NISTIR 8377

User Guide for NIST Media Forensic Challenge (MFC) Datasets

Haiying Guan
Andrew Delgado
Yooyoung Lee
Amy N. Yates
Daniel Zhou
Timothee Kheyrkhah
Jon Fiscus

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8377

National Institute of Standards and Technology
U.S. Department of Commerce

NISTIR 8377

User Guide for NIST Media Forensic Challenge (MFC) Datasets

Haiying Guan
Andrew Delgado
Yooyoung Lee
Multimodal Information Group
Information Technology Laboratory

Amy N. Yates
Image Group
Information Technology Laboratory

Daniel Zhou
Timothee Kheyrkhah
Jonathan Fiscus
Multimodal Information Group
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8377

July 2021

U.S. Department of Commerce
Gina M. Raimondo, Secretary

National Institute of Standards and Technology
James K. Olthoff, Performing the Non-Exclusive Functions and Duties of the Under Secretary of Commerce for Standards and Technology & Director, National Institute of Standards and Technology

National Institute of Standards and Technology Interagency or Internal Report 8377
Natl. Inst. Stand. Technol. Interag. Intern. Rep. 8377, 41 pages (July 2021)

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.8377

Abstract

More than 300 individuals from 150 organizations across 26 countries and regions use the NIST released Media Forensic Challenge (MFC) datasets for their research. The MFC datasets were created for use in the DARPA MediFor (Media Forensics) program. Since their release, multiple questions have been fielded regarding the dataset properties, including contents, metadata definitions, usage, data repurposing, etc. For example: what do the datasets contain? What are the definitions of the different kinds of metadata? How does one label the data with the reference information to build the training data for machine learning algorithms? How would one modify/extract the data for their own research purposes? This document serves as a user guide for the MFC datasets, including those used in the Nimble Challenge (NC). This guide includes: 1) a description about MFC datasets including background, evolution history, and the dataset summary by the evaluation tasks; 2) user access and permissions of MFC datasets; 3) an introduction to the MFC data by providing a simple example of a manipulation journal graph and its detailed corresponding MFC dataset reference files; 4) an introduction to a flexible subset selection approach, "Selective Scoring," to sample the test probes from the entire test set for the particular task evaluation; 5) information to help users gain a deeper understanding of the metadata by presenting two commonly used approaches to illustrate the manipulation operation statistic histogram distributions, and 6) a general template of the NIST MFC evaluation dataset to facilitate the future dataset generation.

Key words

Media Forensics, NIST Media Forensic Challenge (MFC) Evaluation, Journaling Tool (JT), Manipulation journal graph, Image Manipulation Detection, Image Manipulation Localization, Manipulation Localization Reference Mask, JPEG 2000, Manipulation Reference Ground-truth, Localization Mask, DARPA MediFor (Media Forensic) program.

Acknowledgments

The authors gratefully acknowledge the members of the DARPA Media Forensics (MediFor) Program[1], PAR Government[2], University of Colorado (UC) Denver[3], and Air Force Research Lab in the MediFor project. PAR Government conducted this work under DARPA sponsorship via Air Force Research Laboratory (AFRL) contract FA8750-16-C-0168. NIST conducted this work under NIST Interagency Agreement Number 1505-774-08-000, IRB Number ITL-0018-774-01.

Grateful thanks go to Matthew Turek, Neil Johnson, David Doermann, Rajiv Jain, Mark Kozak, Joe Austin, Eric Robertson, Jeff Smith, and Jeff Carlo for their instructions, strong supports, and contributions.

Special thanks go to Diane E. Ridgeway for her work on this report.

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.

[1] www.darpa.mil/program/media-forensics
[2] www.pargovernment.com
[3] artsandmedia.ucdenver.edu

Table of Contents

List of Tables

List of Figures

Glossary

Nimble Challenge (NC) – The name of NIST media forensic challenge kickoff dataset in 2016 and the challenge evaluation in 2017.

Media Forensic Challenge (MFC) – In 2018 the Nimble Challenge was renamed to the Media Forensic Challenge and became the evaluation series that supported the DARPA MediFor Program's performer evaluations from 2018-2020.

Open Media Forensic Challenge (OpenMFC) – The successor of the MFC media forensic evaluation series, supported by NIST and open to public participation.

Manipulation Journal –This is an automatically generated manipulation history graph log of media file manipulations with automatic output manipulation masks from a detector algorithm. Each journal tracks the media manipulations and software according to NIST manipulation data collection guidelines.

Journaling Tool (JT) – A software application developed by PAR Government that is used to create a graph representation of the image and video media manipulation steps performed by the manipulator. The Journaling Tool is a unified framework for data and metadata collection, annotation, and generation of automated manipulations designed according to data collection requirements. The journaling tool supports three major functions: (1) recording manual manipulations; (2) automatically generating the manipulation journals given a journal graph and input media files; (3) automatically extending the existing journals given manipulation operations with the parameters. The intent of journaling is to capture a detailed provenance graph for the media manipulations and to collect the data and metadata information from each manipulation step and manipulated media file to form the reference file for various media forensic tasks. Journaling Tool is publicly available as an open-source package on github (https://github.com/PAR-Government/media-journaling-tool). It is implemented in Python with a detailed user guide, refer to [1] for details.

Manipulation Journal Graph – A Directed Acyclic Graph (DAG) documenting both the media manipulation history and associated metadata.

Manipulation Operation – Standardized manipulation technique name of the operation used to generate the target image from the source image.

Manipulation Reference Ground Truth Mask – An image where each pixel indicates whether the associated pixel in the test media has been manipulated or not. If the media was manipulated, then the reference mask is a composite mask which aggregates all manipulations' masks along the path from a base image to the test media in the final node. The composite mask has the same dimension as the test image. Note: the reference mask in NC2017 is a composite mask in Portable Network Graphics (png) format, while the reference mask in MFC18 and later is represented in JPEG2000[4] format. Each channel of the JPEG2000 represents a manipulation operation mask aligned with the test image. The reference mask is aligned to the test media for uniformity over all operations including seam carving and cropping, for which the mask describes pixels removed.

[4] en.wikipedia.org/wiki/JPEG_2000

1. Introduction

"Seeing's not always believing."[5] In the past several years, the rapid growth and advancement in media generation and falsification techniques, such as the use of Generative Adversarial Networks (GAN) to create Deepfakes, has overturned an age-old saying. Combined with advanced technologies in computer graphics, media editing, and tampering technologies, it has become increasingly easy to create very realistic computer-generated images or tampered images which can skew public perception of fact. More and more, people will benefit by knowing the authentication information of a given media (image or video). Media forensics systems perform automated image/video manipulation detection, fact verification, and other related tasks as well, as construct the phylogeny graphs describing the manipulation history of images. The market demands[6] for automatic media forensic technologies coming from different application domains are soaring[7].

The NIST Media Forensics Challenge (MFC) supported the evaluation of the Defense Advanced Research Projects Agency (DARPA) Media Forensics (MediFor) Program[8] performers' systems 2017-2020. NIST's follow-on effort, OpenMFC [9][13], seeks to provide the datasets, build a research platform and inspire research worldwide with leaderboard evaluation, and advance the state-of-the-art of media forensics. MFC and OpenMFC provide the community with rich resources such as datasets [2], benchmark evaluations [10][11][12][13], and the latest system performance reports [3][4][5][6]. Both evaluation programs have received widespread attention. In response, NIST publicly released the datasets generated from the 2016 and 2017 Nimble Challenges and 2018 Media Forensic Challenge [2]. More than 300 individuals from 150 organizations in 26 countries and regions worldwide use these datasets for research. Several documents about the datasets [2], data collection tool [1], and evaluation plans [7][8][9] have been published.

This document provides a detailed description about:

(1) how MFC evaluation data and its references are structured in the datasets (Section 3 and Section 4).
(2) how to analyze the statistical features of a MFC dataset (Section 2 and Appendix E).
(3) how to use NIST dataset reference files to design and implement special evaluations (Section 5).
(4) how to repurpose the datasets to obtain the training data and the ground-truth labels from the metadata or features recorded in the reference files for the training of machine learning algorithms.

This guide is intended to cover currently released data, provide a baseline for future data releases, and encompass a variety of users' research and development needs, including:

(1) direct usage for NIST MFC or OpenMFC challenge evaluation defined by NIST evaluation plans [7][8][9].
(2) partial usage for specific tasks or new tasks or evaluations that are not directly defined by MFC challenges.
(3) indirect usage for machine learning algorithm modeling or other applications.

[5] https://www.thesunchronicle.com/tribune/seeing-s-not-always-believing/article_ce37deb8-09d6-5fe1-97bd-5a12ee26078a.html
[6] https://www.computer.org/publications/tech-news/research/social-media-verification-assistant
[7] https://www.fastcompany.com/40551971/how-darpa-is-fighting-deepfakes
[8] https://www.darpa.mil/program/media-forensics

The examples set forth in this guide draw upon non-sequestered, publicly releasable data from the media forensics challenges from 2016 through 2020. NIST MFC sequestered datasets from 2017 to 2020 contain similar data with the same structure, which continue to be utilized for evaluations and are not discussed herein.

2. Dataset Organization

2.1. Dataset description

To facilitate media forensic research and support the DARPA MediFor program, NIST collaborated with the DARPA MediFor data collection and manipulation teams (PAR Government[9] and the University of Colorado Denver[10]), to build a series of datasets for yearly benchmark evaluations:

- Nimble Challenge (NC, the former name of MFC) 2016 Kickoff dataset
- Nimble Challenge 2017 datasets
- MFC18 datasets
- MFC19 datasets
- MFC20 datasets

The NC16 Kickoff dataset was generated by NIST and had a simple structure and self-explainable reference files. This dataset was released to DARPA performers and the public after the MediFor 2016 Kickoff meeting.

NC17, MFC18, MFC19, and MFC20 datasets were designed by NIST and used for the DARPA MediFor year-round evaluations. Since 2017, PAR Government and UC Denver (also called the MediFor data collection and manipulation teams) collected the imagery (images and videos) and performed the manipulations. NIST used this data collection to generate the datasets and administer benchmark evaluations of media forensics technologies. The evaluations were designed to substantiate the integrity of a media object's representation, content, and provenance. NIST published an evaluation plan [7][8][9] that defines evaluation tasks, rules for participation, protocols for implementation, metric approaches, and analysis procedures. The MFC evaluation participants included the researchers participating in the DARPA MediFor program and a small number of external researchers interested in media forensic.

In general, the datasets have a similar structure and include the following items:

- Original high-provenance images or videos
- Manipulated images or videos
- The reference ground-truth information for detection, localization, verification, and provenance

This user guide focuses on the structure, data and metadata descriptions, and the usage of the MFC datasets.

Since NC17, two groups of datasets were provided for each year's evaluation: *open datasets*, which included development datasets and Evaluation Part 1 (EP1) datasets, and *sequestered datasets*, including EP2 and EP3 datasets.

Open datasets have been released to the public and can be used for take-home evaluations. Sequestered datasets were used for container evaluations and remained to reserve for future research. This user guide is limited to the open datasets.

[9] www.pargovernment.com
[10] artsandmedia.ucdenver.edu

2.2. NIST MFC dataset creation protocol

One of the key evaluation dataset design philosophies was to separate the performance teams (who focus on developing manipulation detection and localization systems) from the dataset generation teams (who focus on the data collection and manipulation). The data generation teams are top-level experts and artists experienced in image and video editing and manipulation. They use different image manipulation software (such as Adobe Photoshop, GIMP, etc.) to generate state-of-the-art manipulations that cover all commonly used manipulation operation types and mimic real-world manipulations. Media forensic algorithm developers advise the dataset generation team, but do not provide data for the evaluations or directly participate in its design or manipulation. In this way, the MFC performers are kept from creating evaluation datasets to avoid problems such as oversimplifying the complexity of the real-world manipulations, underfitting the algorithm by only focusing on a certain manipulation operation, overfitting the algorithm by knowing the testing data. The goal is to motivate MFC performers to concentrate on real-world forensic applications and deliver a functional, effective, and operational media forensic system prototype.

2.3. NIST MFC dataset summary

Figure 1 shows the evolution of NIST evaluation datasets over the past five years. Each manipulated image in the Nimble Challenge 2016 kickoff dataset, collected and manipulated by NIST, only contained one of the following major single manipulation operations: *Clone*, *Remove*, and *Splice*. The NC2017 dataset was the first to use the manipulation Journaling Tool (JT) to collect metadata and record the steps of the manipulation ground-truth. The Automatic Journaling Tool (AutoJT) automatically created journals and generated manipulated images to reduce the human workload. MFC18 used more manipulation journals and contained more operation types to build the evaluation datasets. The Extended Journaling Tool (ExtendJT) was used to automatically extend the existing human journals to generate additional manipulated images for MFC18. MFC19 used a more extensive set of journals and contained more advanced manipulations such as GAN and green-screen manipulation. MFC20 covers auto journals and extended journals generated for special studies such as compression and social media laundering.

Table 1 shows the overall summary of each NIST open evaluation dataset from 2017 to 2020. The open evaluation dataset, also known as the "Evaluation Part 1 (EP1)" dataset, was released to the DARPA MediFor performers to support each evaluation cycle. NC16 Kickoff, NC17 EP1, and MFC18 EP1 are publicly available at the time of this writing.

Table 1. NIST MFC Evaluation datasets: Part 1 summaries.

Evaluation Task	NC17 Open Eval.	MFC18 Open Eval.	MFC19 Open Eval.	MFC20 Open Eval.
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

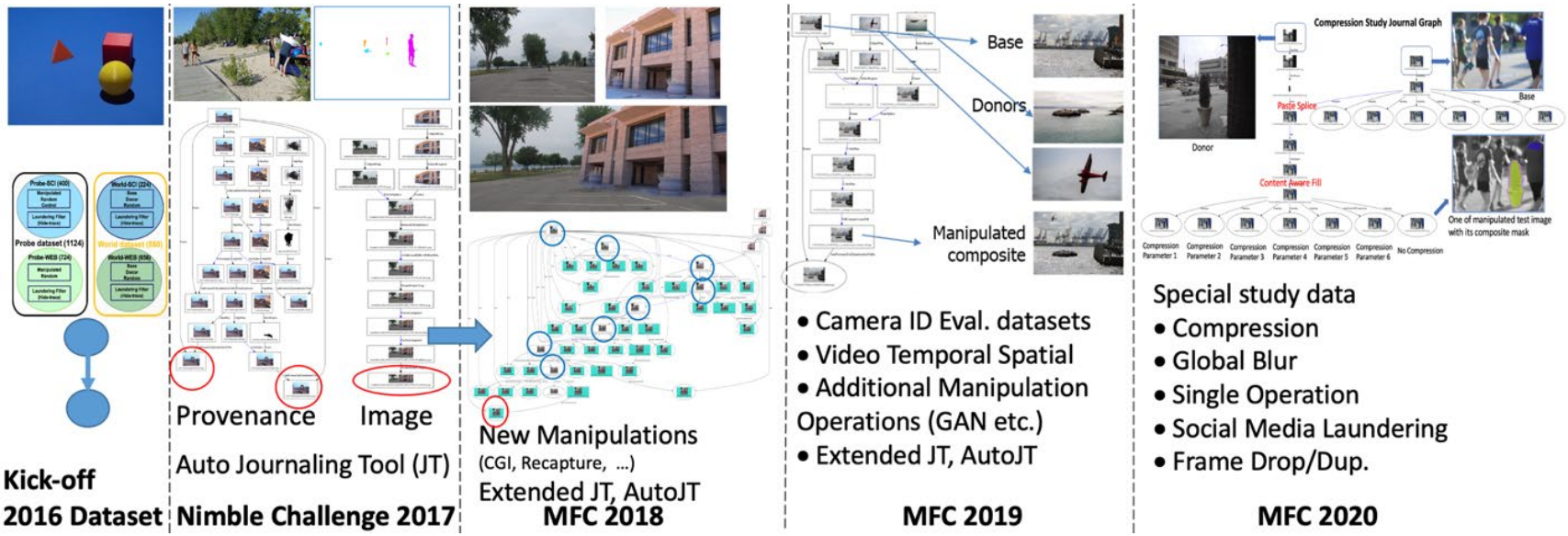


Figure 1. NIST evaluation dataset evolution.

2.4. NIST MFC releasable datasets categorized by the evaluation tasks

The MFC evaluations designed evaluation tasks to measure the performance of different media forensic systems. Each task has at least one open dataset (EP1) for each year. Some tasks have multiple development datasets that were used to collect feedback for future evaluation dataset design. Please refer to Appendix A: Challenge Task Definitions or the evaluation plans [7][8] for the task definitions.

The following tables identify datasets and content created for each evaluation task. Table 2 shows the twelve releasable development and evaluation datasets for the image manipulation detection and localization task. Table 3 shows the eight releasable development and evaluation datasets for the video manipulation detection and localization task.

Table 2. NIST MFC Image Manipulation Detection and Localization releasable datasets.

IMDL datasets	Dev./Eval.	Image Probe (K)	Image Journal	Date	Availability
Kick Off (NC16)	Dev.	1.1	-	07/2016	Y
NC17 Dev	Dev.	3.5	394	03/2017	Y
MFC18 Dev1	Dev.	5.6	117	01/2018	Y
MFC18 Dev2	Dev.	38	432	02/2018	Y
MFC19 Dev1	Dev.	3.6	56	02/2019	N
MFC GlobalBlur	Dev.	5.8	235	09/2019	N
MFC Compression	Dev.	17	75	11/2019	N
NC17 EP1	Eval.	4	406	06/2017	Y
MFC18 EP1	Eval.	17	758	03/2018	N
MFC18 GAN FULL	Eval.	1.3	267	04/2018	Y
MFC19 EP1	Eval.	16	1383	03/2019	Y
MFC20 EP1	Eval.	20	2536	03/2020	N

Table 3. NIST MFC Video Manipulation Detection and Localization releasable datasets.

VMDL datasets	Dev./Eval.	Video Probe	Video Journal	Date	Availability
NC17 Dev	Dev.	212	23	03/2017	Y
MFC18 Dev1	Dev.	116	8	01/2018	Y
MFC18 Dev2	Dev.	231	36	02/2018	Y
NC17 EP1	Eval.	360	45	06/2017	Y
MFC18 EP1	Eval.	1028	113	03/2018	Y
MFC18 GAN	Eval.	118	-	06/2018	Y
MFC19 EP1	Eval.	1530	163	03/2019	Y
MFC20 EP1	Eval.	1421	217	03/2020	N

Table 4 shows the four releasable evaluation datasets for the splice manipulation detection and localization task. Table 5 shows the eighteen releasable development and evaluation datasets for the camera verification task. The MFC18, MFC19, and MFC20 evaluations have six subsets for six subtasks each year. The datasets were built around June in 2018, 2019, and 2020 respectively. The number of probe pairs, cameras, and journals for each subset is shown in the table. In total, there are eighteen releasable datasets for the camera verification task.

Because the Provenance Filtering (PF) and Provenance Graph Building (PGB) tasks share the same set of probe images each year, the probe count and journal count values for both PF and PGB datasets are the same every year. PF and PGB datasets share the same statistics as shown in Table 6.

All datasets in Table 4, Table 5, and Table 6 are not available for download yet. However,

they are releasable for new tasks in the future OpenMFC evaluation.

Table 4. NIST MFC Splice Manipulation Detection and Localization releasable datasets.

SMDL datasets	Dev./Eval.	Probe Pair (K)	Image Journal	Date
Kick Off (NC16)	Dev.	89.6	-	07/2016
NC17 EP1	Eval.	330	156	06/2017
MFC18 EP1	Eval.	18	381	03/2018
MFC19 EP1	Eval.	18	621	03/2019
MFC20 EP1	Eval.	18	1266	03/2020

Table 5. NIST MFC Camera Verification releasable datasets.

CV datasets		MFC18			MFC19			MFC20		
Test	Train	Probe	Cam.	Jour.	Probe	Cam.	Jour.	Probe	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

Table 6. Provenance Filtering (PF)/Provenance Graph Building (PGB) releasable datasets.

PF/PGB datasets	Dev./Eval.	Probe	World Images	Image Journal	Date
NC17 EP1	Eval.	1K	1M	406	06/2017
MFC18 EP1	Eval.	10K	1M	641	03/2018
MFC19 EP1	Eval.	9420	2M	1025	03/2019
MFC20 EP1	Eval.	5927	2M	1572	03/2020

3. User Access, Use, and Permissions

3.1. How to obtain the data

The MFC datasets are available by signing up for an account on the OpenMFC evaluation website (https://mfc.nist.gov/). After signup, the user will get access to the NC16 Kickoff dataset. The NC17 and MFC18 datasets are available to OpenMFC participants only. If the user wants to join the OpenMFC evaluation, they need to sign up and upload a signed Data Use Agreement (Appendix B: MediFor Data Use Agreement) and Evaluation Participation Agreement (Appendix C: Evaluation Participation Agreement). The user would then get the access credentials to the NC17 and MFC18 image and video datasets through the web interface. The OpenMFC 2020-2021 participants will also get the following datasets without the reference ground-truth during the OpenMFC 2020-2021 evaluation process:

- MFC18 GAN image dataset
- MFC18 GAN video dataset
- MFC19 image dataset
- MFC19 video dataset

3.2. Operating system requirements

In general, there are no special operating system requirements for executable programs. The user may use any operating system to download the dataset through the password-protected web interface following the instructions described on the download page.

3.3. Specification of applications programs

There is no special specification of application programs required to access and use any of the files associated with the data. The user can use any image visualization tool or video player to visualize the image and video data. The user can use any text editing tool to view the index and reference files associated with the datasets.

3.4. Statement of inputs required from the user

The NC17, MFC18, and MFC19 datasets are available to OpenMFC participants only. The participants are required to sign the dataset agreement and fulfill the requirements described in the Evaluation Participation Agreement.

3.5. User agreement process

After the user has signed up and uploaded a signed Data Use Agreement and Evaluation Participation Agreement, the NIST team will review and approve them if the correct information was provided. The user will receive an email notification regarding the approval of the agreements. The user can then log in to their OpenMFC web account and obtain the access credentials for the NC17 and MFC18 image and video datasets through the web interface.

4. NIST MFC Dataset Structure and Files

This section explains how the manipulation data and metadata collected from the manipulation journal graph are embedded in the reference files of the NIST MFC dataset. By understanding the design and structure of the evaluation datasets, researchers can effectively use the MFC datasets in their media forensic evaluations. Moreover, researchers can also use the data or metadata extracted from the MFC datasets for training and modeling purposes with machine learning approaches.

4.1. Media Manipulation Journal: An Example

Before describing the dataset contents, we provide a quick introduction to the image collection process and the manipulation metadata annotation method.

(a) Original image (b) Manipulated image

Figure 2. An example of a manipulated image.

Figure 2 shows two images; the left-side image is an original, pristine image or "Base Node"; the right-side image is the manipulated image. The major manipulations are cloning the umbrella and splicing a polar bear.

A Directed Acyclic Graph (DAG) is used to record the manipulation history, which we call a journal [1]. A node in a DAG represents a media file instance such as an image, video, or audio file. An edge in a DAG, referred to as a link in this report, represents an operation that altered the source node's media to produce the destination node's media. In the general sense, the link represents a function that consumes the source and produces the destination. All metadata associated with the function is maintained with the link, including additional parameters, semantic information, and change analysis. However, it is more accurate to generalize the link as a dependency between source and destination, such that the destination depends directly on the state of the source. The DAG forms a dependency tree and, by nature of its construction, records the sequence of operations used to produce manipulated media from non-manipulated media. Figure 3 shows the manipulation journal of the manipulated image in Figure 2. Two masks for localization are generated by the manipulation journaling tool [2] for the splice of the polar bear and the clone of the umbrella with their own bit plane value, expressed in the two individually colored masks in Figure 7.

There are four types of nodes in a journal: base, donor, final and interim. A base node represents the primary media (original) being unaltered, whereas the donor represents media contributing to an alteration of the base. Base and donor nodes do not have predecessors. Base nodes represent non-manipulated (i.e., high provenance) media that is camera original without any processing after capture. Donor nodes are images with un-specified provenance. Final nodes do not have successors; each final node represents a final product of a sequence of manipulations. All other interim nodes record the state of the media produced by a single manipulation.

Links form the dependencies between each manipulation state of the media. There are two kinds of links: operation and donor. An operation link represents an operation performed on a source node media file to produce a manipulated result. A donor link represents the donation of one media to the alteration of another, such as a paste type operation would require. Although the donor is conceptually a parameter to the paste operation, the link forms the necessary dependency.

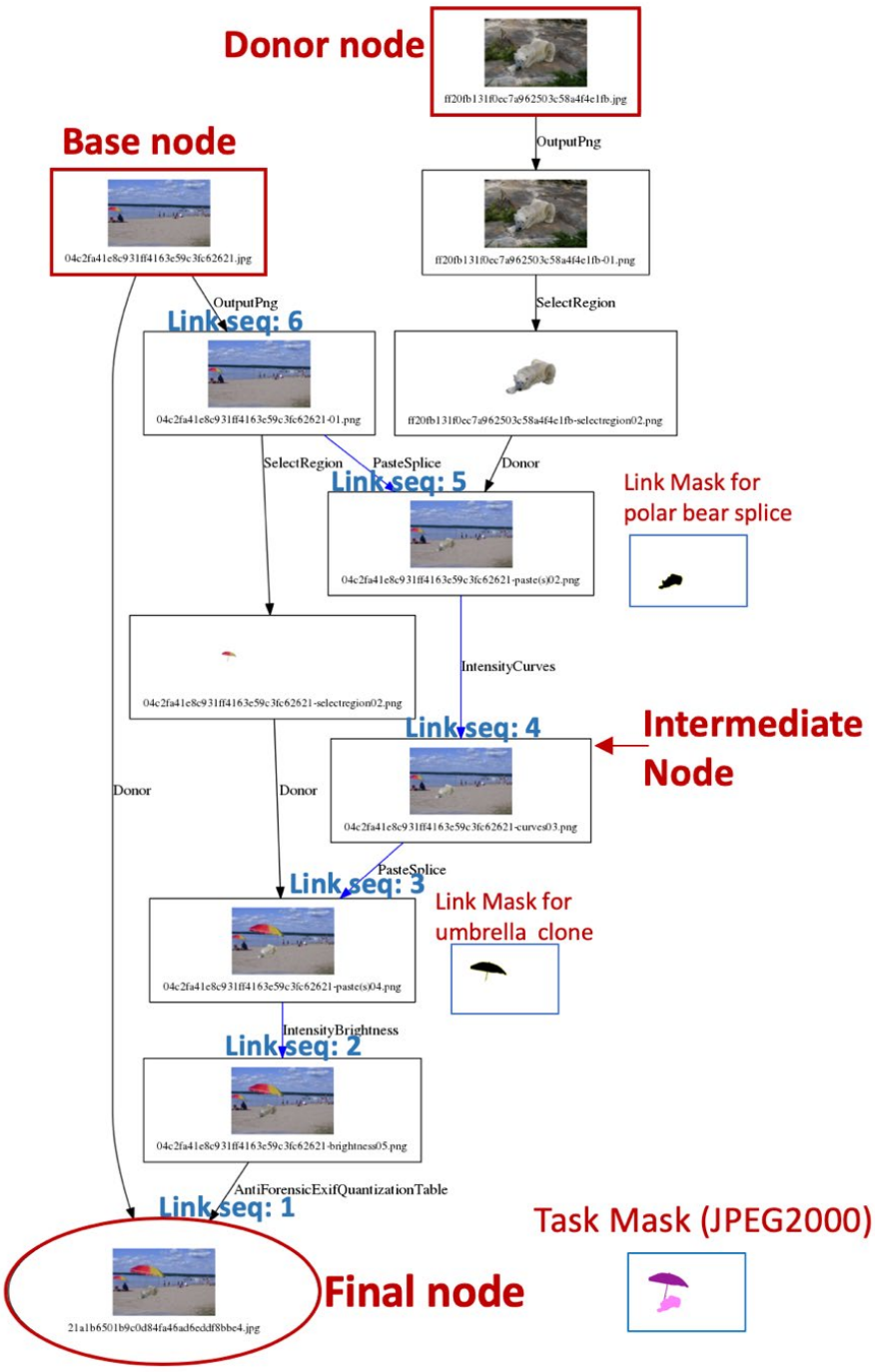


Figure 3. A representative example of an image manipulation journal.

The manipulation reference ground truth mask is an image where each pixel indicates whether the associated pixel in the manipulation media has been manipulated or not. In this example, the final manipulation was done by a series of manipulations, and the two major manipulations were to 'PasteSplice'[11] a polar bear ('Link seq 5' in Figure 3) and 'PasteSplice'

[11] String values in the reference files are surrounded by single quotes ('') in this document. 'PasteSplice' is a manipulation operation defined in *Operation* column in the journal reference file described in Section 4.5.

(or clone) an umbrella ('Link seq 3' in Figure 3). The reference mask is a composite mask which aggregates those two manipulations' masks along the path from a base image to the given test media in the final node (bottom right in Figure 2), which has the same dimension as the final image.

The reference mask in MFC18 and later is represented in JPEG2000 format. Each channel of the JPEG2000 represents a specific operation's local manipulation operation mask aligned with the test image. Given the JPEG2000 reference ground-truth mask and a special operation, it is easy to separate the local masks of the given operation from the rest of other operations, which allows a specific evaluation of this particular operation in the image localization task[7][8].

4.2. Dataset directory structure

The NIST Media Forensic Challenge Evaluation Plan [7][8] Section 3.6 describes the dataset directory structure. We include the directory structure in Figure 4 for the sake of content completion and readers' convenience. README templates of the NIST development and evaluation dataset are shown in Appendix D: The Basic NIST MFC Dataset ReadMe File Example.

Each MFC evaluation dataset contains two data packages: the probe package and the reference package. The probe package contains index files, probe files, world files (for provenance tasks only), and README files. The probe package, marked in black (dark) in Figure 4, contains the data directory and files released to performers for testing detection systems. This package contains testing data only without reference ground truth. The reference package, marked in green (light) containing the complete set with reference ground-truth data, is sequestered during the evaluation process, and is used to evaluate and score given the system responses.

The dataset (<DataSet_Name>) naming convention (e.g., MFC18_EvalPart1_Image) specifies the evaluation program name (MFC or OpenMFC), evaluation year, development or evaluation subset with version number (Dev1 or EvalPart1 etc.) and evaluation task name (Image, Splice, or Video etc.).

The structure described in this document is applicable across all MFC datasets. Special datasets served for the special evaluation task also utilize similar directories and file structures that follow the same convention.

```
<BaseDir/DataSet_Name>
        README.txt
        /probe
                {ImageFileName1}.jpg
                {ImageFileName2}.tif
                …
                {VideoFileName1}.avi
                {VideoFileName2}.gif
                …
        /world
                {ImageFileName1}.bmp
                {ImageFileName2}.png
                …
                {VideoFileName1}.mpg
                {VideoFileName2}.wmv
                …
        /documents
        /indexes
                MFC2018-manipulation-image-index.csv
                MFC2018-manipulation-video-index.csv
                MFC2018-splice-index.csv
                MFC2018-provenancefiltering-index.csv
                MFC2018-provenance-index.csv
        /reference
                /manipulation-image
                        MFC2018-manipulation-image-ref.csv
                        MFC2018-manipulation-image-ref-journalmask.csv
                        MFC2018-manipulation-image-ref-probejournaljoin.csv
                        /mask
                                {ImageFileName1}.png
                                {ImageFileName2}.png
                                …
                /manipulation-video
                        MFC2018-manipulation-video-ref.csv
                        MFC2018-manipulation-video-ref-journalmask.csv
                        MFC2018-manipulation-video-ref-probejournaljoin.csv
                /splice
                        MFC2018-splice-ref.csv
                        MFC2018-splice-ref-journalmask.csv
                        MFC2018-splice-ref-probejournaljoin.csv
                        /mask
                                {ImageFileName1}.png
                                {ImageFileName2}.png
                                …
                /provenancefiltering
                        MFC2018-provenancefiltering-ref.csv
                        MFC2018-provenancefilgering-ref-node.csv
                /provenance
                        MFC2018-provenance-ref.csv
                        MFC2018-provenance-ref-node.csv
```

Figure 4. The MFC dataset directory structure and files.

4.3. Index file

For a given task, the index file defines a system input. The index file naming convention is <DataSet_Name>-<TaskID>-index.csv (e.g., NC2017_EvalPart1-manipulation-image-index.csv). It can be found in the 'indexes' subdirectory. Given an index file, each row specifies a test trial, that is, an image or a video. The test trial is called a 'probe,' which is the subject of the task question prompted to the performer system tested. Taking the corresponding image or video from the 'probe' directory as input, systems perform detection. Section 4.1 in NIST Media Forensic Challenge Evaluation Plan [7][8] defines the index file for the manipulation detection task. The index file contains the following essential columns[12]: *TaskID, ProbeFileID, ProbeFileName, ProbeWidth, ProbeHeight,* and *ProbeFileSize*. Table 7 shows the index file (NC2017_EvalPart1-manipulation-image-index.csv) of the probe example in Figure 2 and Figure 3. *ProbeFileSize,* which is the size of the probe file for input validation, was added to the index columns after NC17, thus it was not included in Table 7.

Table 7. An example of an image index file.

TaskID	ProbeFileID	ProbeFileName	ProbeWidth	ProbeHeight
manipulation	21a1b6501b9c0d84fa46ad6eddf8bbe4	probe/21a1b6501b9c0d84fa46ad6eddf8bbe4.jpg	4928	3264

The index files are CSV formatted files, and the columns are pipe-separated. Two additional columns were added in MFC19 and MFC20 index files: *ProbeFileSize, HPDeviceID*. *ProbeFileSize*, which helps validate the file size. *HPDeviceID* specifies the media source device for camera verification tasks. For the image index file, another additional column, *HPSensorID,* is added to specify the camera sensor ID. For example, for a single device like iPhone (e.g., *HPDeviceID*: 'PAR-X'), it has two cameras: the back camera is the primary camera sensor (e.g., *HPSensorID*: 'PAR-X_primary') and the front camera is secondary camera sensor (e.g., *HPSensorID*: 'PAR-X_secondary'). If the probe image is taken with front camera of the given iPhone, the probe's *HPDeviceID* value is 'PAR-X', and *HPSensorID* is 'PAR-X_secondary'. In MFC evaluation datasets, most of the *HPSensorID* are primary sensors. The header and a probe image index are shown in Figure 5.

```
TaskID|ProbeFileID|ProbeFileName|ProbeWidth|ProbeHeight|ProbeFileSize|HPDeviceID|HPSensorID
manipulation|0018e46f5cc0a0fe50523b636716f474|probe/0018e46f5cc0a0fe50523b636716f474.jpg|3264|2176|3763513344|MK-S860|MK-S860_primary
```

Figure 5. An example of an updated image index file in MFC19.

For the video index file, *FrameCount* and *FrameRate* are added to verify the video content information. The example of a video index is shown in Figure 6.

```
TaskID|ProbeFileID|ProbeFileName|ProbeWidth|ProbeHeight|ProbeFileSize|HPDeviceID|FrameCount|FrameRate
manipulation|001bd1016363c47079a6165535ae7145|probe/001bd1016363c47079a6165535ae7145.mp4|1280|720|31000||645|29.61320
```

Figure 6. An example of a video index file.

4.4. Probe reference file

The probe reference file defines the metadata that supports evaluation for each test trail defined in the index file. Section 3.5.1 in NIST Media Forensic Challenge Evaluation Plan [7][8] defines the probe reference file for the manipulation detection task. The probe reference

[12] In this document, the column names in the MFC index or reference files are in *italic* font.

files name convention is "DataSetID-manipulation-image-ref.csv." The metadata defined in the probe reference file can be classified into two categories: required and optional. The required category includes: *TaskID, ProbeFileID, ProbeFileName, IsTarget, ProbeMaskFileName, ProbeBrowserFileName, BaseFileName, BaseBrowserFileName, JournalName*, etc. They are required in the metadata collection for two major purposes: first, some of the reference metadata columns are ground-truth references for evaluation scoring software to report the system performance. For example, If *IsTarget* is 'Y,' then the testing media is manipulated. *ProbeMaskFileName* is the manipulation image's reference mask image filename, which stores the manipulated regions generated by different operations. The detailed descriptions are given in Section 4.7. Second, other required reference metadata columns record the manipulation-related metadata. For example, *ProbeFileName* defines as a probe image for testing trials. *BaseFileName* defines as an original image, which is the base node image in the manipulation journal graph, as shown in Figure 3.

Table 8 shows the first category columns in the reference file (NC2017_EvalPart1-manipulation-image-ref.csv) of the manipulated probe image in Figure 2 and Figure 3.

Table 8. An example of the required reference columns in the probe reference file.

TaskID	manipulation
ProbeFileID	21a1b6501b9c0d84fa46ad6eddf8bbe4
ProbeFileName	probe/21a1b6501b9c0d84fa46ad6eddf8bbe4.jpg
IsTarget	Y
ProbeMaskFileName	reference/manipulation-image/mask/fff35454245fd5b890547a84cdb1bad3.ccm.png
ProbeBrowserFileName	21a1b6501b9c0d84fa46ad6eddf8bbe4.jpg
BaseFileName	world/04c2fa41e8c931ff4163e59c3fc62621.jpg
BaseBrowserFileName	04c2fa41e8c931ff4163e59c3fc62621.jpg
JournalName	04c2fa41e8c931ff4163e59c3fc62621
ProjectDescription	added a polar bear chilling under a beach umbrella
ProjectType	image

The second category regards the probe feature, and it shows if the probe contains a certain operation or has a certain property. They collect information about the major manipulations applied to the probe, which can be used to support factor-based system performance analysis. For example, the *FaceManipulations* column indicates if the given probe contains any face manipulations or not. *SeamCarving* column indicates if any seam carving techniques process the given probe or not. *AudioSplice* indicates if the given video probe contains audio splice manipulation or not. The reference data can also adapt to newly designed tasks or questions to answer special study questions. For example, if the evaluation task detects all manipulated media, we use the *IsTarget* column as a ground-truth for evaluation package to report system performance. *IsTarget* is the reference (ground truth) column indicating binary classification for the manipulation task. The value 'Y' of *IsTarget* means the media is manipulated. If the evaluation task is to detect if the media is manipulated by Generative Adversarial Network (GAN) technologies. *VideoTaskProbeDesignation* is added after MFC18, which is used to determine which evaluation tasks (three evaluation tasks: video manipulation detection, video manipulation temporal localization, and video manipulation spatial localization) are suitable

for the given operation link. *IsGAN* is a metadata column describing such a type of manipulation. Thus, *IsTarget* == 'Y' and *IsGAN* == 'Y' means at least one of the manipulations is performed with a GAN. *IsTarget* == 'Y' and *IsGAN* == 'N' means no manipulation used GAN technology.

In MFC20 image datasets, we also use an *OMIT* column to define if the probe is included in a special evaluation subset or not. Its values are 'Y' or 'N'. That is, if we want to evaluate the system performance on a subset of testing media instead of the whole evaluation test set, we can exclude the probes (*OMIT* == 'Y') with certain conditions from the testing set without rebuilding the whole test set. For example, to test the compression effects on the detection system in MFC18 evaluation datasets, we save it in a different image format with different compression parameters for the same original or manipulated image. It creates multiple "replicate" trials in the test set. If we prefer to put only one trail for the evaluation, we only assign the value of the desired test probe's *OMIT* column to 'N' (*OMIT* == 'N'). In addition, we assign the values of its other "replicate" *OMIT* column to 'Y' (*OMIT* == 'Y'), which means we excluded its "replicate" probes from the testing set for the evaluation.

4.5. Journal reference file

The journal reference file naming convention is in "DataSetID-manipulation-image-ref-journalmask.csv" format. Unlike the index file or the probe reference file, the row in the journal reference file defines a link in the manipulation journal graph. The journal graph link defines the manipulation operation properties and features. The journal reference file contains the following columns: *JournalName*, *StartNodeID*, *EndNodeID*, *Operation*, *Color*, *Purpose*, and *OperationArgument*. *JournalName*, *StartNodeID*, and *EndNodeID* together specify a unique link in a journal in the dataset. For example, the 'IntensityCurves' link (the label 'Link seq 2' in Figure 3) in the journal graph example in Figure 3 is defined by this link's StartNodeID (value: 04c2fa41e8c931ff4163e59c3fc62621-paste(s)02) and this link's EndNodeID (value: 04c2fa41e8c931ff4163e59c3fc62621-curves03). The link is described by *Operation* (value: 'IntensityCurves,' that is, the manipulation type), *Color* (value: (102, 255, 0), that is, the color was assigned to a link in a journal, which is used to color the mask region generated by the link's operation), *Purpose* (value: 'None,' it is a semantic concept description in some cases, e.g., if the *Operation* is 'add' or 'remove'). All links of a journal define the full journal graph. Table 9 shows the journal reference file (NC2017_EvalPart1-manipulation-image-ref-journalmask.csv) of the previous example in Figure 3. The second data row in Table 9 defines the 'IntensityCurves' link in the journal graph example in Figure 3. Similarly, the first data row in Table 9 defines the 'PasteSplice' operation, which is to paste a bigger umbrella to the image, its operation region mask color (0, 255, 204) is teal color in the reference mask file (reference/manipulation-image/mask/fff35454245fd5b8 90547a84cdb1bad3 .ccm.png) of this probe (Figure 7). The fifth data row in Table 9 defines another 'PasteSplice' operation, which is to paste a polar bear to the image. Its operation region mask color, (255, 102, 0), is orange color. Each row defined in the journal reference file is a unique link of a journal without duplications.

Since MFC19, the following new columns are added in the video journal reference file to support video temporal and spatial detection localization tasks: *VideoTime*, *VideoFrame*, *AudioTime*, *AudioFrame*, *FrameTimeAdjustment*, and *VideoTaskDesignation*. *VideoTime* represents where the manipulation took place in the video channel. *VideoFrame* represents where the manipulation took place using the frame representation. *AudioTime* represents where the manipulation took place in the audio channel. *AudioFrame* represents where the manipulation took place using the frame representation. In some cases, the video frames are

slightly off from the audio time when the video was saved. *FrameTimeAdjustment* is used to adjust the frame and time inconsistency. *VideoTaskDesignation* is used to indicate the availability of the spatial mask of the video in the spatial localization task or the availability of the audio/video segment in the temporal localization task. In the Journaling Tool (JT), each link has a metadata describing collected data, and each operation has a descriptor indicating if a spatial or temporal reference data is collected effectively. Each test probe in the index file has the *VideoTaskDesignation* value to indicate if the final reference data is valid. In some cases, even with the intent of an operation to create spatial mask, overall manipulations along all the paths from the given probe to the base video may fail to produce an acceptable final reference data due to subsequent manipulations. For spatial masks, there are two forms of validity: automatic generated reference data's availability and human inspection, where human inspection is the result of manual validation. The major purpose of *VideoTaskDesignation* is to indicate if the test probe can be used for the spatial localization task or the temporal localization task.

Table 9. The journal reference file of the journal graph in Figure 3.

	index	JournalName	StartNodeID	EndNodeID	Operation	Color	Purpose	OperationArgument
0	18629	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-curves03	04c2fa41e8c931ff4163e59c3fc62621-paste(s)04	PasteSplice	0 255 204	add	man-made object
1	18630	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-paste(s)02	04c2fa41e8c931ff4163e59c3fc62621-curves03	IntensityCurves	102 255 0	None	None
2	18631	04c2fa41e8c931ff4163e59c3fc62621	ff20fb131f0ec7a962503c58a4f4e1fb-selectregion02	04c2fa41e8c931ff4163e59c3fc62621-paste(s)02	Donor	None	None	None
3	18632	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-01	04c2fa41e8c931ff4163e59c3fc62621-selectregion02	SelectRegion	None	None	man-made object
4	18633	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-01	04c2fa41e8c931ff4163e59c3fc62621-paste(s)02	PasteSplice	255 102 0	add	other
5	18634	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-01	OutputPng	None	None	None
6	18635	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-brightness05_01	Donor	None	None	None
7	18636	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-paste(s)04	04c2fa41e8c931ff4163e59c3fc62621-brightness05	IntensityBrightness	255 0 204	None	None
8	18637	04c2fa41e8c931ff4163e59c3fc62621	ff20fb131f0ec7a962503c58a4f4e1fb-01	ff20fb131f0ec7a962503c58a4f4e1fb-selectregion02	SelectRegion	None	None	other
9	18638	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-selectregion02	04c2fa41e8c931ff4163e59c3fc62621-paste(s)04	Donor	None	None	None
10	18639	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-brightness05	04c2fa41e8c931ff4163e59c3fc62621-brightness05_01	AntiForensicExifQuantizationTable	None	None	None
11	18640	04c2fa41e8c931ff4163e59c3fc62621	ff20fb131f0ec7a962503c58a4f4e1fb	ff20fb131f0ec7a962503c58a4f4e1fb-01	OutputPng	None	None	None

4.6. Probe manipulation history reference file

The third MFC dataset reference file is the probe manipulation history reference file, which defines an ordered path from the given probe trace back to the base image. Table 10 shows the probe manipulation history reference file (NC2017_EvalPart1-manipulation-image-ref-probejournaljoin.csv) of the previous example. If the table defined in this file is merged with the table defined in the journal reference file (Table 9) on *JournalName*, *StartNodeID*, and *EndNodeID*, a path from the given probe trace back to the original base media (image/video) with the sequence order is obtained. All operations for the given probe are defined with the rows/links associated with it. The link operation's mask color defines the operation's region color in the final composite localization mask reference (for NC2017) file. For example, the highlighted row in Table 10 is merged with the first data row in Table 9 since they have the same *JournalName*, *StartNodeID*, and *EndNodeID*. The information from Table 10 shows the link is the 4th link from the final probe to the original base. The information from Table 9 shows that the 4th link operation is 'IntensityCurves,' and the composite mask region color of this operation is (102, 255, 0).

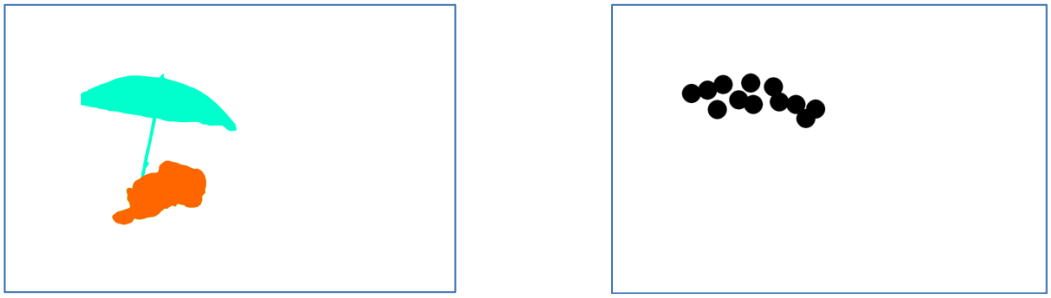
If two probe images pass through/share the same link in the same image journal, two rows will be defined in the probe manipulation history reference file which contain the same link, but different *ProbeFileID* values for two probes respectively (that is, the same *JournalName*, *StartNodeID*, and *EndNodeID*, but with different *ProbeFileID*). If merging the probe manipulation history table with the journal reference table, the sequence order number and the *BitPlane* number of the two probes could be different in general because they belong to two different paths of the different probes.

Table 10. An example of the probe manipulation history file.

ProbeFileID	JournalName	StartNodeID	EndNodeID	Sequence
21a1b6501b9c0d84fa46ad6eddf8bbe4	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-brightness05	04c2fa41e8c931ff4163e59c3fc62621-brightness05_01	1
21a1b6501b9c0d84fa46ad6eddf8bbe4	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-paste(s)04	04c2fa41e8c931ff4163e59c3fc62621-brightness05	2
21a1b6501b9c0d84fa46ad6eddf8bbe4	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-curves03	04c2fa41e8c931ff4163e59c3fc62621-paste(s)04	3
21a1b6501b9c0d84fa46ad6eddf8bbe4	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-paste(s)02	04c2fa41e8c931ff4163e59c3fc62621-curves03	4
21a1b6501b9c0d84fa46ad6eddf8bbe4	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-01	04c2fa41e8c931ff4163e59c3fc62621-paste(s)02	5
21a1b6501b9c0d84fa46ad6eddf8bbe4	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621	04c2fa41e8c931ff4163e59c3fc62621-01	6

4.7. Probe reference mask file

Given a test image, the evaluation masks for each manipulation task performed on the image are condensed into a final composite mask (NC2017) or a JPEG2000 container in which each link mask is a bit plane (after MFC18).



(a) NC2017 ground-truth reference composite mask (b) A sample system output mask result

Figure 7. Examples of the image ground-truth reference mask and system output mask.

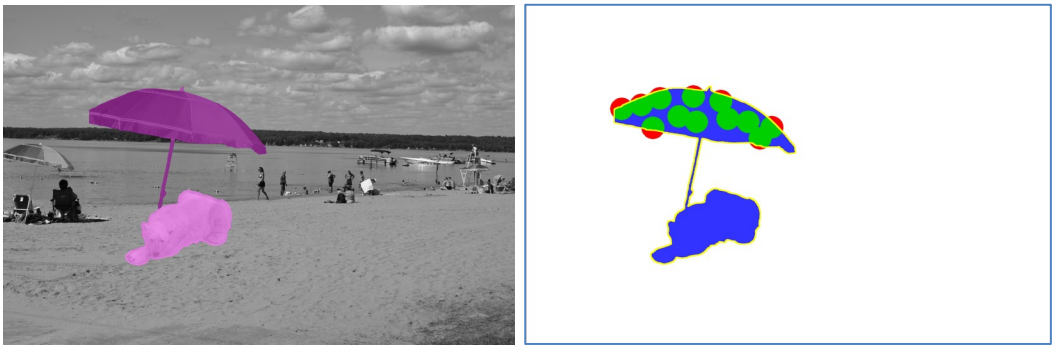
NC2017 reference composite mask is in Portable Network Graphics (png) format with color regions. Figure 7 (a) shows the example of the NC2017 ground-truth reference composite mask for the test probe image (reference/manipulation-image/mask/fff35454245fd5b890547a84cdb1bad3.ccm.png) in png format of the final node image in Figure 3. The localization region mask records the two splices: one is 'PasteSplice' (or 'Clone'), a bigger umbrella (the teal color region), and the other is 'PasteSplice,' a polar bear (the orange color region). The polar bear mask overlay/cover partial umbrella mask in the bottom in the ground-truth reference mask in png format because the png format only supports one layer. Figure 7 (b) on the right shows a sample system output mask. If the evaluation task is to detect all manipulated pixels regardless of manipulation type, then the ground-truth covers every manipulated region (all pink colors as shown in Figure 8 (a) left image). The Matthew Correlation Coefficient (MCC) is used for the localization performance measure, and the MCC value of the system output mask based on the left reference mask is 0.541.

The image localization ground-truth reference mask file is recorded in the *ProbeMaskFileName* column of the probe reference file (MFC*_EvalPart1-manipulation-image-ref.csv) MFC18 and later is updated using JPEG2000 format. JPEG2000 is an image coding system that offers an extremely high level of scalability and accessibility. The standard supports precisions as high as 38 bits/sample. The reference mask file in JPEG2000 format contains multiple channels/layers. Each layer corresponds to a specific operation defined in the *Operation* column in the journal reference file (MFC*_EvalPart1-manipulation-image-ref-journalmask.csv), marked by a distinct color defined in the corresponding *Color* column in the

journal reference file. To specify the channel of the given probe's given link, *BitPlane* is introduced into the probe manipulation history reference file (MFC*_EvalPart1-manipulation-image-ref-probejournaljoin.csv) to record the channel bit plane information, which represents a specific operation's local manipulation operation mask aligned with the probe image.

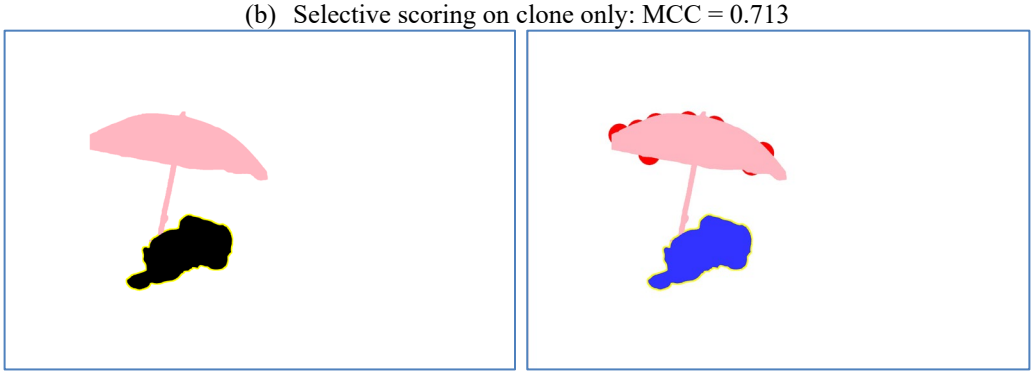
With the mask file in JPEG2000 format, it is easy to separate the local masks of the given operation from the rest of other operations, which allows a specific evaluation of this particular operation in the image localization task [5]. A test image's manipulation reference mask is associated with one or multiple bit planes defined in the corresponding rows of the probe image recorded in the probe manipulation history reference file. The mask in each layer traverses through different transforms defined by each link operation along the path to the final manipulated probe media. For example, a 'PasteSplice' mask may be followed by a 'SeamCarve' in one evaluation media and a 'Warp' in another. The final composite mask could be automatically built using the information extracted from the journal reference file, the probe manipulation history reference file, and the JPEG2000 reference mask file.

The evaluation infrastructure also supports selective scoring on an image localization task. For the same example of the final node in Figure 3, as we introduced before, there are two major paste operations in the final manipulated image of Figure 3. The localization region mask records the two kinds of splices: one is to clone a bigger umbrella (the dark pink color region in the left image of Figure 8 (a)), and the other is to splice a polar bear (the light color region in the left image of Figure 8 (a)). The same system output mask, Figure 7 (b), is also used in the following selective scoring examples. If the evaluation task is to selectively evaluate only the in-image clone detection system, then only the clone operation mask should be used (the black region in the left of Figure 8 (b)) as the ground-truth mask for the evaluation. The MCC of the same system output of the selective scoring on the clone is 0.713. If the evaluation task is to selectively evaluate only the out-image splice detection system, then only the splice operation mask should be used (the black region in the left of Figure 8 (c)) and the selective scoring result on splice is 0. It shows that the system only detected the clone operation in this test probe example.



(a) Evaluation on all operations: MCC = 0.541

(b) Selective scoring on clone only: MCC = 0.713



(c) Selective scoring on splice only: MCC = 0

Figure 8. An example of NC2017 composite mask for selective scoring evaluation.

Appendix E: The Two Approaches for MFC Dataset Histogram Report illustrates two approaches to obtain the dataset manipulation operation histogram distribution statistics.

5. User Customized Data Selection for Special Evaluation

In real media forensic applications, some of the manipulation detection systems are designed to detect general manipulations while others are designed to detect a particular type of manipulations (such as 'Crop' or 'PasteSplice' etc.). We have designed the dataset infrastructure as described in Section 3, which provides a mechanism to trace manipulation operation information for each manipulated probe. Combining the infrastructure with the evaluation scoring package, MediScore[13], the MFC, and OpenMFC evaluation programs support both the general evaluation and the special evaluations. To evaluate the overall system performance on all kinds of manipulations, all probe image/videos in the datasets are used in the evaluation. To evaluate the special manipulations, the selective scoring approach is proposed to select the relevant probes to evaluate special forensic systems without regenerating the whole datasets. Thus, both evaluation requirements are fulfilled by a single evaluation dataset efficiently.

5.1. Subset data selection for customized evaluation task

As introduced above, besides NIST-defined evaluation tasks, performers may like to use NIST data and evaluation software packages to report customized task performance on special detection systems. For example, if we design an evaluation task to answer a special study question, such as "which system performs the best to detect images with paste splice manipulations", we may evaluate the paste splice detection system performance only on

[13] https://github.com/usnistgov/MediScore

'PasteSplice' operations, not other operations. We then select all probes from the whole evaluation dataset using the NIST reference data defined in three reference files. We can do a four table join (the *index* table described in the spreadsheet of the index file, the *reference* table described in the probe reference file, the *journalmask* table described in the journal reference file, and the *probejournaljoin* table described in the probe manipulation history reference file) to obtain all testing probes and the manipulations applied in each of the testing probes. We only select the test probes which contain 'PasteSplice' operations and create a 'PasteSplice' subset with some randomly selected non-manipulated images to build a subset for the 'PasteSplice' selective scoring evaluation. Thus, other manipulation image probes are not presented in the evaluation set, the selective scoring evaluation only focuses on the 'PasteSplice' operation and reports the system performance on 'PasteSplice' detection.

5.2. Example results on the selective scoring evaluation

Figure 9 shows image manipulation detection systems performance on a full evaluation dataset. All probe images in the test set are used for the evaluation.

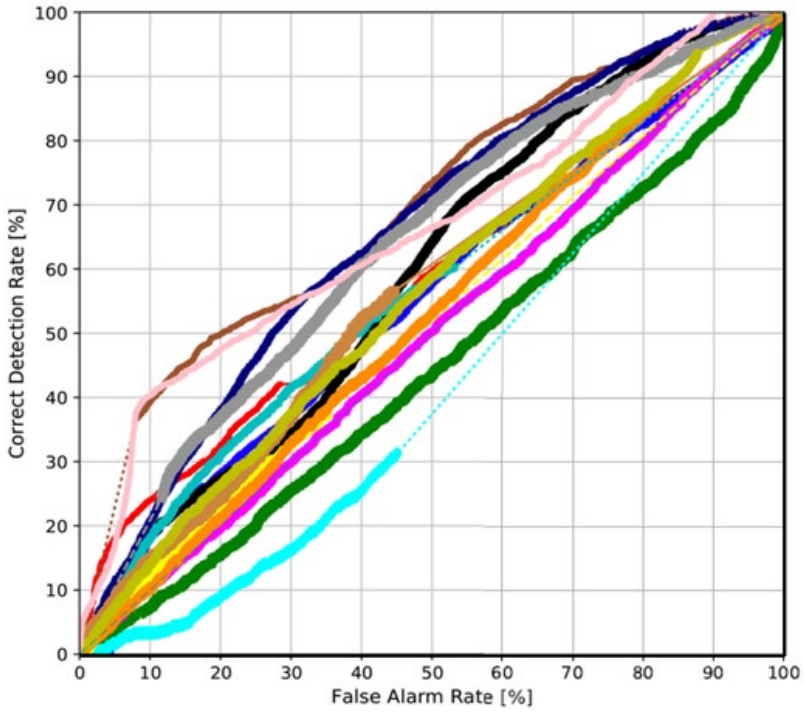


Figure 9. ROC curves of the manipulation detection systems on the full evaluation set.

In the evaluation, there are systems designed specifically for a certain type of manipulation operation, e.g., face manipulation, crop, etc. The evaluation program reports the system performance on those special operations. However, the target probes for such operation are relatively very small compared to the whole dataset size. In some cases, the target probes are less than 5 percent or even smaller than the size of the whole dataset. If we evaluate the system performance based on the whole dataset, then there are a small percent of target test probes and a large percent of nontarget test probes. The ROC curves of all systems visually are nearer to the diagonal line instead of showing the performance difference among systems.

To resolve this issue, we use a selective scoring approach. For example, in the NC2017 evaluation, there are systems designed specifically for the 'Crop' manipulation operation. To report systems' performance on the 'Crop' operation, that is, in terms of a special evaluation

task: “which system performs the best to detect the crop operation?” With the dataset infrastructure, using the approach described above, we can report those performances by following those steps: firstly, using the NIST dataset’s index file and three reference files, do a four-table join (*index* table, *reference* table, *journalmask* table, and *probejournaljoin* table defined in index spreadsheet and three reference spreadsheets), we obtain all testing probes and the manipulations applied in each of the testing probe. Then we only select the test probes which contain the ‘crop’ operation and create a ‘crop’ subset for selective scoring evaluation. Next, using the selected evaluation subset, the selective scoring evaluation reports the best system on the ‘crop’ operation only, which indicates the system’s performance on the selected ‘crop’ detection. NIST MFC scoring package, MediScore, already implemented it and supported such tasks with a selective scoring command line. Please refer to the software user manual for the detail on how to use it.

Figure 10 shows the selective scoring results on the same set of image manipulation detection systems on the ‘Crop’ subset selected from the same full dataset used in Figure 9. In other words, the selective scoring infrastructure selects only the probes with the ‘crop’ manipulation operation as the selective scoring evaluation test set, which is selected from the full test set. Figure 10 demonstrates that the two systems outperform the rest of the systems at detecting the ‘Crop’ operation.

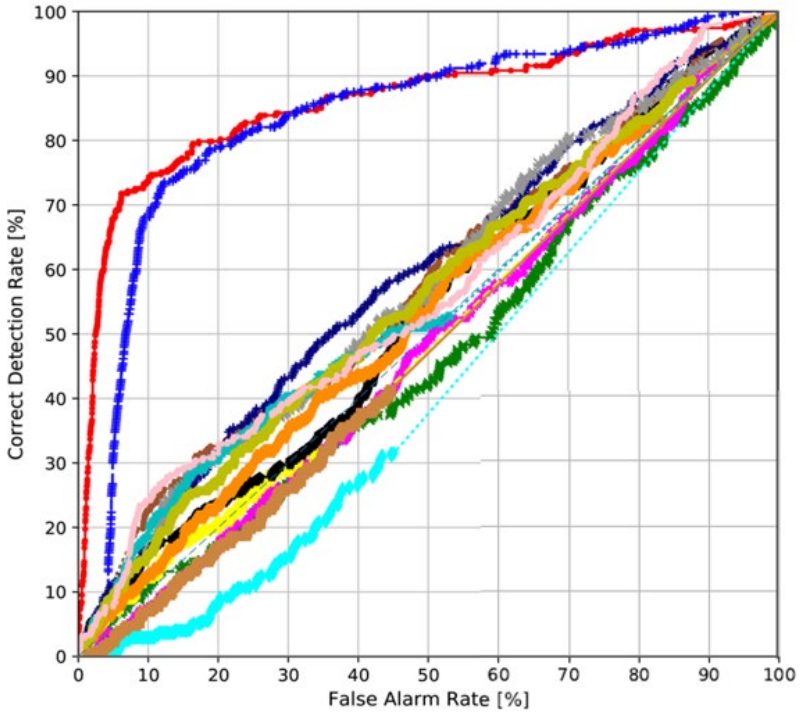


Figure 10. Selective scoring ROC curves for the same set of detection systems on the subset of test images with the ‘Crop’ operation.

6. Conclusions and Future Works

The ongoing spread of deep fake media continues to create urgent interest in consented, tested data for research and development. Through both DARPA and NIST-sponsored media forensics evaluations, the NIST MFC datasets have filled this need and continue to garner

questions and interest from many researchers across the globe. This document has described their structure and use as generated for these evaluation challenges.

In addition, the MFC datasets offer convenient and flexible resources for new efforts involving machine algorithm training. Researchers seeking to extract the data according to their own requirements for their own purpose can find value in the extensive preparation and documentation. The OpenMFC team foresees specific utility for the usage of the existing MFC datasets for the following research areas:

- *Training data generation* - NIST MFC data contains original unmanipulated data with camera model and camera sensor information; the *IsTarge* values defined in the probe reference file aids researchers in identifying the unmanipulated base image (value: 'N') from the manipulated image (value: 'Y') and generate the training data; The *ProbeFileName, BaseFileName, and ProbeMaskFileName* values defined the triple of the manipulated image file, the original image file, and the corresponding manipulated pixel mask file. All probe images with the *SeamCarving* value equal to 'Y' are manipulated images with seam carving technology, etc.
- *Evaluation of new operations and parameters* - MFC's whole manipulation journals, could be used independently or used to update the existing Journaling Tool (JT)[1], Extended Journaling Tool (ExtendedJT), and Automatic Journaling Tool (AutoJT) to generate the manipulated images with different kinds of manipulation operations and parameters themselves.

Media forensic researchers may work independently or collaborate with the NIST OpenMFC team to explore current or future datasets for evaluations.[14]

If the user has any questions on the MFC or OpenMFC datasets, please email them to mfc_poc@nist.gov. Technical support is available from the NIST team.

References

[1] Robertson, E., Guan, H., Kozak, M., Lee, Y., Yates, A., Delgado, A., Zhou, D., Kheyrkhah, T., Smith, J. and Fiscus, J. (2019), Manipulation Data Collection and Annotation Tool for Media Forensics, IEEE computer vision and pattern recognition conference 2019, Long Beach, CA. Available at https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=927817.

[2] Guan, H., Kozak, M., Robertson, E., Lee, Y., Yates, A., Delgado, A., Zhou, D., Kheyrkhah, T., Smith, J. and Fiscus, J. (2019), MFC Datasets: Large-Scale Benchmark Datasets for Media Forensic Challenge Evaluation, IEEE Winter Conference on Applications of Computer Vision (WACV 2019), Waikola, HI. Available at https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=927035.

[3] Fiscus, J. and Guan, H. (2020), Media Forensics Challenge Evaluation Overview, ARO Sponsored Workshop on Assured Autonomy, Workshop Talk. Available at https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=930628.

[4] Fiscus, J., Guan, H., Lee, Y., Yates, A., Delgado, A., Zhou, D., Joy, D. and Pereira, A. (2017), MediFor Nimble Challenge Evaluation 2017, Evaluation Presentation. Available at https://www.nist.gov/system/files/documents/2017/07/31/nist2017mediaforensicsworkshop_20170726.pdf.

[14] Email: mfc_poc@nist.gov

[5] Fiscus, J., Guan, H., Delgado, A., Kheyrkhah, T., Lee, Y., Zhou, D. and Yates, A. (2018), 2018 MediFor Challenge, Evaluation Presentation. Available at https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=928264.

[6] Fiscus, J., Guan, H., Lee, Y., Yates, A., Delgado, A., Zhou, D., Kheyrkhah, T., and Jin, X. (2020), NIST Media Forensic Challenge (MFC) Evaluation 2020 - 4th Year DARPA MediFor PI meeting, Evaluation Presentation. Available at https://www.nist.gov/publications/nist-media-forensic-challenge-mfc-evaluation-2020-4th-year-darpa-medifor-pi-meeting.

[7] Yates, A., Guan, H., Lee, Y., Delgado, A., Zhou, D., and Fiscus, J. (2018), Media Forensics Challenge 2018 Evaluation Plan, Evaluation Plan, Available at https://www.nist.gov/system/files/documents/2018/10/30/mfc2018evaluationplan-clean3_werb.pdf.

[8] Yates, A., Guan, H., Lee, Y., Delgado, A., Zhou, D., Kheyrkhah, T. and Fiscus, J. (2019), Media Forensics Challenge 2019 Evaluation Plan, Evaluation Plan. Available at https://www.nist.gov/system/files/documents/2019/03/12/mfc2019evaluationplan.pdf.

[9] Yates, A., Guan, H., Lee, Y., Delgado, A., Kheyrkhah, T., Fontana, P. C., and Fiscus, J. (2020), Open Media Forensics Challenge (OpenMFC) 2020 Evaluation Plan, Evaluation Plan. Available at https://www.nist.gov/publications/open-media-forensics-challenge-2020-evaluation-plan.

[10] Fiscus, J., Guan, H., Lee, Y., Yates, A., Delgado, A., Zhou, D., Joy, D., and Pereira, A., Nimble Challenge 2017 Evaluation, Evaluation Website. Available at https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation.

[11] Fiscus, J., Guan, H., Delgado, A., Kheyrkhah, T., Lee Y., Zhou, D. , and Yates, A., NIST Media Forensic team, Media Forensics Challenge 2018, Evaluation Website. Available at https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2018.

[12] Fiscus, J., Guan, H., Lee, Y., Yates, A., Delgado, A., Zhou, D., and Kheyrkhah, T., Media Forensics Challenge 2019, Evaluation Website. Available at https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2019-0.

[13] Guan, H., Lee, Y., Kheyrkhah, T., Fontana, P. C., and Fiscus, J. (2020), Open Media Forensics Challenge (OpenMFC) 2020 Evaluation, Evaluation Website. Available at https://mfc.nist.gov.

[14] Lee, Y., Yates, A., Guan, H., Delgado, A., Zhou, D., Kheyrkhah, T., and Fiscus, J., 2018 Multimedia Forensics Challenges (MFC18): Summary and Results, NIST Interagency/Internal Report (NISTIR) Number 8324, Available at https://doi.org/10.6028/NIST.IR.8324

Appendix A: Challenge Task Definitions

This section provides a brief overview of the MFC evaluation task definitions. Please refer to [7][8][9][14] for details.

A.1 Image Manipulation Detection and Localization (IMDL)

The Image Manipulation Detection and Localization (IMDL) task is to detect if an image was manipulated and, if so, then to spatially localize the manipulated region. For detection, an IMDL system provides a confidence score for each trial with higher numbers indicating the image was more likely to be manipulated. For the localization evaluation, the system provides a mask and its bit plane that indicate the manipulated region(s) with a manipulation type.

A.2 Video Manipulation Detection and Localization (VMDL)

The Video Manipulation Detection and Localization (VMDL) task is to detect if a video was manipulated and if so, to localize the manipulated region spatially and temporarily. For detection, a VMDL system provides a confidence score for each trial with higher numbers indicating the video was more likely to be manipulated. For the localization evaluation, the system provides a video mask file that indicates the manipulated region(s) with a manipulation type.

A.3 Splice Manipulation Detection and Localization (SMDL)

The Splice Manipulation Detection and Localization (SMDL) task is to detect if a region of a given image (i.e., the donor) had been spliced into another image (i.e., the probe) and, if so, then to localize the region(s) of the donor and probe images that were used for the splice operation. Like the IMDL task, an SMDL system provides a confidence score along with two masks: one with the region(s) of the donor that was copied and another with the region(s) of the probe that was pasted from the donor.

A.4 Camera Verification (CV)

The Camera Verification (CV) task is to determine if a camera fingerprint from an image matches a claimed camera fingerprint, given a collection of camera device IDs. This task supports both image and video probes, and the dataset consists of three training sets (image, video, and multimedia) and two testing sets (image and video). This yields a total of six training-testing conditions with the composition of the training sets, and the testing sets system provides a confidence score indicating how likely the image or video was captured with the claimed camera.

A.5 Provenance Filtering (PF) and Provenance Graph Building (PGB)

The Provenance Filtering (PF) task is to, given a probe image, return up to N images (a potential pool of related images) from a large collection of images (called the world dataset).

The Provenance Graph Building (PGB) task is to, given a probe image, retrieve the related images from the world dataset and to construct the relationships among the retrieved images and to build the provenance phylogeny graph.

Appendix B: MediFor Data Use Agreement

DATA USE AGREEMENT FOR EVALUATION MEDIA

This Data Use Agreement ("Agreement"), which takes effect as of the date of signature of both parties ("Effective Date"), is entered into by and between _____ ("Media Recipient") and the Information Technology Laboratory of the National Institute of Standards and Technology("Media Provider"). The purpose of this Agreement is to provide Media Recipient with access to Evaluation Media ("EM") for use in the MediFor Research project.

Definitions

Evaluation Media ("EM") – Media, both still and video imagery, to be provided to the Media Recipient by the Media Provider for training, testing and analysis purposes. The EM will be assembled from suitably licensed web resources and High Provenance (HP) imagery where the capturer of the imagery is known. All HP media will be public domain or under Creative Commons zero license (CC0): https://creativecommons.org/choose/zero/.

Supporting Documentation – Any documentation accompanying the EM including descriptions of the manipulation and details related to image provenance.

1. Responsibilities of Media Provider. Media Provider agrees to:

    a. Provide EM to the Media Recipient for test and evaluation in accordance with the definition above and the goals of the MediFor project.

    b. Notify Media Recipient when provided EM is approved for public release.

2. Responsibilities of Media Recipient. Media Recipient agrees to:

    a. Use or disclose the EM and Supporting Documentation only for the purposes permitted by this Agreement;

    b. Use appropriate safeguards to prevent use or disclosure of the EM and Supporting Documentation other than as permitted by this Agreement;

    c. Report to Media Provider any use or disclosure of the EM or Supporting Documentation of which Media Recipient becomes aware that is not permitted by this Agreement, including the presence of prohibited identifiers;

    d. Require any of Media Recipient's subcontractors, agents, or any other party, that Media Recipient grants access to the EM or Supporting Documentation to agree to the same restrictions and conditions on the use and/or disclosure that apply to Media Recipient under this Agreement;

    e. Not use the information present in the EM or Supporting Documentation, alone or in combination to identify or contact any individuals whomay be depicted in any fashion in the provided imagery; and

    f. Not release or redistribute EM or supporting documentation prior to public release approval outside the MediFor project without the express written approval of Media Provider or the Government.

3. Disclosure or Release of EM.

    a. Any EM approved for public release shall be restricted only by the terms of the license under which it has been made available.

b. The license applied to the EM and Supporting Documentation will follow the most restrictive license of the source imagery or the Creative Commons zero license (CC0).

c. EM will only be licensed for public release after testing and evaluation is complete and results have been published to protect the blindness of test material. The descriptions of the manipulations will also be provided when release is authorized.

4. Term & Termination The Agreement shall remain in force until:

a. Term. The term of this Agreement shall commence as of the Effective Date and terminate 5 years from Effective Date.

b. Termination by Media Recipient. Media Recipient may terminate this Agreement at any time by notifying the Media Provider and returning or destroying any provided EM which has not yet been approved for public release.

c. Termination by Media Provider. Media Provider may terminate this Agreement at any time by providing thirty (30) days prior written notice to Media Recipient and requesting the return or destruction of any provided EM which has not yet been approved for public release.

Media Provider	**Media Recipient**
By:_____	By: _____
Name:_____	Name: _____
Title:_____	Title: _____
Date: _____	Date: _____

Appendix C: Evaluation Participation Agreement

NIST Open Media Forensics Challenge 2020 Evaluation (OpenMFC2020) Registration

To participate in the NIST Open Media Forensics Challenge 2020 Evaluation (OpenMFC2020), this registration form must be filled out in its entirety and uploaded to the OpenMFC server. For help, contact mfc_poc@nist.gov.

General terms of participation in MFC2020:

- Participation in the MFC2020 evaluation is voluntary and open to all who find the task of interest and are willing and able to abide by the rules of the evaluation. These evaluations are collaborative efforts where the Government provides access to datasets and an objective evaluation of the technology, and registered sites provide intellectual effort by fully participating in the evaluation cycle. To fully participate a registered site must:
  - (1) become familiar with, and abide by, all evaluation rules;
  - (2) develop/enhance an algorithm that can process the required evaluation datasets;
  - (3) submit the necessary files to NIST for scoring; and
  - (4) attend the evaluation workshop (if one occurs) and openly discuss the algorithm and related research with other evaluation participants and the evaluation coordinators.

  When a site fails to meet these four requirements, the functioning collaboration is weakened and so registrations by the site for future Multimodal Information Group technology evaluations will not be accepted until the site is committed to fully participate.
- Use and distribution of the evaluation data (both source and reference data) is governed by the terms of a license agreement with the data provider(s).
- The site agrees to not publicly compare its results with the results of other participants until the other participant's results are published outside of the MFC Workshop Venue.
- Sites are free to do what they wish with their own results.

NIST serves to coordinate the Open Media Forensics Challenge evaluations. The reported results are not to be construed, or represented, as endorsement of any participant's system, or as official findings on the part of NIST or the U.S. Government.

Site: _____

Site principal investigator:

Name: _____ E-mail: _____

Main contact:

Name: _____

E-mail: _____

Phone: _____

Main contact signature: _____ Date: _____

By signing and returning this form to NIST, you are registering your site as a participant in the NIST 2020 Media Forensics Challenge evaluation and acknowledging that you have read and will abide by the protocols described in this form and the official Evaluation Plan.

Appendix D: The Basic NIST MFC Dataset ReadMe File Example

```
Media Forensics Challenge (MFC) <Year> <Task> Dataset <Dataset_Name>
Date

1. Introduction
This dataset is a release of development resources built by the NIST MFC
Program. This release is ONLY being released for Program-internal
discussions.

The data consists of test material derived from * produced by *. This data
set was generated from a collection of * images.

All base images have been collected by *. (Brief introduction about how
the data was collected, generated, etc.).

The dataset is structured similarly to the MFC image dataset.

2. Directory Structure
      ReadMe.txt  - This file
      /probe      - Directory of images to be analyzed for various
                    manipulations
      /world      - Directory of images that simulate a real-world
                    collection of images
      /indexes    - Directory of index files indicating which images
                    should be analyzed
      /reference  - Directory of subdirectories for each evaluation task,
                    containing the metadata including the reference
                    masks, and the journal files
      /documents  - Directory of required documents

3. System Input Files
The index files are pipe-separated CSV formatted files. The index file for
the Manipulation task will have the columns:
      TaskID        Detection task
                      (e.g., "manipulation")
      ProbeFileID   Label of the probe image
                      (e.g., 00003e6a1efc7022da825396dc680343)
      ProbeFileName Full filename and relative path of the probe image
                      (e.g., /probe/00003e6a1efc7022da825396dc680343.jpg)
      ProbeWidth    Width of the probe image
                      (e.g., 4000)
      ProbeHeight   Height of the probe image
                      (e.g., 300)
      ProbeFileSize File size of probe
                      (e.g., 2500)

4. Reference Files
There are no reference files for performer team's distribution. Reference
files are used by evaluation team and may be released to public after
evaluation.

The reference files are pipe-separated CSV formatted files. The reference
file for the manipulation task will have the columns:
      TaskID        Detection task
                      (e.g., "manipulation")
      ProbeFileID   Label of the probe image
```

```
                    (e.g., 001f9af3165a39c9e42aee922f874326)
      ProbeFileName Full filename and relative path of the probe image
                    (e.g., /probe/001f9af3165a39c9e42aee922f874326.jpg)
      IsTarget      If the image is manipulated ('Y') or not ('N')
                    (e.g., 'Y')
      BaseFileName  Full filename and relative path of the base image of
                    the given probe
                    (e.g., /world/d247cf38f1ee6c03f605d251b44b6bfd.jpg)
      HPDeviceID    Camera device ID (not camera model ID) provided the
                    data collection team.
                    If "UNDEF", the data is unknown, or not provided for
                    training.
                    (e.g., MK-NEX5T)
      HPSensorID    Camera sensor ID (HPDeviceID_primary or HPDeviceID_secondary)
                    provided by the data collection team.
                    (e.g., IPhoneX6_primary)
5. File Naming
The image files in this release will be named <randomString/MD5>.<extension>.

6. Distribution
THIS DATA IS PROVIDED "AS IS" for use in the MFC Program.  With regard to
this data, NIST/(Other Organization) MAKES NO EXPRESS OR IMPLIED WARRANTY
AS TO ANY MATTER WHATSOEVER, INCLUDING MERCHANTABILITY, OR FITNESS FOR A
PARTICULAR PURPOSE.

7. Contacts
If you have any questions about this dataset, please contact the following
people:

        POC Name  (To be updated based on the personnel of datasets)
        POC Email (To be updated based on the personnel of datasets)
------------------------------------------
Change Log
Creation Date - README created by *
Updated  Date - README updated by * on *
```

Appendix E: The Two Approaches for MFC Dataset Histogram Report

A dataset may have many features/factors such as manipulation types, image/video formats, and their compression parameters, manipulation software and their functions (such as content-aware removal, seam carving, etc.), the manipulators and their skill sets, the metadata or image features of the original image/video used for manipulation, etc. All factors could affect system performance in some cases or to some extent. The distribution of those factors of the evaluation dataset in the high factor dimensional space is important information that both the dataset generation and the system development teams would like to know.

Taking the manipulation type/operation as a factor, its histogram distribution provides an overview picture of a dataset's manipulations. In this section, two statistical approaches to visualize the histogram distribution of the manipulation operations is introduced: the unique operation link counts and the probe count.

E.1 Journal operation link count histogram

The journal operation links in the journal graph record all the operations in the journal without duplications. As we described in Section 4.5, each row in the journal reference file defines a unique link in the test dataset. We use the journal reference file to obtain the journal operation link count histogram: for each manipulation operation defined in the journal reference file, we count how many times it appears in the files. Each count is a unique link. For example, for a particular operation - blur, each count in the blur histogram bin defines a unique operation to a given image, and a given region. The count is not duplicated for the same blur operation with the same region and the same given image using this approach.

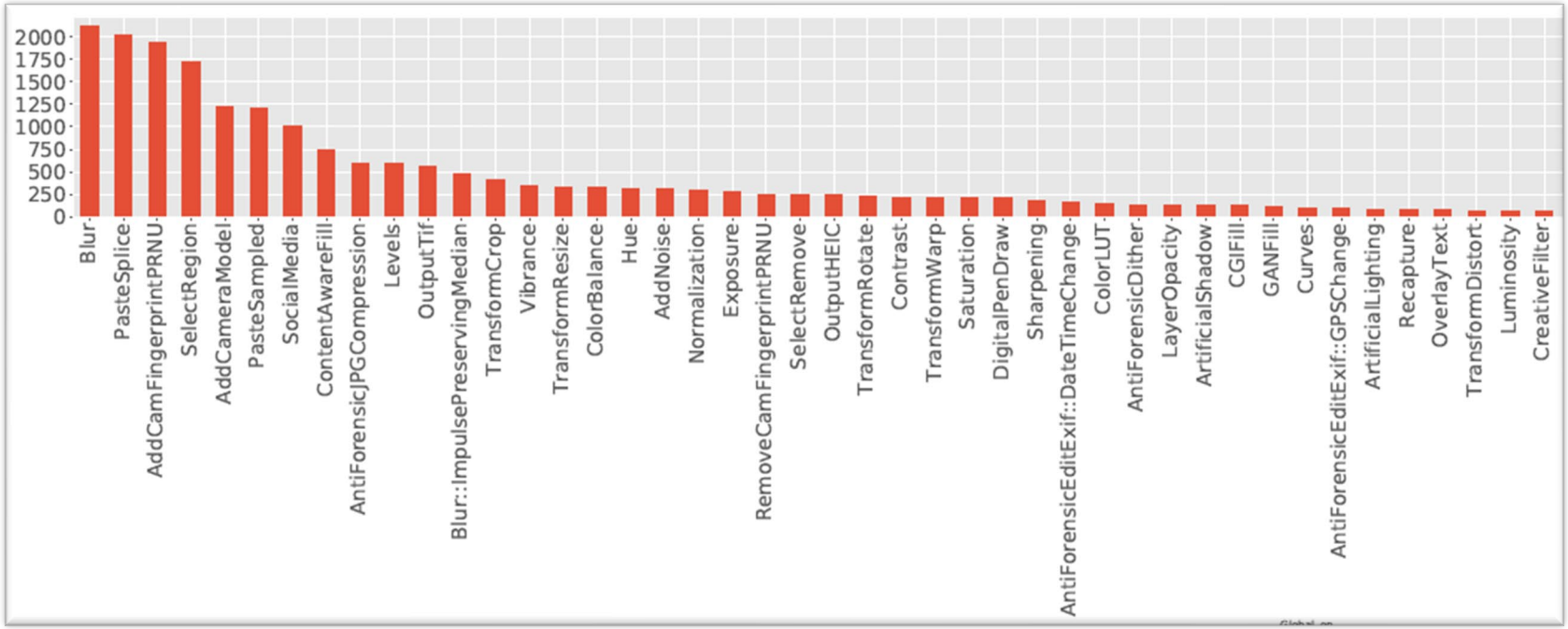


Figure 11. An example of journal operation link count histogram for MFC20 EP1 Image dataset.

Taking MFC20 EP1 Image dataset as an example, which is the latest dataset that contains the most of manipulation operations, Figure 11 shows a visualization of a partial histogram of journal operation link count histogram generated using the journal reference file. The whole histogram is too long to be visualized clearly here. In the figure, the histogram bin counts represent the number of unique manipulation operations taken from the journal reference file.

E.2 Probe operation count histogram

Another way to show histogram distribution is to count the number of manipulation operation links based on the number of the probes, which means, if there are two probes that come from a single journal and two probes that share the same manipulation operation link, this link is counted twice in histogram bin of this link's manipulation operation. The histogram could be obtained using the following approach: firstly, to obtain all metadata associated with a probe, we join four tables together: the index file table (which defines the desired probes in the dataset), the probe reference file table (which defines the reference data of all the probes in the dataset), the journal reference file table (which defines all the links in the journal graph with manipulation operations), and the probe manipulation history reference file table (which defines the path of the given probe with all links related to the probe). That is, for any probe defined in the index file, all the links along the path from the probe to the base image are obtained in the final joined table. With this approach, we can also obtain all manipulation operations given a probe. Secondly, we count how many times it appears in a probe's path for each type of manipulation operation, and we add all counts given all probes to obtain the total count for the given manipulation operation. As previously discussed, if two probes in the same journal share the same operation link, then that link contributes twice to the operation's histogram bin. Furthermore, one link could be counted multiple times when multiple manipulated probe images in a journal share the same link. Finally, for each histogram bin corresponding to an operation, the probe count represents the number of the probes in the whole dataset that contain this operation.

Figure 12 shows a partial histogram of the probe operation link count histogram for the MFC20 EP1 image dataset. Again, the entire histogram is too long to be visualized here. It is shown that the probe manipulation operation distributions counts are much higher than the unique journal operation link count due to the link duplication counts.

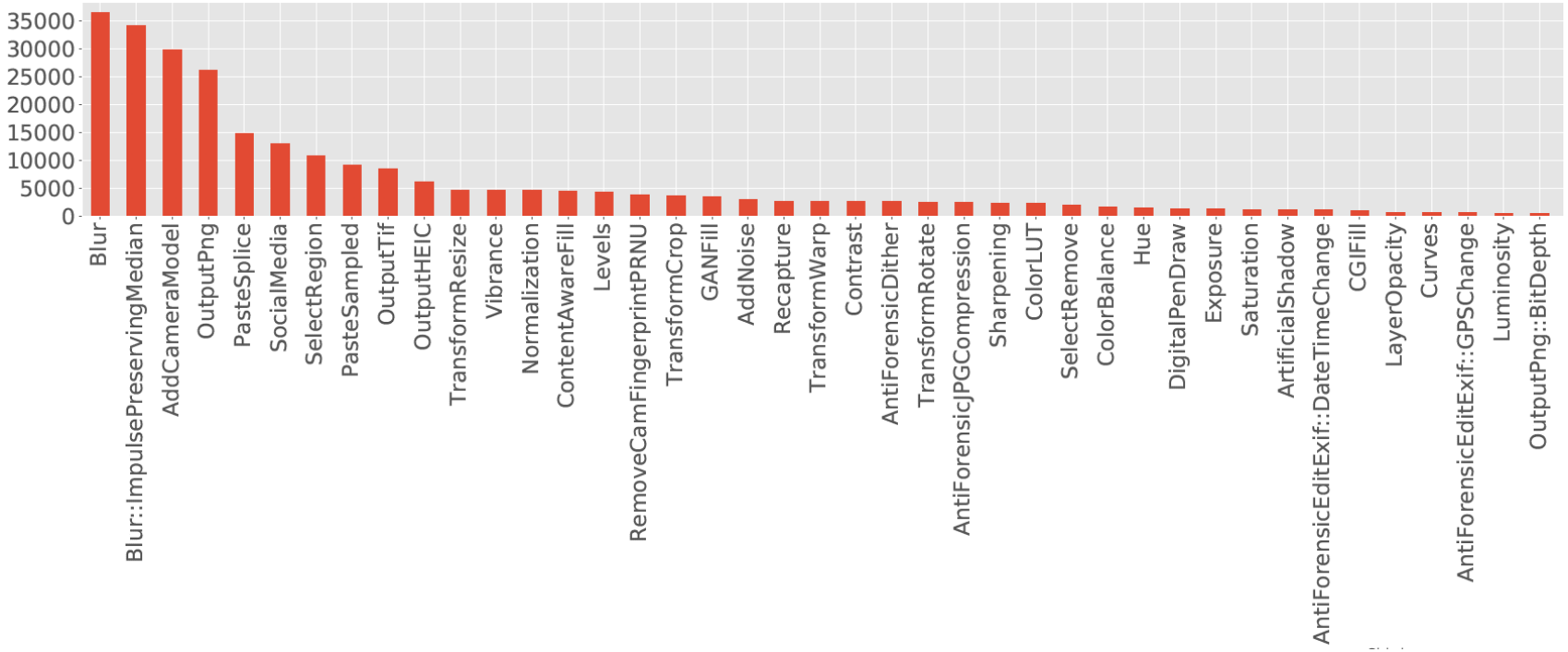


Figure 12. An example of probe operation link count histogram for MFC20 EP1 Image dataset