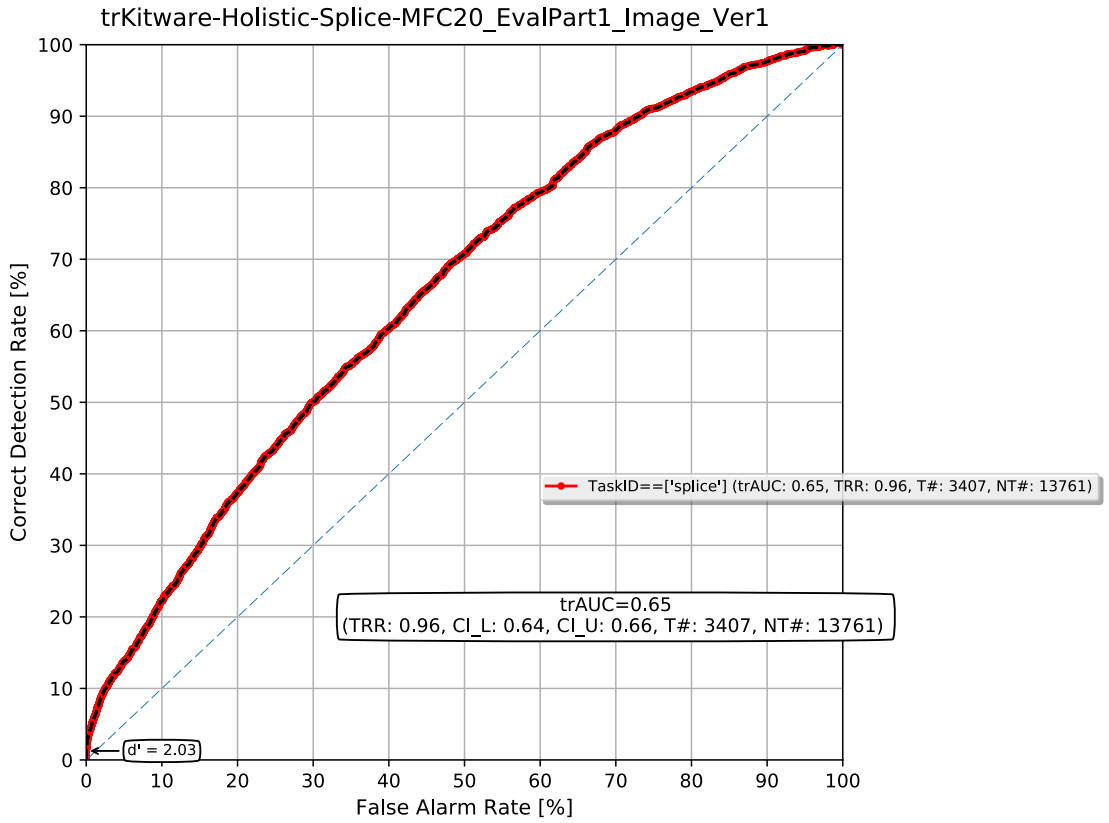
MFC Image Splice Evaluation Open Dataset Summary

NIST Splice Data Sets	Probe Pair	Image Journal	Date
NC17 EvalPart1	329K	156	06/2017
MFC18 EvalPart1	18K	381	03/2018
MFC19 EvalPart1	18K	621	03/2019
MFC20 EvalPart1	**18K**	**1266**	**03/2020**

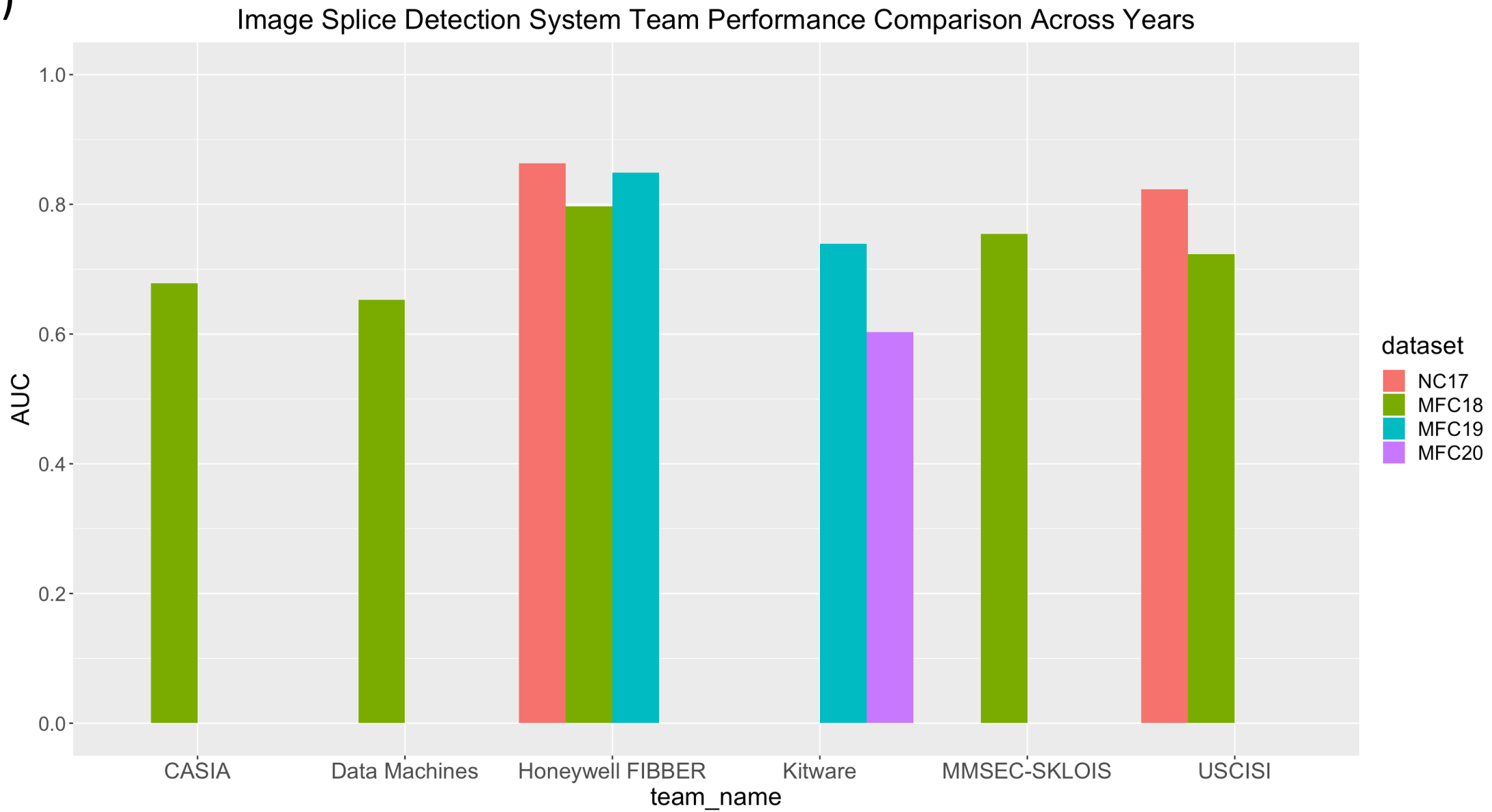
Image Splice Manipulation Detection Container Results

- 18K probe images
- 1 teams on detection system as 04/09/2020:
- Team ID: Kitware
- System ID: Kitware-Holistic-Splice-MFC20_EvalPart1_Image_Ver1
- FullData
  - AUC = 0.603
  - CD@0.05FA = 0.126
  - TRR = 1.0
- OptIn:
  - AUC = 0.653
  - CD@0.05FA = 0.139
  - TRR = 0.96

Figure: TA1 systems, Splice, MFC20 EP1, OptIn (TRR =0.96)



Splice Manipulation Detection System - Team Performance Comparison Across Years (Full Data)



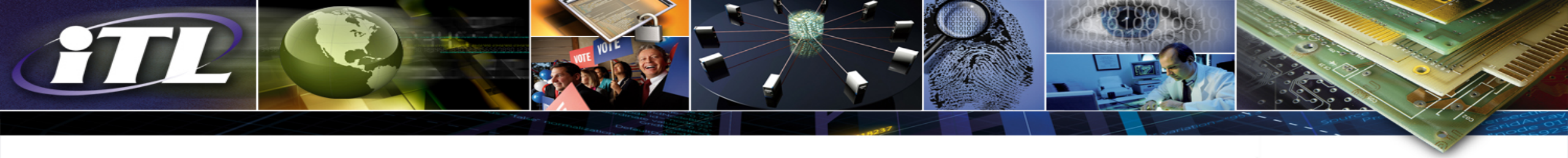
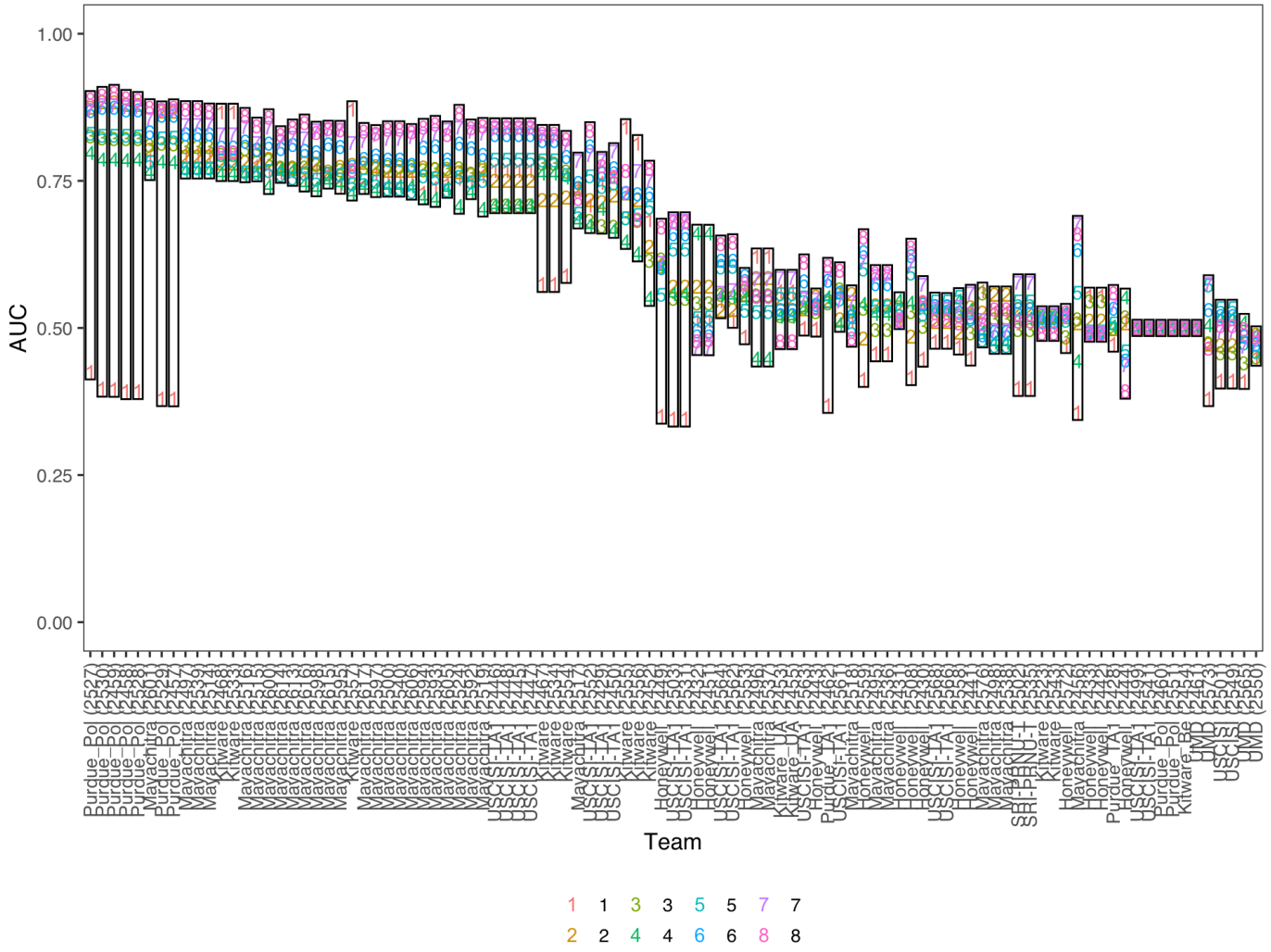


Image Manipulation and Splice Detection and Localization Analysis

Image Detection

At least 200 probes

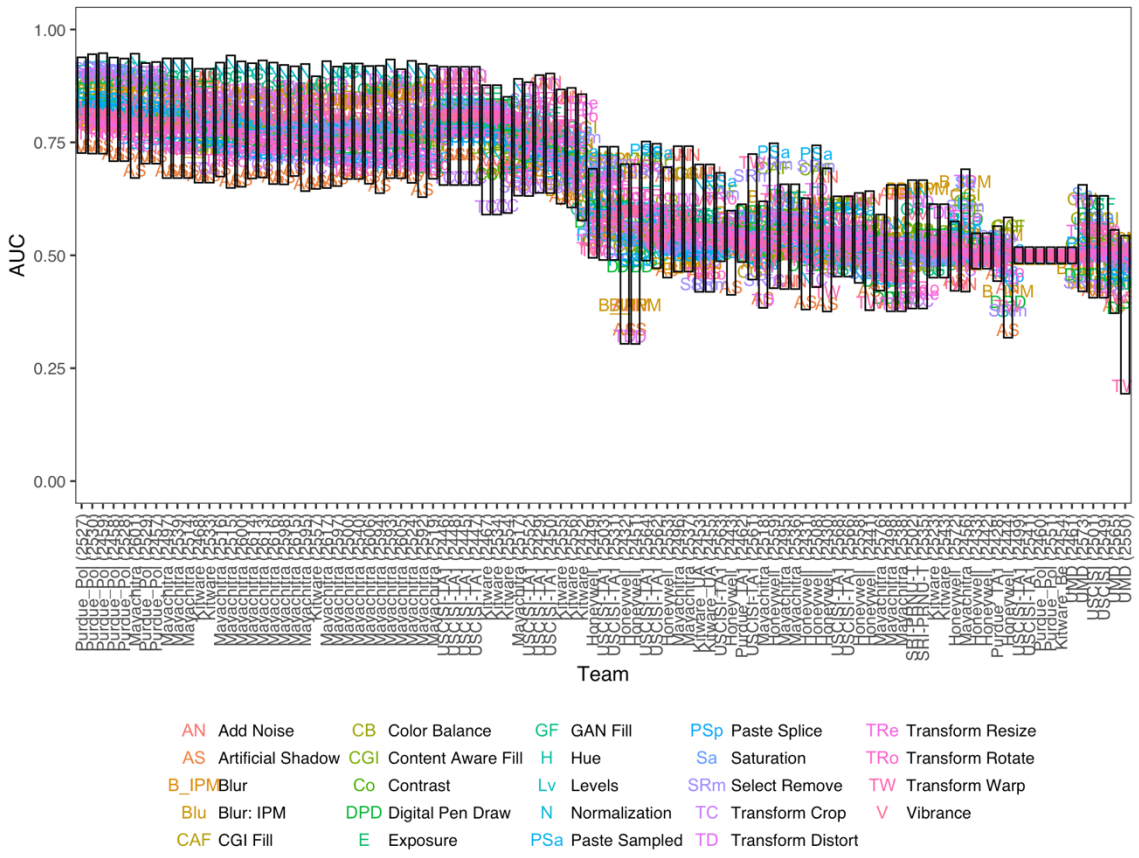
Effect of Manipulation Count on Detection



Lower = Better

Manipulation Count	Mean Rank
7	2.76
8	2.95
6	3.65
2	4.51
3	5.11
5	5.11
1	5.95
4	5.97

Effect of Operation on Detection



Operation	Mean Rank
Exposure	8.43
Hue	8.65
Tr Distort	8.92
Add Noise	9.01
Normalization	9.74
Sel Remove	10.02
GAN Fill	10.04
Blur IPM	10.44
Saturation	10.57
Color Balance	10.61
Vibrance	11.05
Paste Sample	12.06

Operation	Mean Rank
Levels	12.07
Contrast	12.31
C Aware Fill	12.89
Tr Resize	13.11
CGI Fill	13.17
Tr Warp	13.43
Tr Rotate	14.73
Blur	15.29
Dig Pen Draw	15.75
Paste Splice	17.52
Tr Crop	19.08
Art Shadow	21.11

Detection Teams

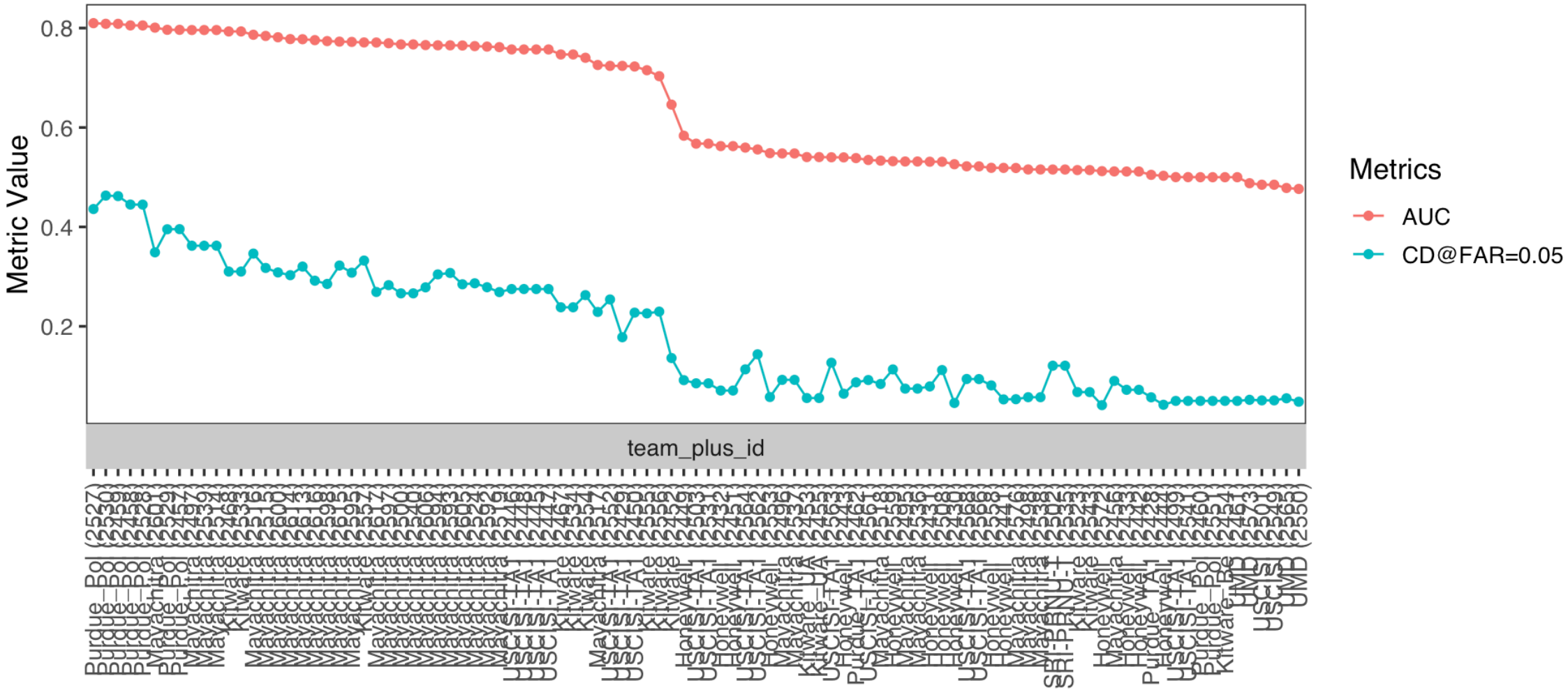
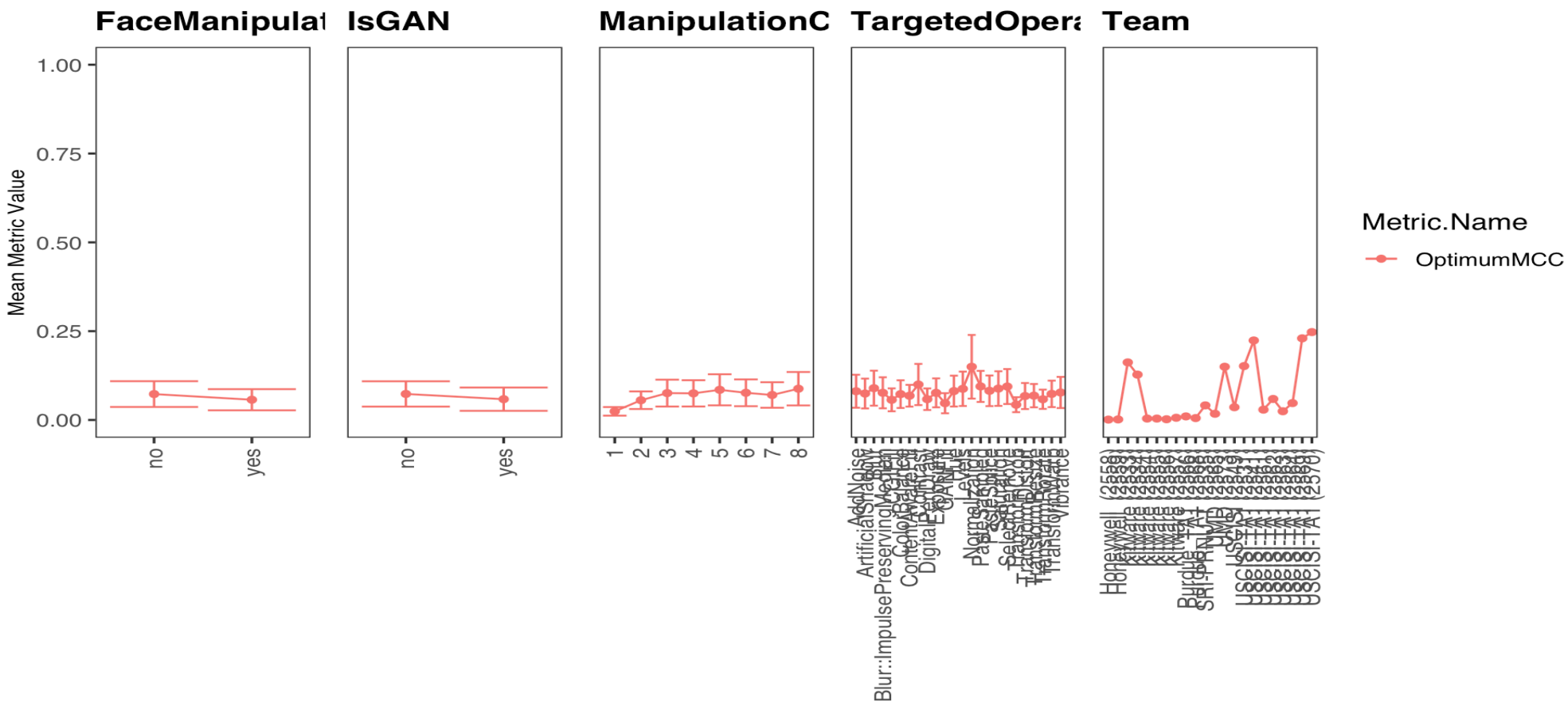
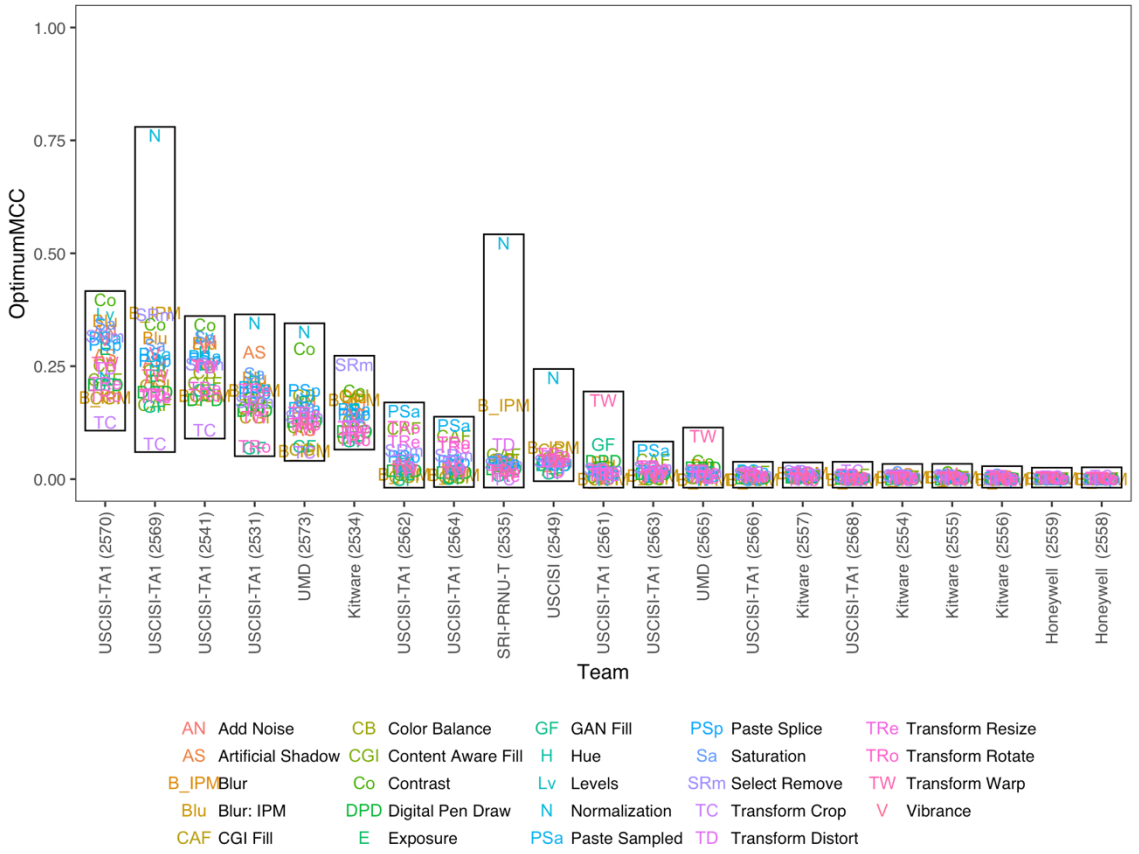


Image Localization

At least 200 probes



Effect of Operation on Localization



Operation	Mean Rank
Paste Sample	7.33
Sel Remove	7.57
Blur	8.76
Levels	9.14
Saturation	9.62
Paste Splice	9.67
Hue	9.76
Tr Resize	11.29
Vibrance	11.86
Add Noise	11.90
C Aware Fill	11.95
Contrast	12.05

Operation	Mean Rank
Art Shadow	12.10
Exposure	12.76
Color Balance	13.52
Tr Distort	13.86
Tr Rotate	13.90
Tr Warp	14.19
Blur IPM	15.00
Tr Crop	15.17
Normalization	15.64
Dig Pen Draw	15.86
CGI Fill	17.14
GAN Fill	19.95

Localization Teams

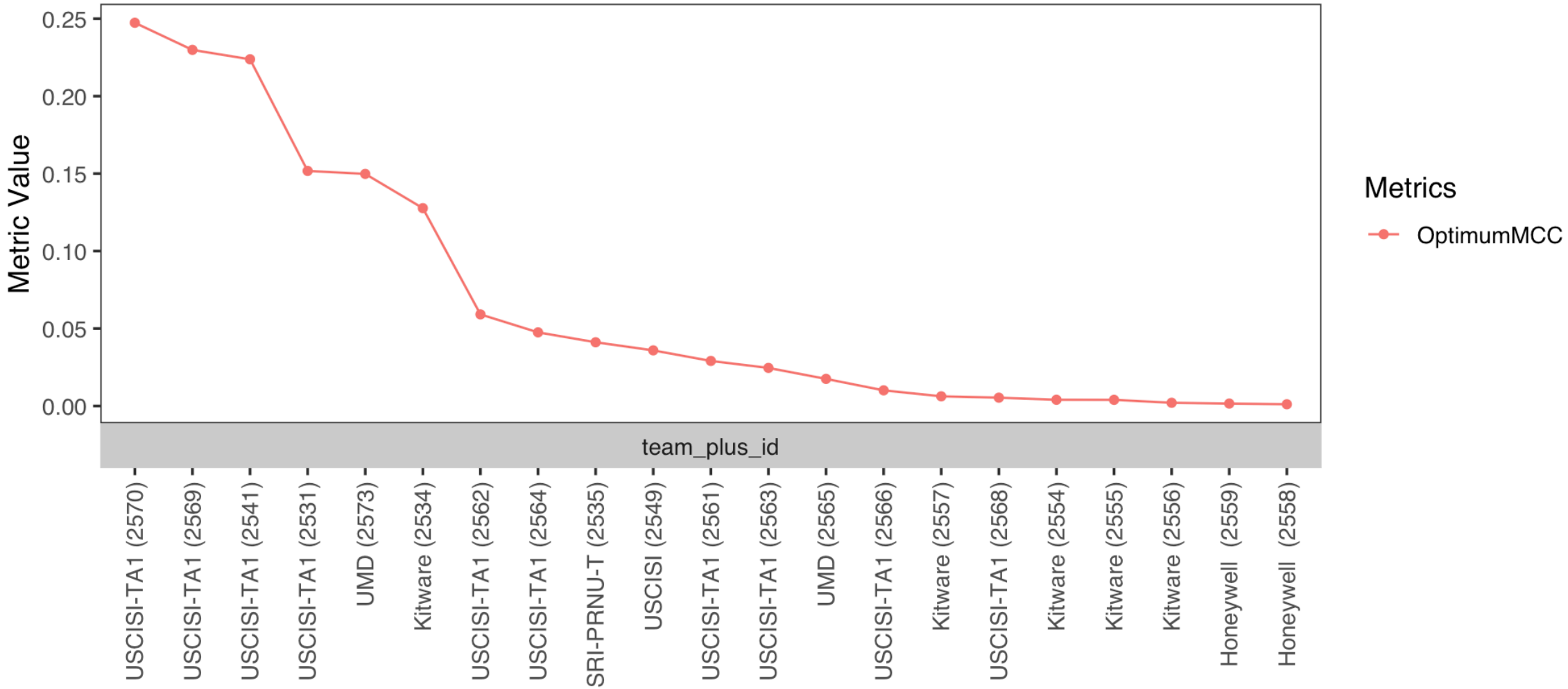


Image Special Studies

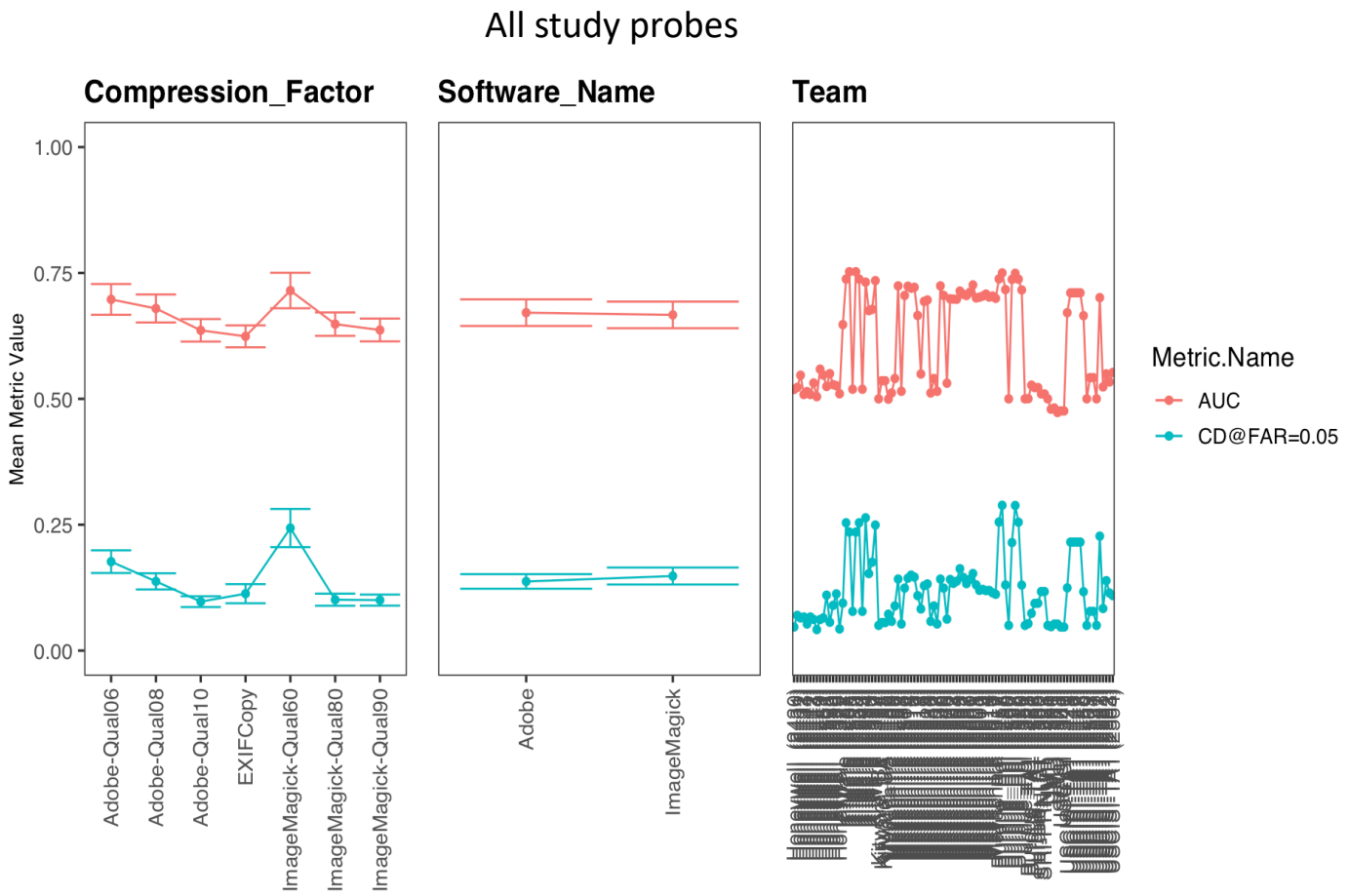
- Image Studies
  - Compression
  - Global Blur
  - Single Operation
  - Social Media Laundering - Image

Study Condition Definition

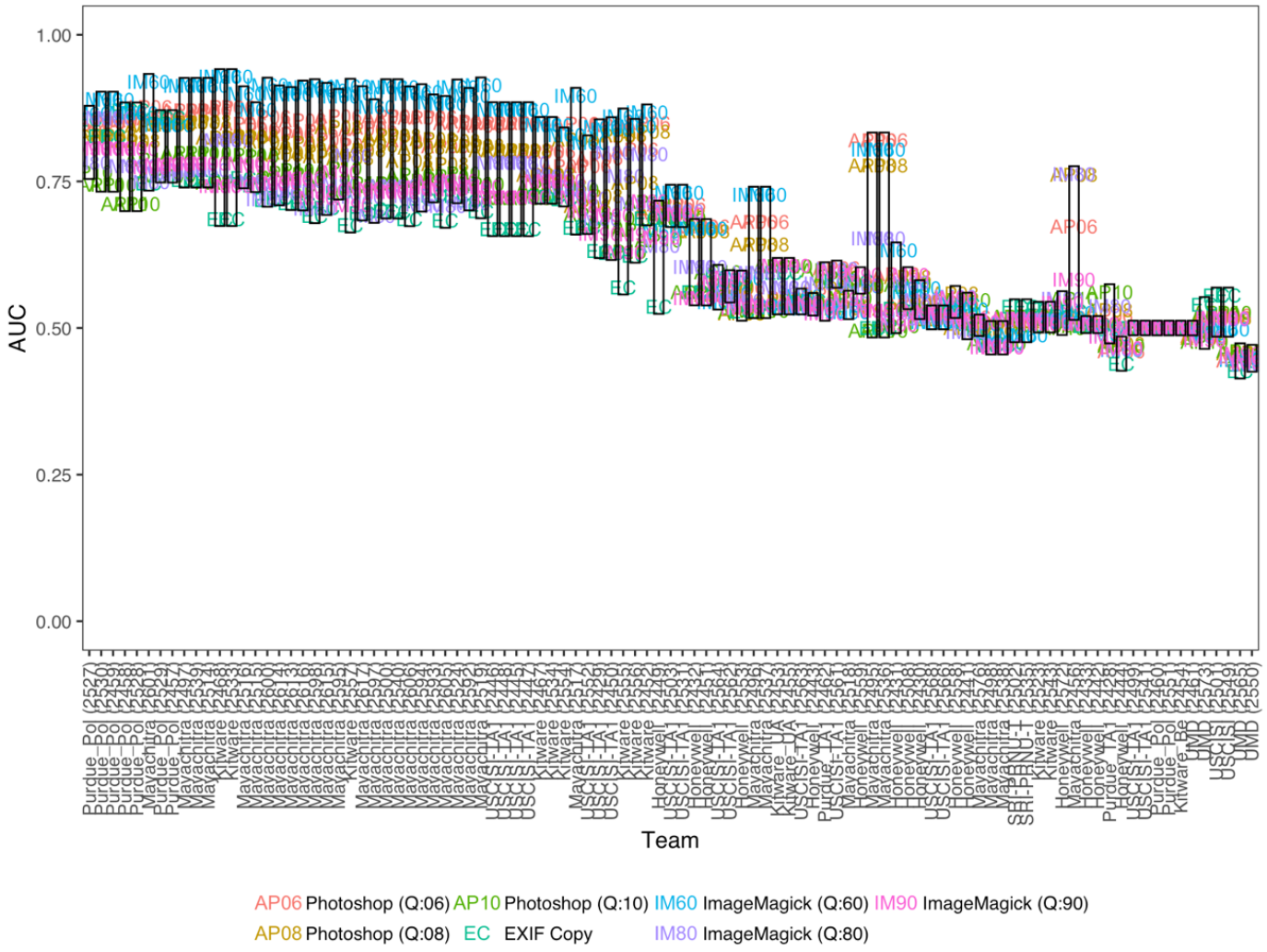
- Manipulation Detection (MD)
  - Target = Any manipulated image
  - Non-Target = HP media
- Operation-Only Detection (OOD)
  - Target = Only image with operation of interest; no other operations are present
  - Non-Target = HP media

Image Manipulation Detection Special Study – Compression Results (MD)

- 2 software approaches:
  - Adobe Photoshop
  - ImageMagick
- 3 compression levels for each approach:
  - Adobe (YesRGB): 6, 8, 10
  - ImageMagick: 60, 80, 90



Compression Factor (MD)



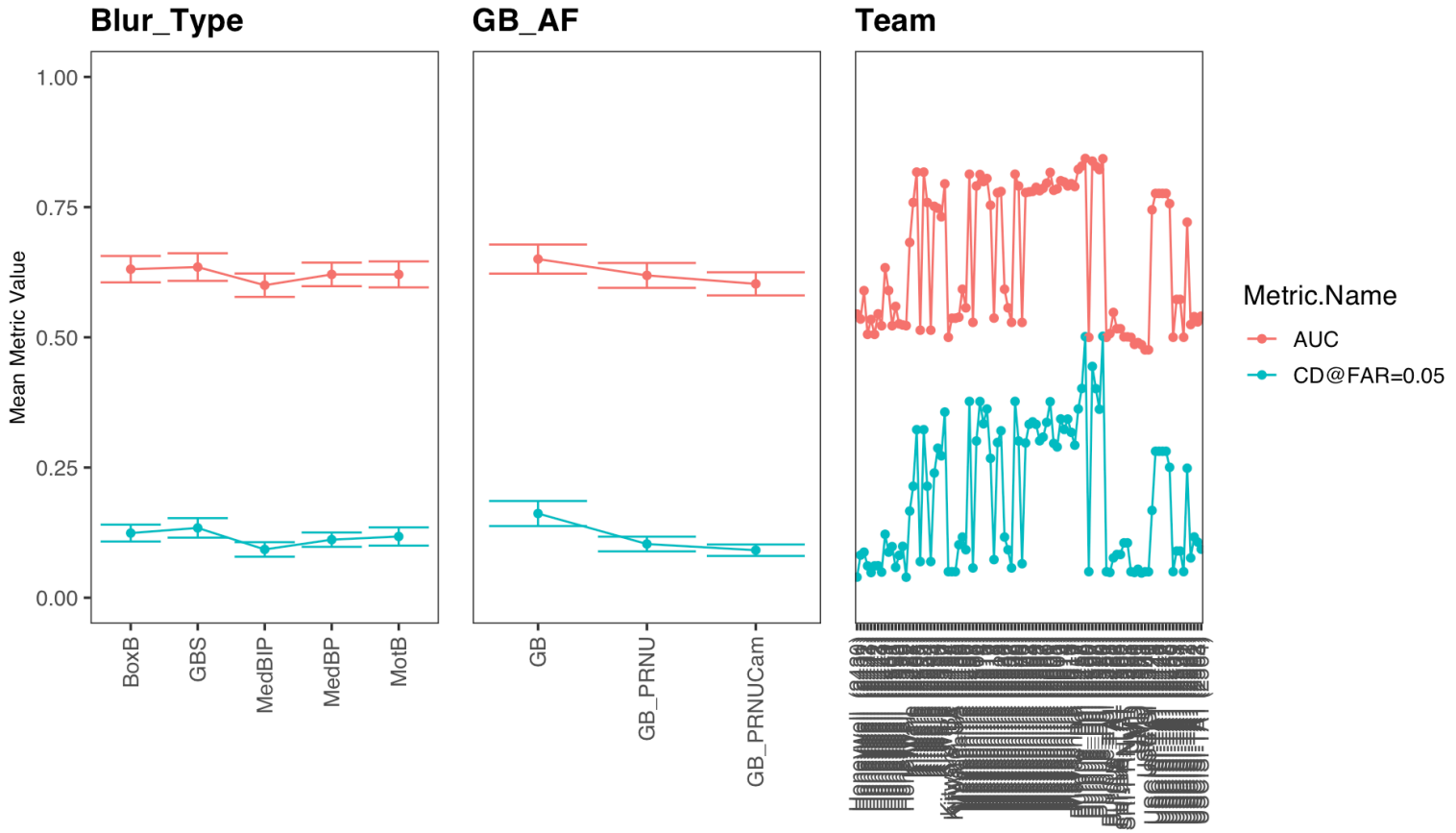
Lower = Better

Compression Factor	Mean Rank
AP: Qual 6	2.74
IM: Qual 60	2.77
AP: Qual 8	3.52
AP: Qual 10	4.54
IM: Qual 80	4.58
IM: Qual 90	4.64
EXIF Copy	5.22

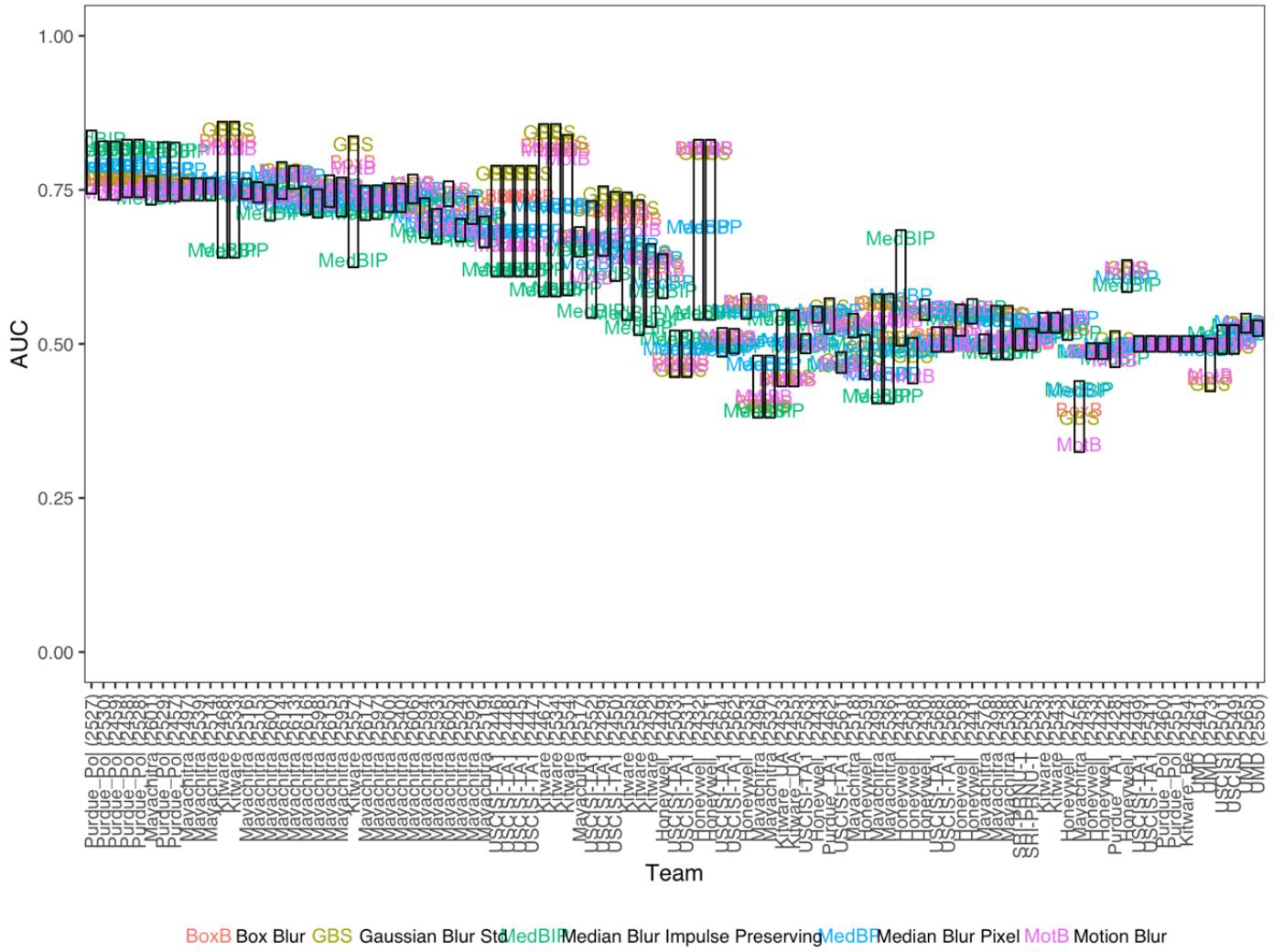
Image Manipulation Detection Special Study – Global Blur Results (MD)

- Blur types:
  - Gaussian Blur Std (GBS)
  - Median Blur Pixel (MedBP)
  - Median Blur Impulse Preserving (MedBIP)
  - Motion Blur (MotB)
  - Box Blur (BoxB)
- Anti-Forensics
  - Global Blur (GB)
  - GB + PRNU
  - GB + PRNU + Camera Mod

All study probes



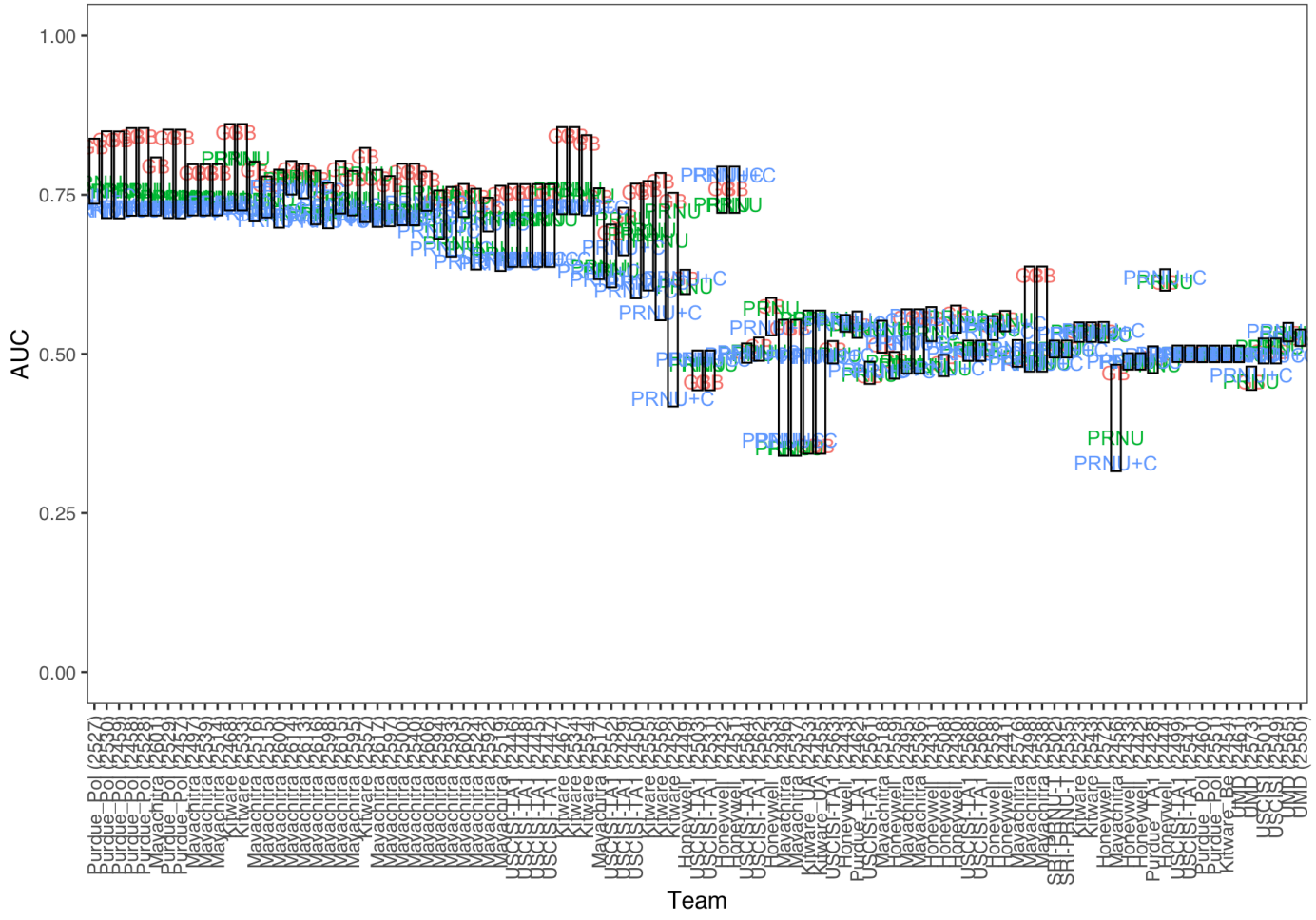
Blur Type (MD)



Lower = Better

Blur Type	Mean Rank
GBS	2.32
BoxB	2.65
MedBP	3.08
MotB	3.20
MedBIP	3.75

Global Blur Anti-Forensics (MD)



Lower = Better

Anti-Forensics	Mean Rank
GB	1.44
GB+PRNU	2.02
GB+PRNU +Cam	2.54

Image Manipulation Detection Special Study – Social Media Laundering (Image) Results (MD)

- 5 Scenarios
  - Scenario 1: Facebook (Synthetic)
    - mobile upload, mobile download
  - Scenario 2: Facebook (Synthetic)
    - desktop upload, mobile download
  - Scenario 3: Facebook (Synthetic)
    - mobile upload, desktop download
  - Scenario 4: Facebook (Synthetic)
    - desktop upload, desktop download
  - Scenario 5: Instagram (Synthetic)
    - mobile upload, mobile download

All study probes

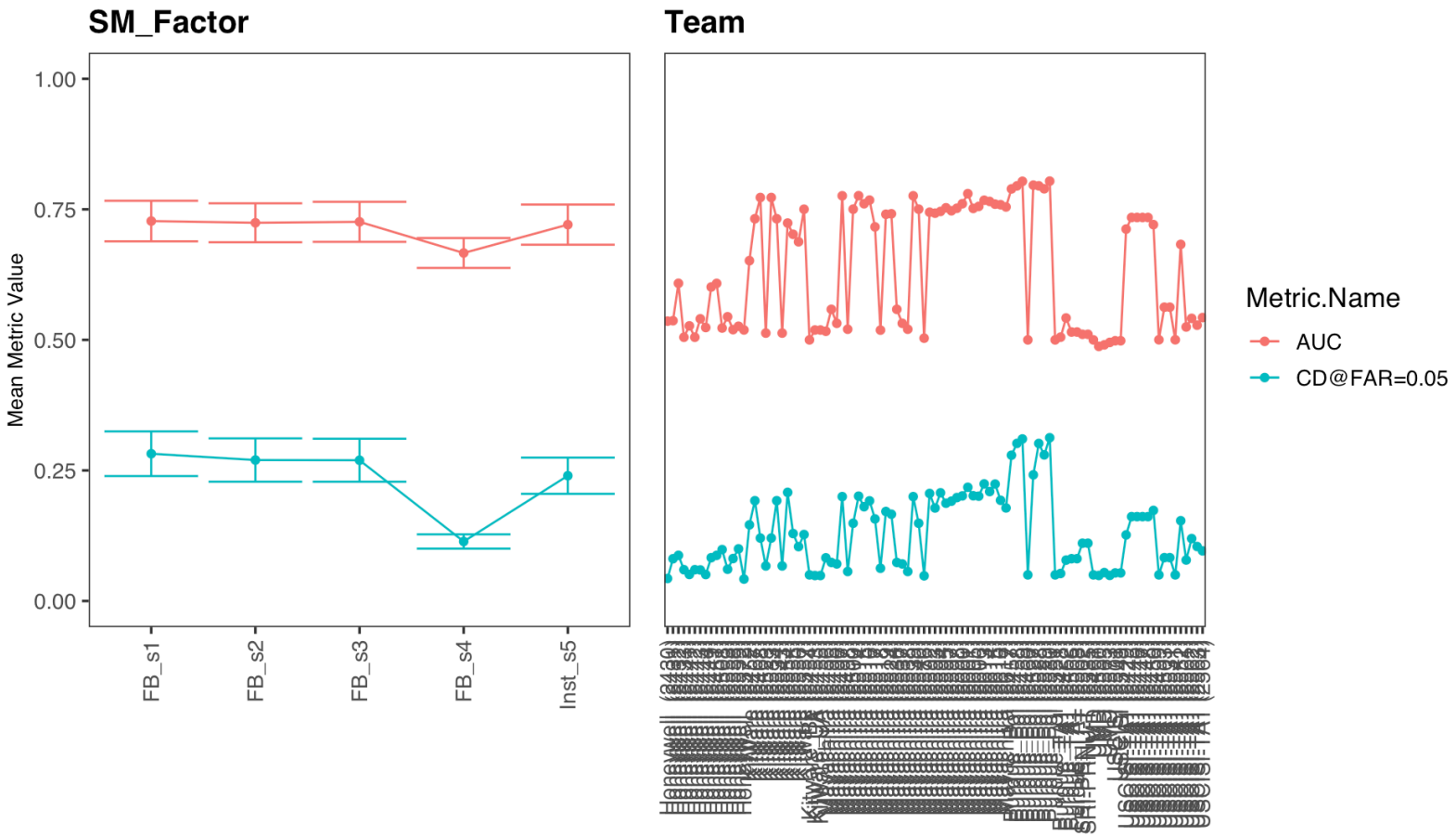
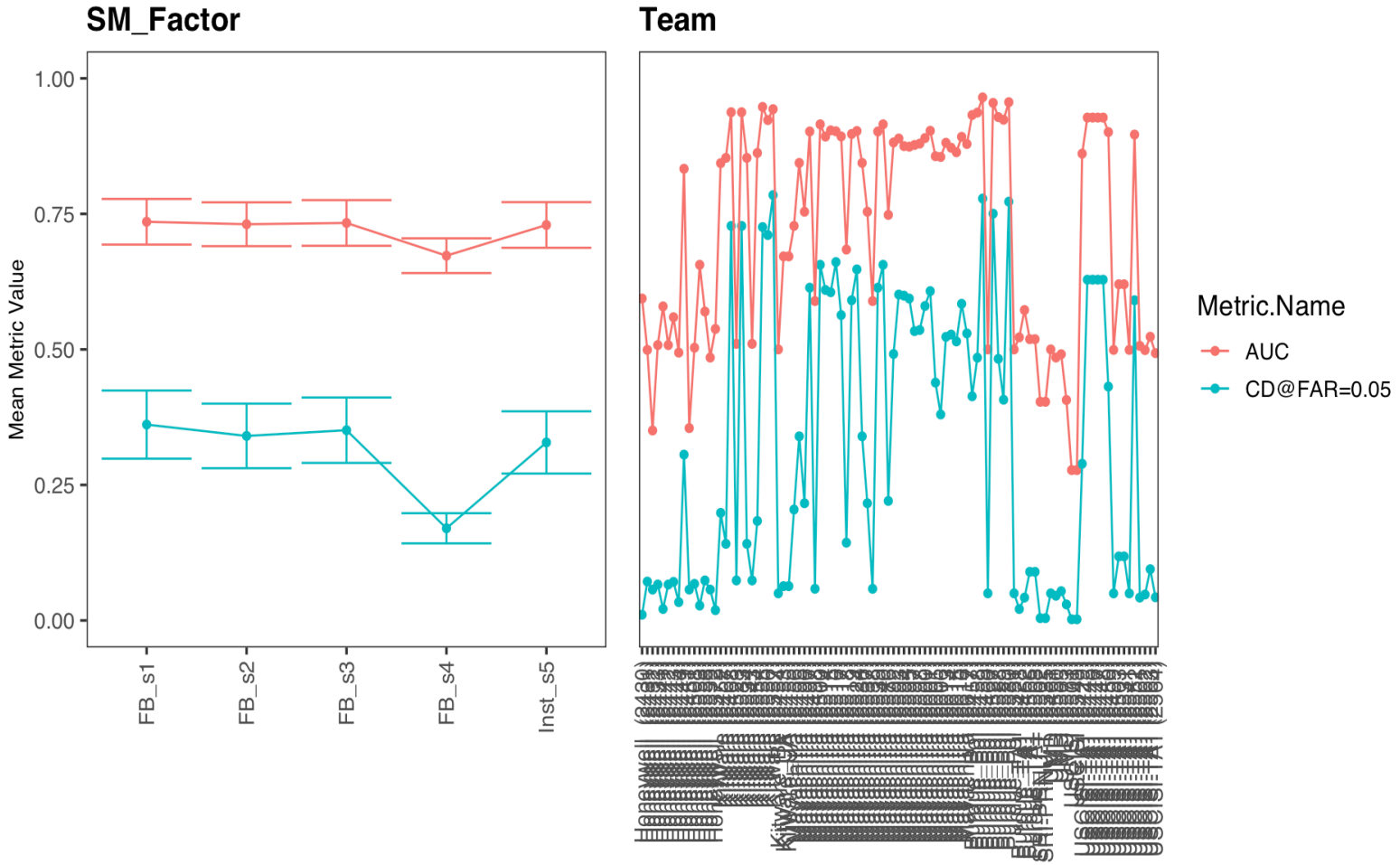
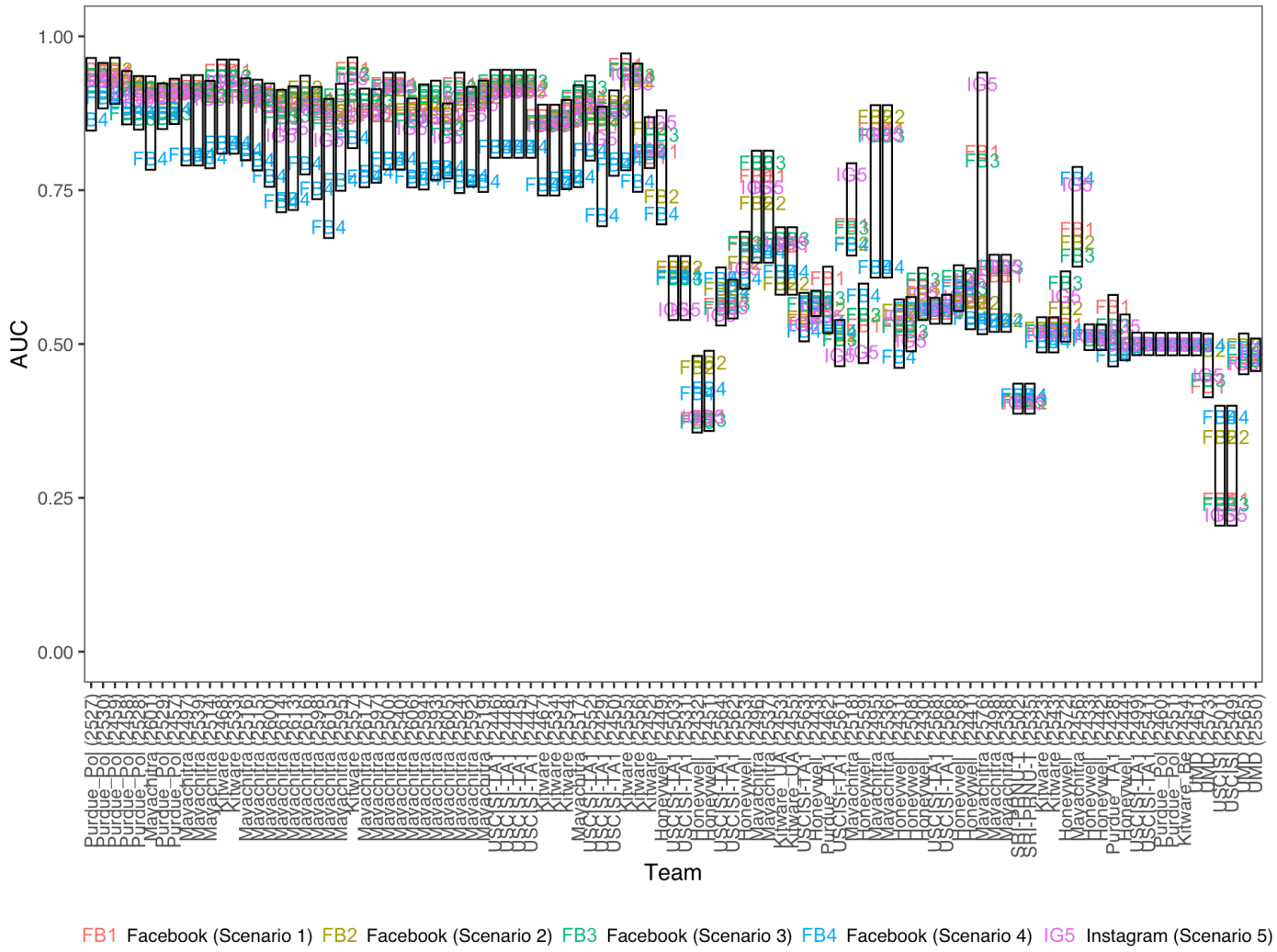


Image Manipulation Detection Special Study – Social Media Laundering (Image) Results (OOD)

- 5 Scenarios
  - Scenario 1: Facebook (Synthetic)
    - mobile upload, mobile download
  - Scenario 2: Facebook (Synthetic)
    - desktop upload, mobile download
  - Scenario 3: Facebook (Synthetic)
    - mobile upload, desktop download
  - Scenario 4: Facebook (Synthetic)
    - desktop upload, desktop download
  - Scenario 5: Instagram (Synthetic)
    - mobile upload, mobile download



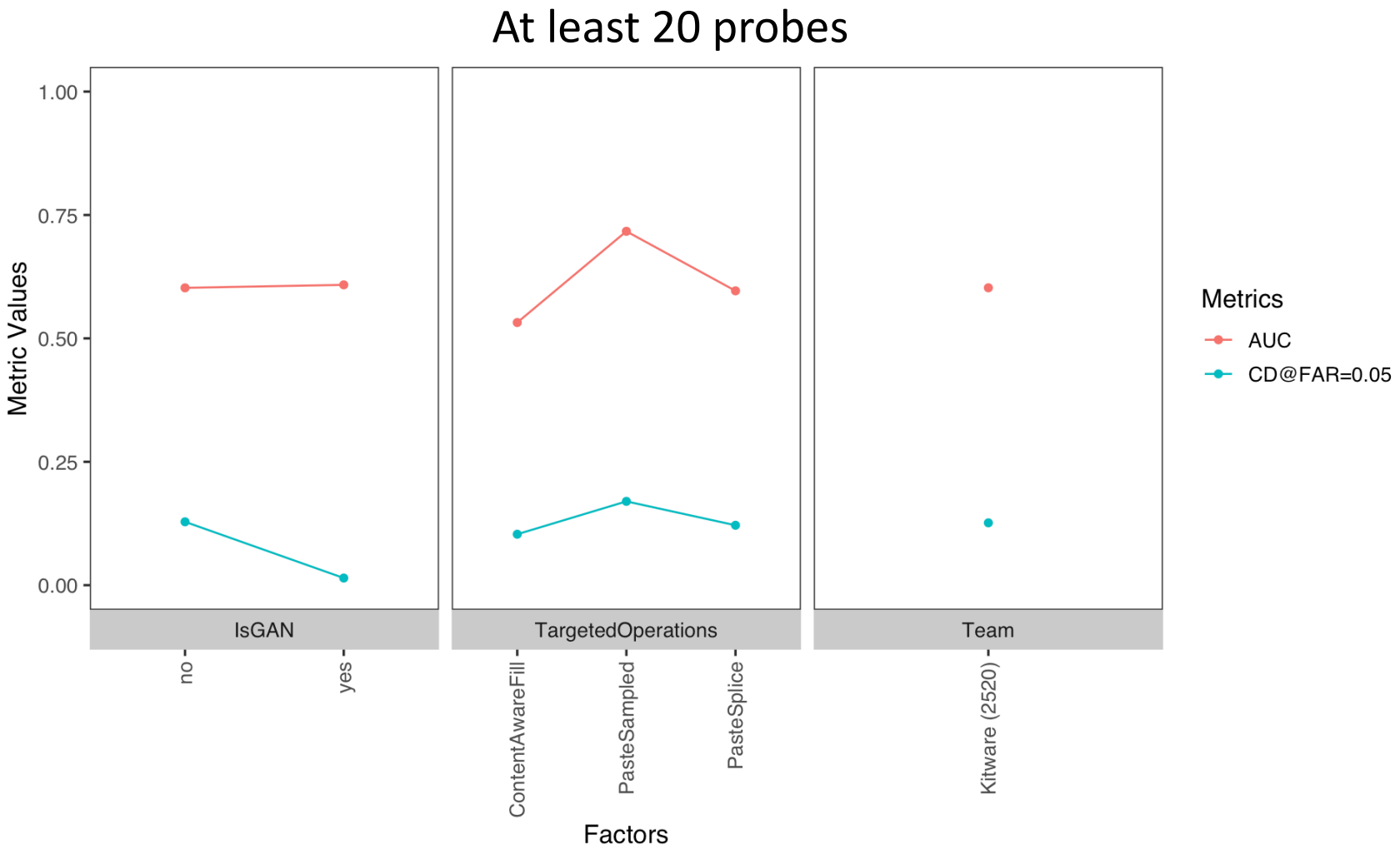
Social Media Factor (MD)

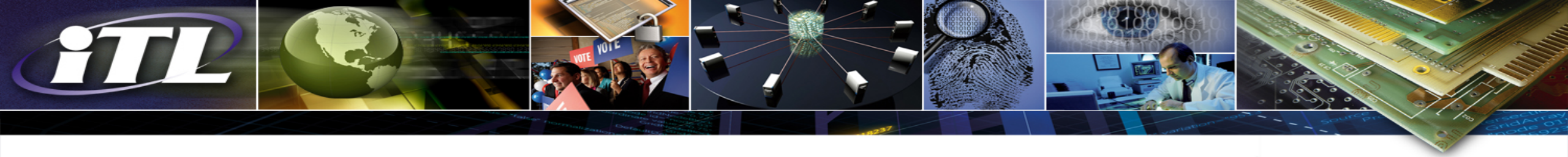


Lower = Better

SM Factor	Mean Rank
FB 1	2.41
FB 2	2.61
FB 3	2.61
IG 5	3.33
FB 4	4.04

Splice Detection





MFC20 Video Evaluation Results Deep Dive

Jonathan Fiscus (Co-PI), Dr. Haiying Guan (Co-PI), Dr. Yooyoung Lee,
Dr. Amy Yates[+], Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah,
Dr. Xiongnan Jin

Multimodal Information Group, [+] Image Group
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology (NIST)

April 21-25, 2020

Video Detection and Localization Outline

- Task definition
- Evaluation dataset
- MFC20 Video Detection and Localization results
- MFC20 Video Detection and Localization result analysis

Video Manipulation Detection and Temporal Localization

- Video Detection metrics
  - Receiver Operating Characteristic (ROC)
  - Area Under the Curve (AUC)
  - Correct Detection (CD) at False Alarm Rate (FAR) of 5%
- Video Temporal Localization
  - Metrics: Matthew Correlation Coefficient (MCC)

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

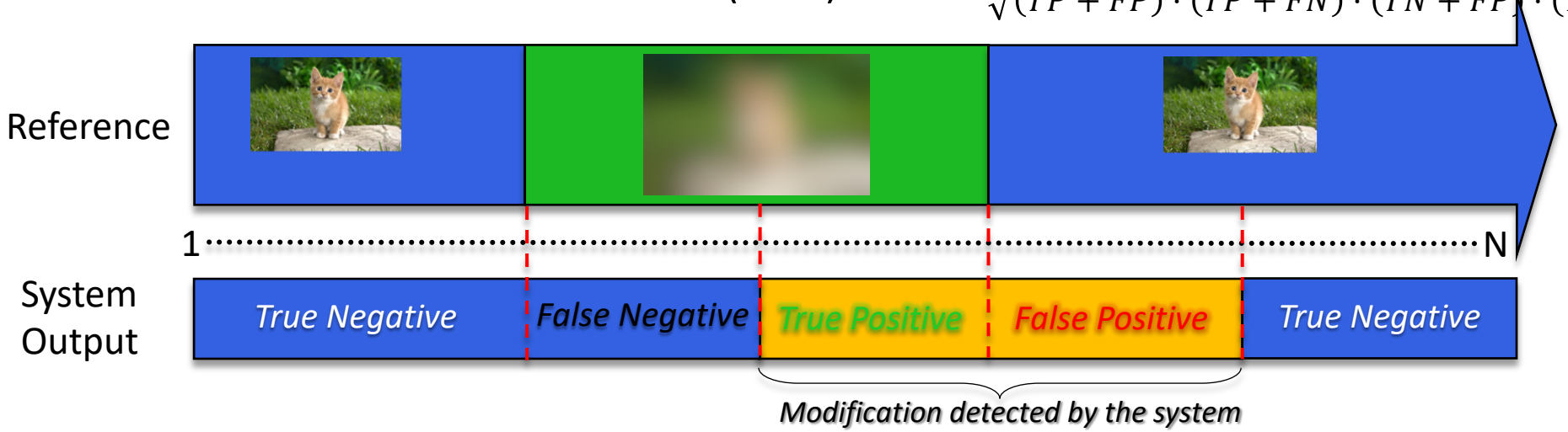


Figure: Video Temporal Detection and Localization

MFC Video Evaluation Open Dataset Summary

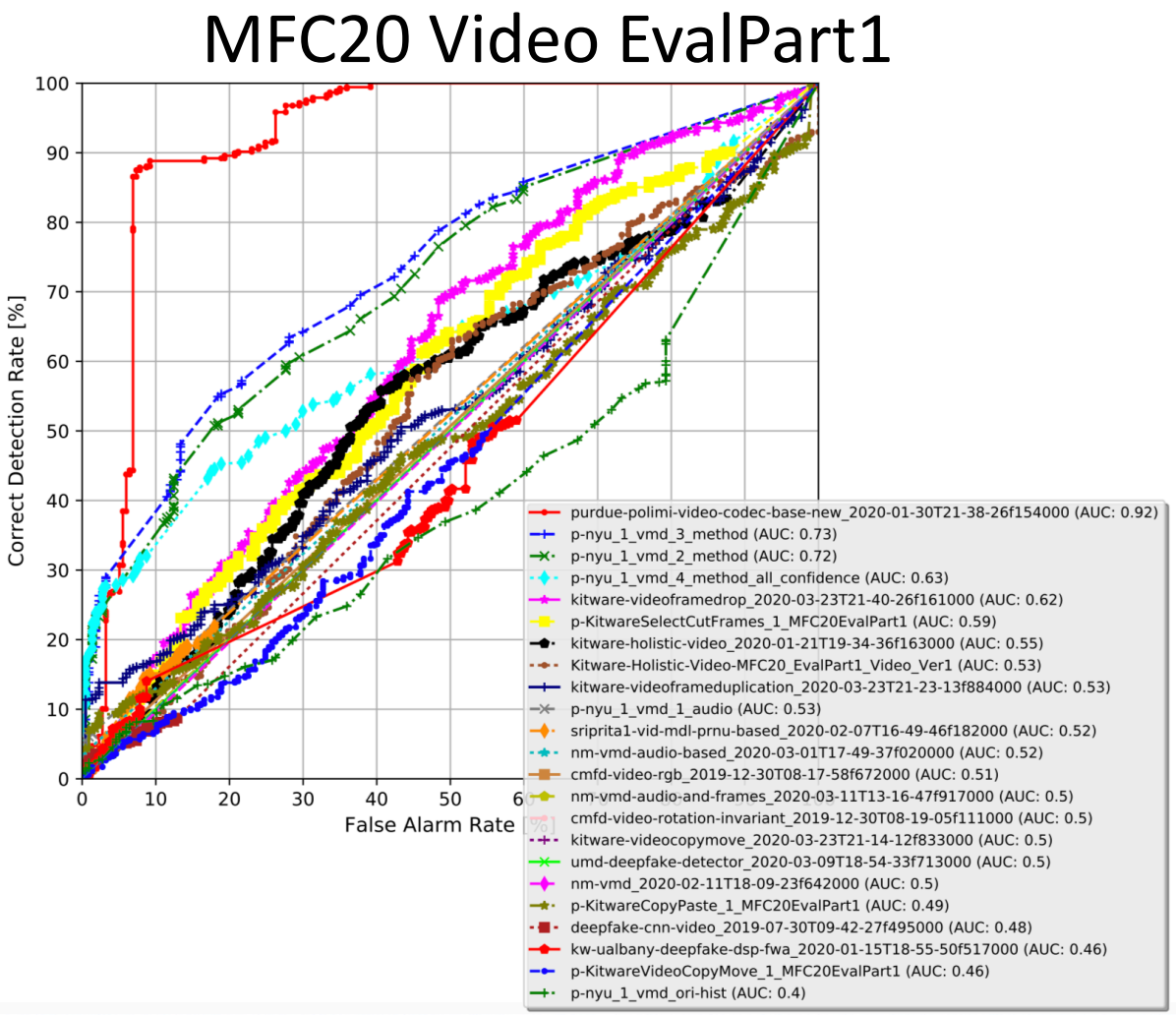
NIST Data Sets	Video Probe	Video Journal	Date
NC17 EvalPart 1	360	47	06/2017
MFC18 EvalPart1	1K	114	03/2018
MFC19 EvalPart1	1.5K	163	03/2019
MFC20 EvalPart1	**2K**	**217**	**03/2020**

Video Detection Task Participation

- 2K Probes:
- 9 teams:
  - Kitware_UAlbany
  - Kitware
  - NYU
  - Purdue_Polimi
  - SRI-PRNU-TA1
  - UMD
  - UNIFI
  - USCISI-TA1.1
  - USCISI-TA1.2
- Two Evaluation Conditions:
  - Video Only
  - Video with MetaData
- Two Special Collections
  - Frame Drop/Duplication
  - Social Media Laundering - Video

Video Detection Performance: Video Only, Full Data

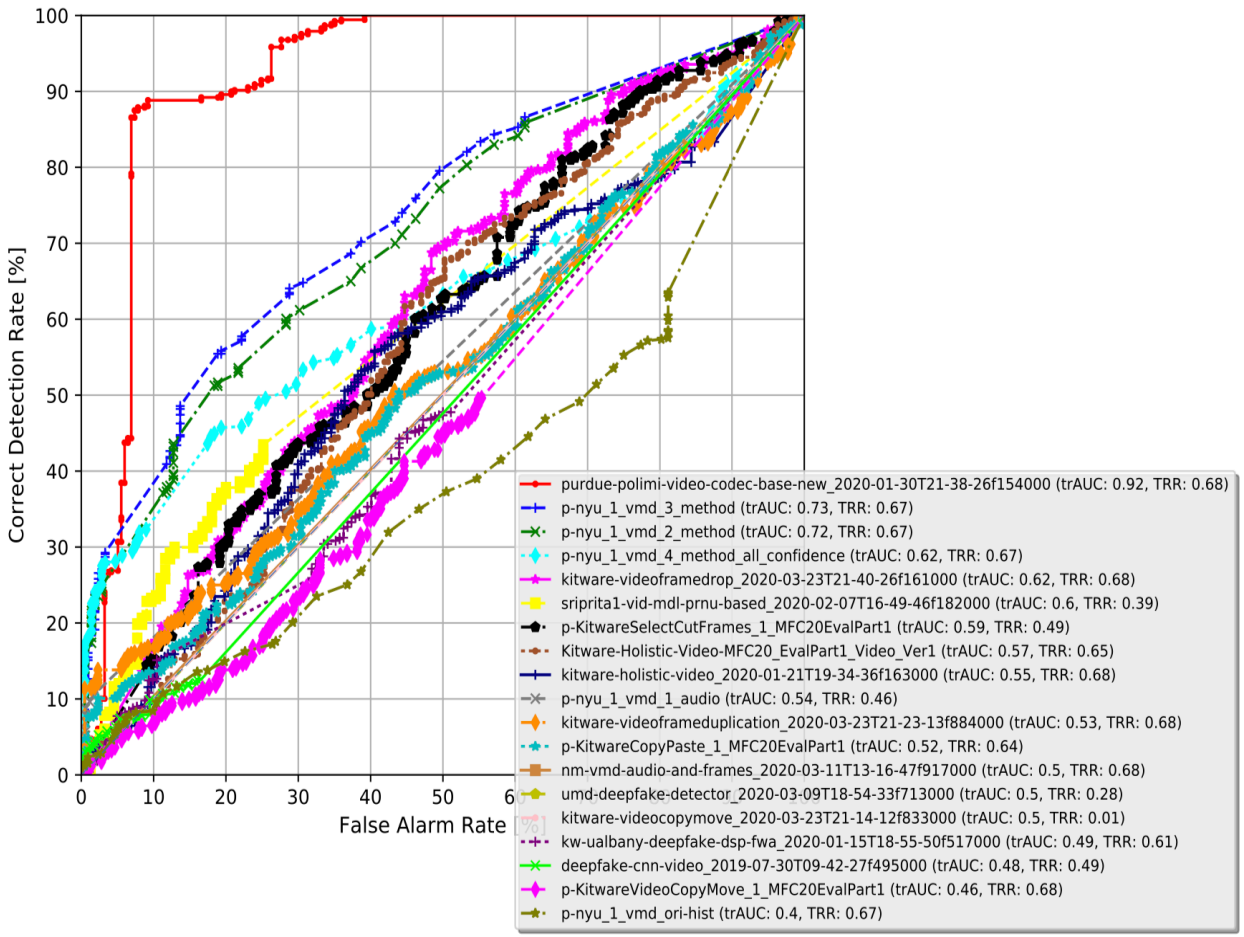
- 23 video detection systems:
- Highest AUC & CD@0.05FA:
  - AUC = 0.921; (CD@0.05FA = 0.269)
  - Team ID: Purdue_Polimi
  - System ID: purdue-polimi-video-codec-base-new_2020-01-30T21-38-26f154000



Video Detection Performance: Video Only, Opt-In

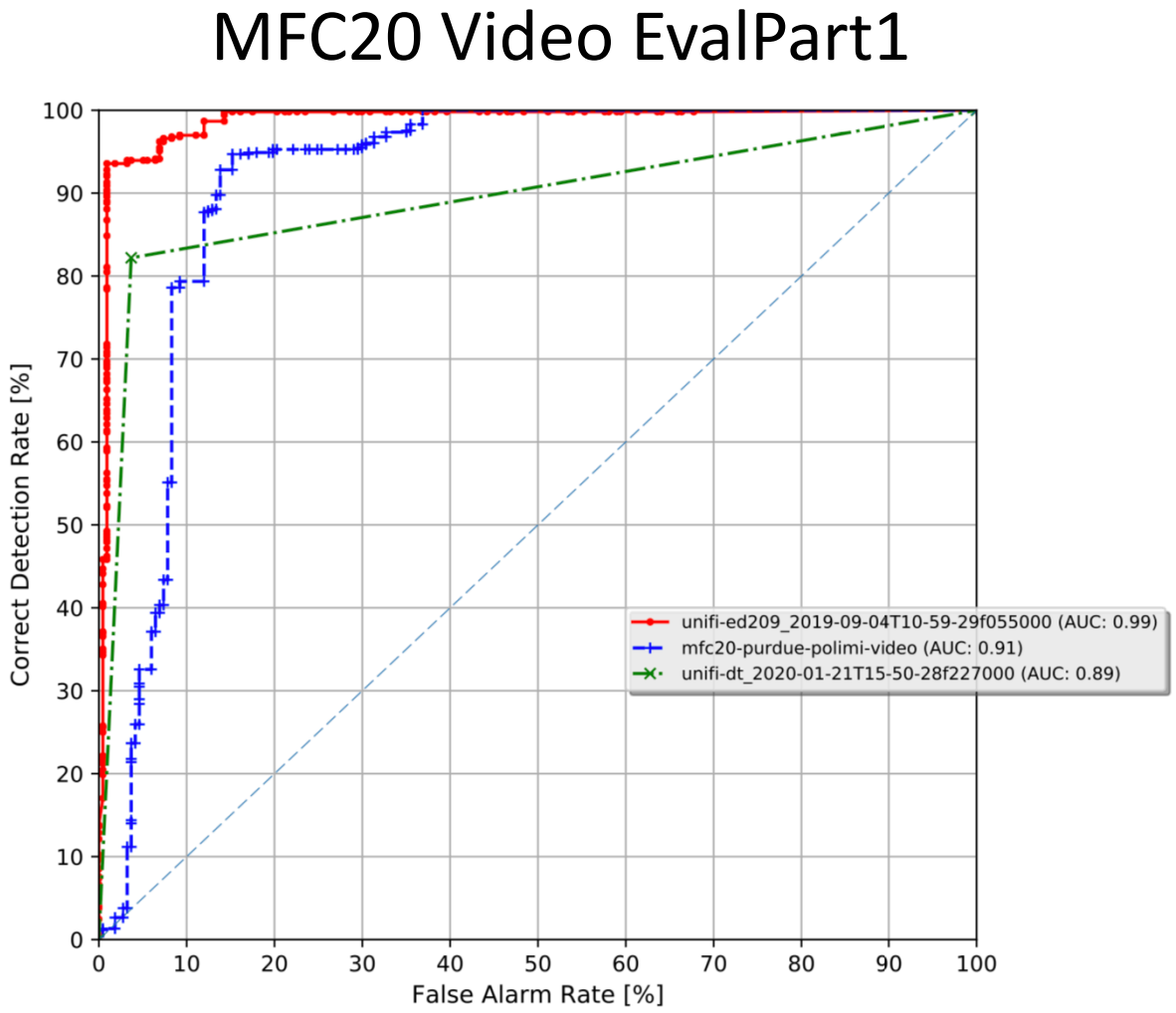
- 19 video detection systems
- Highest AUC & CD@0.05FA
  - AUC = 0.92; (CD@0.05FA = 0.269)
  - TRR = 0.68
  - Team ID: Purdue_Polimi
  - System ID: purdue-polimi-video-codec-base-new_2020-01-30T21-38-26f154000

MFC20 Video EvalPart1



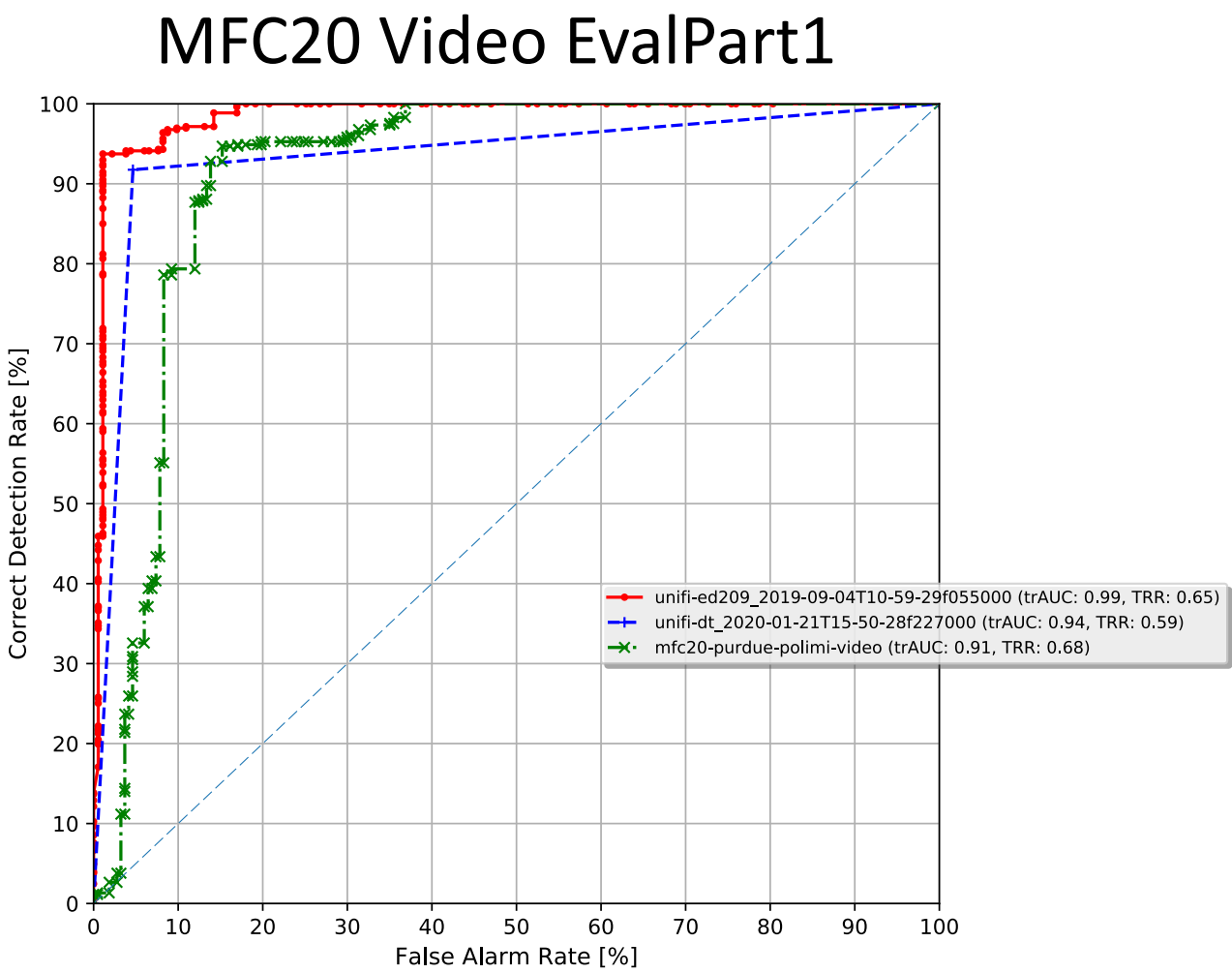
Video Detection Performance: Video + Metadata, Full Data

- 2 teams:
  - Purdue_Polimi
  - UNIFI
- Highest AUC & CD@0.05FA:
  - AUC = 0.987; (CD@0.05FA = 0.939)
  - Team ID: UNIFI
  - System ID: unifi-ed209_2019-09-04T10-59-29f055000



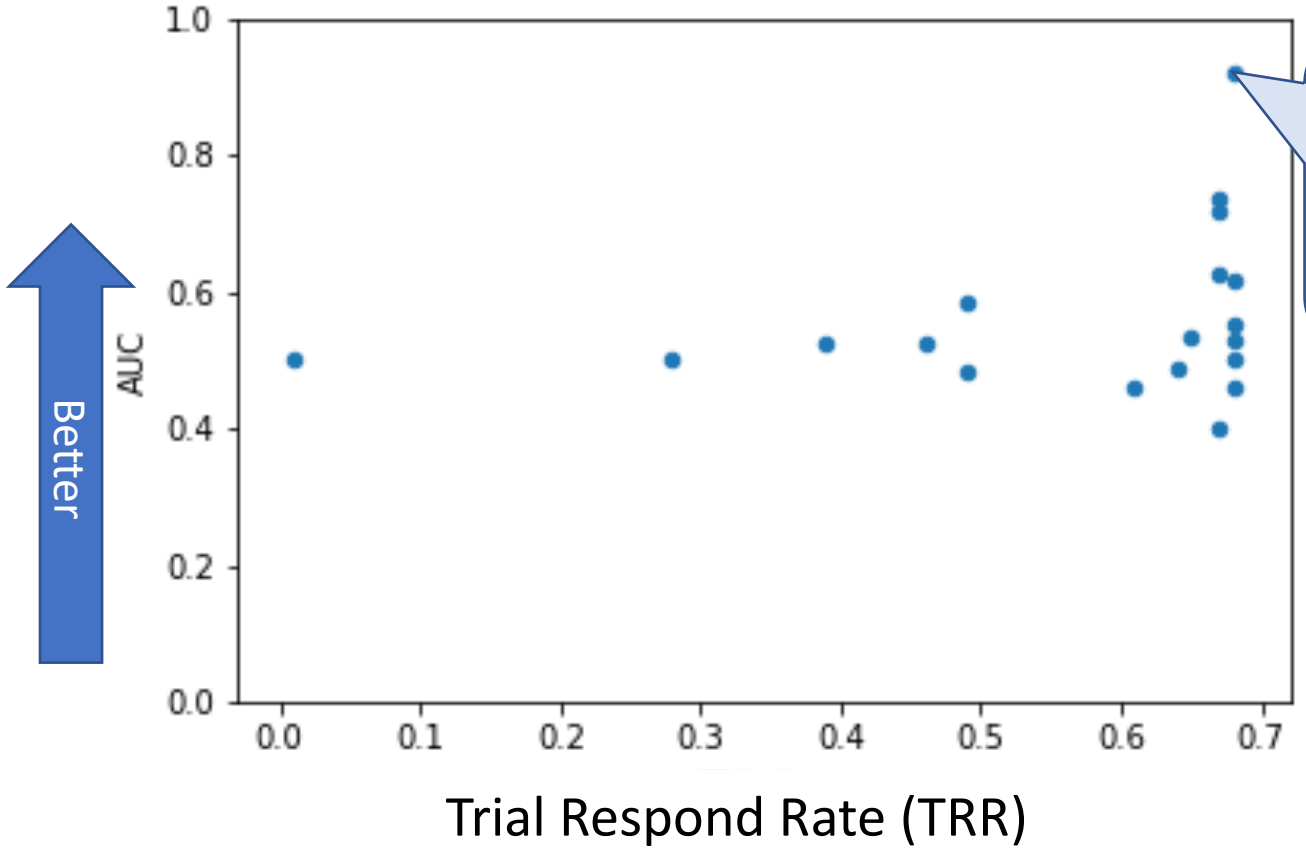
Video Detection Performance: Video + Metadata, Opt-In

- 2 teams:
  - Purdue_Polimi
  - UNIFI
- Highest AUC & CD@0.05FA:
  - AUC = 0.991; (CD@0.05FA = 0.919)
  - Team ID: UNIFI
  - System ID: unifi-ed209_2019-09-04T10-59-29f055000



Video Manipulation Detection Container Results

- Opt In on MFC20 EP1, Video Only condition

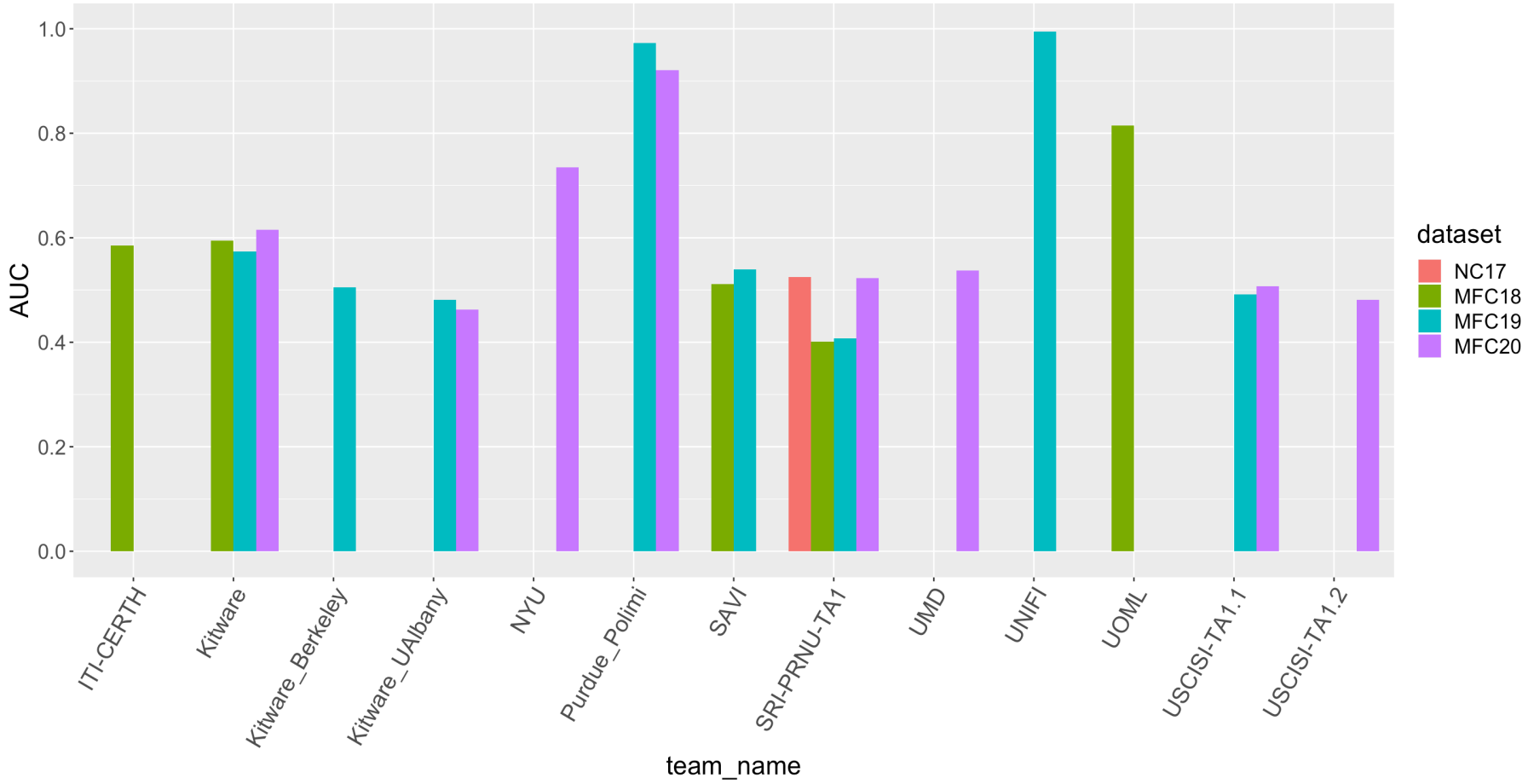


Highest AUC = 0.921:
Team ID: Purdue_Polimi
System ID: purdue-polimi-video-codec-base-new_2020-01-30T21-38-26f154000
CD@0.05FA: 0.269
TRR: 0.68

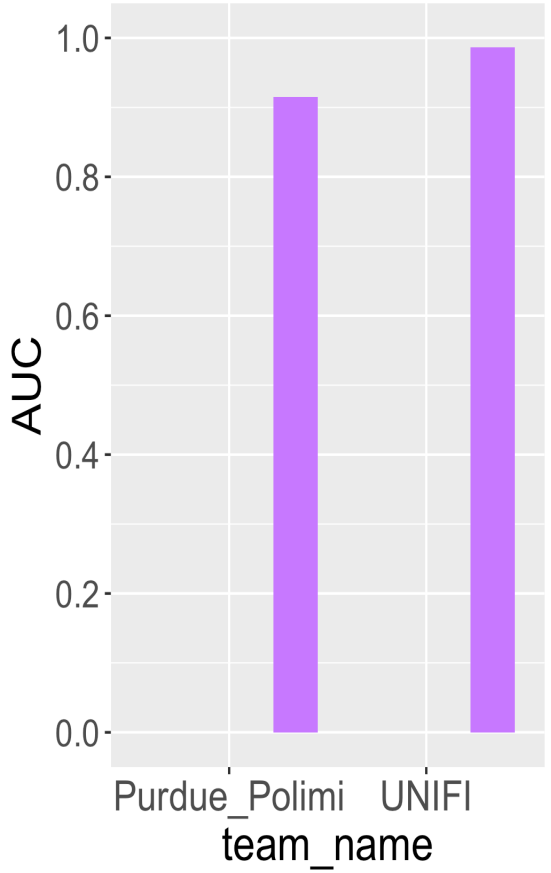
Figure: Video detection, Video Only condition, Opt In system Area Under the Curve (AUC) vs. Trial Response Rate (TRR) on MFC20 EP1 Video dataset (each point is an analytic system)

Historical Video Detection Performance (Full Data)

Video Only

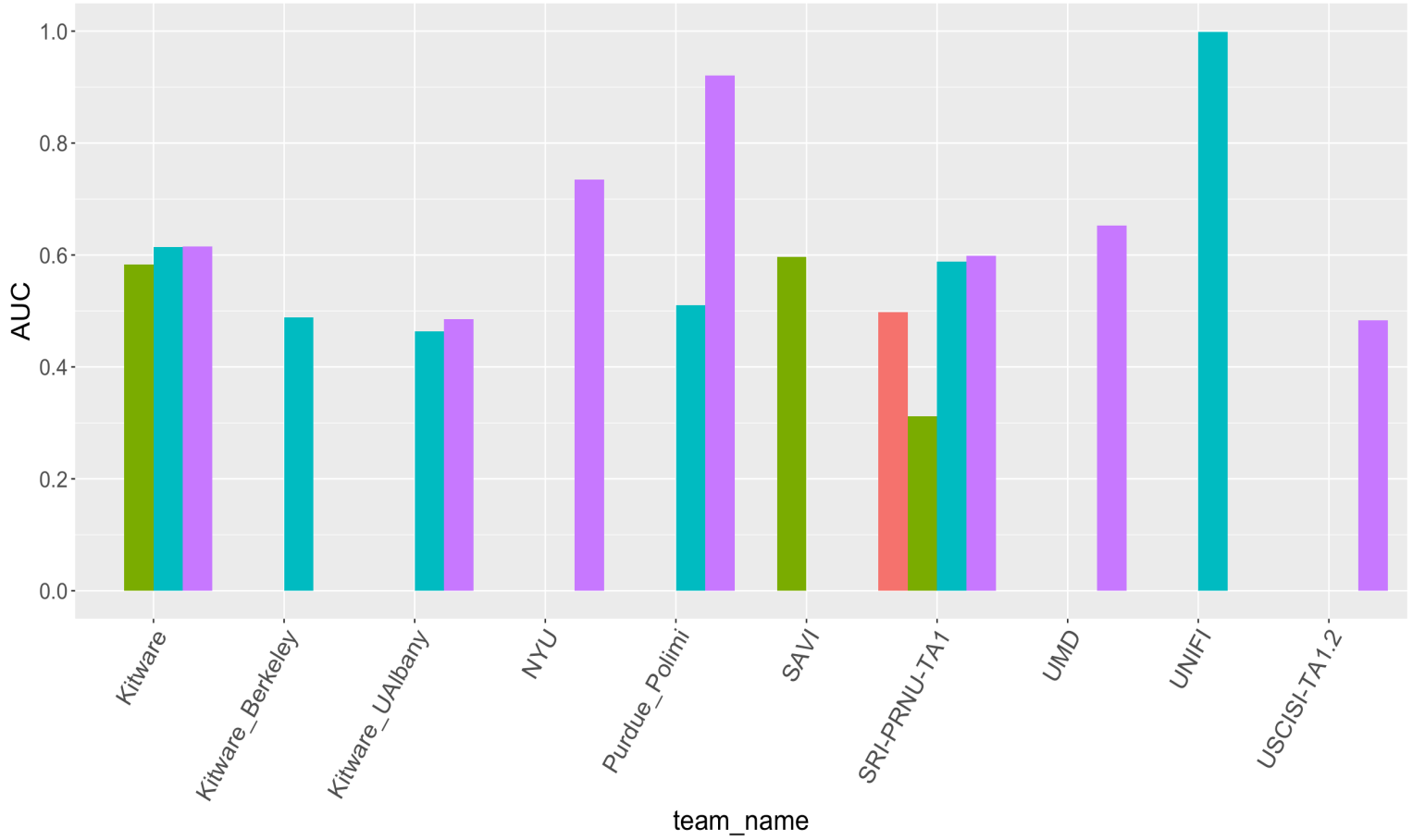


Video + Metadata

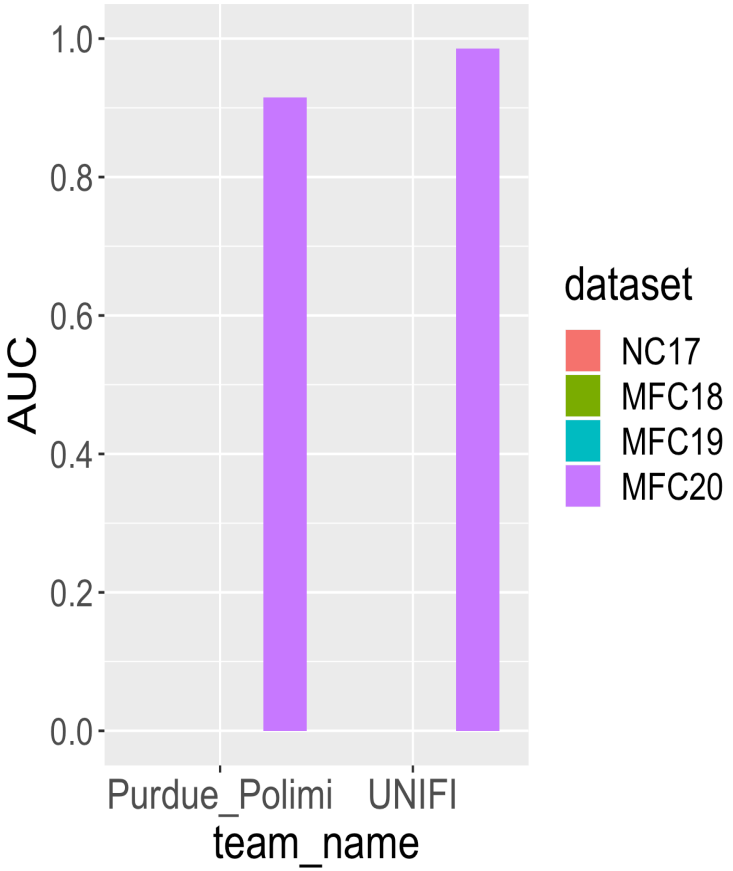


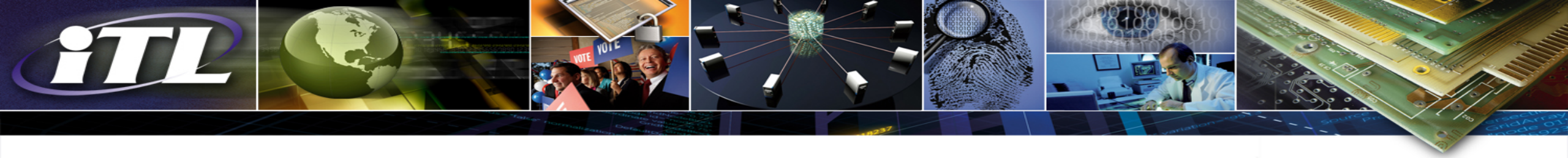
Historical Video Detection Performance (OptIn)

Video Only



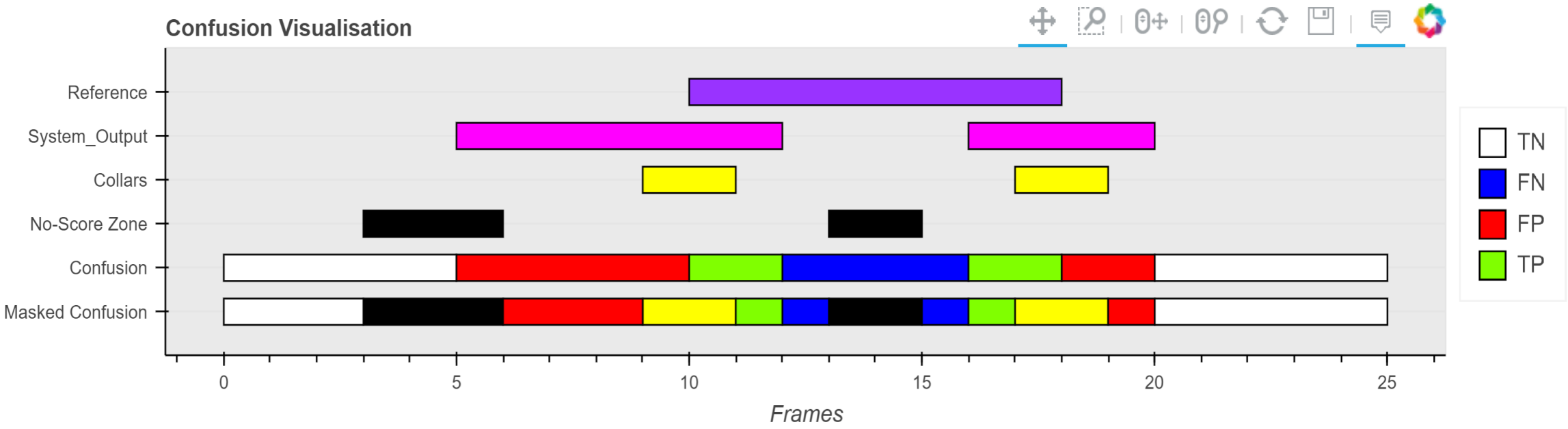
Video + Metadata





Video Temporal and Spatial Localization

Temporal Localization Scoring Visualization



Score: $MCC = 1 \rightarrow$ Perfect system
$MCC = 0 \rightarrow$ Random system
$MCC = -1 \rightarrow$ Inverted system

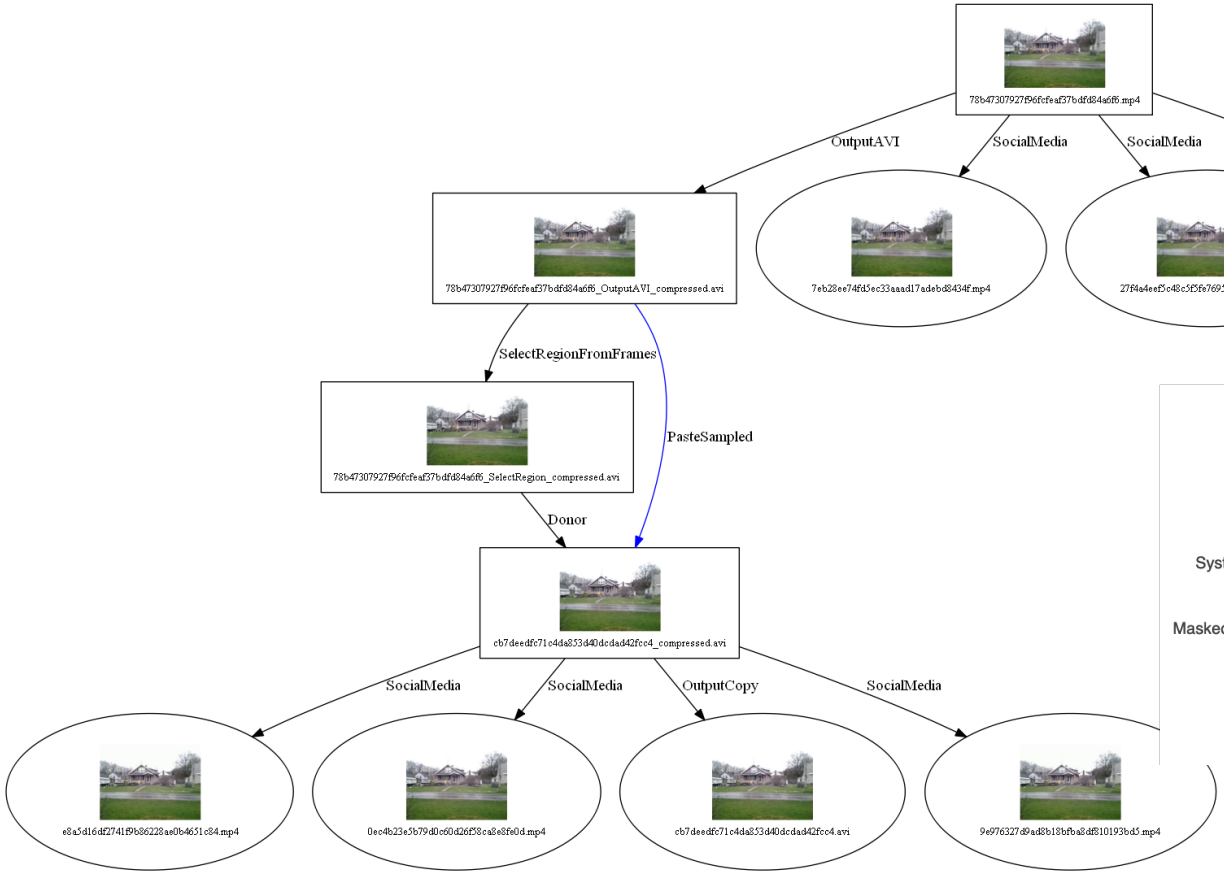
Example: $TP = 2, TN = 8, FP = 4, FN = 2$ $\quad MCC = \frac{8}{\sqrt{2880}} \approx 0.15$

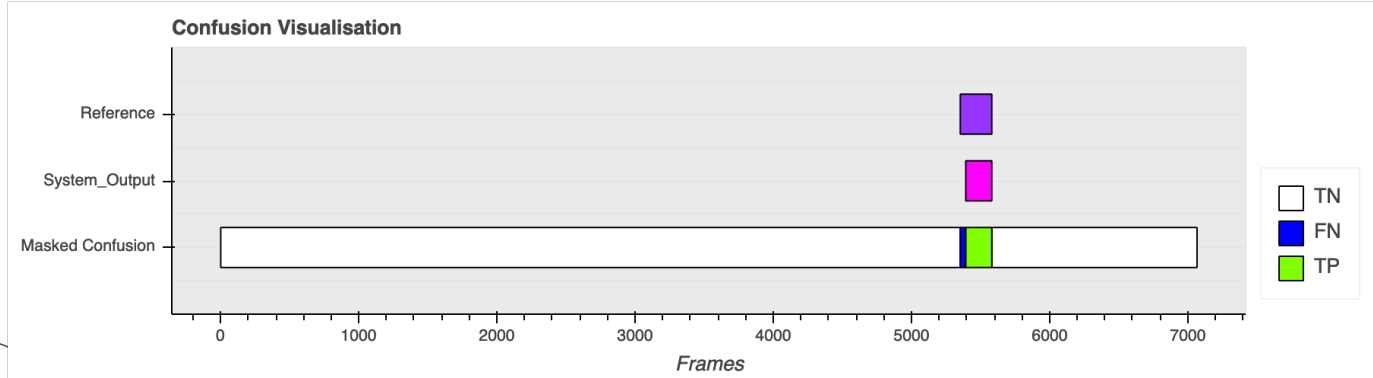
Video Temporal Localization Results

- Dataset
  - MFC20-EvalPart1-Video-LocSubset-Ver1
  - video subset with 117 probs
- 7 teams
  - Kitware
  - Purdue_Polimi
  - SRI-PRNU-TA1
  - UMD
  - UNIFI
  - USCISI-TA1.1
  - USCISI-TA1.2
- Condition: Video Only

- Full Data Highest MCC
  - MCC = 0.003
  - Team ID: Kitware
  - System ID: kitware-videoframeduplication_2020-03-23T21-23-13f884000
- OptIn Data Highest MCC
  - MCC = 0.004
  - Team ID: SRI-PRNU-TA1
  - sriprita1-vid-mdl-prnu-based_2020-02-07T16-49-46f182000
  - TRR = 0.44

A Video Example (House) in MFC20 Evaluation





Temporal Localization MCC = 0.90788

Video Spatial-Temporal Localization (VSTL)

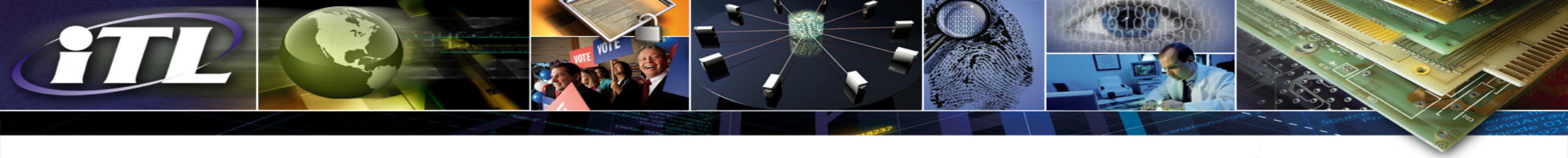
- Task Definition

  Video spatial localization task is to determine the spatial edits of a particular video if it is determined (correctly or incorrectly) to be manipulated.

- Metrics

  Optimum Matthew Correlation Coefficient (MCC)

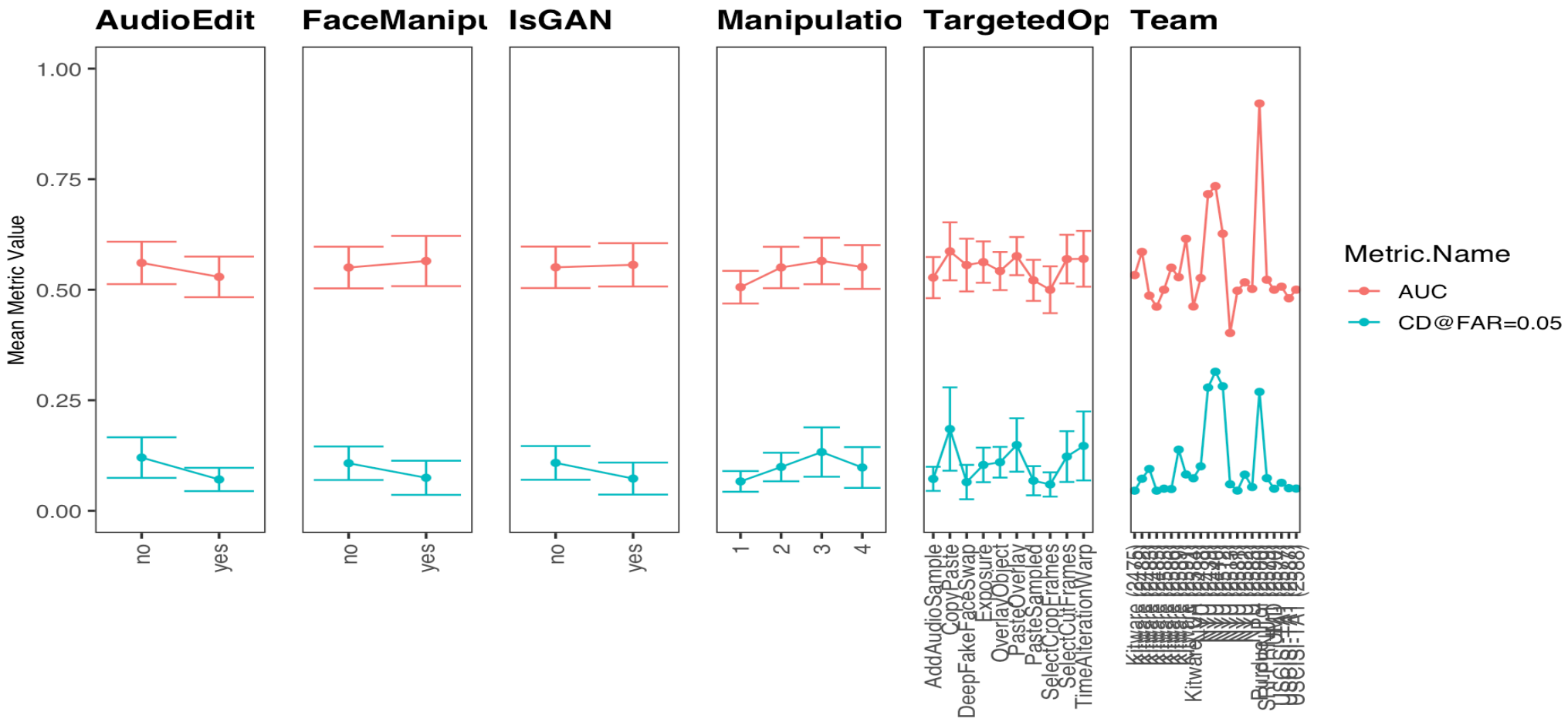
- Status
  - One container submitted that is capable of VSTL
  - Scoring under way



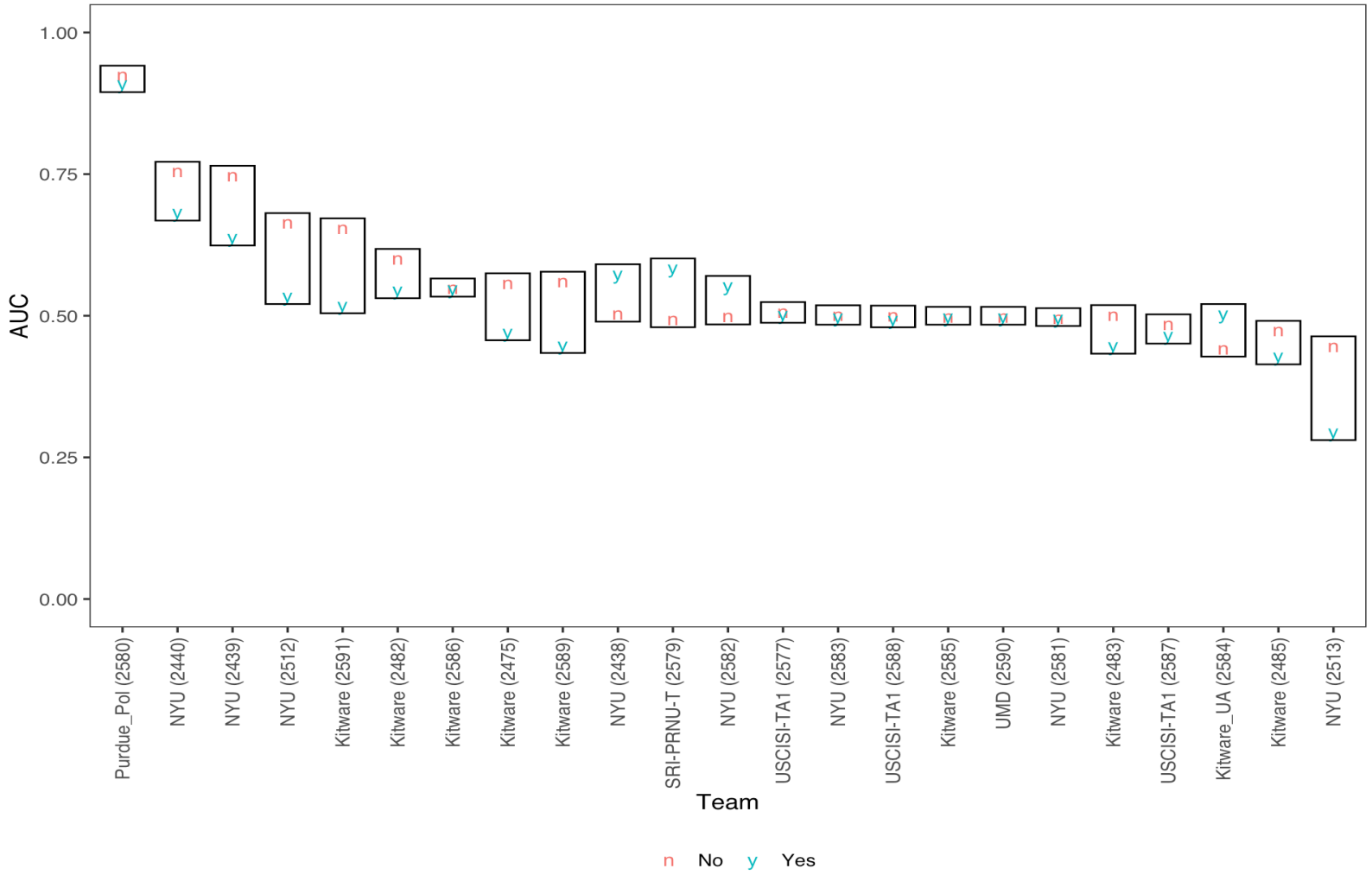
Video Manipulation Detection Analysis

Video Detection

At least 20 probes



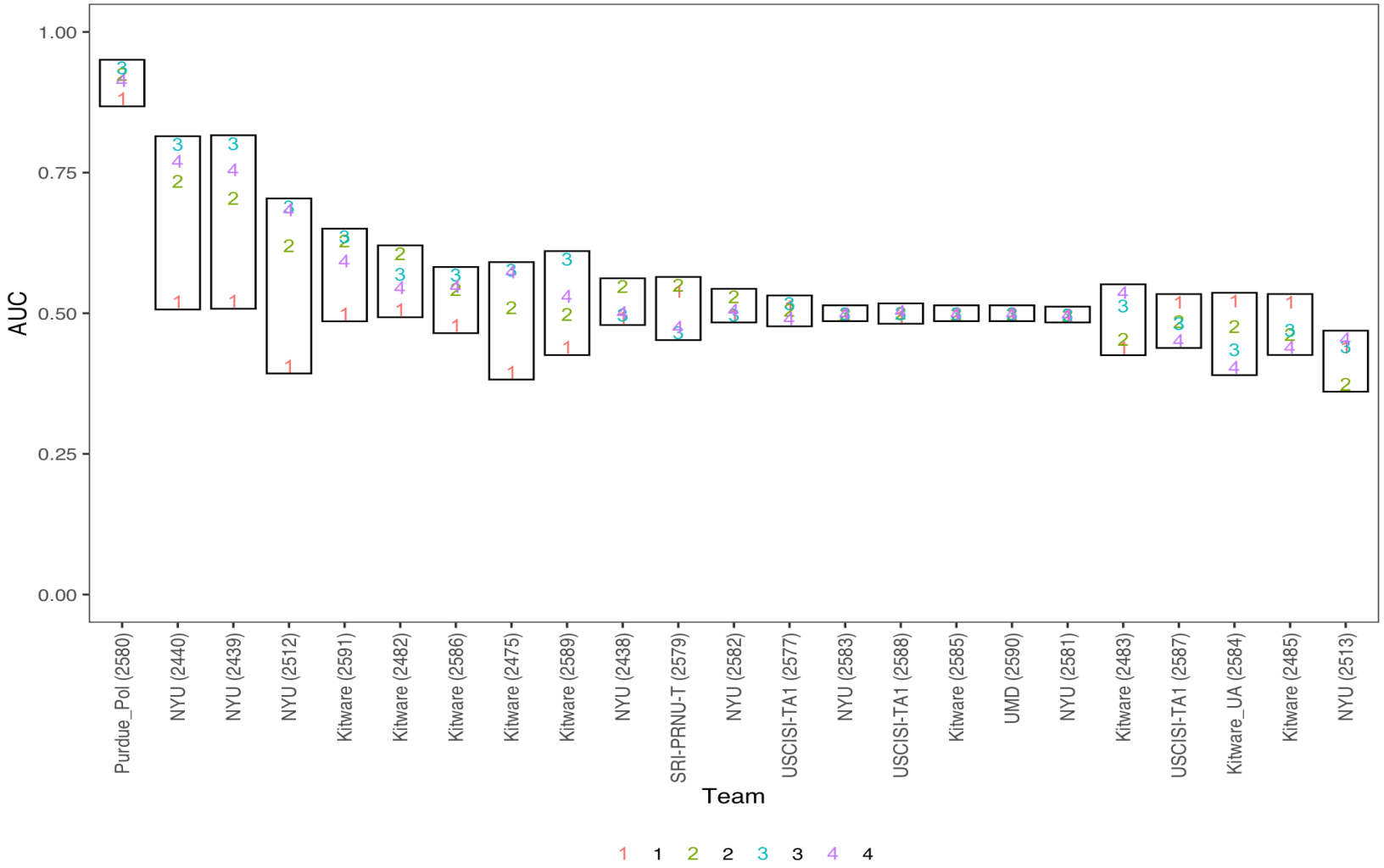
Effect of Audio Edit on Detection



Lower = Better

Audio Edit	Mean Rank
no	1.24
yes	1.76

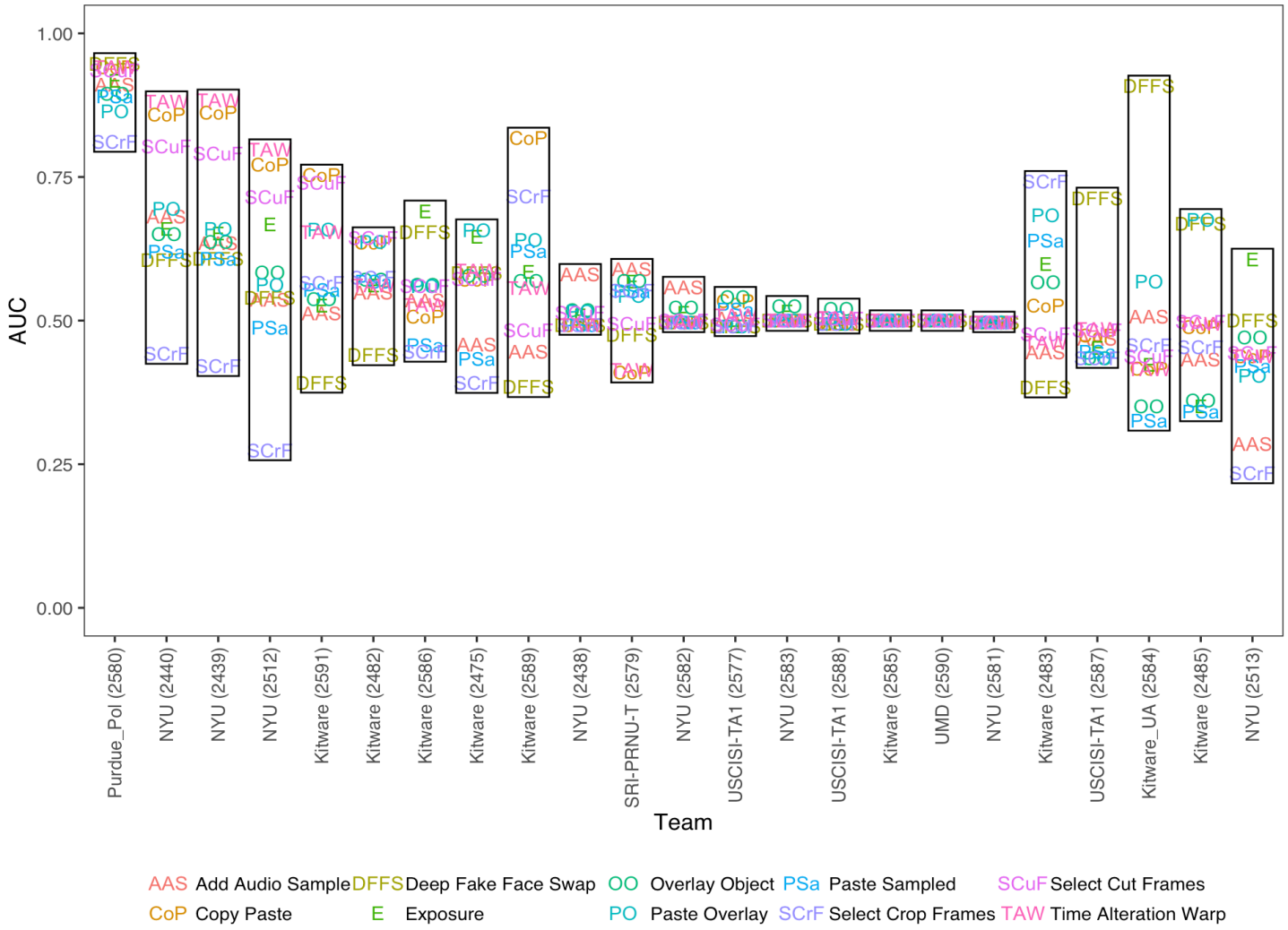
Effect of Manipulation Count on Detection



Lower = Better

Manipulation Count	Mean Rank
3	2.02
2	2.39
4	2.48
1	3.11

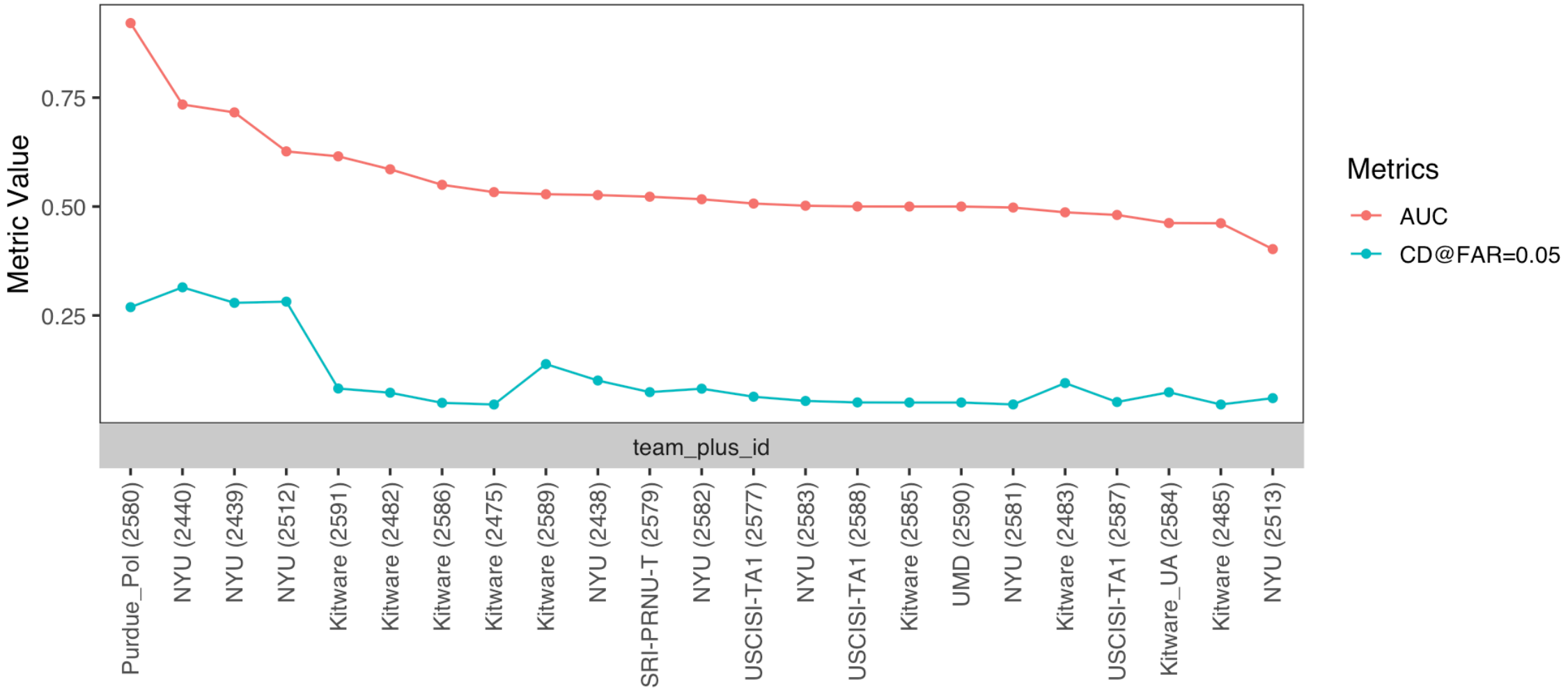
Effect of Operation on Detection



Lower = Better

Operation	Mean Rank
Sel Cut Frames	4.52
Overlay Object	4.80
Exposure	4.98
Paste Overlay	4.98
Copy Paste	5.00
Time Alt Warp	5.02
Add Audio Smpl	6.02
DF Face Swap	6.07
Paste Sampled	6.67
Sel Crop Frames	6.93

Detection Teams



Video Special Studies

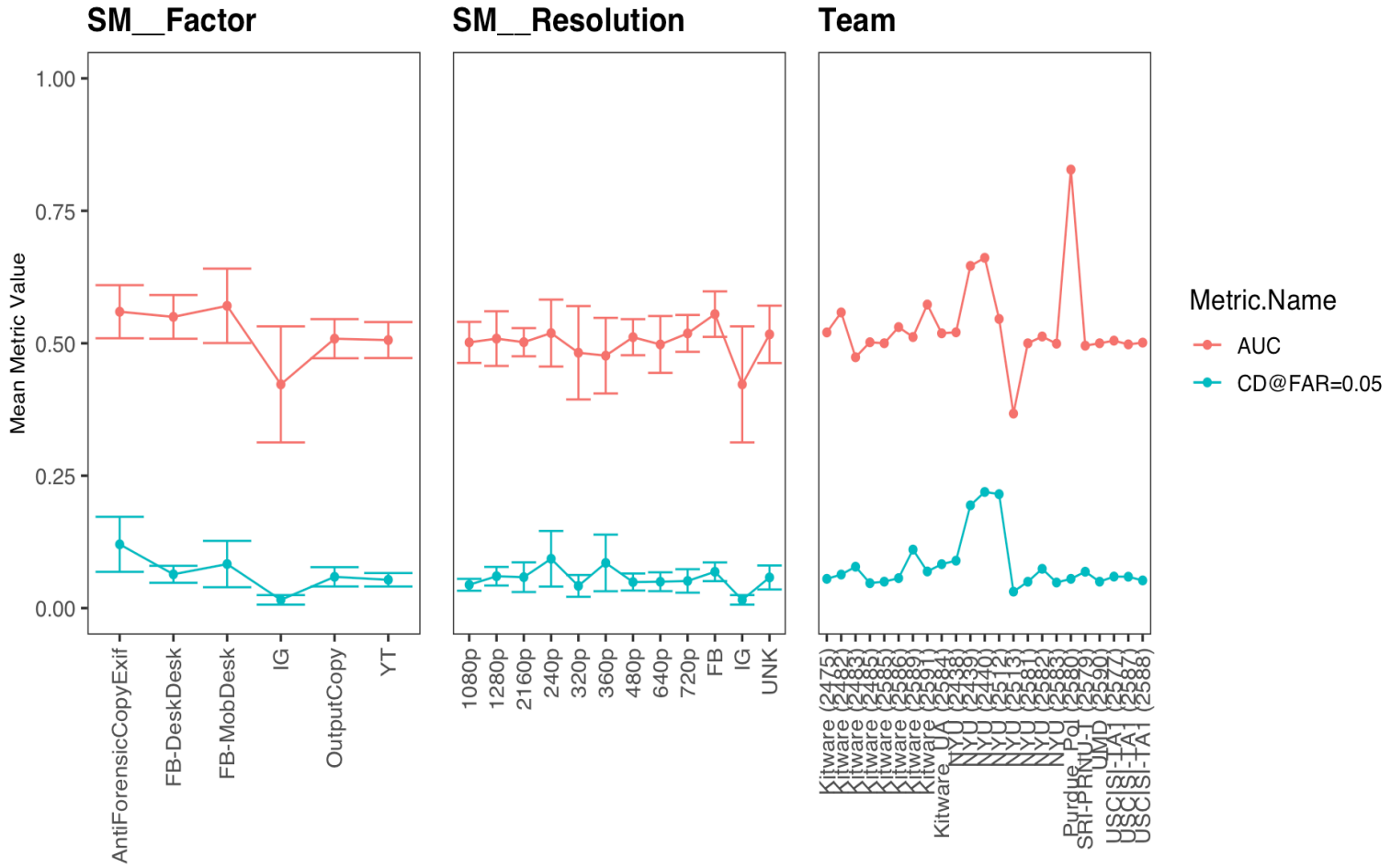
- Video Studies
  - Frame Drop/Duplication
  - Social Media Laundering - Video

- Study Condition Definition
  - Manipulation Detection (MD)
    - Target = Any manipulated video
    - Non-Target = HP media
  - Operation-Only Detection (OOD)
    - Target = Only video with operation of interest; no other operations are present
    - Non-Target = HP media

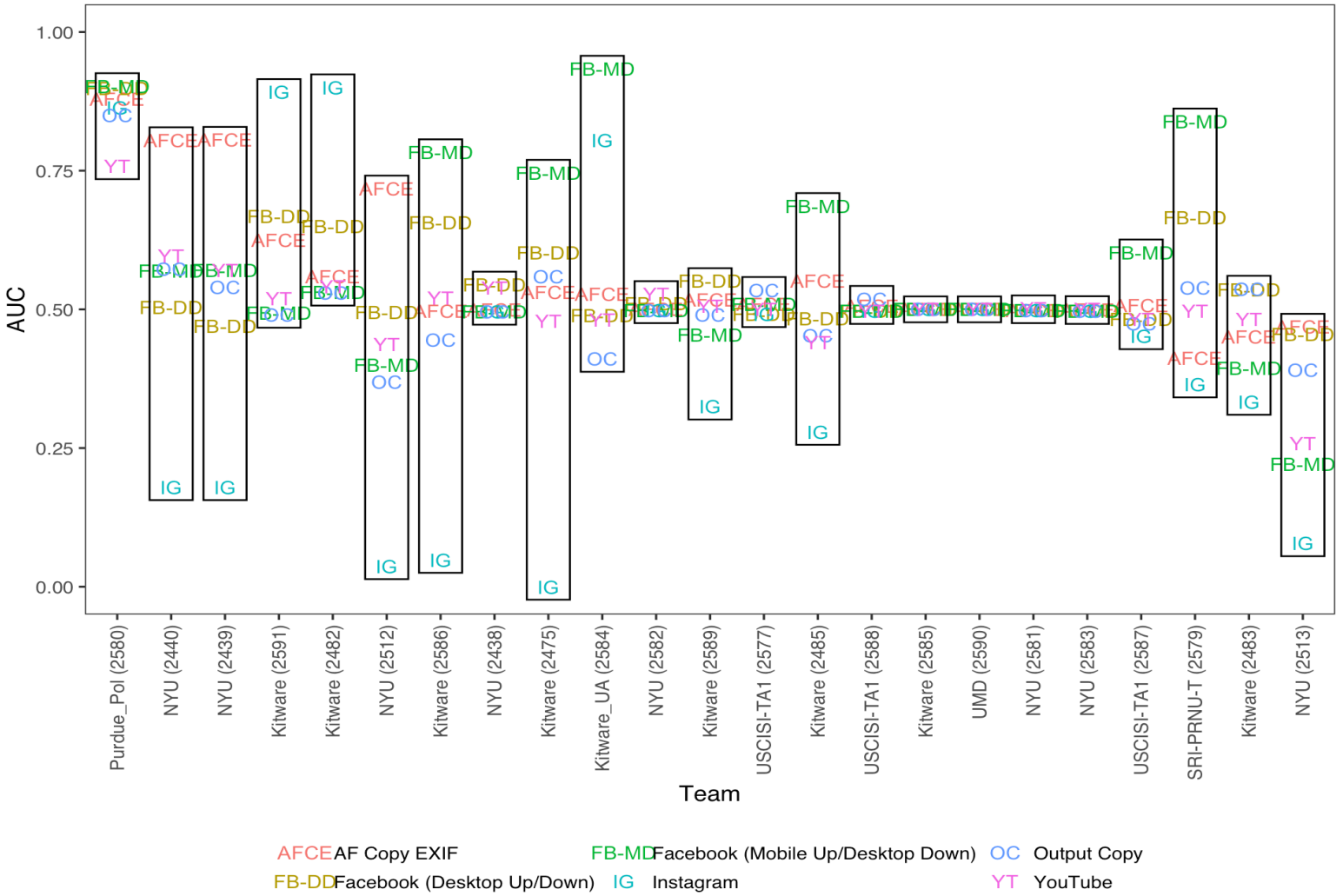
Video Manipulation Detection Special Study – Social Media Laundering (Video) Results (MD)

- 6 Scenarios
  - Facebook (DeskDesk) (Synthetic)
    - desktop upload, desktop download
  - Facebook (MobDesk) (Synthetic)
    - mobile upload, desktop download
  - Instagram (Synthetic)
    - desktop upload, desktop download
  - YouTube (Actual/Manual)
    - desktop upload, desktop download
    - Multiple resolutions
  - Anti-Forensic Copy EXIF
    - Original compression
    - Copy original EXIF
  - Output Copy
    - No compression
    - No copying EXIF

All study probes



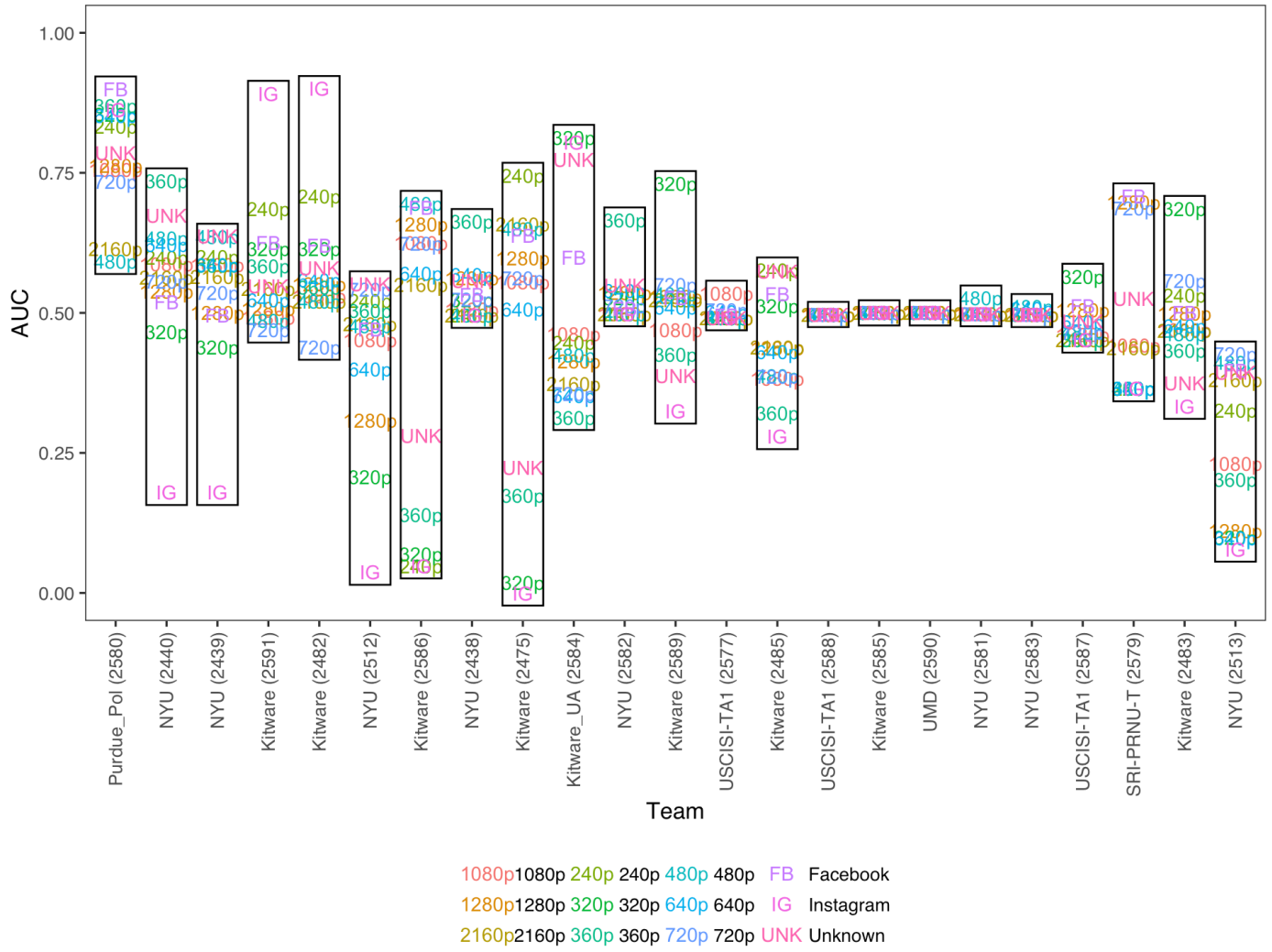
Social Media Factor (MD)



Lower = Better

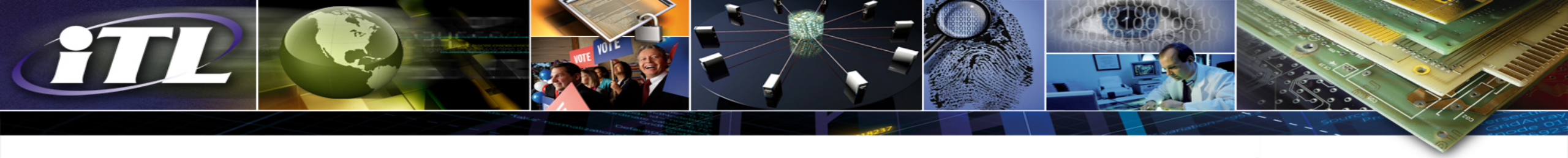
SM Factor	Mean Rank
AF Copy EXIF	2.74
FB-DeskDesk	2.98
FB-MobDesk	3.22
YT	3.37
Output Copy	3.93
IG	4.76

Resolution (MD)



Lower = Better

Resolution	Mean Rank
FB	4.87
Unknown	5.50
240p	6.02
480p	6.22
640p	6.26
720p	6.30
1280p	6.37
320p	6.43
360p	6.98
1080p	7.04
2160p	7.41
IG	8.59



MediFor Challenge Evaluation 2020 (Provenance Tasks)

Jonathan Fiscus (Co-PI), Dr. Haiying Guan (Co-PI), **Dr. Yooyoung Lee**, Dr. Amy Yates[+], Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah, Dr. Xiongnan Jin

Multimodal Information Group, [+] Image Group
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology (NIST)

April 21-25, 2020

Provenance Outline

- Task definition and performance measure
- Provenance evaluation datasets
- Results and analyses
- Summary

Provenance Filtering (PF)

- Searching for a potential pool of related images from a large collection of datasets (called the world dataset)
- Given an image (probe), the goal is to return up to N images of the predicted relevant images from the world dataset
- System output
  - JSON file that contains N filtered images including a confidence score that indicates how likely the filtered image is related with respect to the probe image.

PF Performance Measure

- The recall of first n images from the world dataset sorted by 'confidence score'

$$recall = \frac{|\{relevant\} \cap \{retrieved\}|}{|\{relevant\}|}$$

MFC20 primary metric: Recall@300

Provenance Graph Building (PGB)

- Constructing the relationships among the retrieved images along with finding the ancestor and descendent sequences
- Given a probe image (e.g., base, donor, intermediate, or final modified images), the goal is to build a provenance phylogeny graph that describe the relationships among the images with the manipulation sequences.
- System Output
  - JSON file that contains both nodes and links with the two types of confidence scores
    - Node: how likely the retrieved image (node) is presented in the provenance graph with respect to the probe image
    - Link: how likely the two nodes (between a source node and a target node) have the relationship (link) in the provenance graph

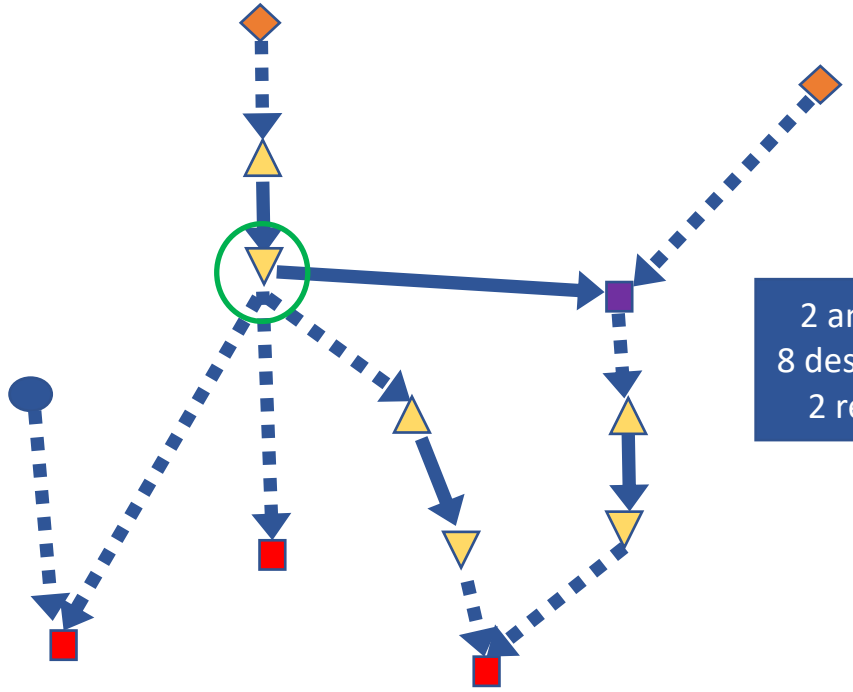
Provenance Graph Building (Eval Condition)

- Full graph
  - All images related to the probe image are evaluated with the ancestors and descendants' sequences
- Subset (direct path) graph
  - The subset of the related images (node set) is restricted to ancestors and descendants of the probe image and ***only directed paths related to the probe image*** are evaluated

Example of Full vs Subset Reference Graph

Probe: node marked in green circle

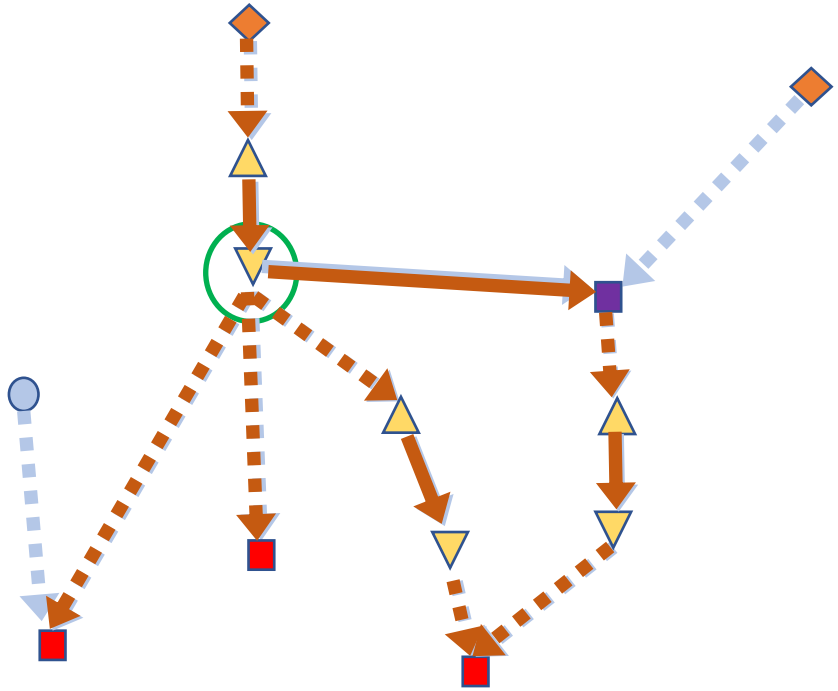
Full Reference Graph



2 ancestors
8 descendants
2 relatives

Subset (Direct Path) Reference Graph

2 ancestors
8 descendants



PGB Performance Measures

- Evaluation metrics: Graph Similarity and Generalized F-measure
  - Given the system output provenance graph, G_s, the set of nodes (or vertices) of the system output provenance graph is V_s while the set of links (or edges) is E_s

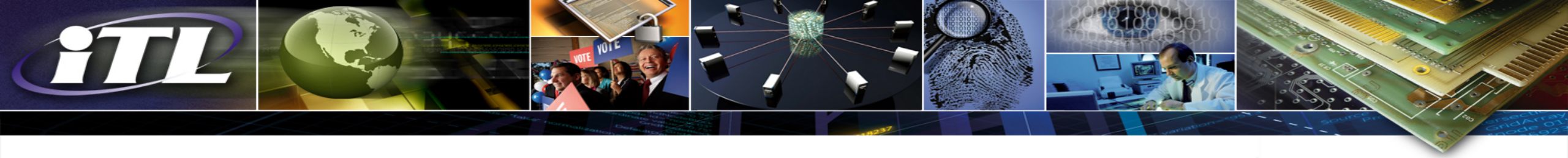
  - Nodes overlap: $\boxed{\mathrm{sim}_{\mathrm{NO}}(G_r, G_s) = 2\frac{|V_r \cap V_s|}{|V_r|+|V_s|}}$

  - Links overlap: $\boxed{\mathrm{sim}_{\mathrm{LO}}(G_r, G_s) = 2\frac{|E_r \cap E_s|}{|E_r|+|E_s|}}$

  - **Nodes and links overlap**: $\boxed{\mathrm{sim}_{\mathrm{NLO}}(G_r, G_s) = 2\frac{|V_r \cap V_s|+|E_r \cap E_s|}{|V_r|+|V_s|+|E_r|+|E_s|}}$

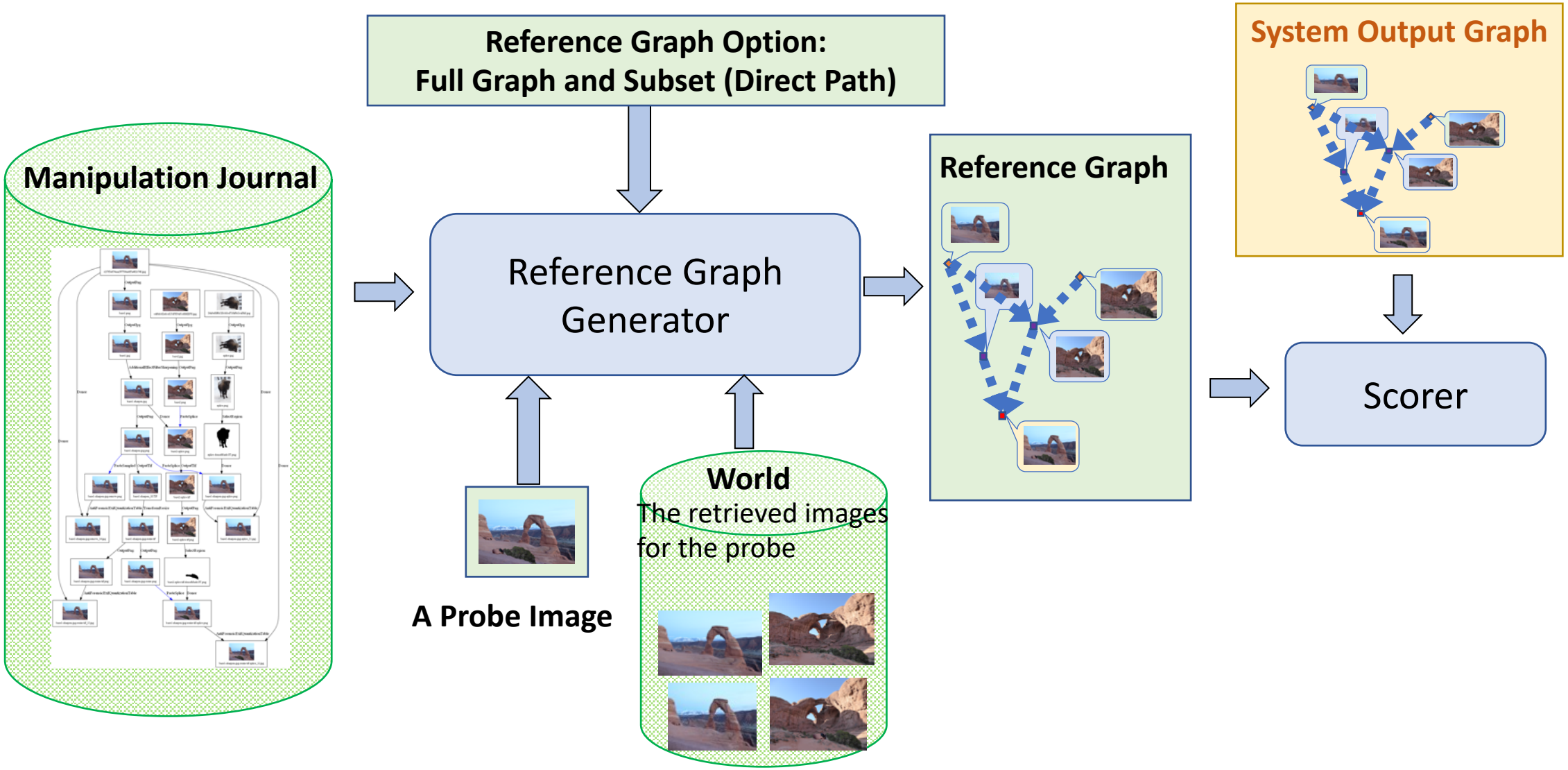
- Graph color code in HTML measure output

	Green (correct detections)	Red (false alarms)	Gray (missed detections)
Node	Correctly located images	Falsely located images	Omitted images
Link	Correctly linked images	Falsely linked images	Omitted links



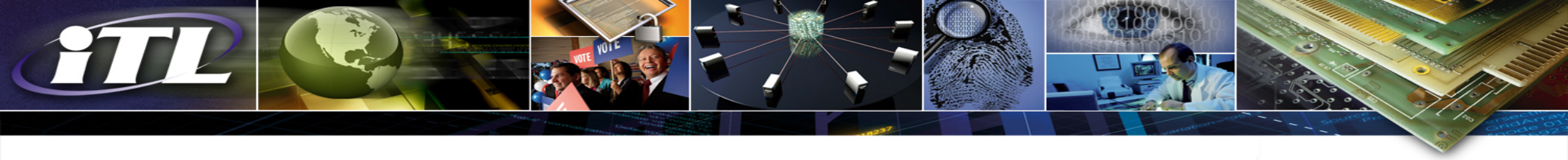
Provenance Eval Datasets

Provenance Graph Test Data and Reference Generation



Provenance Dataset

NIST Data Sets	Image Probe	Image Journal	World	Date
NC17 EvalPart1	1000	406	1M	06/2017
MFC18 EvalPart1	10,000	641	1M	03/2018
MFC19 EvalPart1	9420	1025	2M	03/2019
MFC20 EvalPart1	**5926**	**1571**	**2M**	**02/2020**



Provenance Results and Analyses

Provenance Filtering Overview

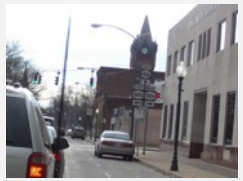
System Input

Probe Image

World Image Set





⋮

System Output

Retrieving a set of N images with confidence score

Algorithm


0.9


0.8


0.8


0.7

⋮

Performance Measure

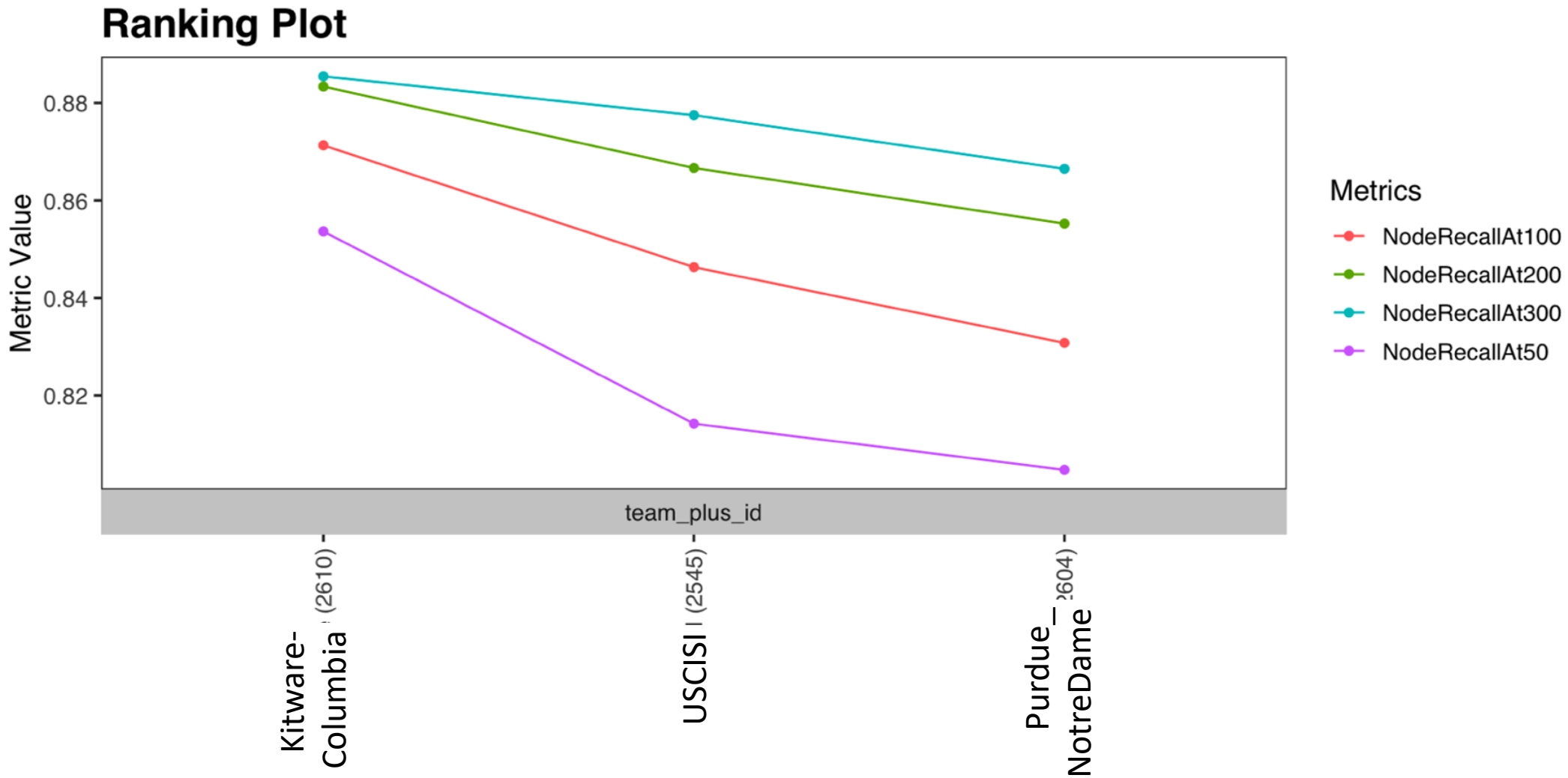
The recall of first n images from the world dataset sorted by 'confidence score'

$$recall = \frac{|\{relevant\} \cap \{retrieved\}|}{|\{relevant\}|}$$

Recall@50, Recall@100
Recall@200, **Recall@300**

MFC20 Provenance Filtering Results

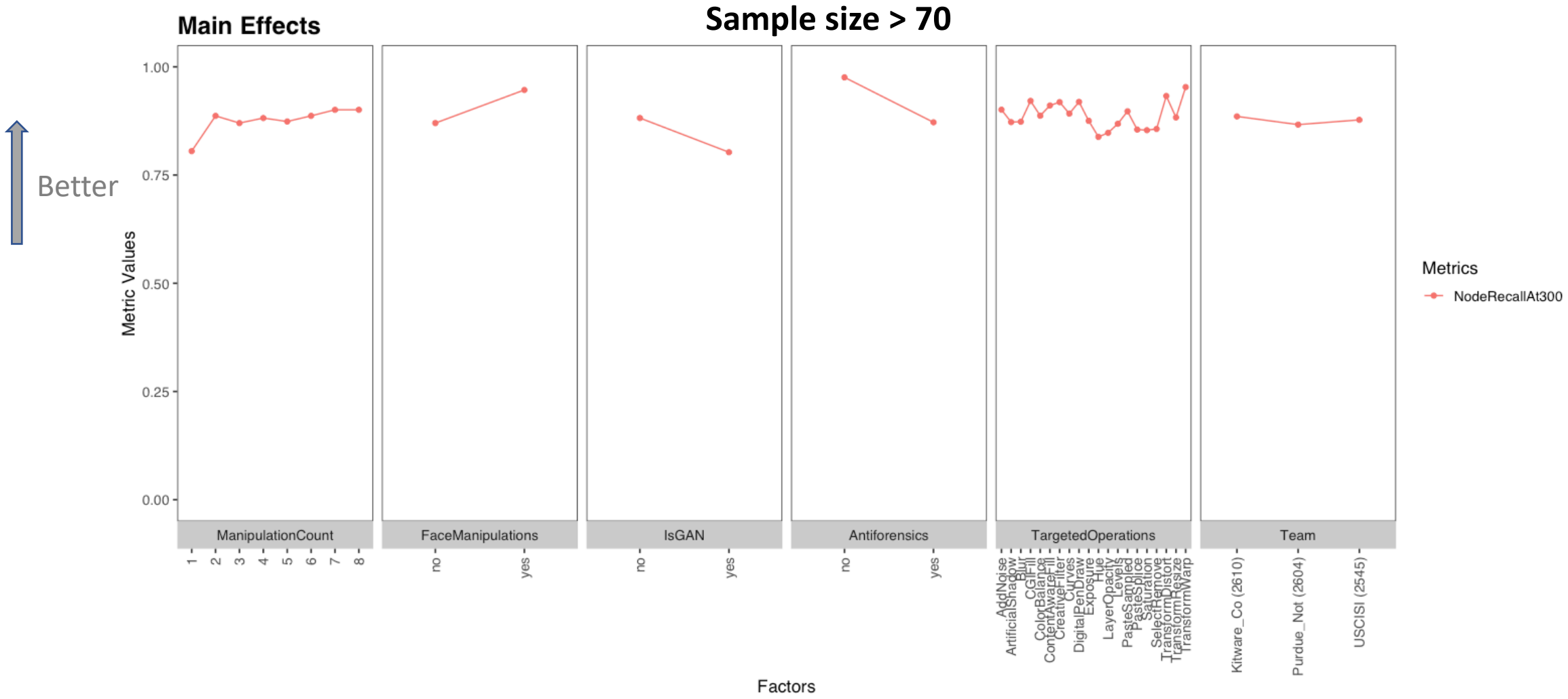
Better

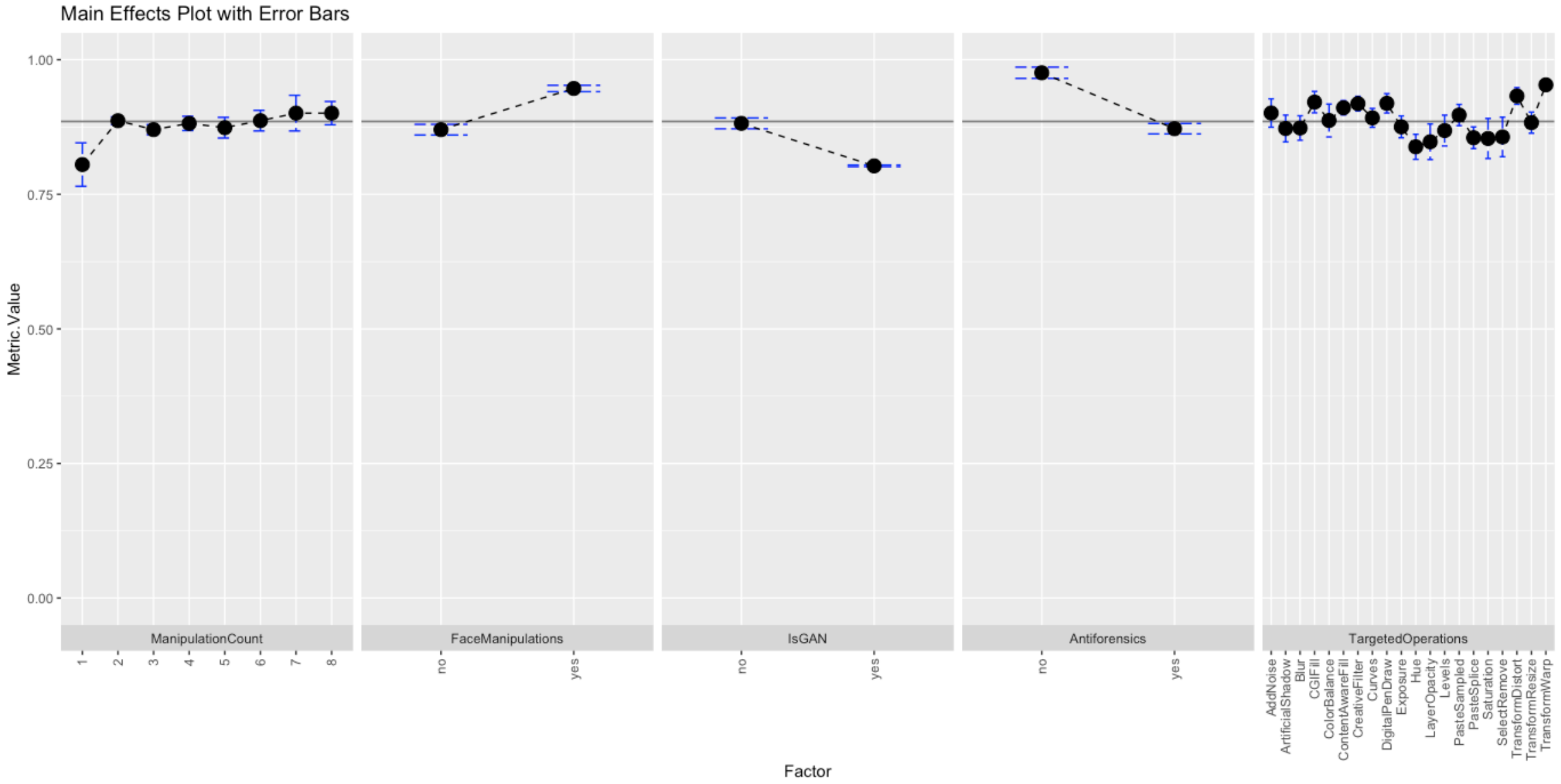


Provenance: Factor of Interest

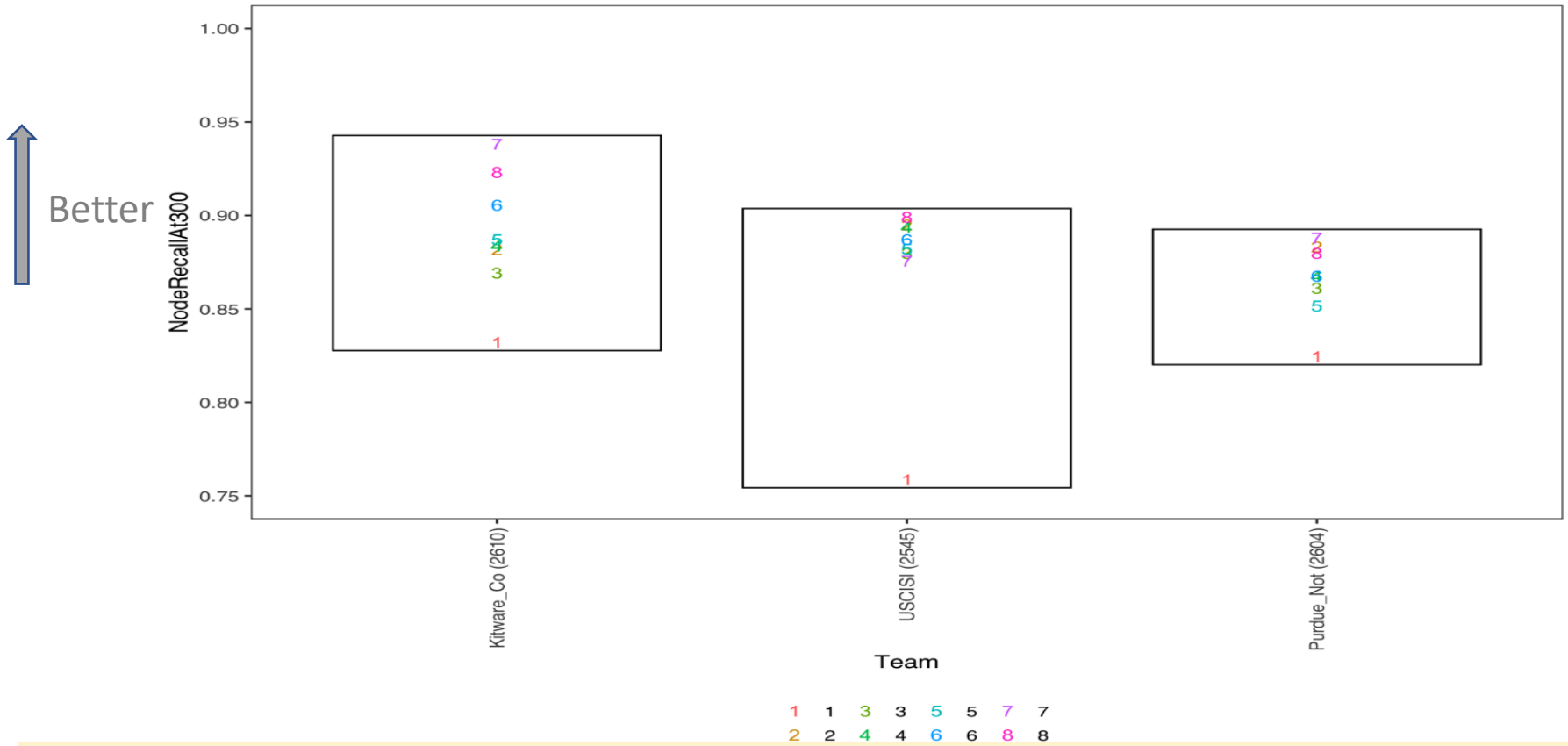
- 6 factors from the reference and journal information
  - Manipulation Count (1 to 8)
  - Face Manipulation (yes/no)
  - GAN (yes/no)
  - Antiforensics (Y: **After** Antiforensics, N: **Before** Antiforensics)
  - TargetedOperations (20 different manipulation operations)
  - Team systems (Kitware_Columbia, Purdue_NotreDame, USCISI)
- Removed factor settings are less than 70 target trials

What are the important factors for PF?




Main Effects Plot with Error Bars
Metric.Value
1.00
0.75
0.50
0.25
0.00
ManipulationCount
1
2
3
4
5
6
7
8
FaceManipulations
no
yes
IsGAN
no
yes
Antiforensics
no
yes
TargetedOperations
AddNoise
ArtificialShadow
Blur
CGIFill
ColorBalance
ContentAwareFill
CreativeFilter
Curves
DigitalPenDraw
Exposure
Hue
LayerOpacity
Levels
PasteSampled
PasteSplice
Saturation
SelectRemove
TransformDistort
TransformResize
TransformWarp
Factor

Manipulation Count



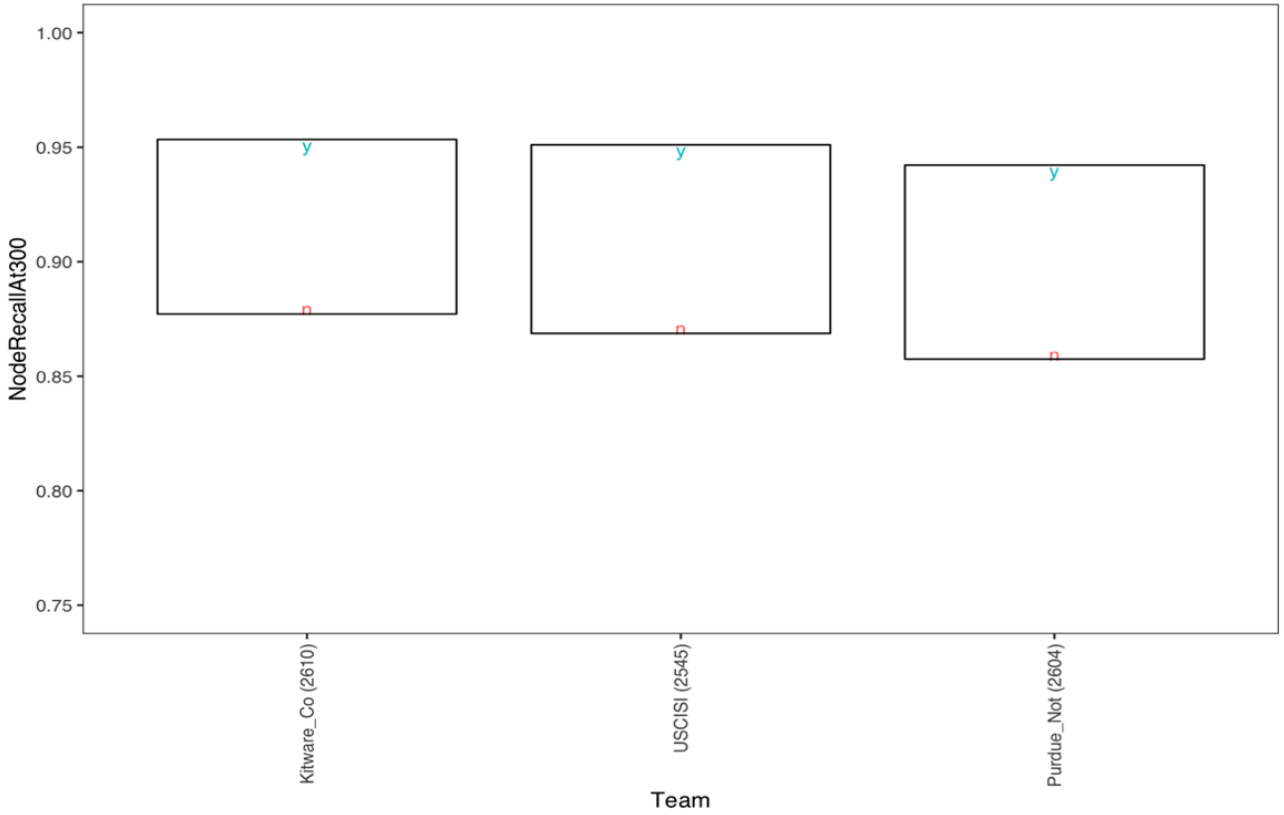
Factor	Avg.Rank
8	2.000000
7	3.000000
2	3.333333
6	3.666667
4	4.333333
5	5.333333
3	6.333333
1	8.000000

Observations

- For the PF task, single manipulation is harder to retrieve across the three systems
- USCISI has larger effect on the manipulation count followed by Kitware-Columbia

Face Manipulation

Better



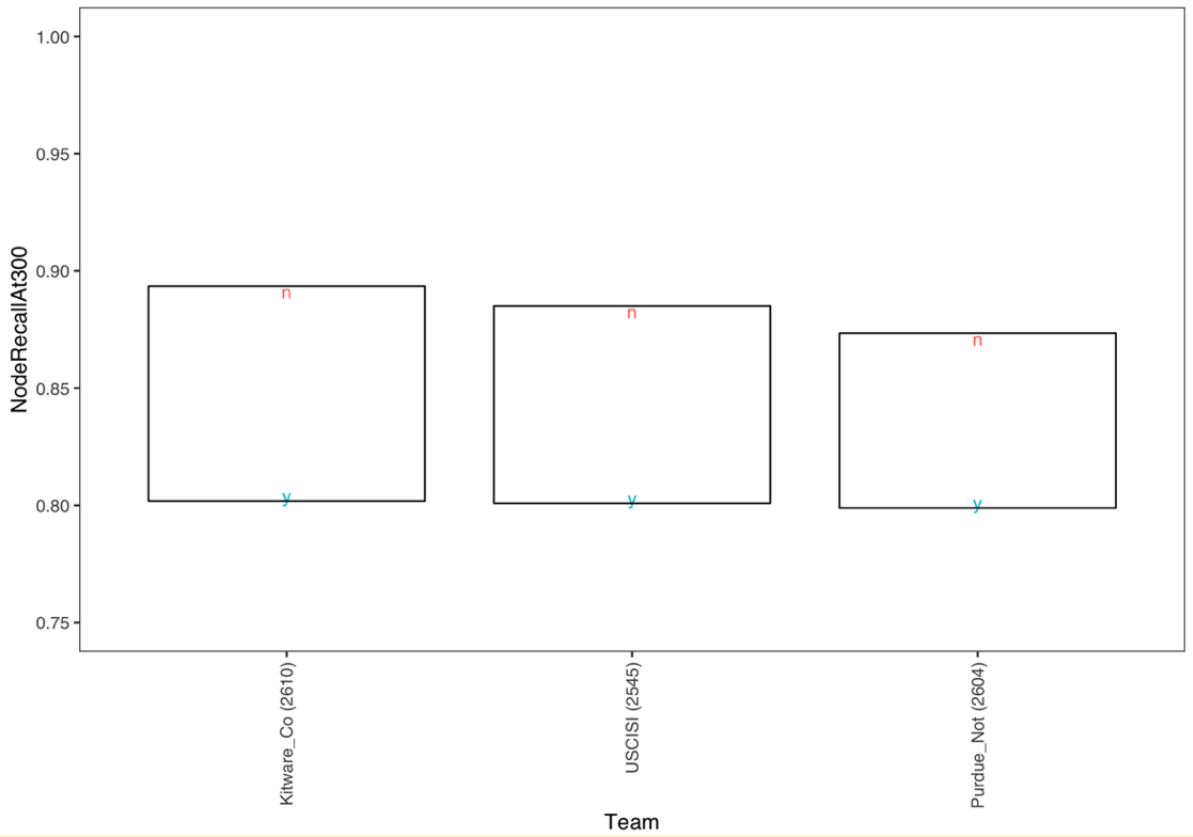
Y: Face manipulation
N: Non-face manipulation

Observations

- The face manipulations are easier to retrieve compared to the non-face manipulations across the three systems
- Face Manipulation has larger effect on all three systems

GAN

Better



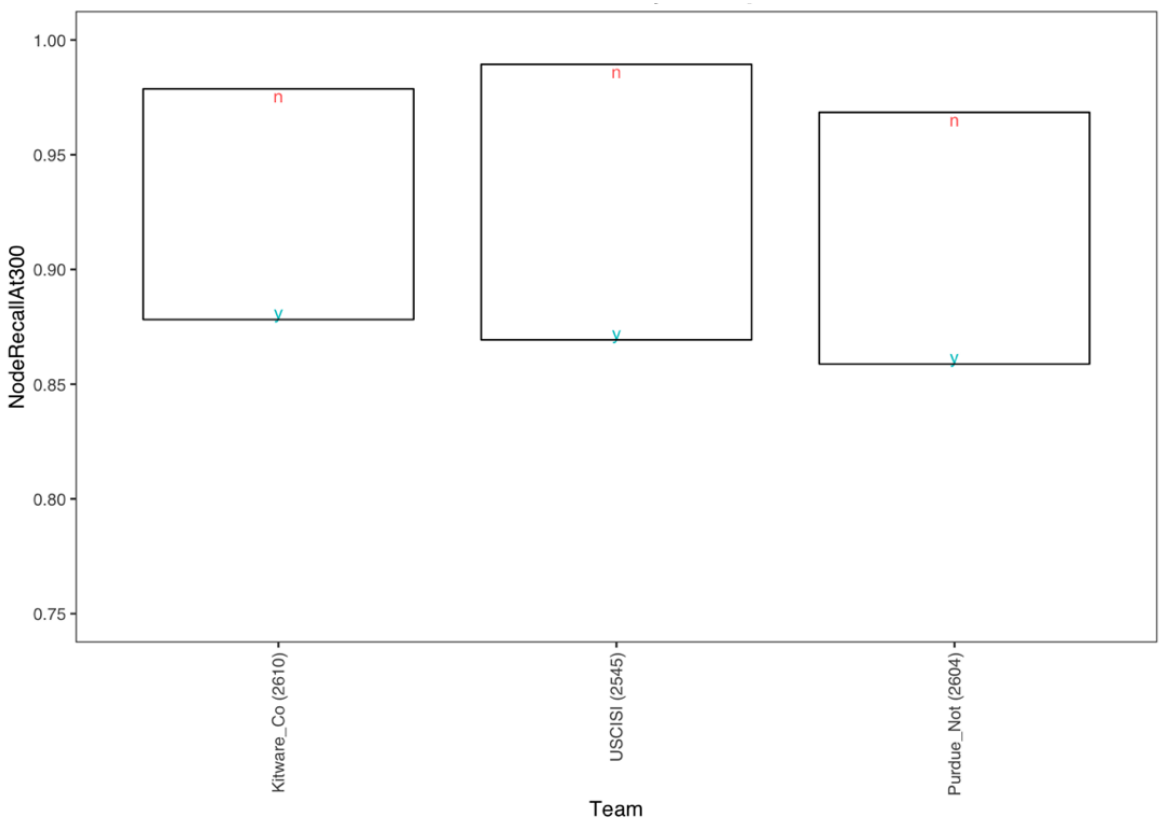
Y: GAN manipulation
N: Non-GAN manipulation

Observations

- The GAN-based manipulations are harder than the Non-GAN manipulations across the three systems
- All the three systems have large effect on the GAN factor

Antiforensics

Better



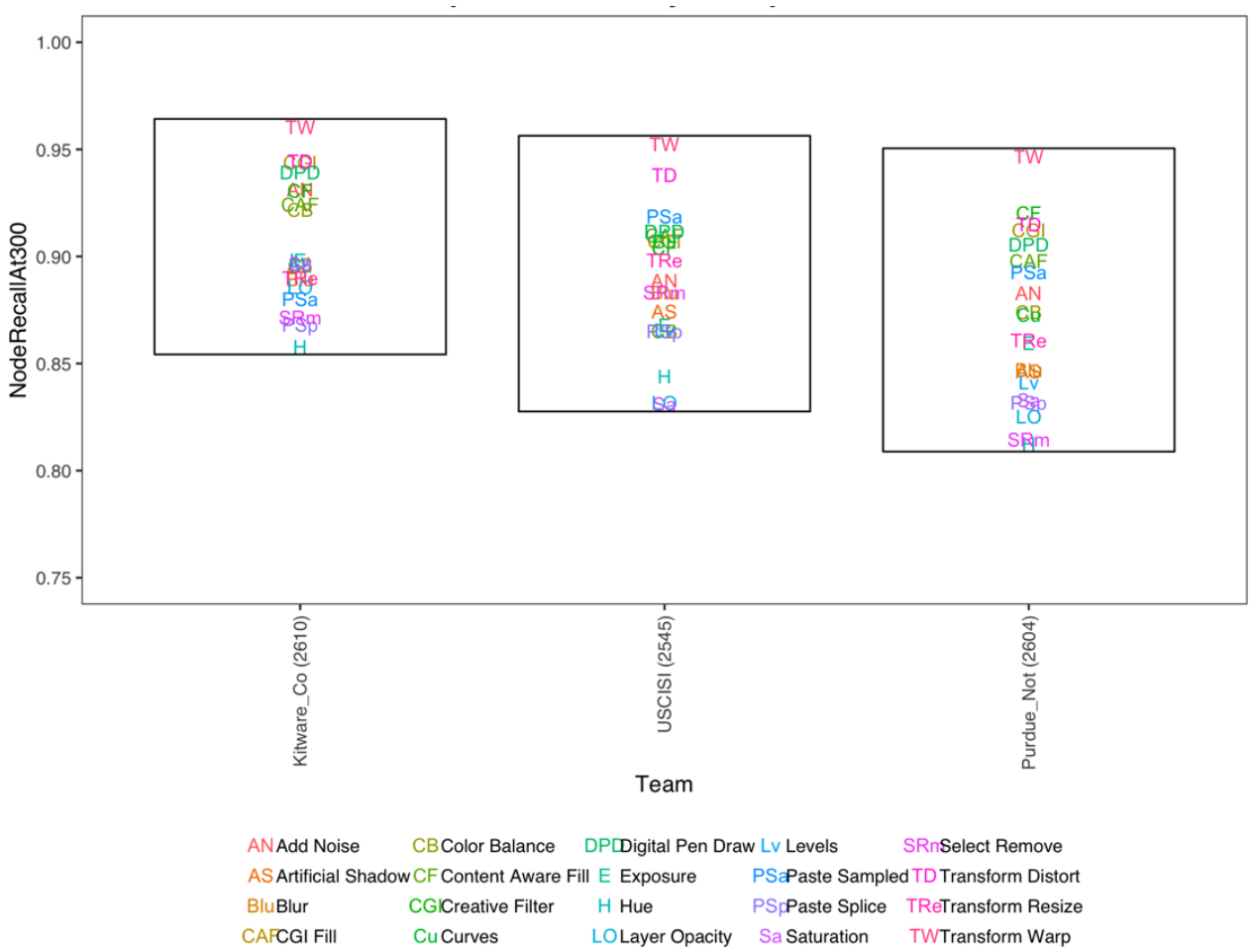
Y: After Antiforensics
N: Before Antiforensics

Observations

- The manipulations before Antiforensics are easier than after Antiforensics across the three systems
- All the three systems have large effect on the Antiforensics factor

Targeted Operations

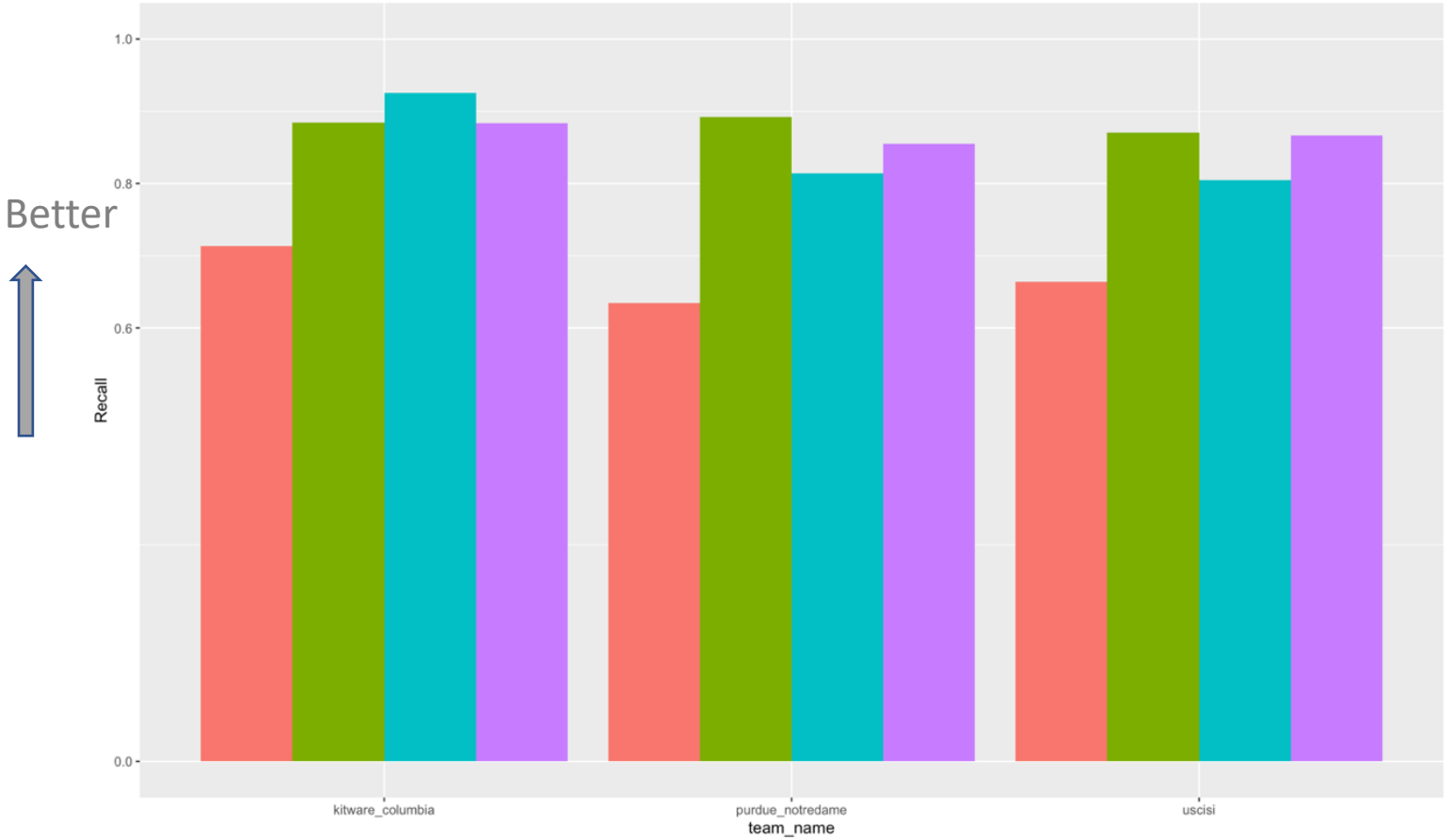
Better



Factor	Avg.Rank
TransformWarp	1.000000
TransformDistort	2.333333
CGIFill	4.333333
DigitalPenDraw	4.333333
CreativeFilter	5.333333
ContentAwareFill	6.000000
AddNoise	7.666667
PasteSampled	9.000000
Curves	10.000000
ColorBalance	11.000000
TransformResize	11.333333
Exposure	11.666667
ArtificialShadow	13.000000
Blur	13.333333
Levels	13.333333
Saturation	15.666667
SelectRemove	16.000000
LayerOpacity	17.666667
PasteSplice	17.666667
Hue	19.333333

- Out of the 20 operations, ”TransformWarp” is easier for the PF task across the three systems

Provenance Filtering Results over Years



Kitware-Columbia

Purdue_ NotreDame

USCISI

- Metric: Recall@200
- Different systems over years

- Different datasets over years

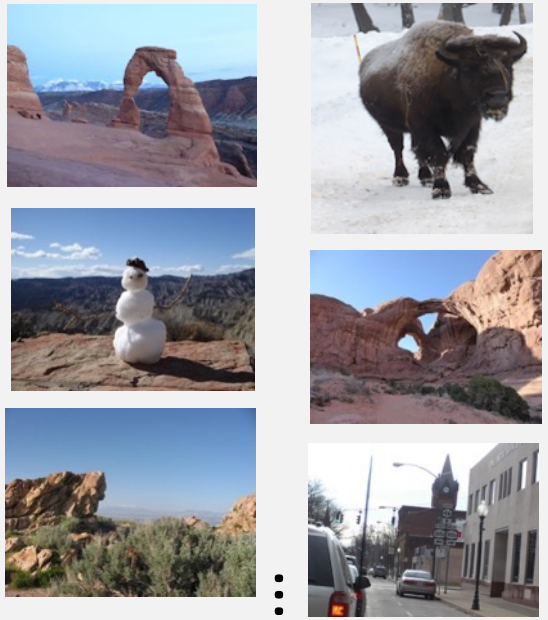
NIST Data Sets	Probe	World
NC17 EvalPart1	1K	1M
MFC18 EvalPart1	10K	1M
MFC19 EvalPart1	9420	2M
MFC20 EvalPart1	5926	2M

Provenance Graph Building Overview

System Input

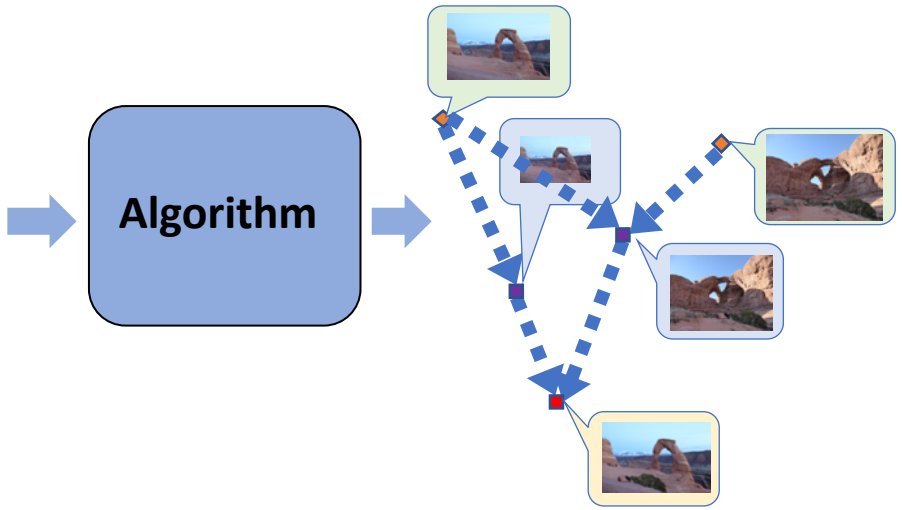
Probe Image

World Image Set



System Output

Constructing a provenance graph that describes the relationships among the images with the ancestor and descendent sequences



Performance Measure

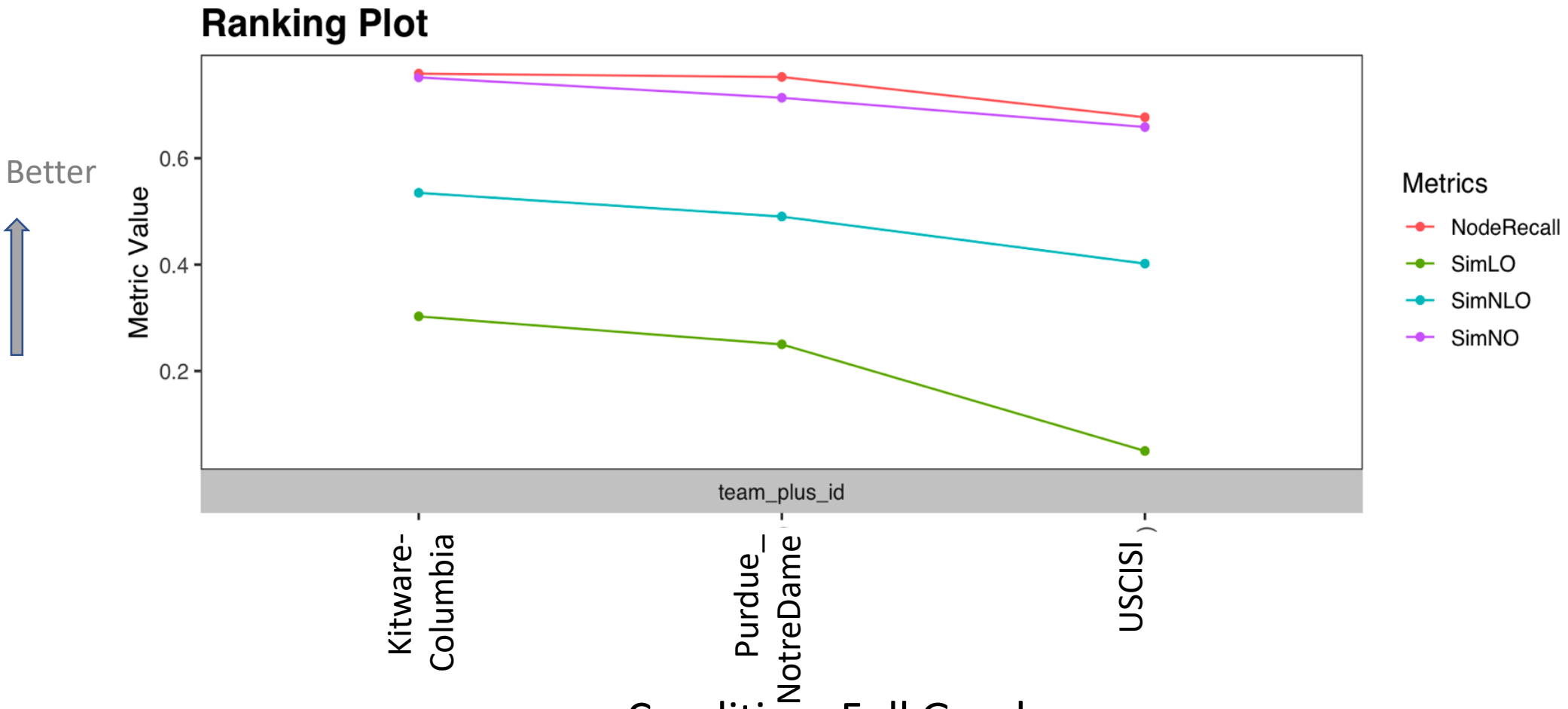
Node and Edge overlap similarity metric of a provenance graph

Generalized F-measure:

- Sim(Nodes Overlap)
- Sim (Links Overlap)
- **Sim(Nodes+Links Overlap)**

- NodeRecall

MFC20 Provenance Graph Building Results

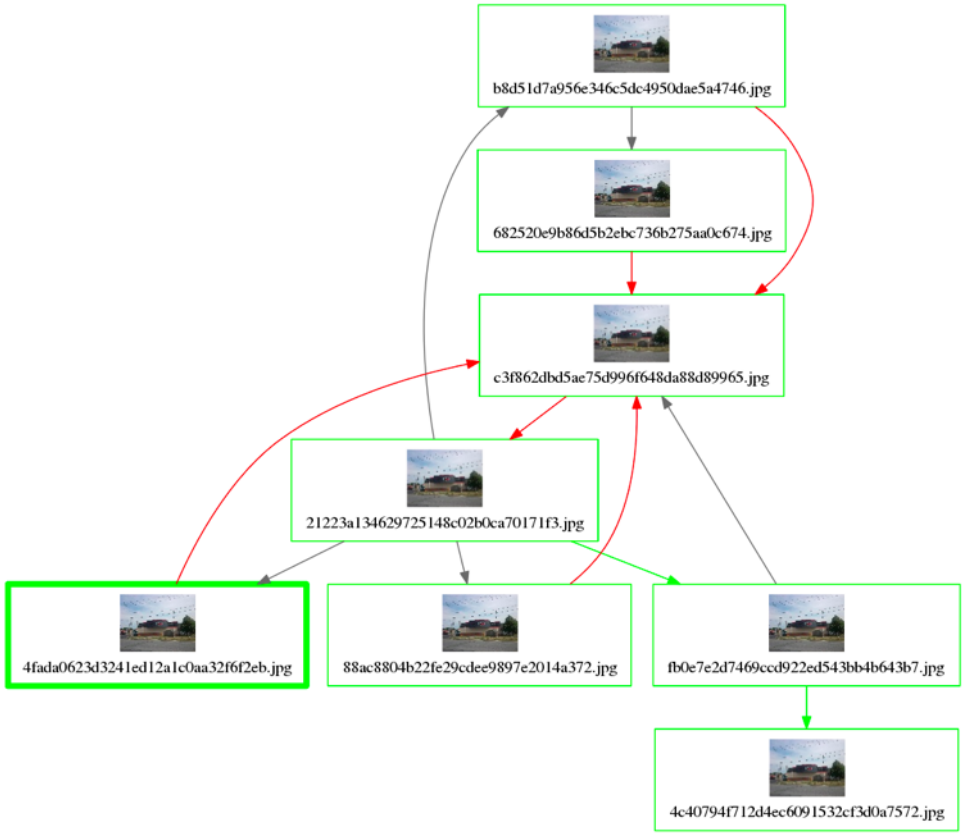


Condition: Full Graph

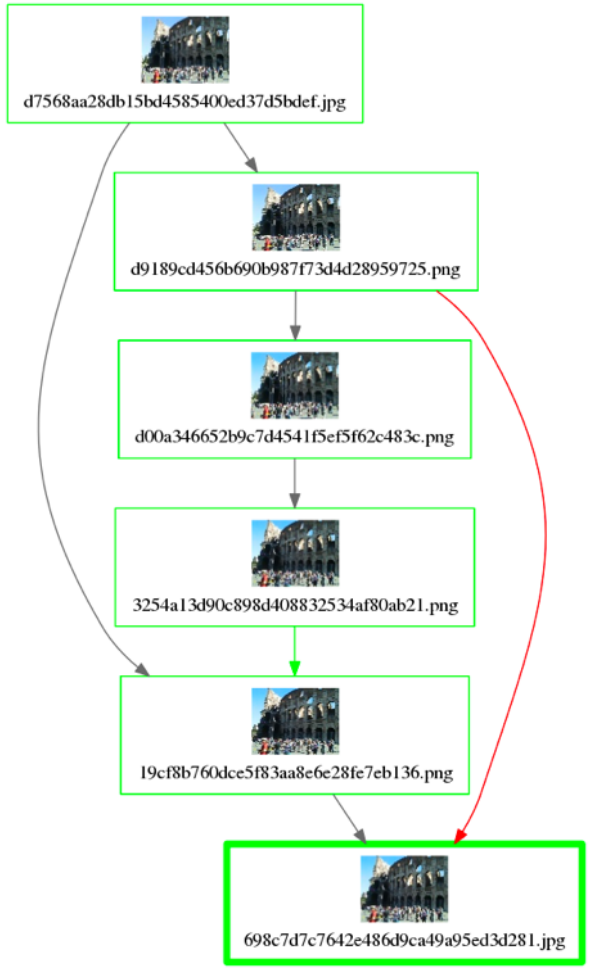
No submission for the subset (direct-path) graph condition for MFC20

System Output Scoring Examples (high scores)

SimNLO:0.73



SimNLO:0.70

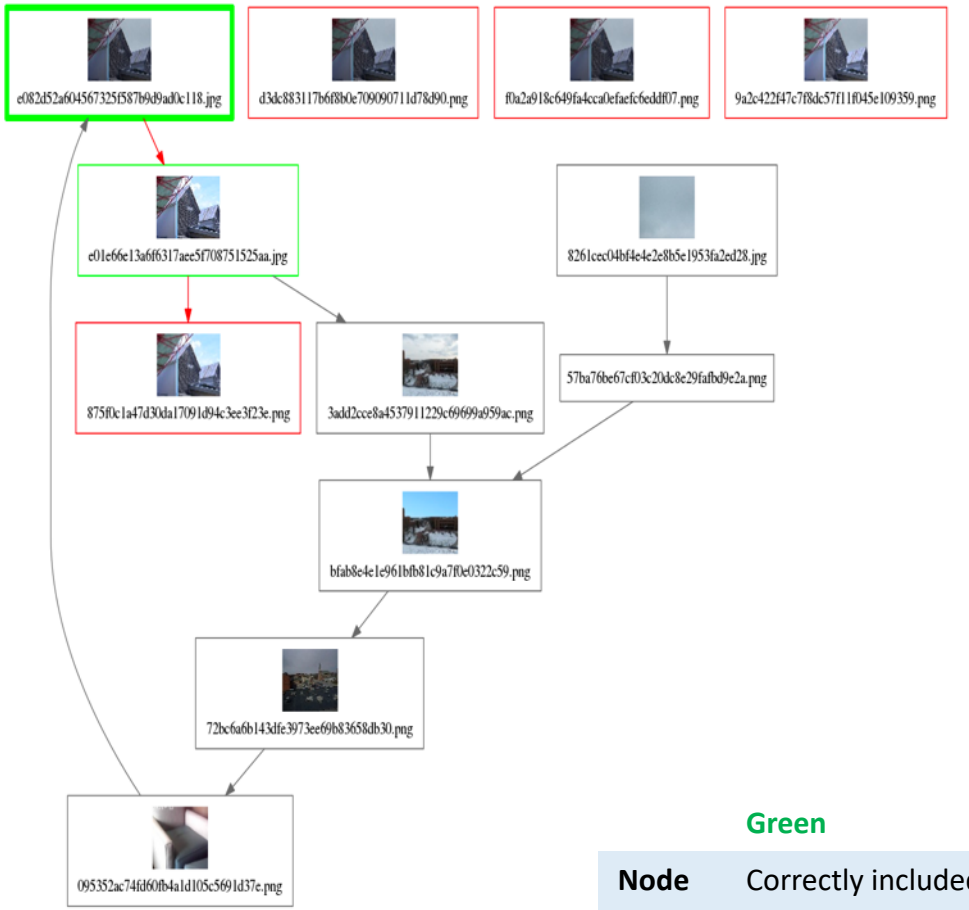


SimNLO:0.67

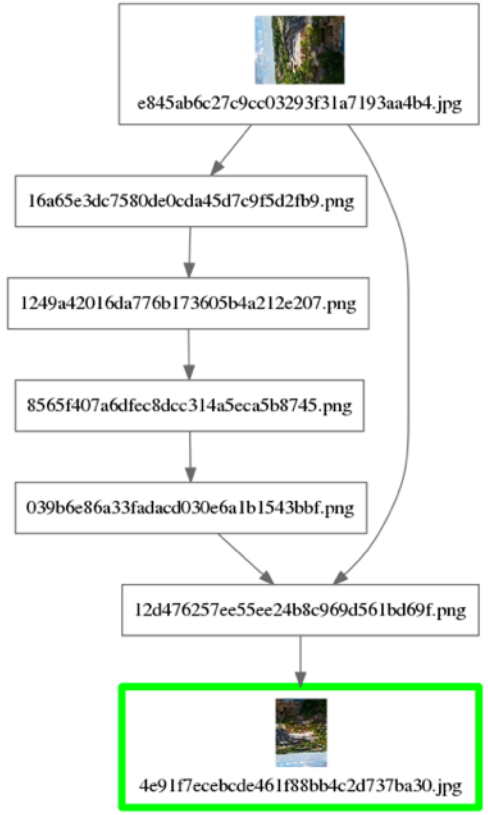
	Green	Red	Gray
Node	Correctly included	False alarm node	Omitted node
Link	Correctly linked	False alarm link	Omitted link

System Output Scoring Examples (low scores)

SimNLO:0.17



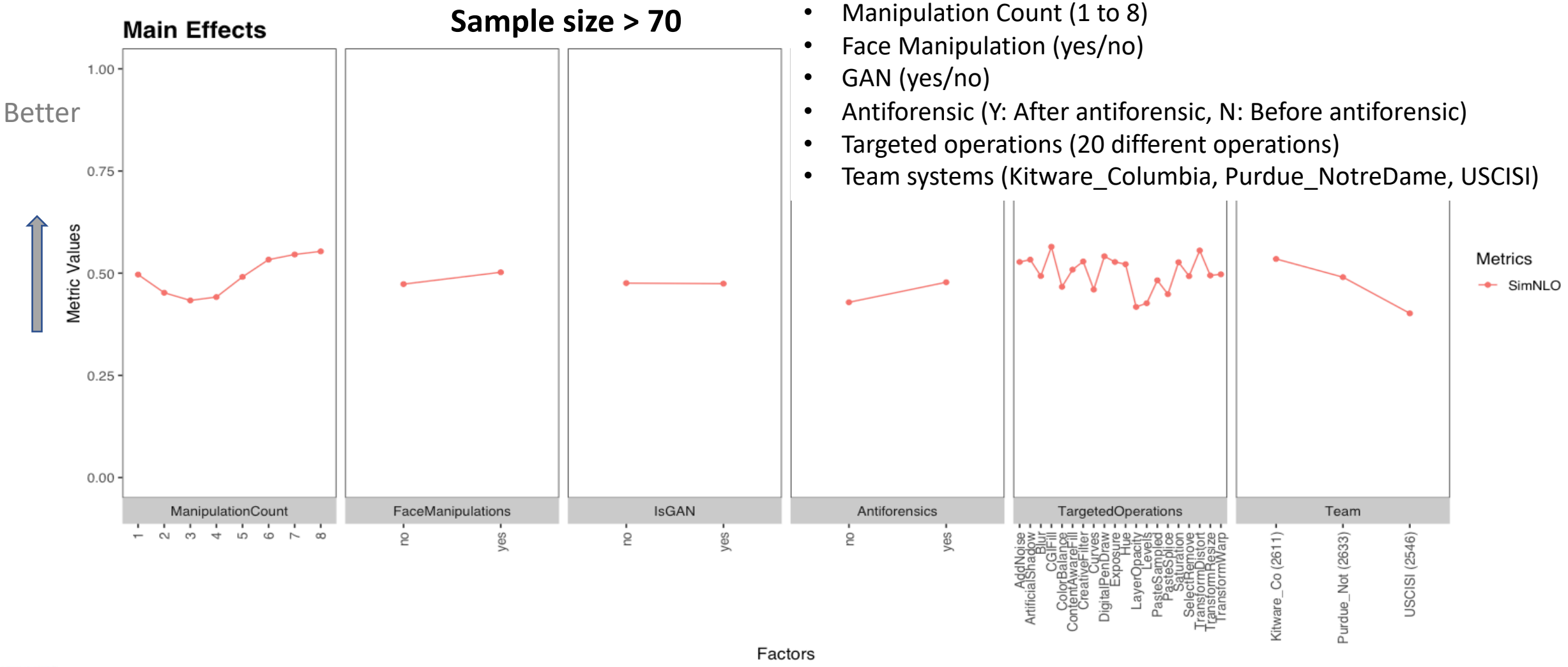
SimNLO:0.13



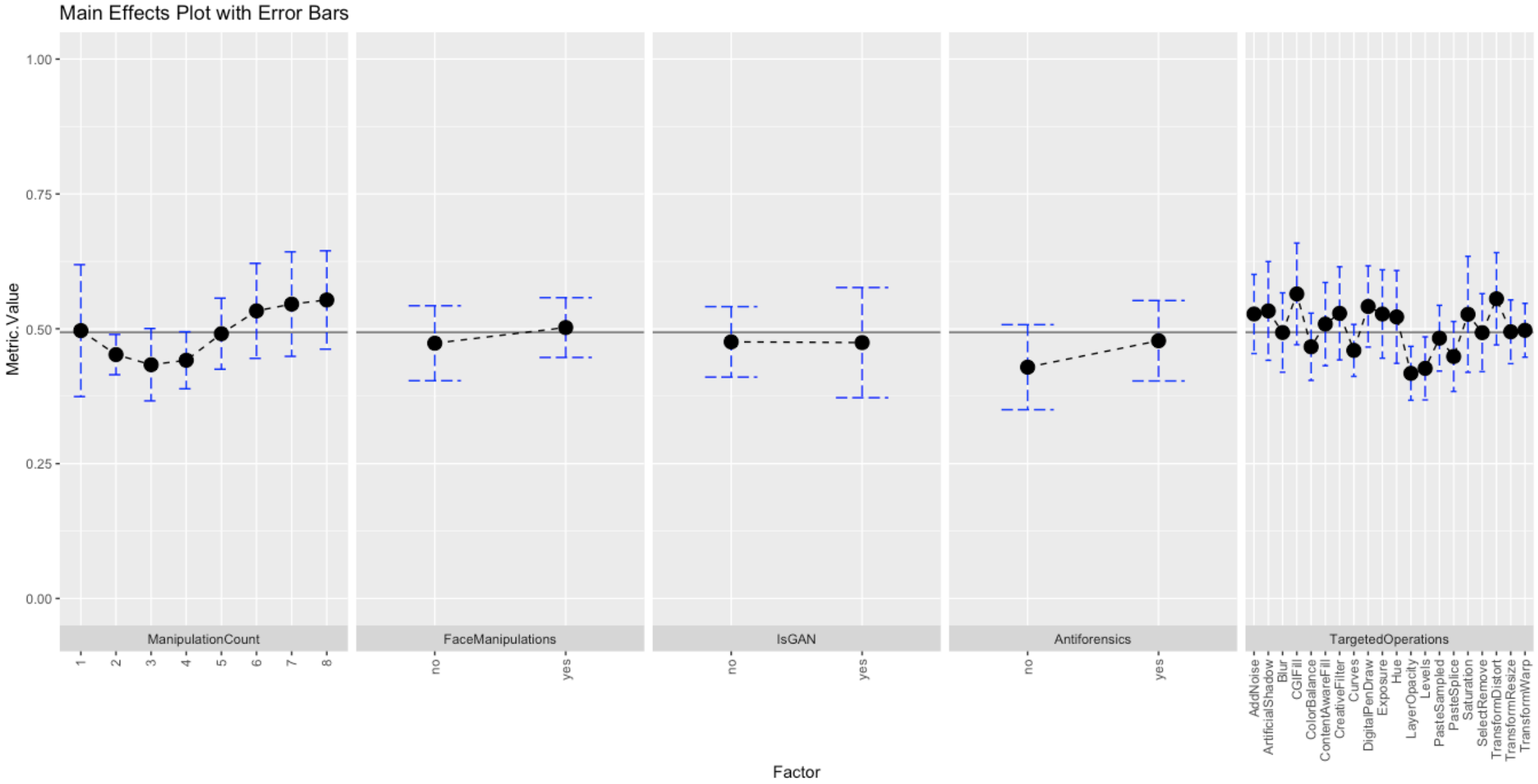
SimNLO:0.10

	Green	Red	Gray
Node	Correctly included	False alarm node	Omitted node
Link	Correctly linked	False alarm link	Omitted link

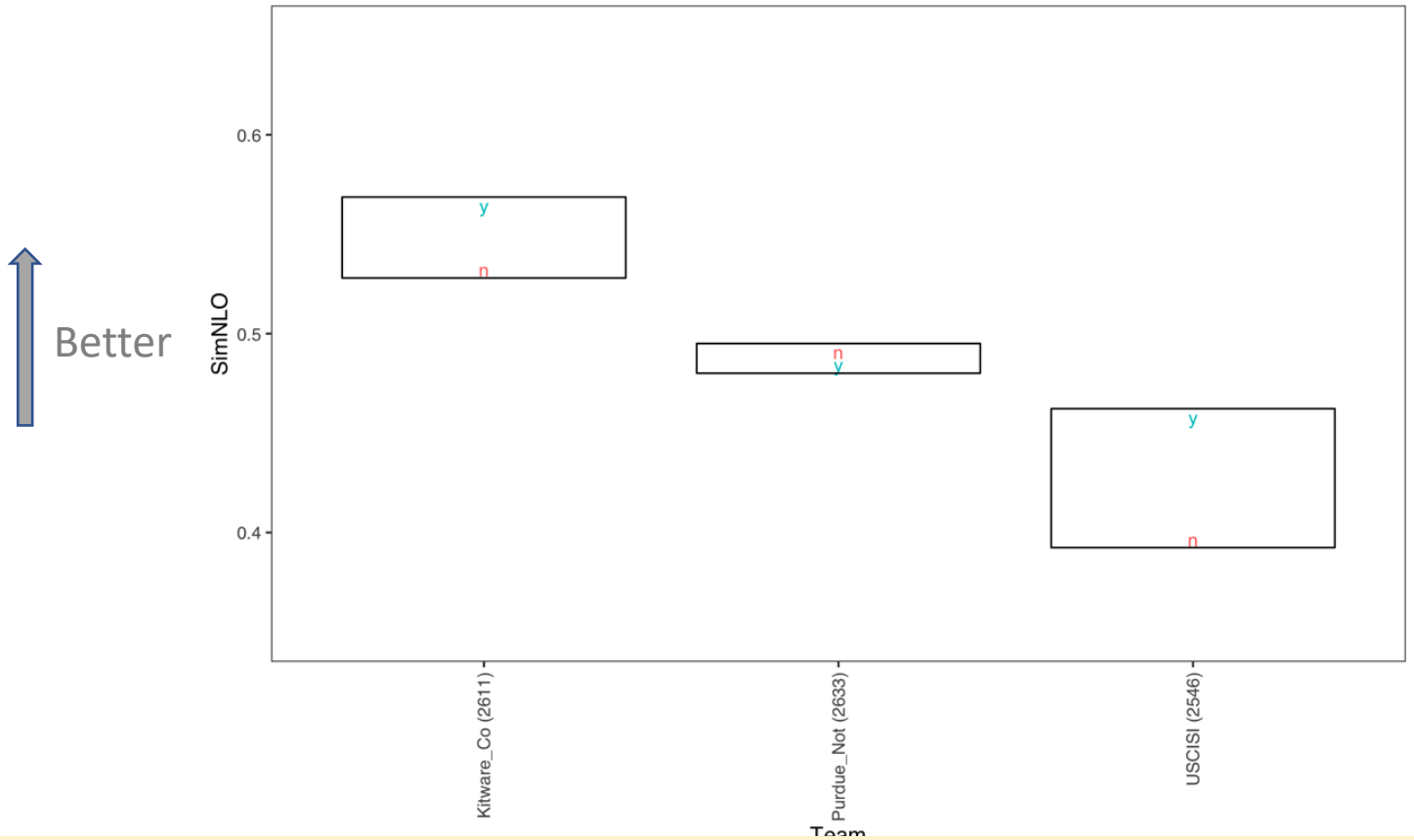
What are the important factors for PGB?



- Manipulation Count (1 to 8)
- Face Manipulation (yes/no)
- GAN (yes/no)
- Antiforensic (Y: After antiforensic, N: Before antiforensic)
- Targeted operations (20 different operations)
- Team systems (Kitware_Columbia, Purdue_NotreDame, USCISI)


Main Effects Plot with Error Bars
Metric.Value
1.00
0.75
0.50
0.25
0.00
ManipulationCount
1
2
3
4
5
6
7
8
FaceManipulations
no
yes
IsGAN
no
yes
Antiforensics
no
yes
TargetedOperations
AddNoise
ArtificialShadow
Blur
CGIFill
ColorBalance
ContentAwareFill
CreativeFilter
Curves
DigitalPenDraw
Exposure
Hue
LayerOpacity
Levels
PasteSampled
PasteSplice
Saturation
SelectRemove
TransformDistort
TransformResize
TransformWarp
Factor

Face Manipulation



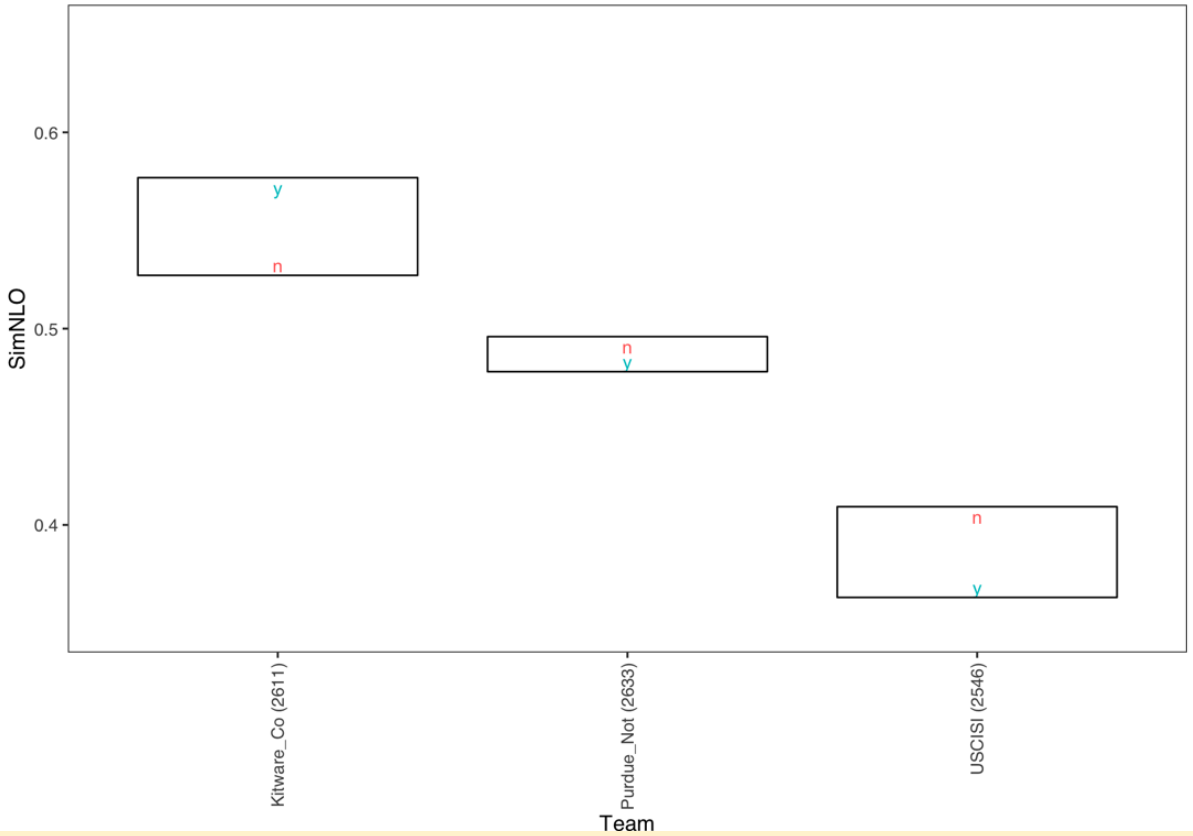
Y: Face manipulation
N: Non-face manipulation

Observations

- The face manipulations are easier for both Kitware-Columbia and USCISI while there is barely any effect on Purdue_Notredam
- USCISI has the largest effect on Face Manipulation followed by Kitware-Columbia

GAN

Better



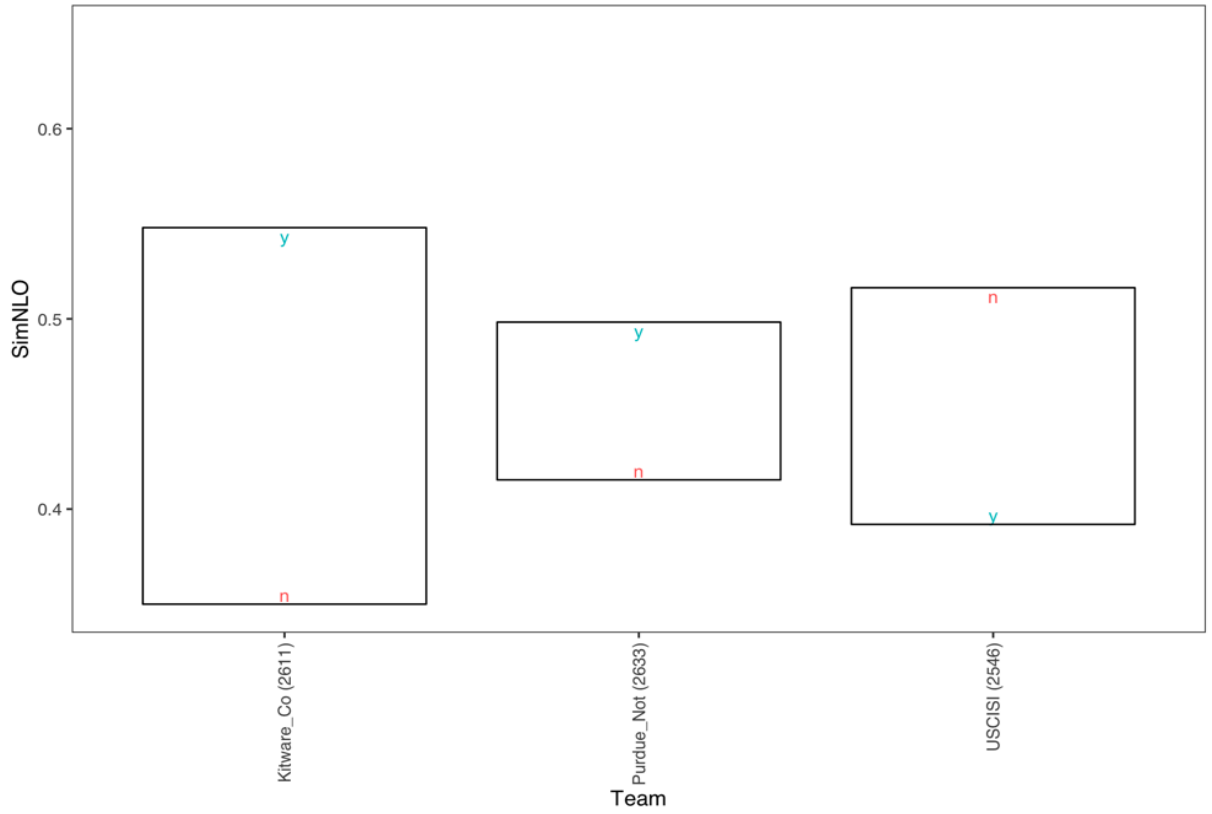
Y: GAN manipulation
N: Non-GAN manipulation

Observations

- The GAN-based manipulations are harder for USCISI and Purdue_Notredam, but easier for Kitware-Columbia
- The GAN factor has higher effect on Kitware-Columbia and USCISI

Antiforensic

Better

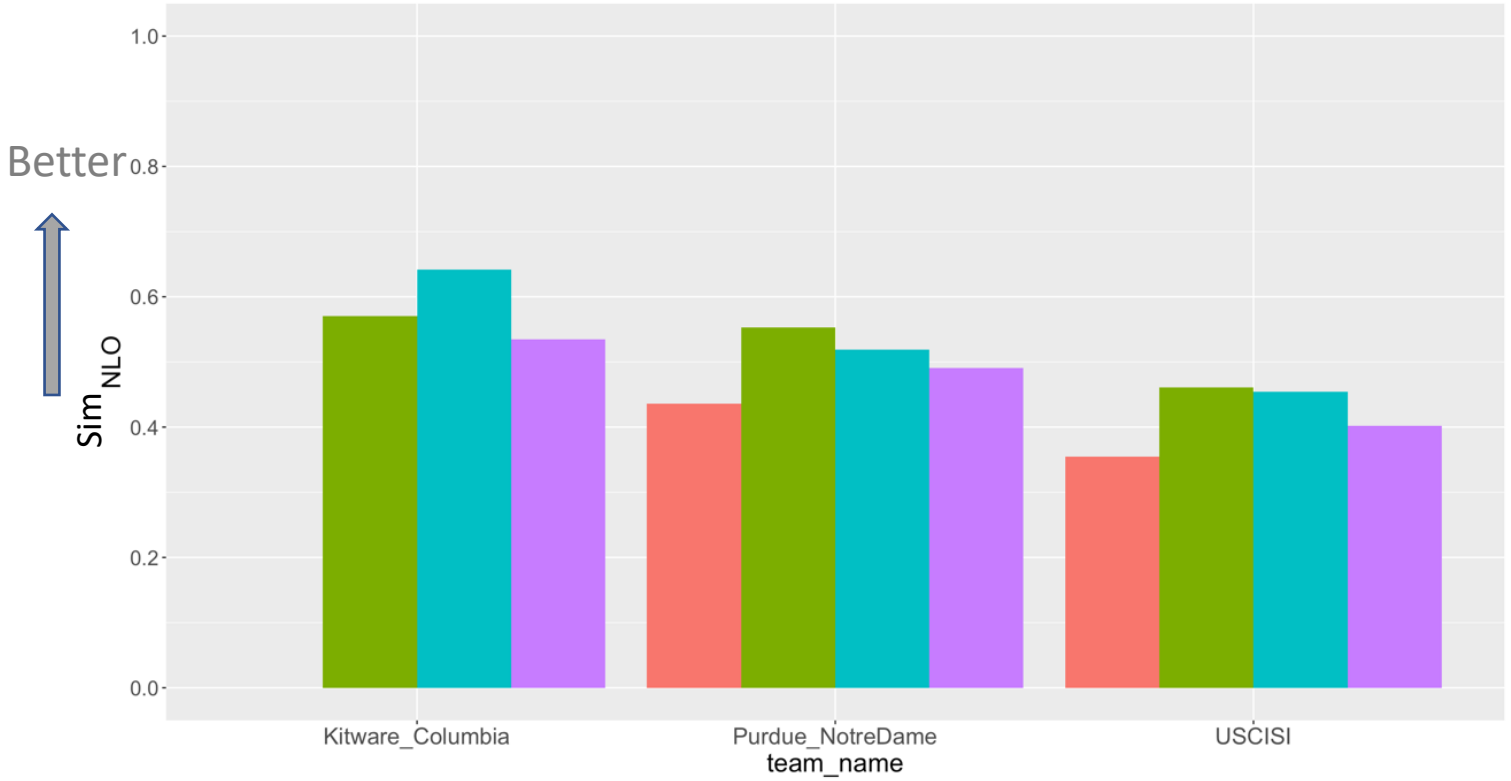


Y: After Antiforensics
N: Before Antiforensics

Observations

- The manipulations after Antiforensics are easier than before Antiforensics for Kitware-Columbia and Purdue_Notredam, but opposite for USCISI
- Kitware-Columbia has the highest effect on the Antiforensics factor

Provenance Graph Building Results over Years



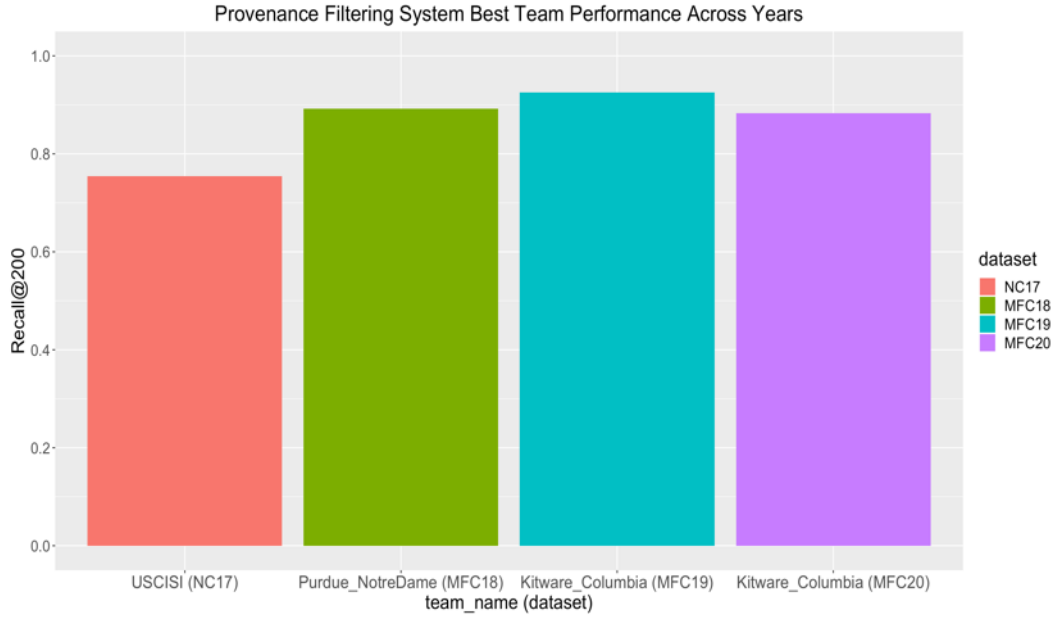
- Metric: Sim(NLO)
- Different systems over years

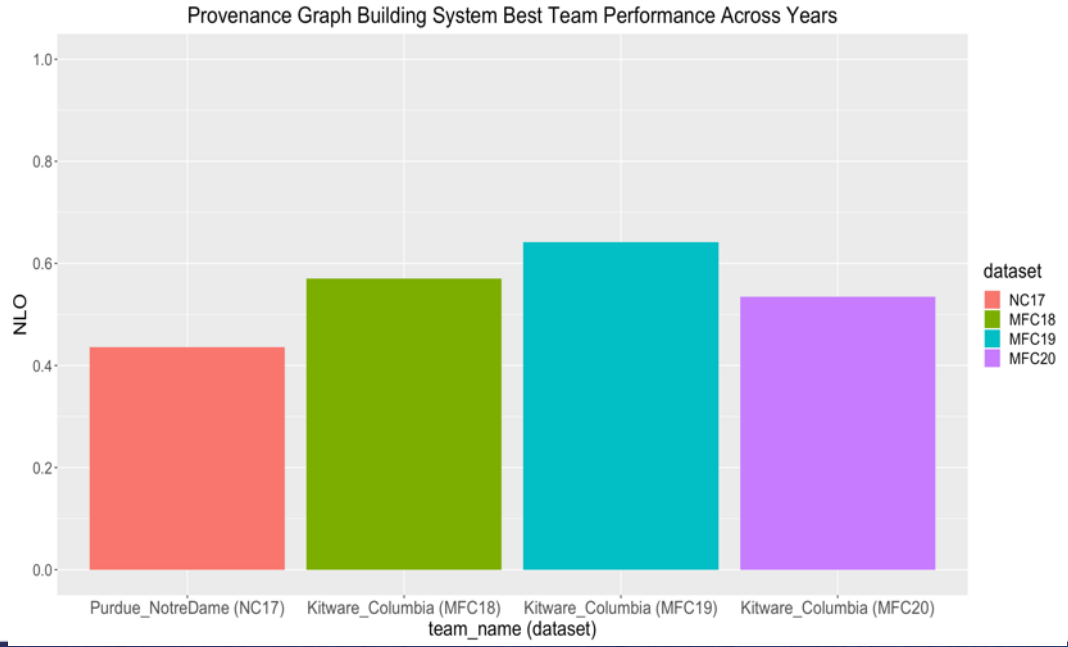
- Different datasets over years

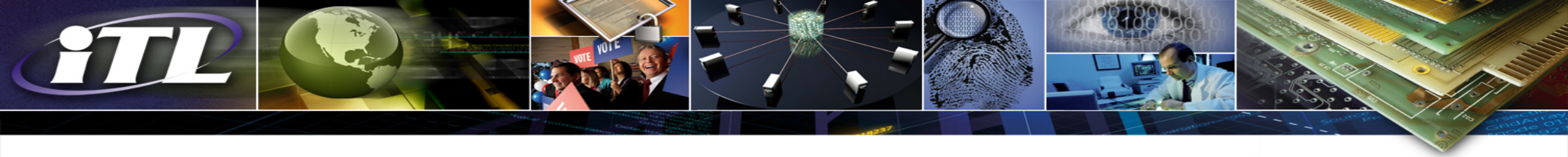
NIST Data Sets	Probe	World
NC17 EvalPart1	1K	1M
MFC18 EvalPart1	10K	1M
MFC19 EvalPart1	9420	2M
MFC20 EvalPart1	5926	2M

Provenance Summary

- Provenance filtering (PF) and graph building (PGB) are a challenging task
- For the MFC20 evaluation, Kitware-Columbia has the highest performance for both PF and PGB
- Important factors given the 6 factors are different between PF and PGB
  - Each factor behaves differently between PF and PGB
  - Antiforensic has larger effect on both PF and PGB performance
- Best system results trend over years







MFC20 Camera Verification Evaluation Results Deep Dive

Jonathan Fiscus (Co-PI), **Dr. Haiying Guan** (Co-PI), Dr. Yooyoung Lee,
Dr. Amy Yates[+], Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah,
Dr. Xiongnan Jin

Multimodal Information Group, [+] Image Group
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology (NIST)

April 21-25, 2020

Camera ID Verification Outline

- Task definition
- Evaluation data
- Evaluation metrics
- MFC20 result

Camera ID Verification Task

- Task: Determine if a probe is from a claimed camera fingerprint.
  - If relevant, determine where the media regions had content changes.

Training Images/Video(s)

System input

System Output

Camera ID Verification Evaluation Features

- Objective: how does system perform:
  - with limited training resources
  - matching sensors cross media modality
- Key features
  - Specify training data
  - Support cross modality on training and testing data among image, video, and multimedia
  - Support localization task

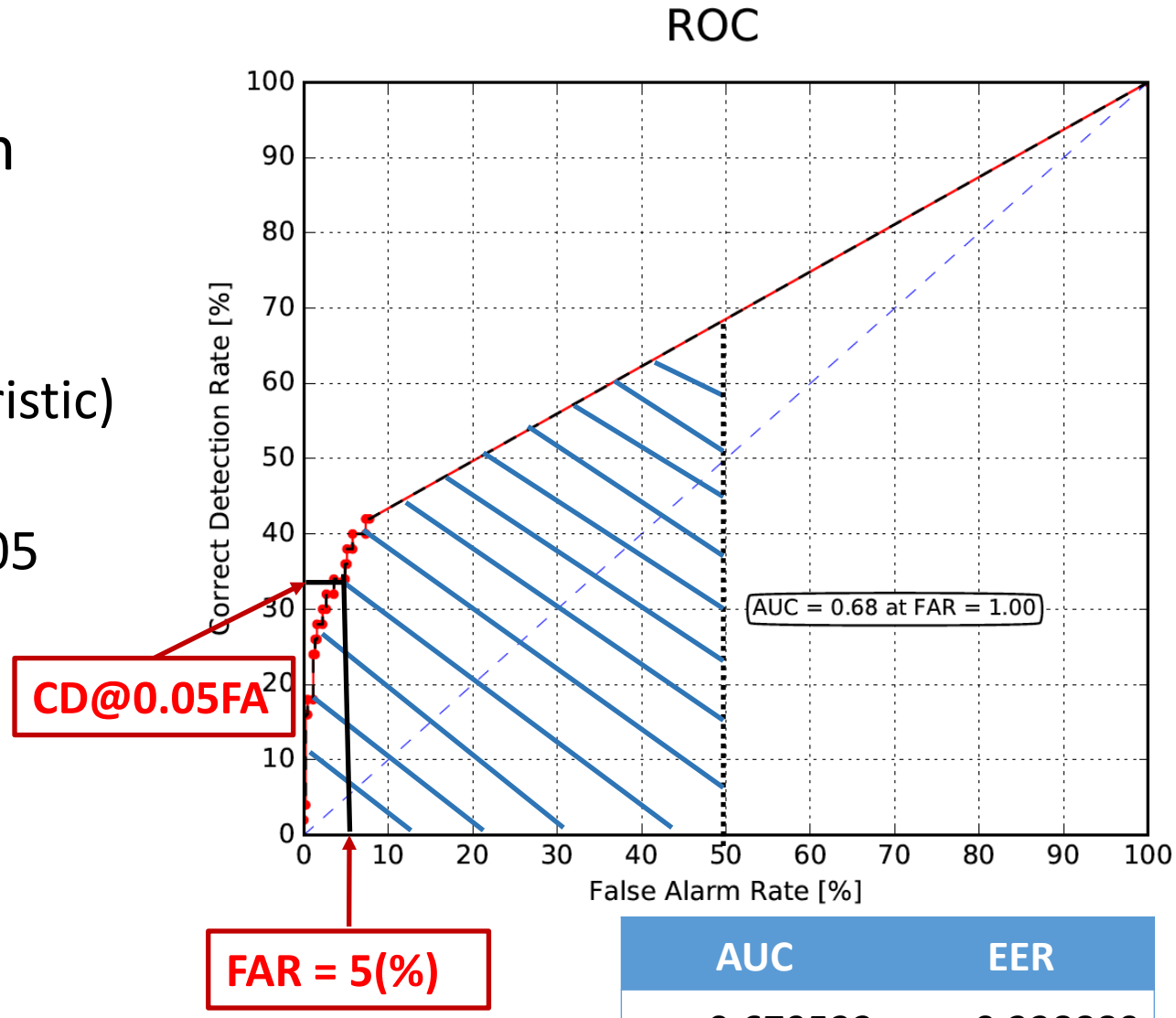
MFC20 Camera ID Verification Datasets

- Six datasets:
  - 3 training modalities (Image, Video, Multimedia)
  - 2 testing modalities (Image, Video)

Test	Train	Probe Pair	Camera	Journal
Image	Image	11288	106	1454
	Video	9346	88	1411
	Multimedia	9346	88	1411
Video	Image	788	35	87
	Video	767	34	87
	Multimedia	767	34	87

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
  - EER (Equal Error Rate)



AUC	EER
0.679533	0.328889

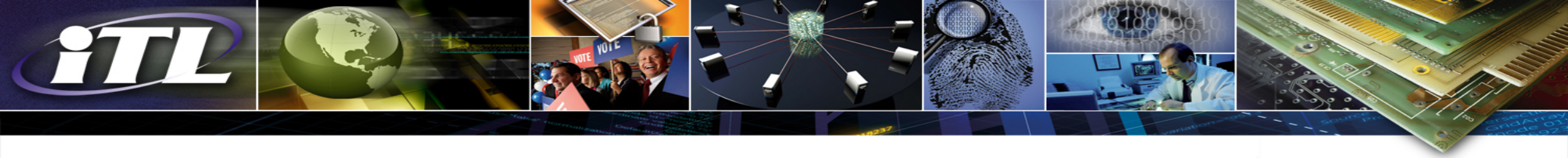
Holistic vs. Opt In Technologies

- Allowing Systems to Respond When/If Appropriate

- Evaluation challenge:
  - Some media forensic systems **determine** a response should not be returned
    - E.g., the video frame size is different with trained model, the image is not with the supported formats, the pixel is saturated etc.

Probe Status	Description
Processed	probe was fully processed
OptOut	the system **determined** a response should not be returned
OptOutLocalization	the system, **determined** a detection response but not a localization response should be returned
NonProcessed	A system failure of some kind occurred and will be scored with low probability

- NIST reports:
  - Holistic performance measures: score all trials
  - Opt In performance measures:
    - Trial Response Rate (TRR) – Percent of processed, NonProcessed, and FailedValidation images
    - Performance measures excluding opt'd out probes

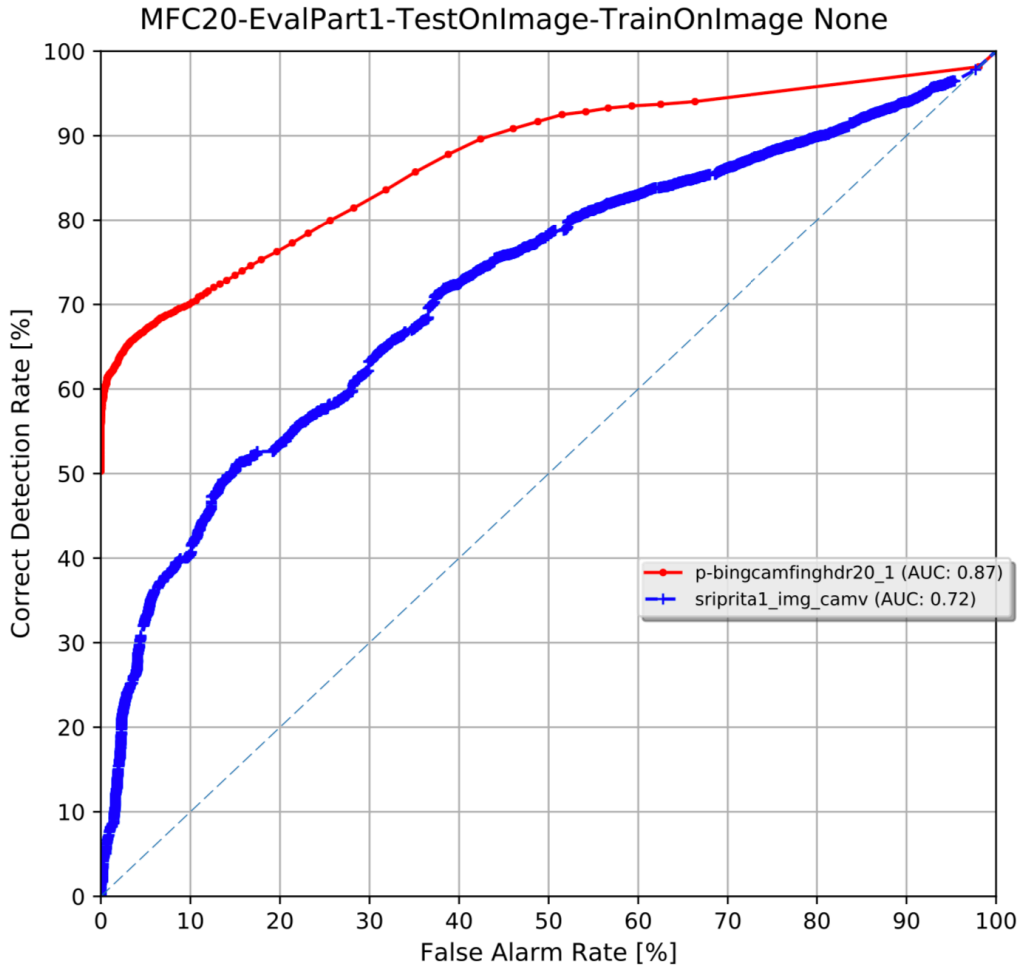


Camera ID Verification
Subtask: Train and Test On Image

MFC20 EvalPart1 Train and Test on Image (Full Data)

- 11288 image camera pairs
- 106 cameras models
- 2 teams:
  - Binghamton
  - SRI-PRNU-TA1
- Highest AUC system:
  - Team ID: Binghamton
  - AUC = 0.872
  - CD@0.05FA = 0.67
  - System ID: p-bingcamfinghdr20_1

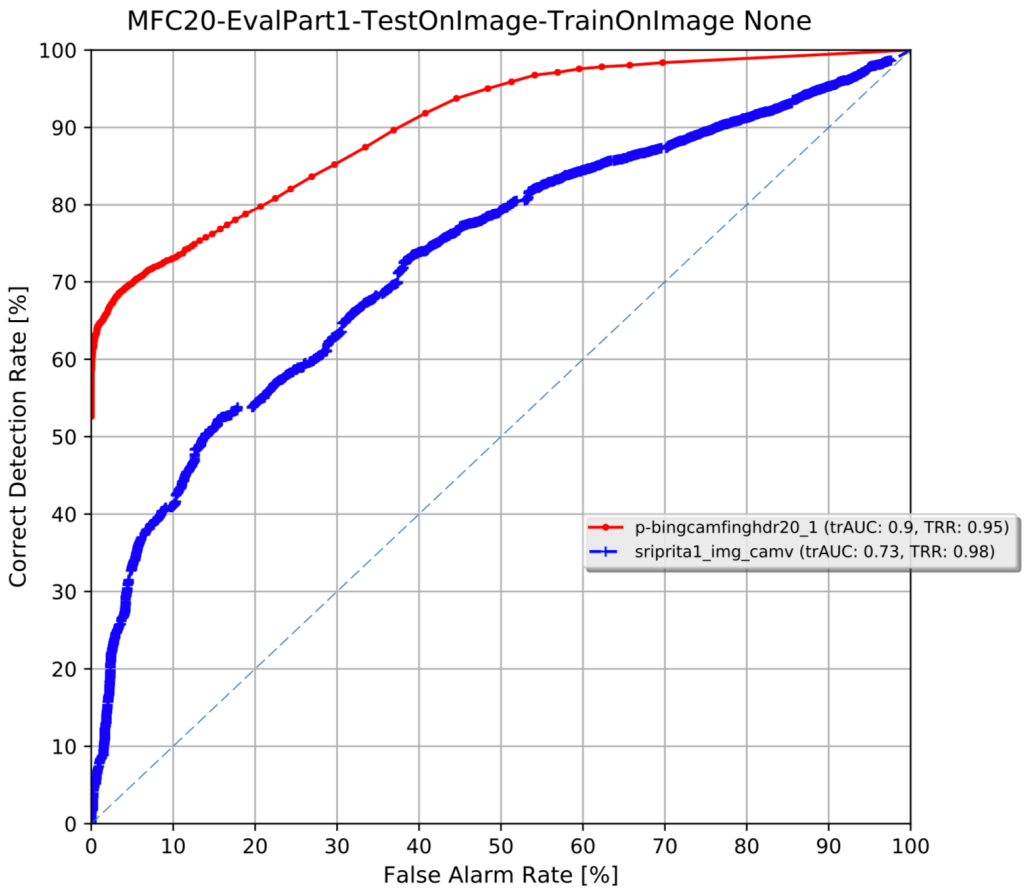
Figure: TA1 system MFC20 EP1, All probes (regardless of Opt In)



MFC20 EvalPart1 Train and Test on Image (Opt In)

- 2 teams:
  - Binghamton
  - SRI-PRNU-TA1
- Highest AUC system (OptIn) :
  - Team ID: Binghamton
  - OptIn TRR = 0.95
  - AUC = 0.902
  - CD@0.05FA = 0.698
  - System ID: p-bingcamfinghdr20_1

Figure: TA1 system MFC20 EP1 (Highest OptIn)



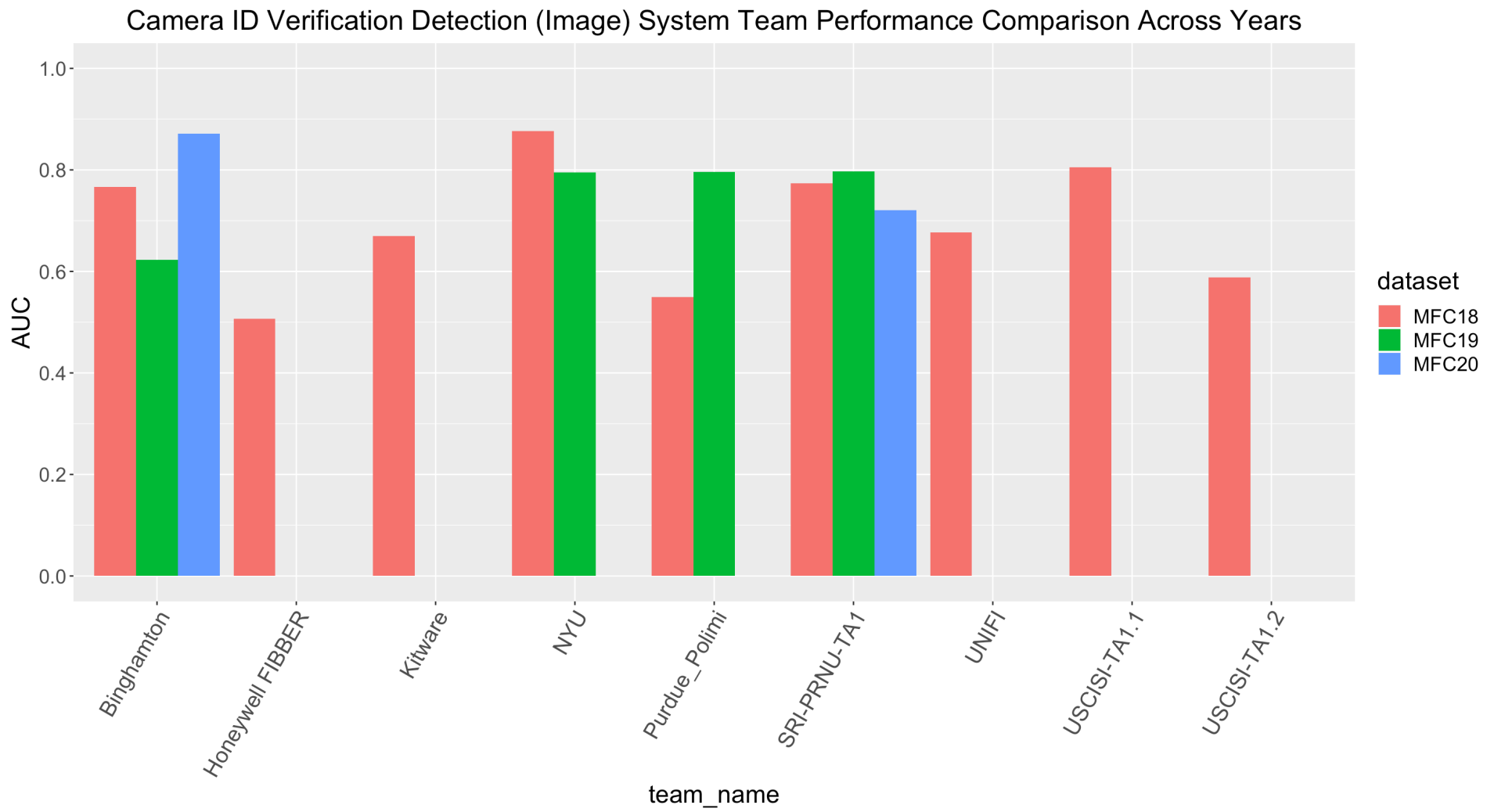
Performance Comparison Across Years
- Camera ID Verification Data Set Summary

- Test on image dataset summary

		MFC18			MFC19			MFC20		
Test	Train	Probe Pair	Cam.	Jour.	Probe Pair	Cam.	Jour.	Probe Pair	Cam.	Jour.
Image	Image	5275	39	452	8804	73	844	11288	106	1454
	Video	3383	25	410	6845	57	802	9346	88	1411
	Multimedia	3383	25	410	6845	57	802	9346	88	1411

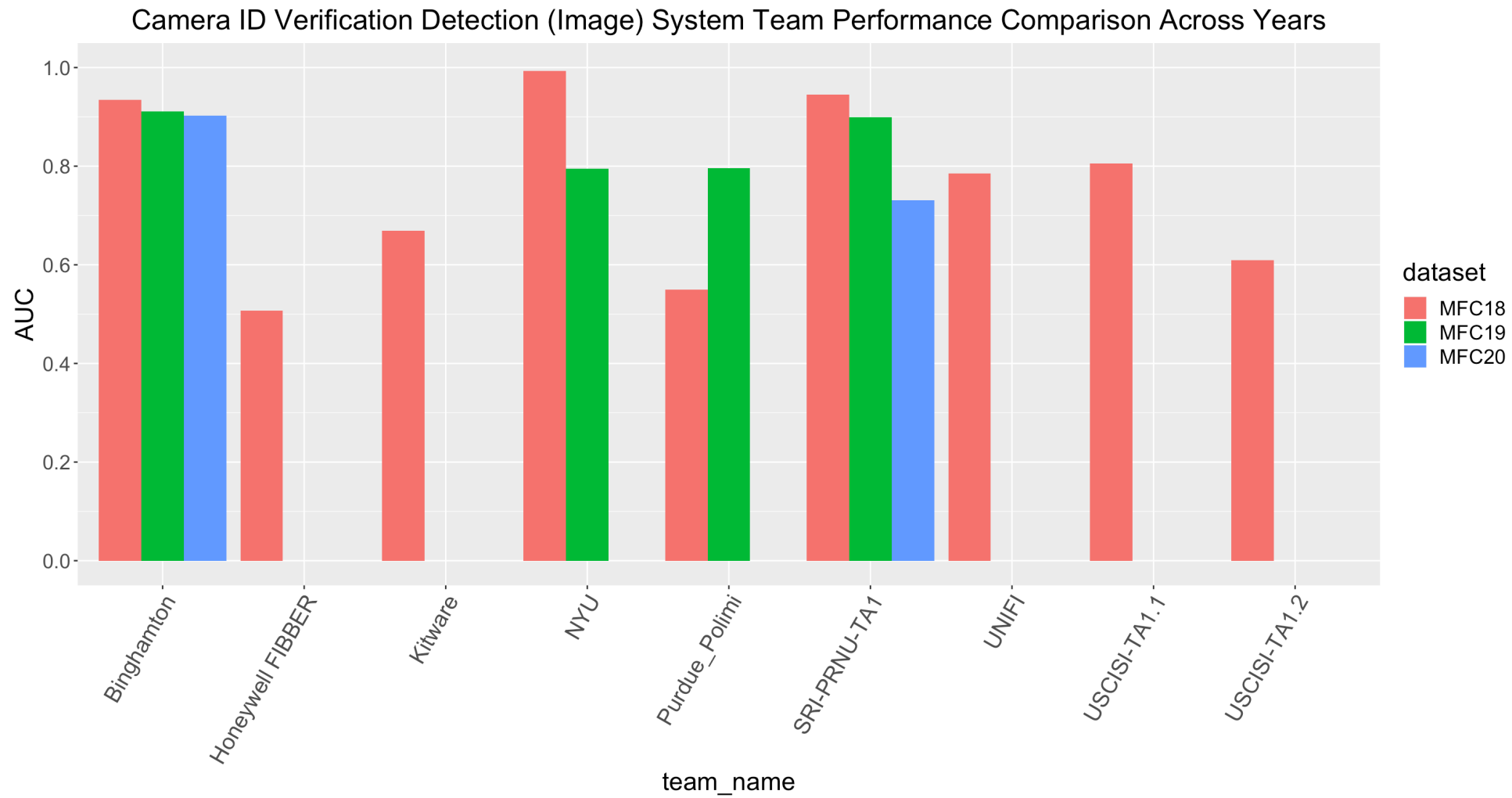
Camera ID Verification Detection: Train and Test on Image

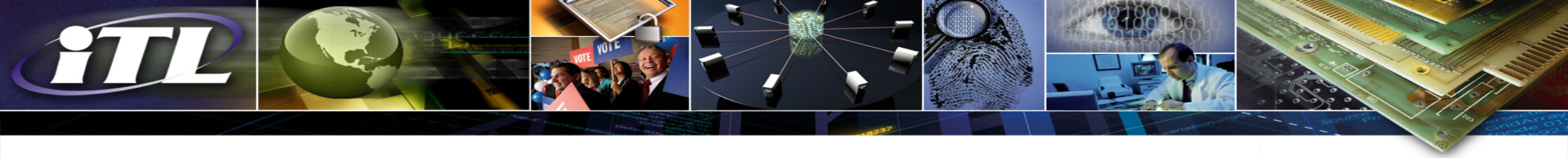
- Team Performance Comparison Across Years (Full Data)



Camera ID Verification Detection: Train and Test on Image

- Team Performance Comparison Across Years (OptIn)



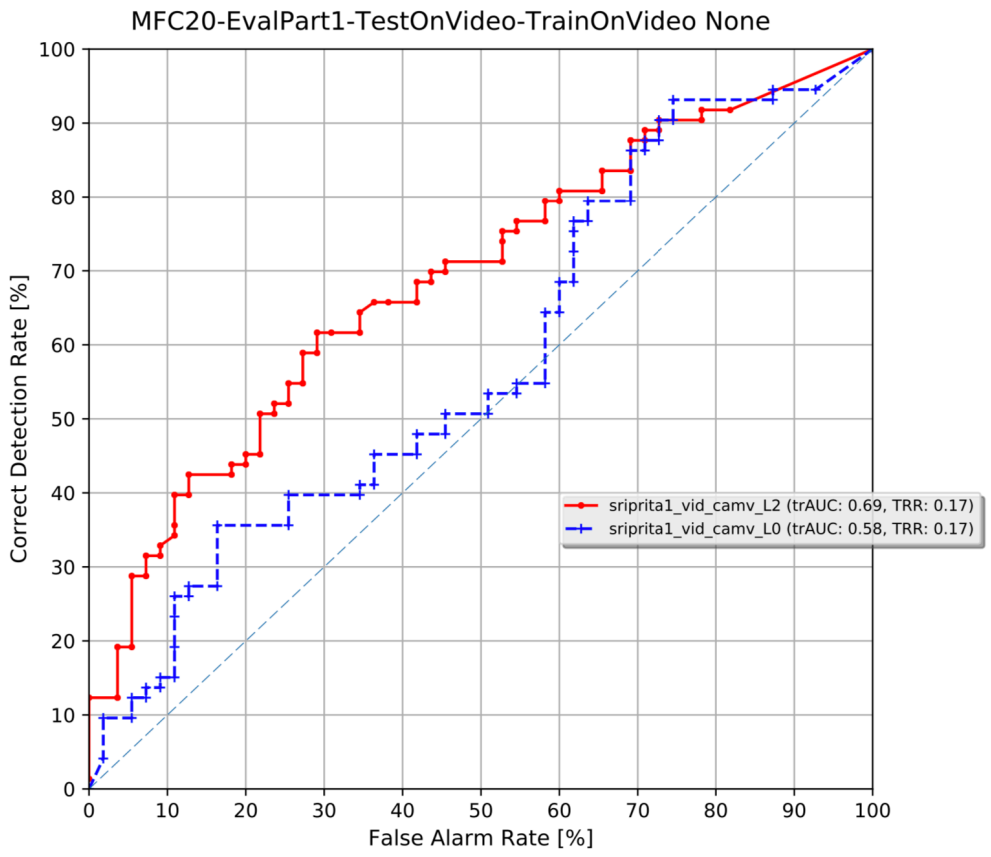


Camera ID Verification
Subtask: Train and Test On Video

MFC20 EvalPart1 Train and Test On Video

- 767 video camera pairs
- 34 cameras models
- 1 team:
  - SRI-PRNU-TA1
- 2 systems:
  - Highest AUC = 0.689 (Opt In TRR = 0.17)
  - Highest CD@0.05FA = 0.192
  - Team ID: SRI-PRNU-TA1
  - System ID: sriprita1_vid_camv_L2

Figure: TA1 system MFC20 EP1, Opt In (TRR = 0.17)



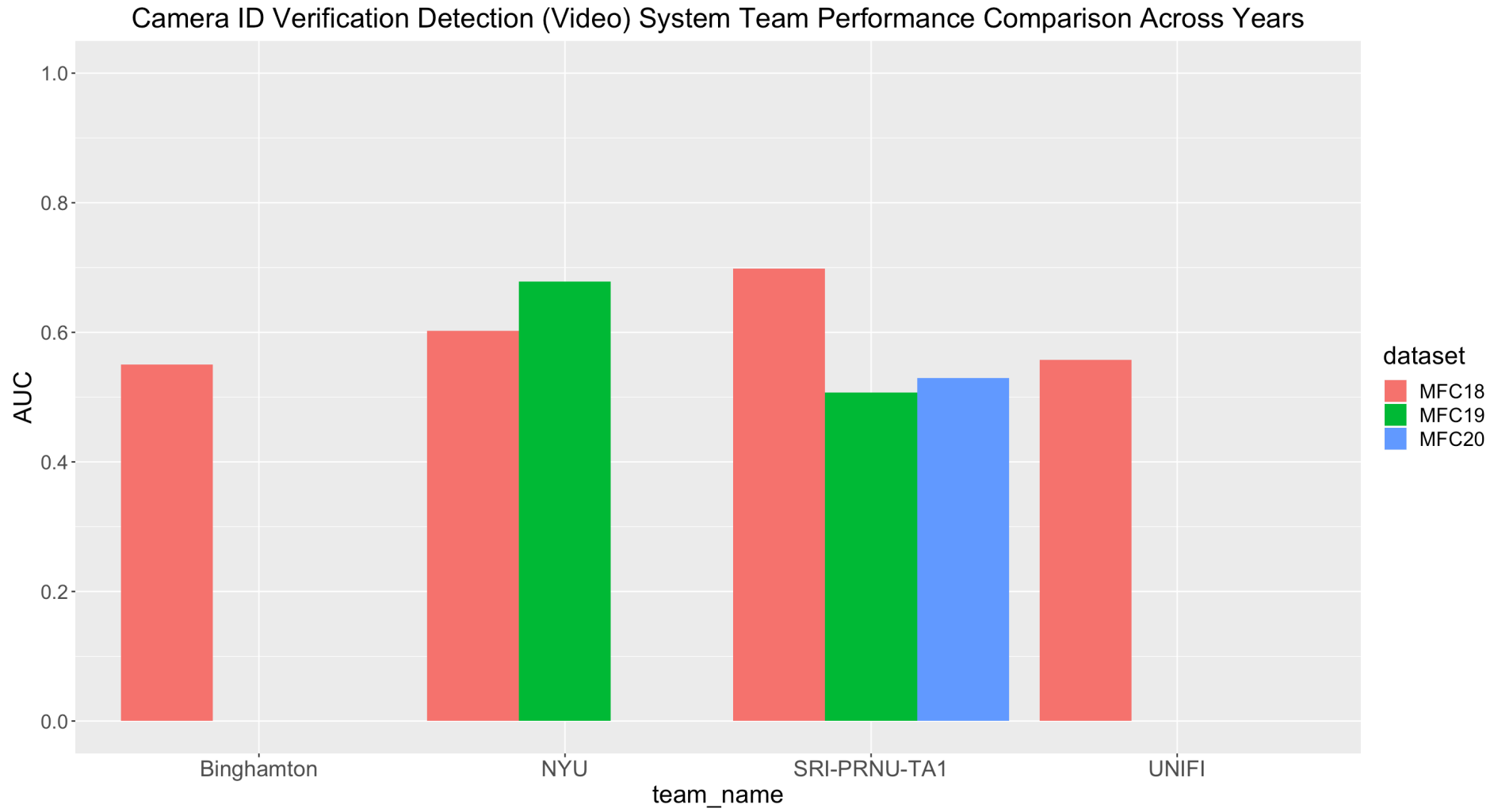
Performance Comparison Across Years
- Camera ID Verification Data Set Summary

- Test on video dataset summary

		MFC18			MFC19			MFC20		
Test	Train	Probe Pair	Cam.	Jour.	Probe Pair	Cam.	Jour.	Probe Pair	Cam.	Jour.
Video	Image	289	11	67	351	23	81	788	35	87
	Video	289	11	67	337	22	81	767	34	87
	Multimedia	289	11	67	337	22	81	767	34	87

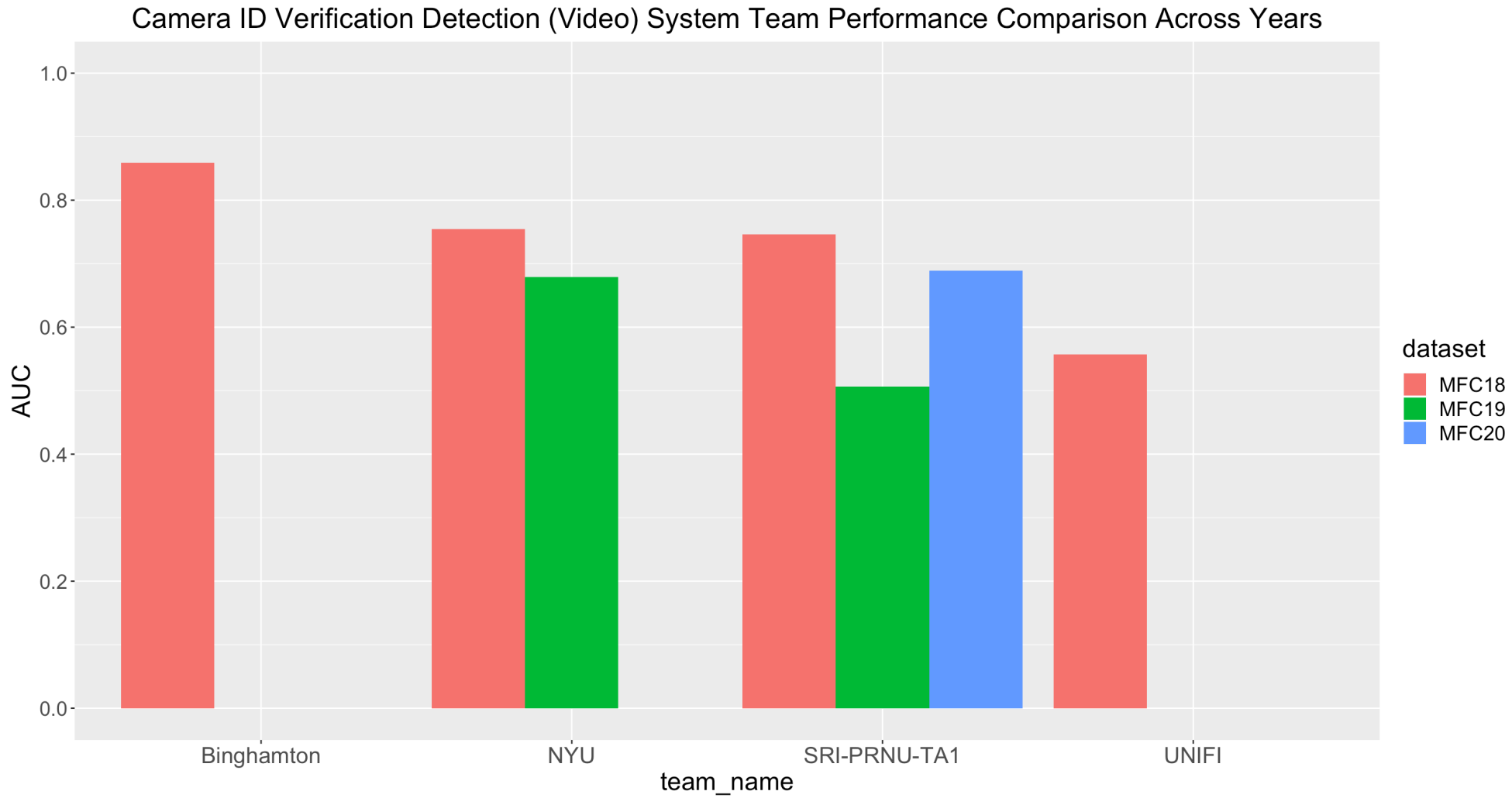
Camera ID Verification Detection: Train and Test on Video

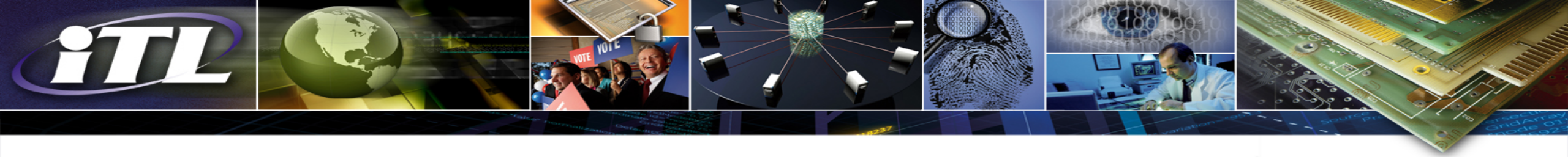
- Team Performance Comparison Across Years (Full Data)



Camera ID Verification Detection: Train and Test on Video

- Team Performance Comparison Across Years (Opt In)





MFC20 Event Verification Evaluation Results Deep Dive

Jonathan Fiscus (Co-PI), **Dr. Haiying Guan** (Co-PI), Dr. Yooyoung Lee,
Dr. Amy Yates[+], Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah,
Dr. Xiongnan Jin

Multimodal Information Group, [+] Image Group
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology (NIST)

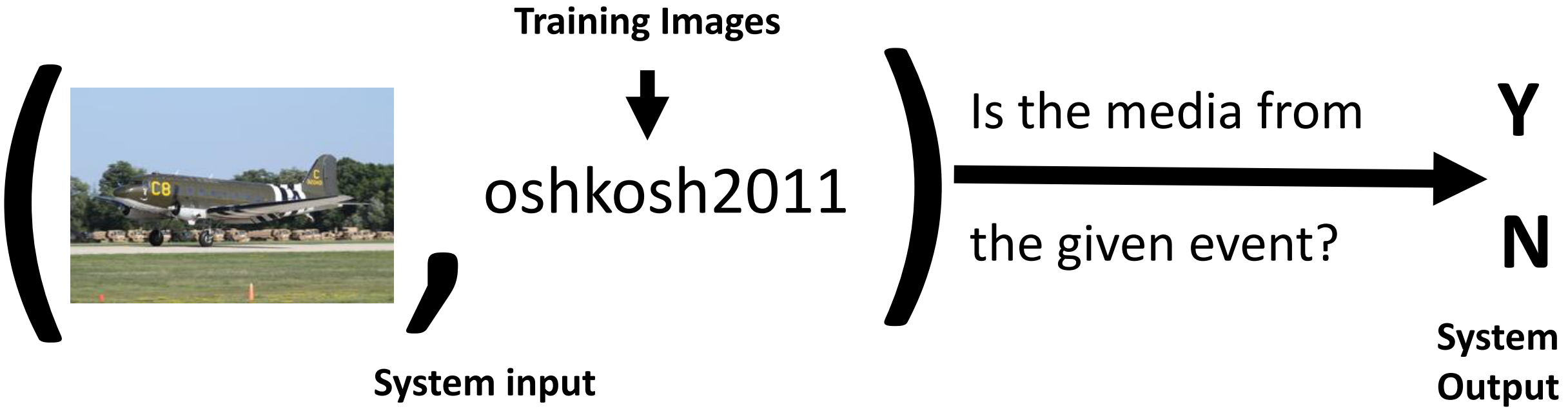
April 21-25, 2020

Event Outline

- Task definition
- Evaluation data
- Evaluation metrics
- MFC20 result

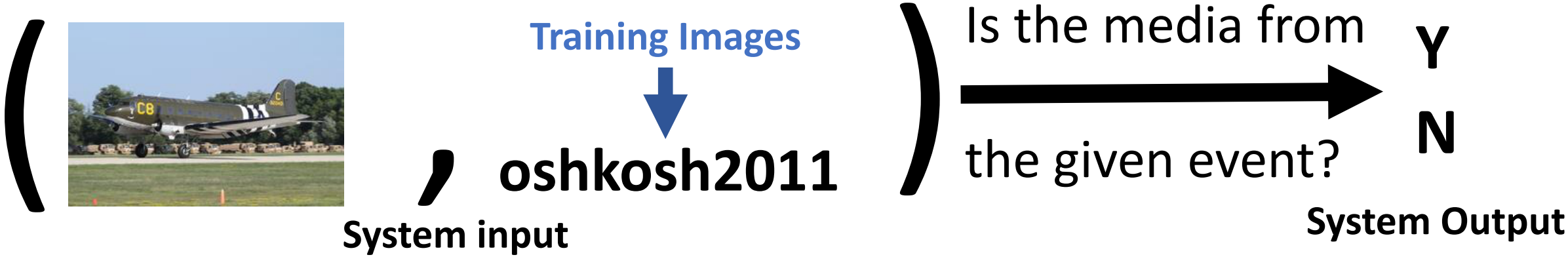
Event Verification Task

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



Event Verification Task

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



MFC20 Event Verification Dataset

- 12 Events
  - 6 hurricane, 3 air show, and 3 others

  hurricane_matthew, **hurricane_sandy**, hurricane_harvey, hurricane_katrina, hurricane_Irma, hurricane_ike, **oshkosh2011**, oshkosh2010, **berlin_air_show**, berlin_marathon, chinese_new_year_london_2014, chicago_blizzard_2011.
- Datasets
  - Training: about 200 per event
  - Testing: about 50 per event

oshkosh2011

oshkosh2010

hurricane_katrina

hurricane_ike

berlin_marathon

chicago_blizzard_2011

Event Verification System Performance

- Test Data
  - 12 Events
  - 2K training images
  - 574 test pairs
- 1 Team: Mayachitra
- Highest AUC = 0.909
  - CD@0.05FA = 0.533
  - System ID: RN50-sysOut-Dev_MFC_18_MFC_19

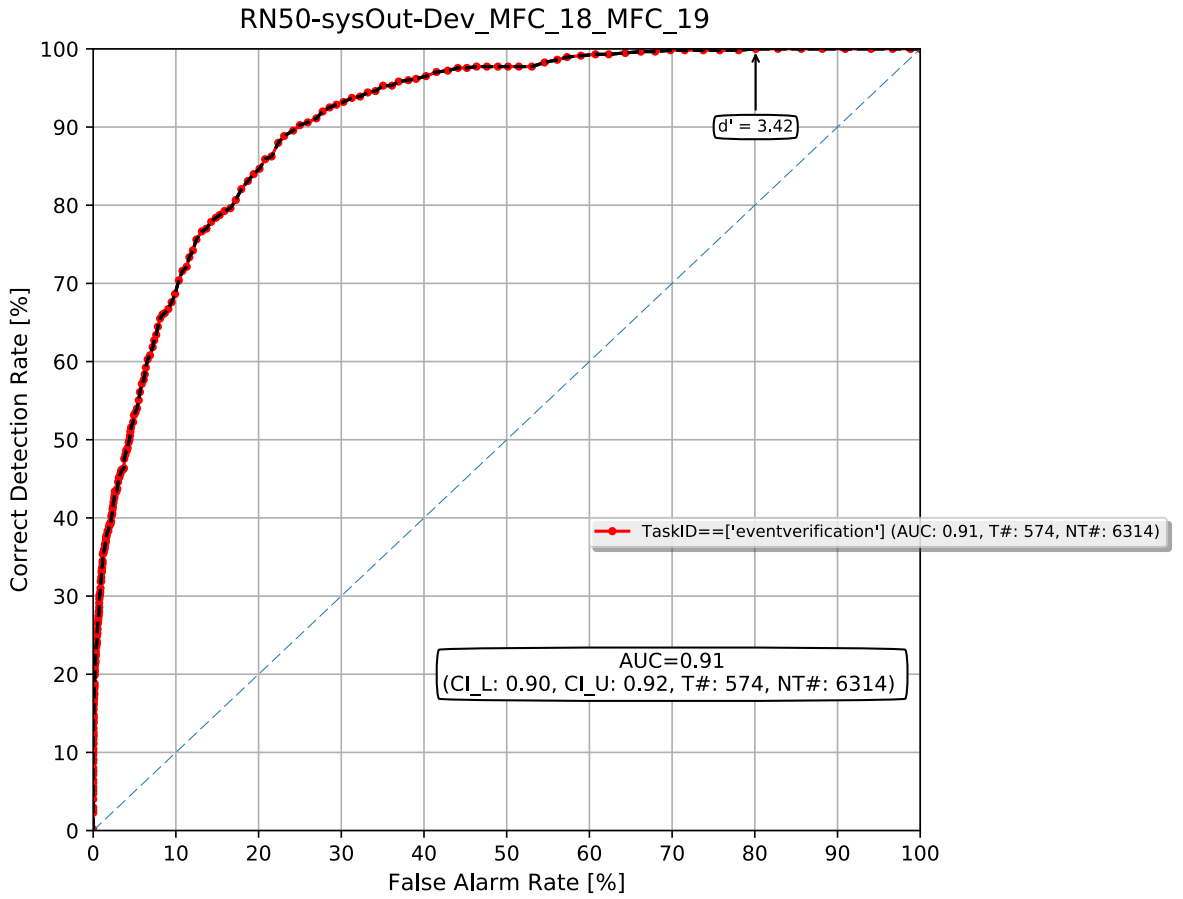
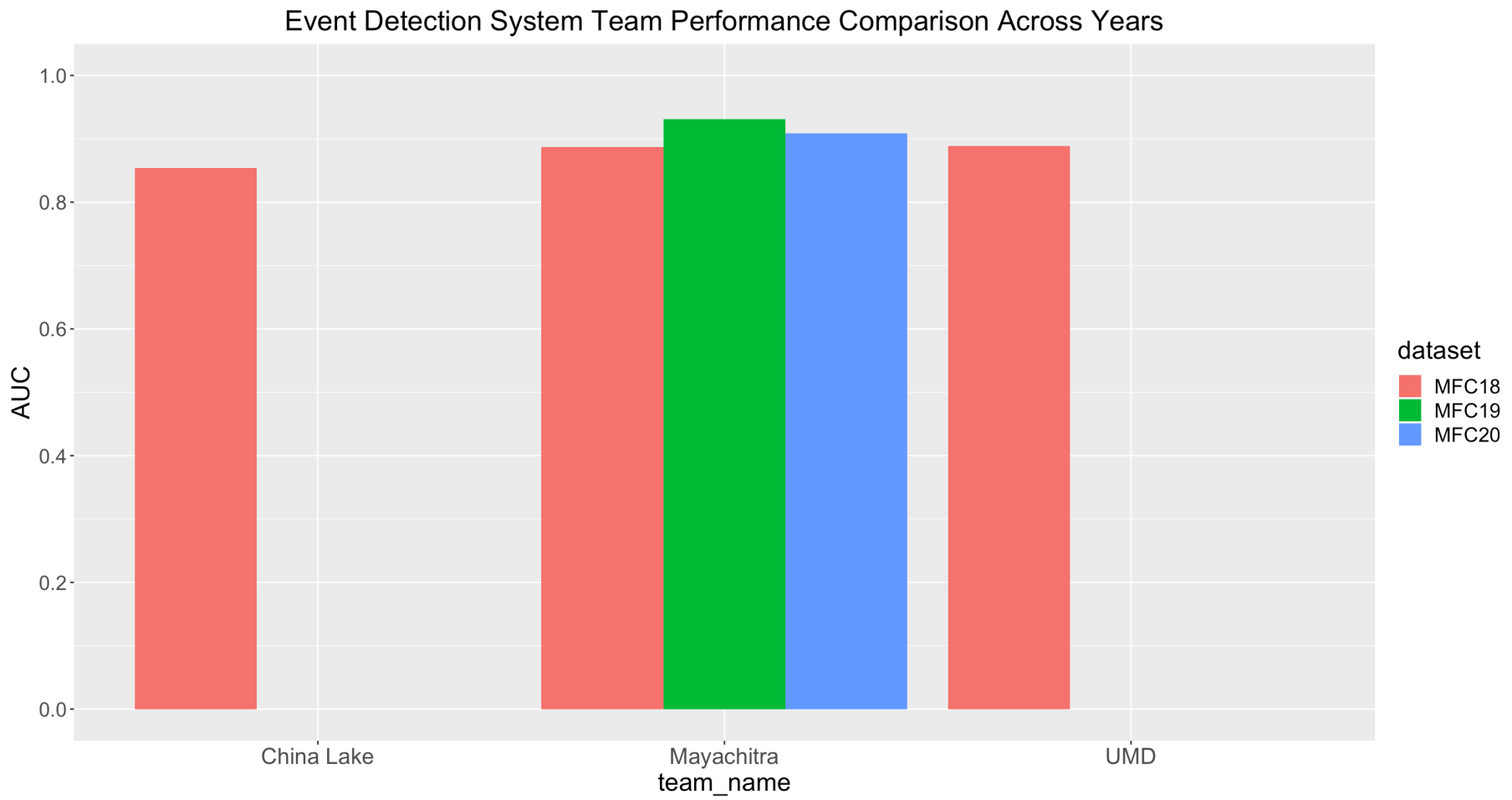
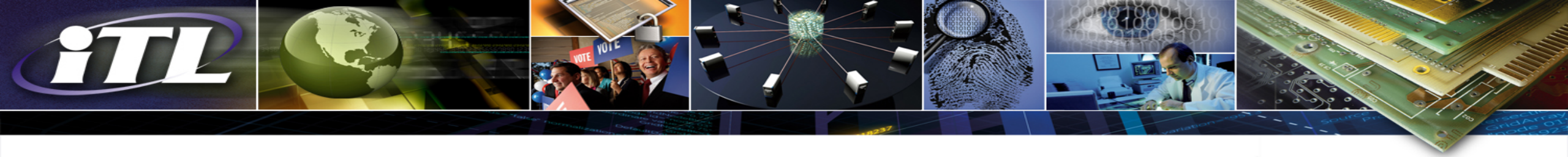


Figure: MFC20 EP1 Event ROC

Event Verification Detection System

- Team Performance Comparison Across Years





Thank You for Your Attention!

- NIST MediFor Team: medifor-nist@nist.gov
- MediFor Confluence: https://mediforprogram.com
- MediScore Git: https://gitlab.mediforprogram.com/jfiscus/MediScore
- MediBrowser: https://medifor.rankone.io/
- NIST MediFor Data: https://mig.nist.gov/MFC2019/Resources.html