AN ADVERSARIAL APPROACH TO SPLICED FORGERY DETECTION AND LOCALIZATION IN SATELLITE IMAGERY

A Thesis

Submitted to the Faculty

of

Purdue University

by

Emily R. Bartusiak

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Electrical and Computer Engineering

May 2019

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF THESIS APPROVAL

Dr. Edward J. Delp, Chair
    School of Electrical and Computer Engineering
Dr. Amy R. Reibman
    School of Electrical and Computer Engineering
Dr. Michael D. Zoltowski
    School of Electrical and Computer Engineering

Approved by:
    Dr. Linda J. Mason
        Head of the Graduate Program

To Curly Top, Funny Guy, Laura Lu, Blondie, PMBA, Bara Boochie, and Flaffa

for your infinite support, unconditional love,
and endless source of happiness and laughter.

ACKNOWLEDGMENTS

I believe strongly in saying “thank-you” – and in saying it loudly – to the people who contribute to my successes and to my learning. I am so thankful for everyone who aided me, both directly and indirectly, in conducting the research discussed in this thesis and in completing my Master’s of Science in Electrical and Computer Engineering at Purdue University.

First and foremost, I want to thank Professor Edward J. Delp for serving as my advisor throughout my pursuit of the Master’s Degree and overseeing this research project. I am grateful for the opportunity to conduct research as a Master’s student, allowing me to apply knowledge gained in coursework and internships to a meaningful project and to continue learning through its development. Furthermore, I am grateful for the opportunity to explore research in the Video and Image Processing Laboratory (VIPER) starting as an Undergraduate student my Sophomore year. The time I spent with VIPER members and its advisor are key factors in my decision to pursue Graduate School. I am so grateful for the advice, mentorship, and opportunities he provided to me throughout the years

Next, I want to thank the other members of my Advisory Committee – Professor Amy R. Reibman and Professor Michael D. Zoltowski – for their support throughout my academic career. I appreciate their guidance as I transitioned from the Bachelor’s program to the Master’s program, the candid conversations with them about my future and their career paths, and their continued support.

Two other professors have served important roles in my advancement as well. I want to thank Professor Carla Zoltowski for sponsoring my Master’s Degree and giving me an opportunity to work with her as a Graduate Teaching Assistant. This experience has given me a chance to develop my leadership, teamwork, communication, and design skills in another environment. I have learned extensively about designing courses, interpreting feedback from students and professors, and creating meaningful opportunities for students to learn. I also want to thank Professor Charles Bouman for his support and encouragement of my graduate studies. Between his Digital Image Processing course and conversations with him whenever we run into each other, I have learned so much from him and received so much encouragement from him, sometimes when I least expected it and needed it the most. I am truly grateful for his support.

Next, I want to express gratitude for my two graduate mentors and co-authors: David Güera and Sri Kalyan Yarlagadda. It has been an honor and a pleasure to conduct research with both of them. I appreciate our meetings, their feedback and criticism, their honesty, and our intellectually stimulating conversations both inside and outside of the lab. I have learned so much from both of them and am grateful for their camaraderie and collaboration.

Now, I want to thank Dr. Khalid Tahboub for his mentorship during my Bachelor's program and encouragement to pursue Graduate School. I am so grateful for the opportunity to work with him in the VIPER lab because he exposed me to fascinating topics related to video and image processing that I would not have been introduced to in my Undergraduate curriculum otherwise. He exemplifies kindness, patience, and intelligence and contributed to my interest in pursuing this topic further.

At this time, I want to thank both the past and the present members of the VIPER lab for their support, inspiration, and friendship. Between dinner in San Diego with past and present members, my Senior Design project for which I received feedback from Jeehyun Choe, meetings at conferences, and meaningful conversations about their research projects and approaches, they have taught me how to be a better researcher and various career paths for the road ahead.

Finally, I want to thank my family for all of their support and encouragement. I am so appreciative of my six cheerleaders who express excitement in my endeavors and my accomplishments and are always willing to visit with me when I need a friend, advice, an outside perspective, or a laugh.

This work is based on research sponsored by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) under agreement number FA8750-16-2-0173. I am grateful for their generous support and contributions that have made this thesis possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Bartusiak, Emily R. MSECE, Purdue University, May 2019. An Adversarial Approach to Spliced Forgery Detection and Localization in Satellite Imagery. Major Professor: Edward J. Delp.

The widespread availability of image editing tools and improvements in image processing techniques make image manipulation feasible for the general population. Oftentimes, easy-to-use yet sophisticated image editing tools produce results that contain modifications imperceptible to the human observer. Distribution of forged images can have drastic ramifications, especially when coupled with the speed and vastness of the Internet. Therefore, verifying image integrity poses an immense and important challenge to the digital forensic community. Satellite images specifically can be modified in a number of ways, such as inserting objects into an image to hide existing scenes and structures. In this thesis, we describe the use of a Conditional Generative Adversarial Network (cGAN) to identify the presence of such spliced forgeries within satellite images. Additionally, we identify their locations and shapes. Trained on pristine and falsified images, our method achieves high success on these detection and localization objectives.

1. INTRODUCTION

1.1 Significance of this Research

Proper communication at both the public and personal level is key to the healthy development and advancement of human civilization. Over the years, the means of communication has evolved. In the present day, the most popular and important technical platform for communication is the Internet. It enabled the creation of many social media systems, instant messaging capabilities, and email services that provide very inexpensive and effective ways to express and share one's ideas with the rest of the world. While effective communication systems for sharing information can help us become more informed and connected as a society, they can also be used to spread misinformation to achieve a nefarious objective. Hence, it is of paramount importance that we verify and authenticate the shared data on these systems.

Ideas manifest in many forms. Of the various ways to express them, including verbal speech, written text, and displayed signage and symbols, images prevail as one most prevalent means of communication. A single picture can convey so much information in such a short period of time, a preference of many in today's fast, quick-paced society. Oftentimes, images shared on the Internet lack context and sources, though. Viewers of such images tend to assume instantaneous understanding and fail to investigate the origins or true messages of these uncontextualized images. Furthermore, editing images has become very easy. Tools such as GIMP and Photoshop can be used to modify images in a number of ways, and they are easily accessible to the general public. The motivation for editing images could be rather harmless: to smooth skin on a portrait that captures a moment forever; to enhance the colors of a scene in nature experienced on vacation when the sky was greyer than desired; to further dramatize a movie scene; or to elicit laughter and provide entertainment. Unfortunately though, image editing is sometimes performed with a more malicious intent. When edited images surface on the Internet, either for an innocuous or malevolent reason, people eager to be the first to publicize news to their network and to promote their own opinions share them without much examination or thoughtfulness in regard to the authenticity of the content. Thus, imagery promoting easily-misinterpreted messages or outright erroneous content proliferates on the Internet.

The combination of these convincing tampered images with the image-sharing practices of Internet users can have detrimental consequences. To address this problem, the forensic community has developed a wide array of tools to detect various kinds of image forgeries [1–3]. Most of the images shared on the Internet originate from consumer and smart-phones cameras, but other types of imagery, such as satellite images, serve very important purposes in business and government applications. Thus, they thus pose new problems and challenges for the forensic community [4, 5].

With the increase in the number of satellites equipped with imaging sensors and the technological advancements made in satellite imaging technology, high resolution images of the ground are becoming popular and easily obtainable. It is now possible to not only access these overhead images from public websites [6] but also to purchase custom satellite imagery of specic locations. Just like any other type of image, satellite images can also be doctored. Although the forensic community is keen on developing tools to tackle forgeries of all types, it has been biased towards imagery captured by consumer cameras and smartphones [7–10]. The nature of acquisition of satellite imagery is quite different from that of images from consumer cameras. Consequently, it is important that forensic tools be developed that specically target satellite imagery.

In recent years, some methods [11–13] to contend with satellite image forgeries have been developed. In [11], Ho et al. proposed an active forensic method based on watermarks to verify the authenticity of satellite images. Watermarks are an effective way of ascertaining whether an image is forged or not, as long as watermarks are utilized in the first place. However, their absence renders such methods ineffective. In [13], Ali et al. proposed a passive method based on machine learning to detect inpainting in satellite images. Yarlagadda et al. [12] proposed a method based on deep learning to detect splicing in satellite images. They employ Generative Adversarial Networks (GANs) [14, 15] to learn a compact representation of pristine satellite images, operating in a one-class fashion. Then, they use that representation to detect splicing of various sizes. Although these methods produce promising results, there is still room for improvement.

In this dissertation, we discuss detection and localization of splicing in satellite images. Splicing refers to the process of replacing pixels in a certain region of an image with pixels from another image to add or remove an object in the original image. We employ a Conditional Generative Adversarial Network (cGAN) to learn a mapping from a satellite image to its splicing mask. The trained cGAN operates on a satellite image of interest and outputs a mask of the same resolution that indicates the likelihood of a pixel belonging to a spliced region. Our cGAN's architecture is an extension of the popular pix2pix [16] network. Differently from [12], we learn a direct mapping from an image to its forgery mask. Moreover, only pristine images are used for training in their approach, but in contrast, we provide both pristine and forged images to train our model. We use the same dataset proposed in [12] to validate our method and report both the detection and localization performances.

This dissertation is organized as follows. First, we examine Conditional Generative Adversarial Networks in Chapter 2. Next, we explain our method for estimating spliced forgery masks of satellite images in Chapter 3. Then, we show experimental results obtained using our proposed method in Chapter 4. Finally, we conclude our work and discuss potential research opportunities for this topic in Chapter 5.

1.2 Contributions of this Thesis

In this thesis, we developed a Conditional Generative Adversarial Network (cGAN) to detect and localize spliced forgeries in satellite images. We demonstrated high performance of the method on the used dataset.

1.3 Publication Resulting from the Thesis

- **E. R. Bartusiak**, S. K. Yarlagadda, D. Gera, F. M. Zhu, P. Bestagini, S. Tubaro, E. J. Delp. Splicing Detection And Localization In Satellite Imagery Using Conditional GANs. *IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR)*, March 2019. San Jose, CA.

2. BACKGROUND INFORMATION

2.1 Goals

We investigate the following two specific objectives in this paper: forgery detection and localization. *Detection* refers to the goal of determining if an RGB satellite image **I** has been modified via splicing. It is a binary classification problem where images can be considered *forged*, if they have been modified, or *pristine*, if not. *Localization* refers to the image segmentation goal of identifying each pixel in a forged image that belongs to the spliced entity, otherwise known as the *forgery*. These goals are defined in a similar manner to those outlined in [12].

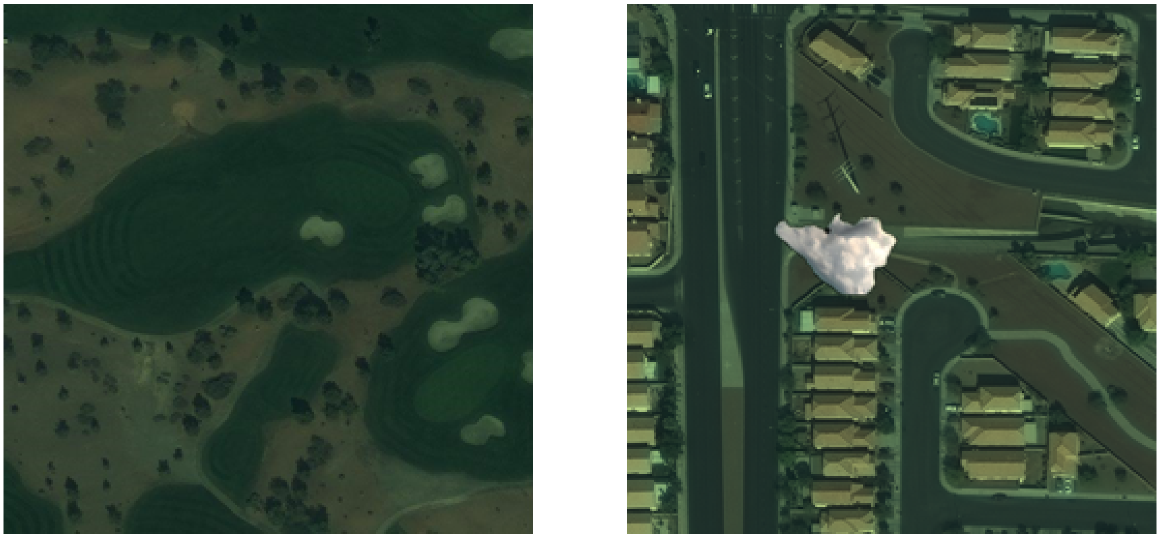
2.2 Forgery Masks

Forgery masks **M** are used to help us visualize and determine the outcomes for these objectives. For an image **I**, a forgery mask **M** of the same dimensions shows the forgery in **I**, if it exists. In other words, for a satellite image $\mathbf{I}(x, y)$ where (x, y) specifies the coordinate location of a pixel in **I**, the corresponding forgery mask **M** is comprised of values defined as

$$\mathbf{M}(x, y) = \begin{cases} 255 & \text{if } \mathbf{I}(x, y) \text{ is forged,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

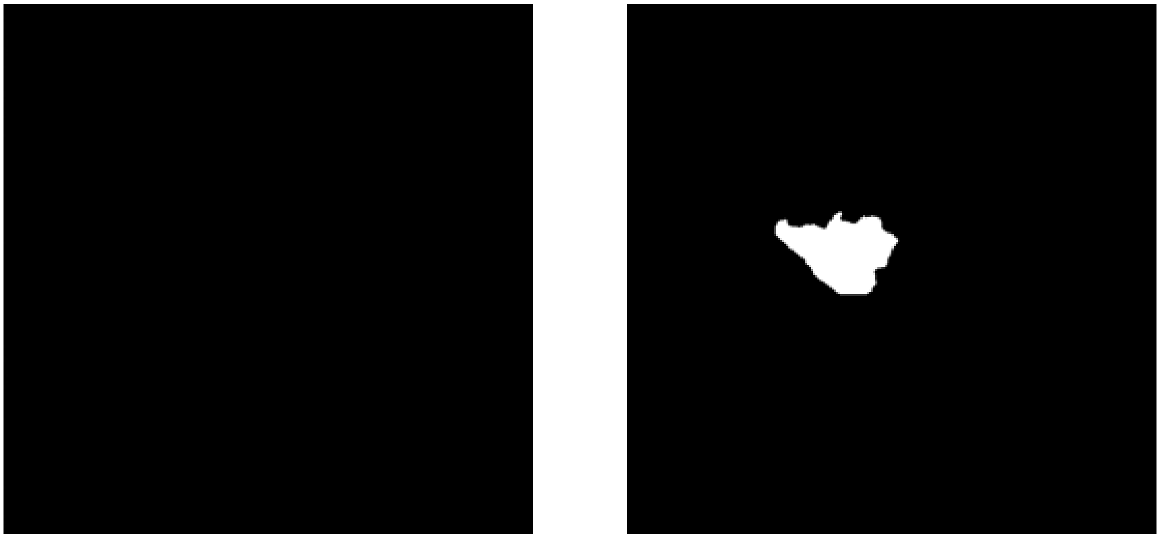
Therefore, the shape, size, and location of a forgery in an image **I** can be ascertained from the mask **M** if it contains white pixel values (i.e., 255). At an extreme, an entirely white mask ($\mathbf{M} \neq 0$) indicates that every pixel in **I** has been manipulated, whereas an entirely black mask ($\mathbf{M} = 0$) represents a pristine image.

Our approach is to train a cGAN to create $\hat{\mathbf{M}}$, an estimate of the forgery mask $\mathbf{M}$. $\mathbf{I}$ is considered doctored if $\hat{\mathbf{M}} \not\approx 0$, meaning that a forgery is detected in it and is comprised of the pixels located at $\{(x, y) : \hat{\mathbf{M}}(x, y) \neq 0\}$. On the other hand, the image $\mathbf{I}$ is considered pristine if no forgery is detected, indicated by $\hat{\mathbf{M}} \approx 0$. Examples of satellite images and their corresponding ground truth forgery masks can be seen in Figure 2.1.



(a) Pristine **I**

(c) Forged **I**



(b) Pristine **M**

(d) Forged **M**

Fig. 2.1. Image - mask $\{\mathbf{I}, \mathbf{M}\}$ pairs. (a) and (c) portray two example satellite images under analysis. (b) and (d) illustrate their corresponding ground truth masks. A pristine image's mask is entirely black, like (b). (d) displays a mask that contains a forgery. The cGAN will try to create mask estimates that resemble masks (b) and (d).

2.3 Conditional Generative Adversarial Networks (cGANs)

We train our cGAN on both pristine and forged images to learn a mapping from an input image $\mathbf{I}$ to a forgery mask $\mathbf{M}$. It consists of two parts: a generator G and a discriminator D. Figure 2.2 shows the overall cGAN architecture. Additional details about the general cGAN concepts reported in this section can be found in [16].

2.3.1 The Generator

The generator G has a 16-layer U-net architecture (8 encoder layers, 8 decoder layers) with skip connections [17]. When G is presented with an image $\mathbf{I}$, it computes an estimated forgery mask $\hat{\mathbf{M}}$, defined as $\hat{\mathbf{M}} = G(\mathbf{I})$. The generator's objective is to create $\hat{\mathbf{M}}$ that is close to the true $\mathbf{M}$. Meanwhile, the discriminator D is trained to differentiate between the true input-mask pairs $\{\mathbf{I},\ \mathbf{M}\}$ and synthesized input-mask pairs $\{\mathbf{I}, \hat{\mathbf{M}}\}$ coming from the generator. In a cGAN, the generator G is coupled to the discriminator D with a loss function. During the course of training, the discriminator forces the generator to produce masks that resemble both the style and the content of the ground truth masks. In learning to mimic the style of these authentic masks, the generator G produces mask estimates that cannot be distinguished from ground truth masks by the discriminator. Meanwhile, in learning to produce mask estimates that are close to the ground truth masks, the generator G better constructs the content of the masks so that they show the correct locations of forgeries. Both of these concentrations improve the generator's performance.

2.3.2 The Discriminator

The discriminator D has an architecture of a 5-layer Convolutional Neural Network (CNN) that implements binary classification on masks. Sometimes, a true image-mask pair $\{\mathbf{I},\ \mathbf{M}\}$ is presented to D. Other times, an image-mask estimate pair $\{\mathbf{I}, \hat{\mathbf{M}}\}$ is presented. In both cases, the image under analysis $\mathbf{I}$ is presented to the discriminator D along with either a true mask $\mathbf{M}$ or a synthesized mask $\hat{\mathbf{M}}$. D divides the input into patches of size 70×70 pixels. It then predicts the likelihood that each patch is pristine, assigning values ranging from zero (0) to one (1) to represent the probability that a patch is pristine. A label of 0 indicates that the discriminator D believes with strong likelihood for the presented image-mask pair to contain a synthesized mask $\hat{\mathbf{M}}$ that came from the generator G. In contrast, a label of 1 indicates that D believes with strong likelihood for the pair to contain a true mask $\mathbf{M}$ that came from the dataset. The values for all of the patches are averaged to determine the classification for the entire input.

The following equations describe the two cases outlined in this paragraph:

$$D(\mathbf{I}, \hat{\mathbf{M}}) = D(\mathbf{I}, G(\mathbf{I})) = 0, \tag{2.2}$$

$$D(\mathbf{I}, \mathbf{M}) = 1. \tag{2.3}$$

2.3.3 The Coupled Loss Function

The generator G and the discriminator D compete in a min-max game, training and improving each other over time. The coupled loss function of the network is described by the following equation:

$$\begin{aligned} \mathcal{L}_{\text{cGAN}}(G, D) = & \mathbb{E}_{\mathbf{I},\mathbf{M}}[\log(D(\mathbf{I}, \mathbf{M}))] + \\ & \mathbb{E}_{\mathbf{I}}[\log(1 - D(\mathbf{I}, G(\mathbf{I})))]. \end{aligned} \tag{2.4}$$

The generator G tries to minimize this equation in order to create mask estimates that are misclassified by D. Concurrently, the discriminator D tries to maximize this equation in order to discern between true and synthesized image-mask pairs.

So far, we have described a network in which the generator G learns to create masks that could be mistaken for real forgery masks by D. However, this does not ensure that the synthesized masks will correctly show forgeries in images. For example, $\hat{\mathbf{M}}$

may "fool" D and be classified as an authentic mask for $\mathbf{I}$ without resembling its ground truth mask. In such a case, $\hat{\mathbf{M}} \not\approx \mathbf{M}$. Therefore, we impose an additional constraint on the generator so that it learns to reconstruct the ground truth masks of training images, i.e., $\hat{\mathbf{M}} \approx \mathbf{M}$. This can be achieved by training G to minimize a reconstruction loss $\mathcal{L}_{\mathrm{R}}$ between $\hat{\mathbf{M}}$ and $\mathbf{M}$.

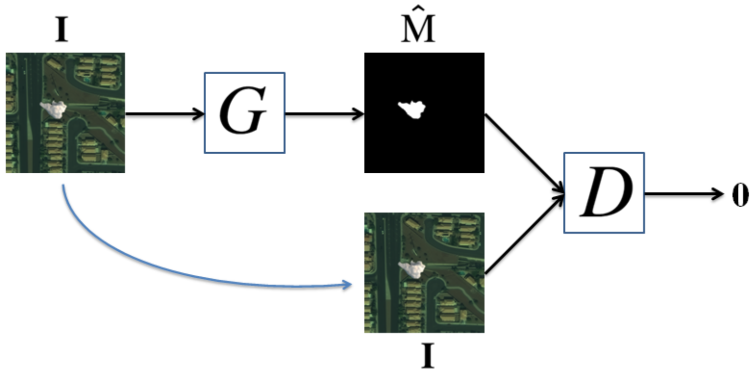
Since our primary task is to classify every individual pixel into two classes (i.e., forged or pristine), we choose $\mathcal{L}_{\mathrm{R}}$ to be a binary cross-entropy (BCE) loss term. This is different with respect to the classic pix2pix method which uses L_1 loss for the loss term $\mathcal{L}_{\mathrm{R}}$. We later verify in our experiments that BCE is indeed a better choice for $\mathcal{L}_{\mathrm{R}}$ over L_1.

The total loss function of the cGAN is denoted as:

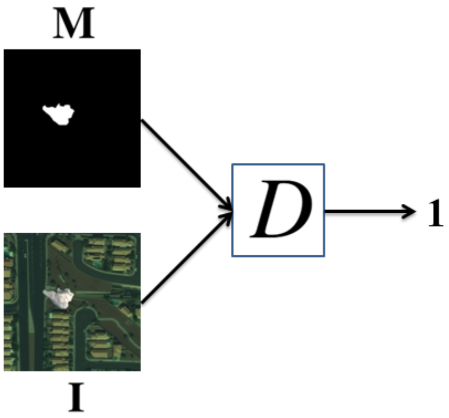
$$\mathcal{L} = \mathcal{L}_{\mathrm{cGAN}} + \lambda \mathcal{L}_{\mathrm{R}}. \tag{2.5}$$

2.3.4 Architecture During Testing

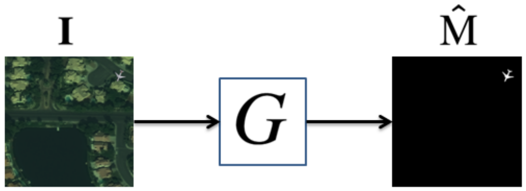
Once training is complete, the generator G is capable of producing masks that are realistic and close to the ground truth masks in the dataset. To analyze new images, the discriminator D is not considered, and the generator G is used to produce mask estimates for new images that it has never seen before during training.



(a) Generator G coupled to discriminator D during training



(b) Discriminator D during training



(c) Generator G after training

Fig. 2.2. cGAN architecture. (a) shows the G-D training configuration, where G produces mask estimate $\hat{\mathbf{M}}$ and presents it to D for evaluation. D attempts to classify non-authentic $\{\mathbf{I}, \hat{\mathbf{M}}\}$ pairs as 0. (b) depicts D during training when presented with a true mask $\mathbf{M}$. It attempts to classify true $\{\mathbf{I}, \mathbf{M}\}$ pairs as 1. The model resembles (c) after training is complete.

3. IMPLEMENTATION

In this chapter, we report the details of our experiments. First, we describe the image dataset. Next, training strategies are discussed.

3.1 Dataset

We utilized the dataset presented in [12] for our experiments. It contains color images of overhead scenes from a satellite and their corresponding ground truth forgery masks. Each image-mask pair is defined as $\{\mathbf{I}, \mathbf{M}\}$ and has resolution 650×650 pixels. The images were adapted from ones originally provided by the Landsat Science program [18, 19] run jointly by NASA [20] and US Geological Survey (USGS) [21].

To create forged images, objects such as airplanes and clouds were spliced into some of the images at random locations. These doctored images fall into one of three size categories (small, medium, or large) based on the approximate dimensions of the forgery they contain relative to the patch dimensions (70×70 pixels) used by the discriminator D to analyze a mask. Small forgeries are approximately 32×32 pixels; medium forgeries are approximately 64×64 pixels; and large forgeries are approximately 128×128 pixels. This information is summarized in Table 3.1. The remaining satellite images were left as pristine.

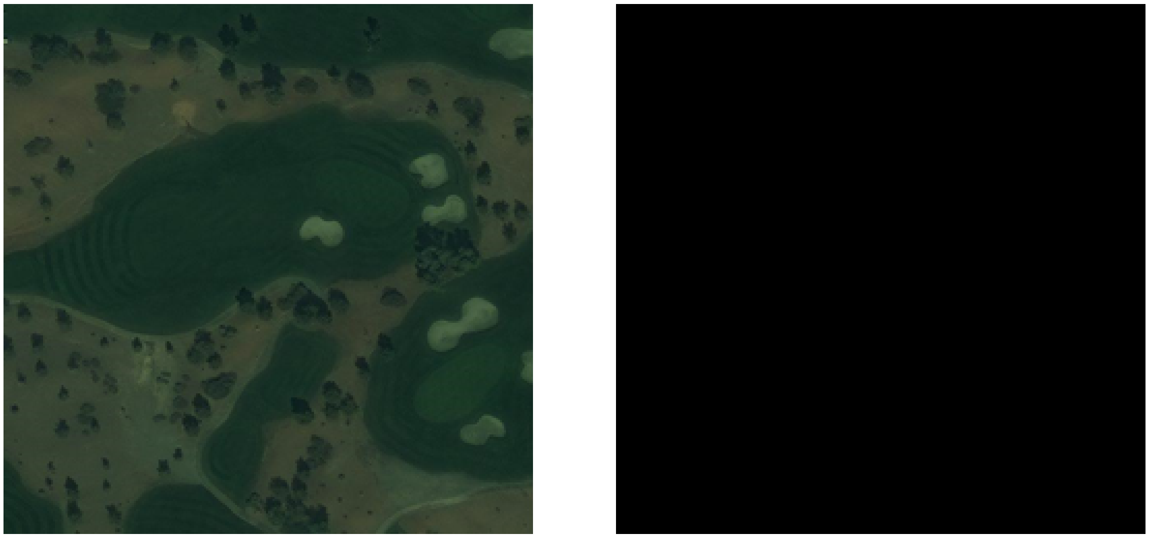
Table 3.1.
Spliced Forgery Object Size Details

Forgery Size	**Approximate Pixel Dimensions**
Small	32×32 Pixels
Medium	64×64 Pixels
Large	128×128 Pixels

For our purposes, pristine and small-forgery samples underwent data augmentation to increase the size of the training dataset. Augmentation methods included rotating pristine and small-forgery $\{\mathbf{I}, \mathbf{M}\}$ pairs by multiples of 90° and flipping them about the vertical and horizontal center axes. This produced our dataset $\mathcal{D}$, which contains 344 total $\{\mathbf{I}, \mathbf{M}\}$ pairs. Of the 344 $\{\mathbf{I}, \mathbf{M}\}$ pairs included in this dataset, 158 pairs contain small forgeries, 32 pairs contain medium forgeries, 31 pairs contain large forgeries, and 123 are pristine. These subsets of $\mathcal{D}$ are denoted as $\mathcal{D}_S$, $\mathcal{D}_M$, $\mathcal{D}_L$, and $\mathcal{D}_P$, respectively. This information is summarized in table 3.2. Examples of images and masks in this dataset are shown in Figures 3.1 through 3.4.

Table 3.2.
Dataset

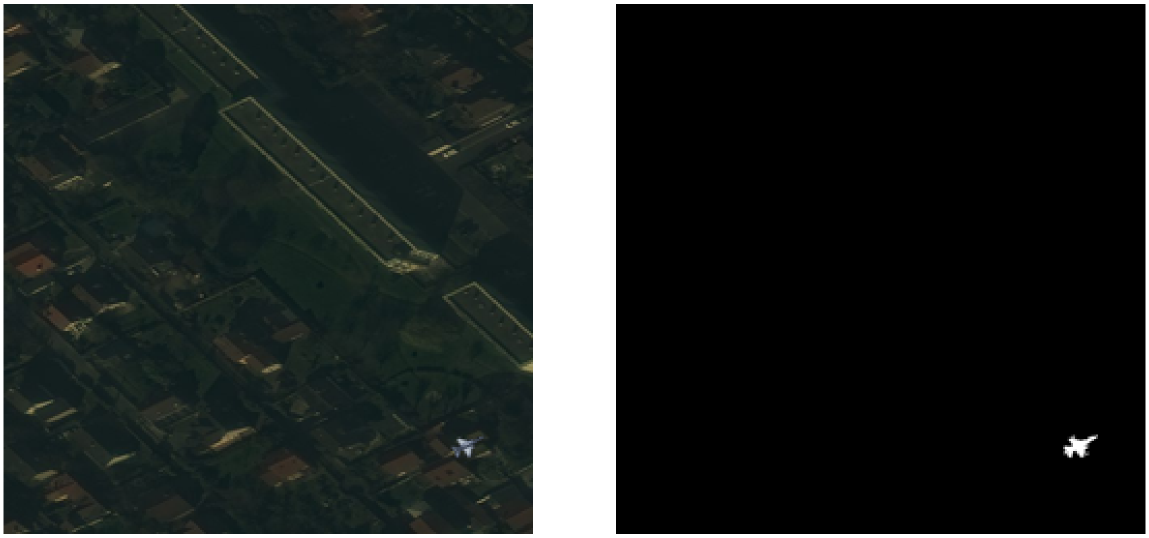
Subset	**# $\{\mathbf{I}, \mathbf{M}\}$ Pairs**
$\mathcal{D}_S$	158
$\mathcal{D}_M$	32
$\mathcal{D}_L$	31
$\mathcal{D}_P$	123



(a) Pristine $\mathbf{I}$

(b) Pristine $\mathbf{M}$

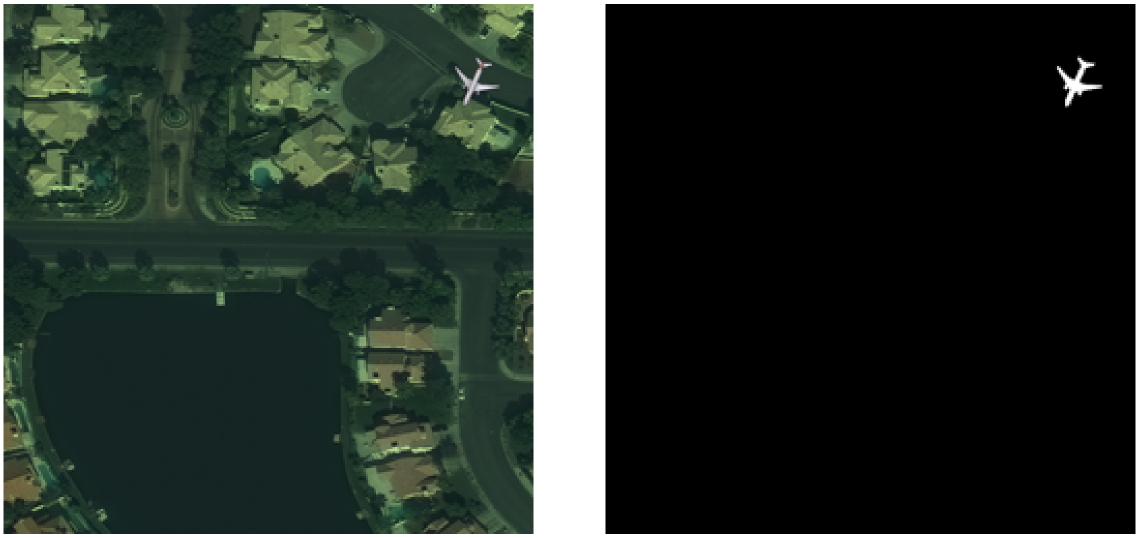
Fig. 3.1. An input image and its ground truth mask (an $\{\mathbf{I}, \mathbf{M}\}$ pair) from $\mathcal{D}_P$.



(a) Small forgery $\mathbf{I}$

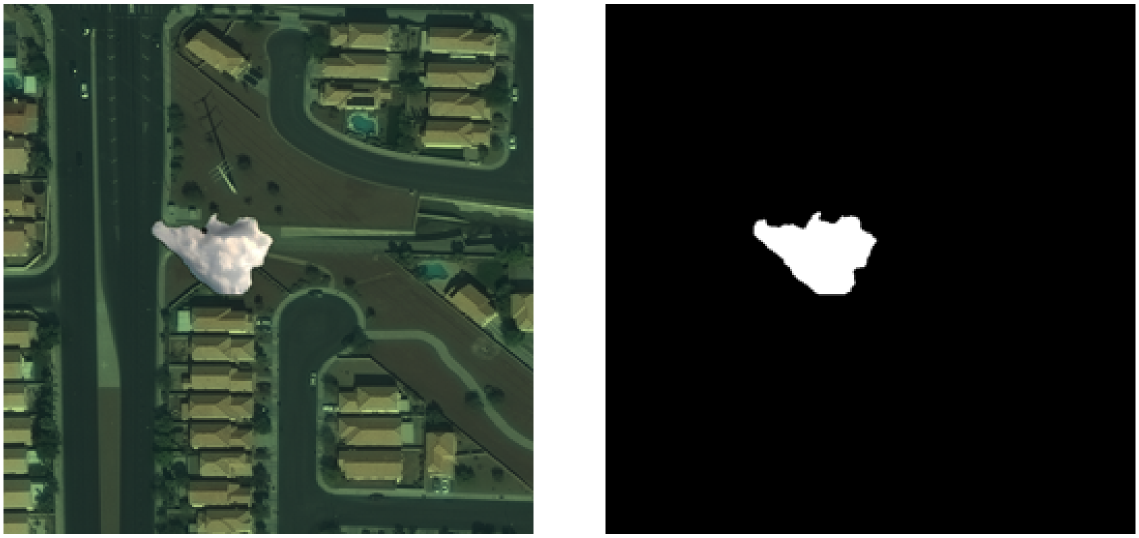
(b) Small forgery $\mathbf{M}$

Fig. 3.2. An input image and its ground truth mask (an $\{\mathbf{I}, \mathbf{M}\}$ pair) from $\mathcal{D}_S$.



(a) Medium forgery $\mathbf{I}$ (b) Medium forgery $\mathbf{M}$

Fig. 3.3. An input image and its ground truth mask (an $\{\mathbf{I}, \mathbf{M}\}$ pair) from $\mathcal{D}_M$.



(a) Large forgery $\mathbf{I}$ (b) Large forgery $\mathbf{M}$

Fig. 3.4. An input image and its ground truth mask (an $\{\mathbf{I}, \mathbf{M}\}$ pair) from $\mathcal{D}_L$.

3.2 Training Strategies

The dataset $\mathcal{D}$ was split into three subsets for training, validation, and testing purposes. The training dataset $\mathcal{D}_{train}$ contains 128 $\mathcal{D}_S$ pairs and 90 $\mathcal{D}_P$ pairs. The validation set $\mathcal{D}_{validation}$ has 32 $\mathcal{D}_S$ pairs and 18 $\mathcal{D}_P$ pairs. The final dataset $\mathcal{D}_{test}$ consists of 32 $\mathcal{D}_M$, 31 $\mathcal{D}_L$, and 15 $\mathcal{D}_P$ pairs. This information is summarized in Table 3.3. By creating disjoint training/validation and evaluation datasets, we observe how well a trained model extends to new forgery sizes. It was hypothesized that small forgeries might pose the biggest challenge to the network, so they compose the training and validation sets.

Table 3.3.
Train-Validate-Test Dataset Split

Set	**# $\mathcal{D}_S$ pairs**	**# $\mathcal{D}_M$ pairs**	**# $\mathcal{D}_L$ pairs**	**# $\mathcal{D}_P$ pairs**
$\mathcal{D}_{train}$	128	0	0	90
$\mathcal{D}_{validation}$	32	0	0	18
$\mathcal{D}_{test}$	0	32	31	15

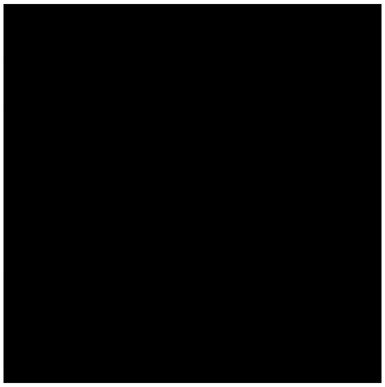
The cGAN was trained for 200 epochs using the Adam optimizer with an initial learning rate of 0.0002. The reconstruction loss coefficient λ was set to 100. After training, the model that performed the best on $\mathcal{D}_{validation}$ was selected to use for testing.

4. EXPERIMENTAL RESULTS

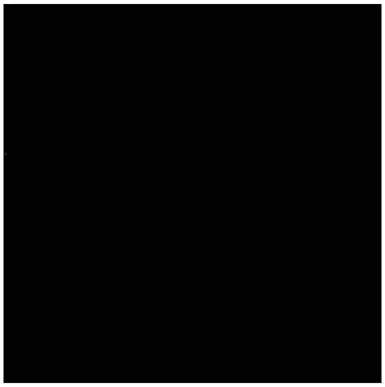
4.1 Visual Analysis of Results

We performed both visual and numerical analysis of the results to determine the effectiveness of our proposed method. Figures 4.1 through 4.12 contain examples of mask estimates produced by G and their corresponding ground truth masks. These selections showcase the different conditions under which the method performed. For example, the location of the forged object varies in these images. Sometimes it is spliced over houses or buildings, and other times it is spliced over a field or a road. The examples also show the different lighting conditions of the images under analysis. Some images are darker than others, causing the spliced object to blend into the background more so than in others. Regardless of the circumstances, these examples show that the model produces mask estimates of both pristine and forged images that very closely resemble the ground truth masks, i.e., $\hat{\mathbf{M}} \approx \mathbf{M}$. Thus, we can clearly see if a forgery is present in an image $\mathbf{I}$ and, if so, its various properties. A numerical analysis of the results further verifies this.

(a) Pristine $\mathbf{I}$

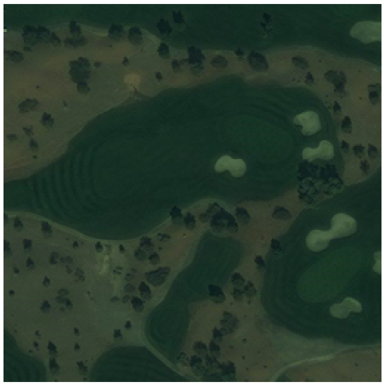


(b) Pristine $\mathbf{M}$

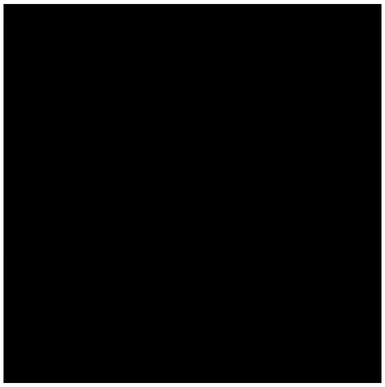


(c) Pristine $\hat{\mathbf{M}}$

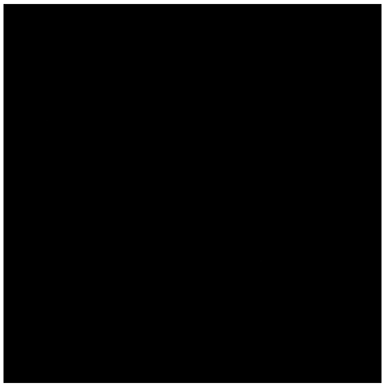
Fig. 4.1. Bright pristine image depicting houses.



(a) Pristine $\mathbf{I}$



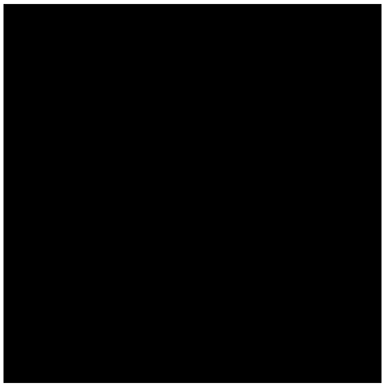
(b) Pristine $\mathbf{M}$



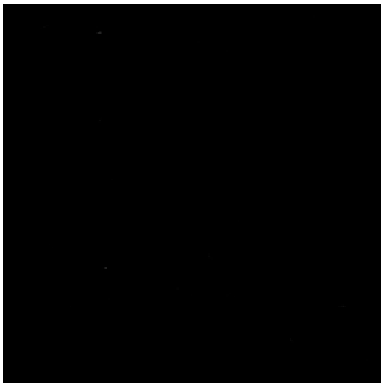
(c) Pristine $\hat{\mathbf{M}}$

Fig. 4.2. Bright pristine image depicting a field.

(a) Pristine **I**

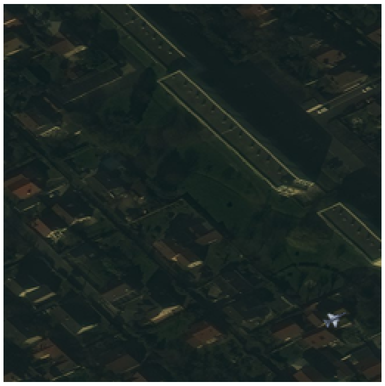


(b) Pristine **M**

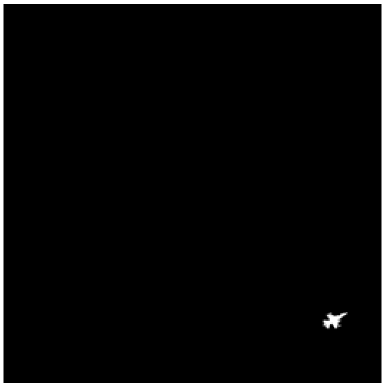


(c) Pristine $\hat{\mathbf{M}}$

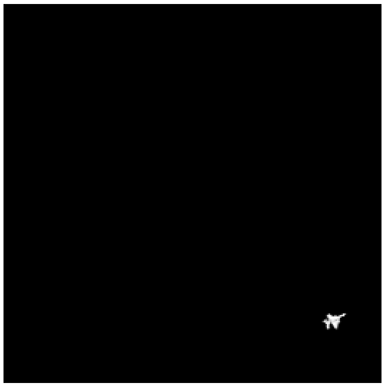
Fig. 4.3. Dark pristine image depicting houses



(a) Small forgery $\mathbf{I}$

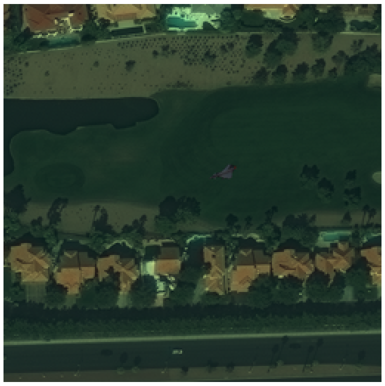


(b) Small forgery $\mathbf{M}$

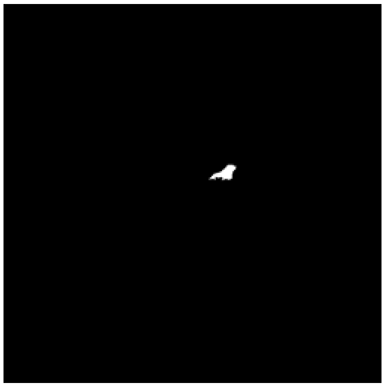


(c) Small forgery $\hat{\mathbf{M}}$

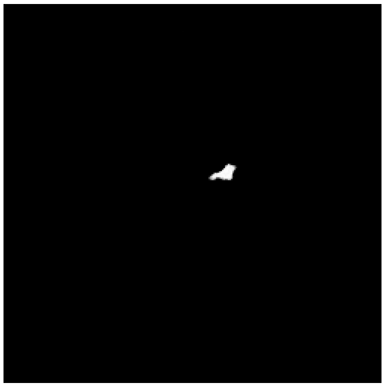
Fig. 4.4. Dark image depicting houses and containing a small-sized plane forgery.



(a) Small forgery $\mathbf{I}$

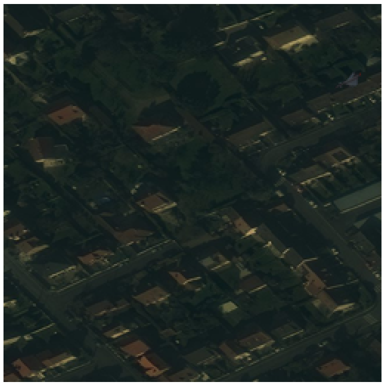


(b) Small forgery $\mathbf{M}$

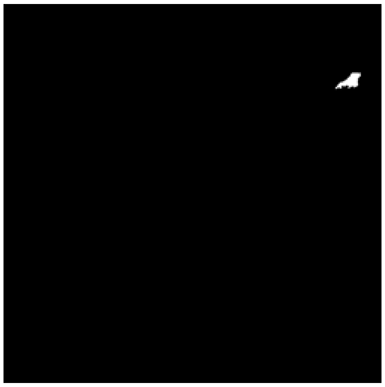


(c) Small forgery $\hat{\mathbf{M}}$

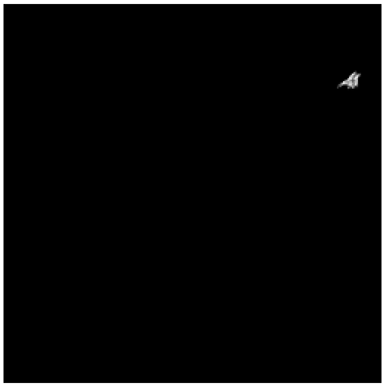
Fig. 4.5. Bright image depicting a field and containing a small-sized plane forgery.



(a) Small forgery **I**

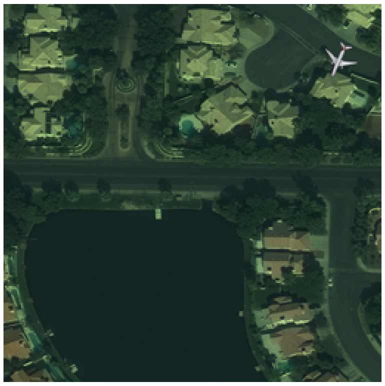


(b) Small forgery **M**

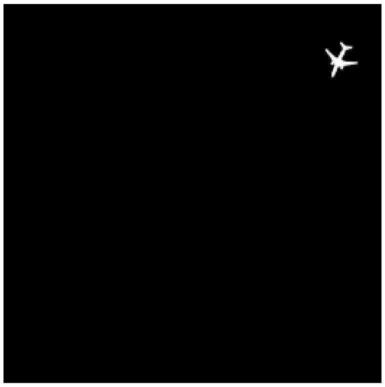


(c) Small forgery $\hat{\mathbf{M}}$

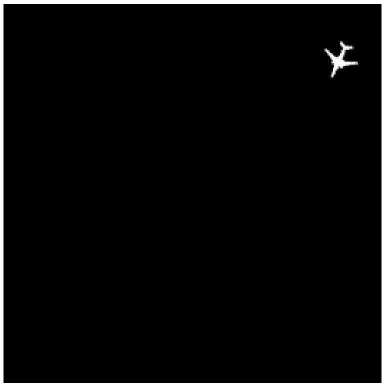
Fig. 4.6. Dark image depicting houses and containing a small-sized plane forgery.



(a) Medium forgery **I**

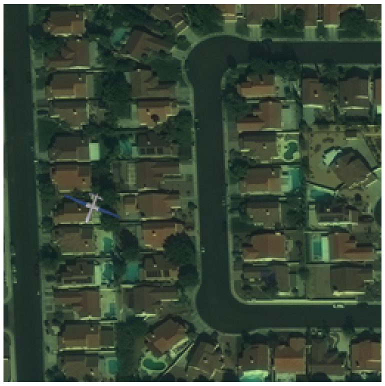


(b) Medium forgery **M**

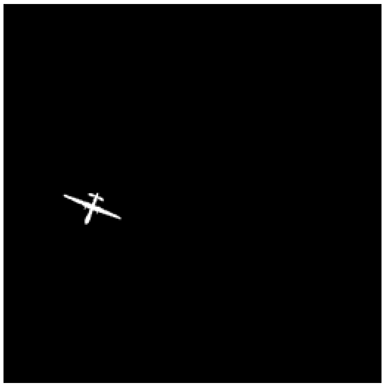


(c) Medium forgery $\hat{\mathbf{M}}$

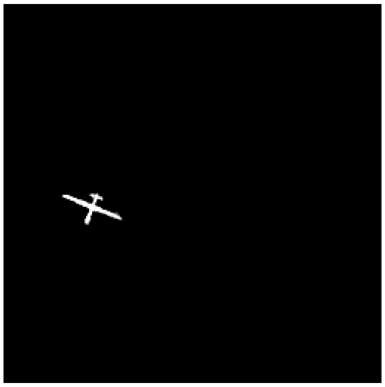
Fig. 4.7. Bright image depicting houses and containing a medium-sized plane forgery.



(a) Medium forgery **I**

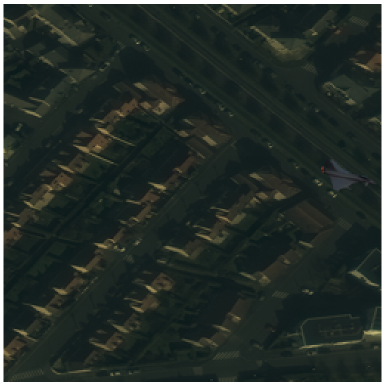


(b) Medium forgery **M**

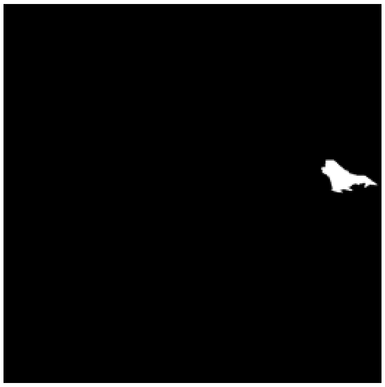


(c) Medium forgery $\hat{\mathbf{M}}$

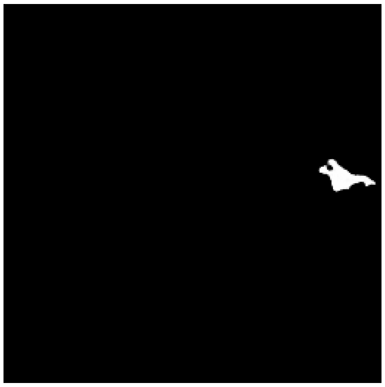
Fig. 4.8. Bright image depicting houses and containing a medium-sized plane forgery.



(a) Medium forgery **I**

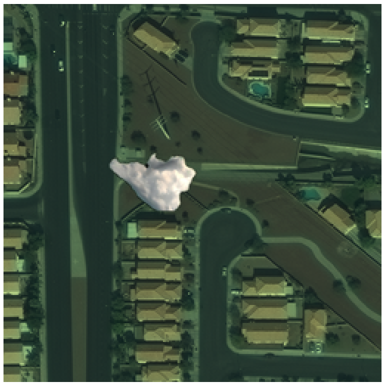


(b) Medium forgery **M**

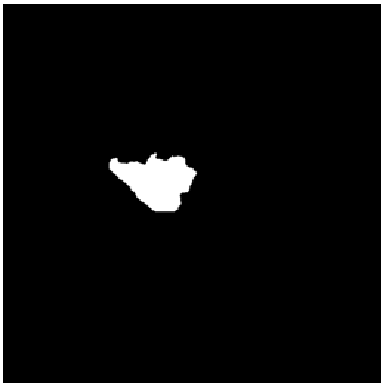


(c) Medium forgery $\hat{\mathbf{M}}$

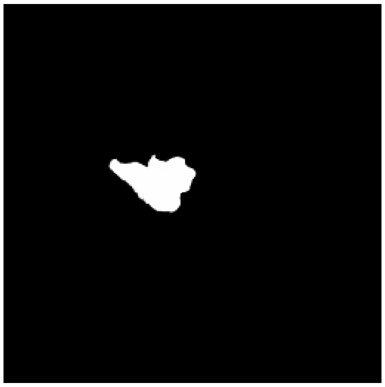
Fig. 4.9. Dark image depicting a road and containing a medium-sized plane forgery.



(a) Large forgery **I**

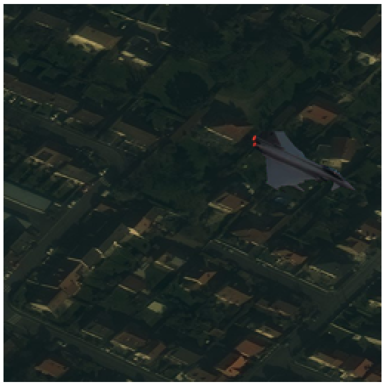


(b) Large forgery **M**

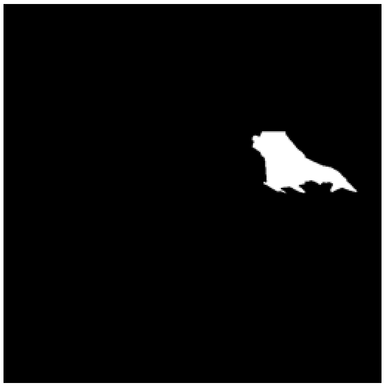


(c) Large forgery $\hat{\mathbf{M}}$

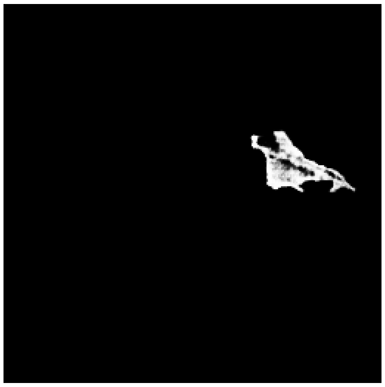
Fig. 4.10. Bright image depicting a field and containing a large-sized cloud forgery.



(a) Large forgery **I**

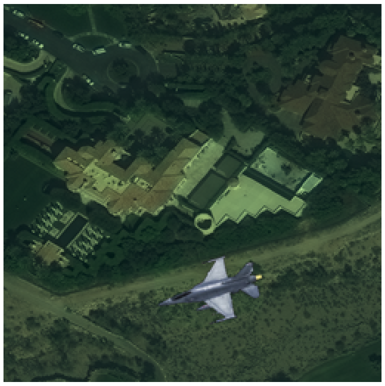


(b) Large forgery **M**

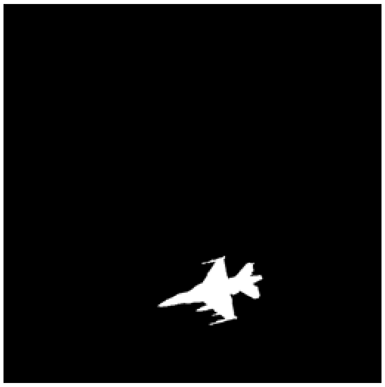


(c) Large forgery $\hat{\mathbf{M}}$

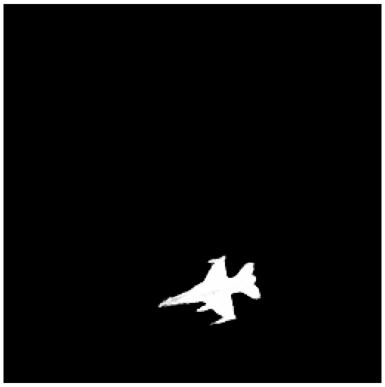
Fig. 4.11. Dark image depicting houses and containing a large-sized plane forgery.



(a) Large forgery **I**



(b) Large forgery **M**



(c) Large forgery $\hat{\mathbf{M}}$

Fig. 4.12. Bright image depicting a field and containing a large-sized plane forgery.

4.2 Numerical Analysis of Results

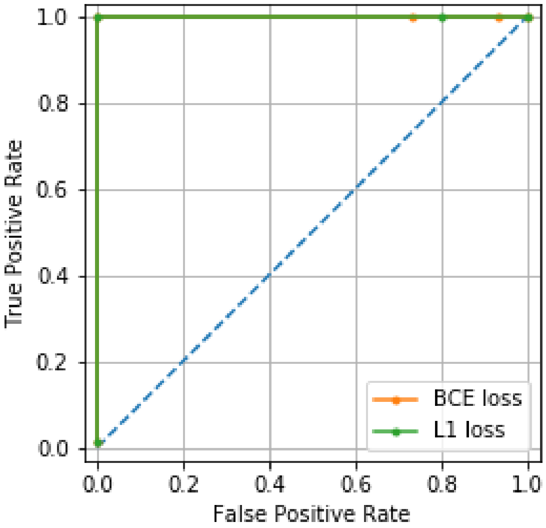
To evaluate forgery detection, each image's classification label (i.e. forged or pristine) must first be determined based on its generated mask estimate $\hat{\mathbf{M}}$. In order to accomplish this, the average pixel value of a mask estimate is employed. This value is defined as

$$\hat{\mathbf{M}}_{avg} = \frac{1}{X \cdot Y} \sum_{x=1}^{X} \sum_{y=1}^{Y} \hat{\mathbf{M}}(x, y), \tag{4.1}$$

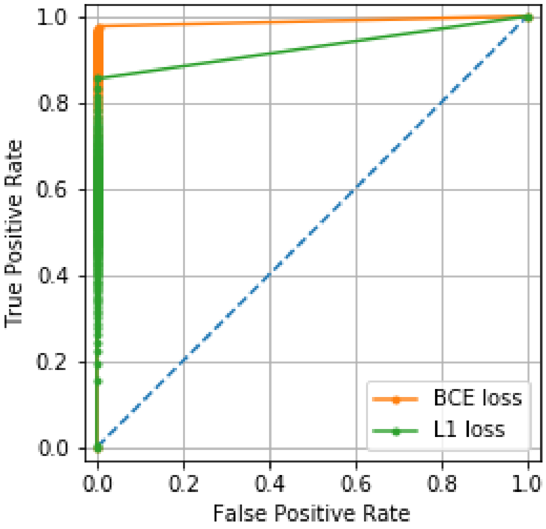
where $X \times Y$ is the image resolution. Then, binary thresholding with threshold T is used to determine whether the image under analysis $\mathbf{I}$ is pristine or forged. As mentioned previously, an image is considered pristine when $\hat{\mathbf{M}} \approx 0$. From a thresholding standpoint, this is achieved when $\hat{\mathbf{M}}_{avg} < T$. Thus, all images that meet this condition are labeled as pristine. Otherwise, they are labeled as forged. Next, labels assigned to each mask estimate are compared to the ground truth labels to determine the performance of this method in regard to detection.

To assess forgery localization, a similar evaluation process occurs; however, only for images in which forgeries are detected. Each mask estimate $\hat{\mathbf{M}}$ is thresholded and then undergoes a pixel-wise comparison to its corresponding ground truth mask $\mathbf{M}$ to evaluate splicing localization.

Figure 4.13 shows the receiver operating characteristic (ROC) curves for our results, which reveal the performance of different thresholds T for both the detection and localization objectives. This figure also illustrates model performances achieved when using BCE loss and L_1 loss for reconstruction. For the detection objective, the areas under the curve (AUC) for both BCE and L_1 loss are 1.000, indicating that it is possible to achieve perfect detection accuracy with thresholding using either BCE loss or L_1 loss for $\mathcal{L}_R$. These results are further verified by the precision-recall (PR) plot in Figure 4.14 for a model using BCE loss. It too indicates that perfect detection is possible with our 2-class model, as its average precision score is also 1.000. On the other hand, a difference in model performance is observed based on the localization objective depending on whether BCE or L_1 is used for $\mathcal{L}_R$. BCE yields a higher AUC value of 0.988 in comparison to L_1, which achieves an AUC of 0.927. The PR curve (again using BCE loss) with an average precision score of 0.953 confirms that localization results are very good. Based on these results, the final model implements BCE loss for reconstruction.

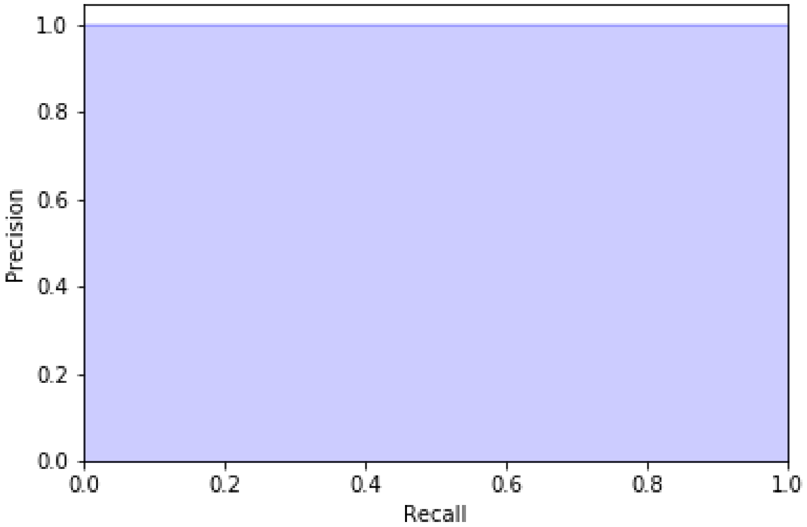


(a) ROC curves for forgery detection depicting comparison between BCE and L_1 loss for $\mathcal{L}_R$

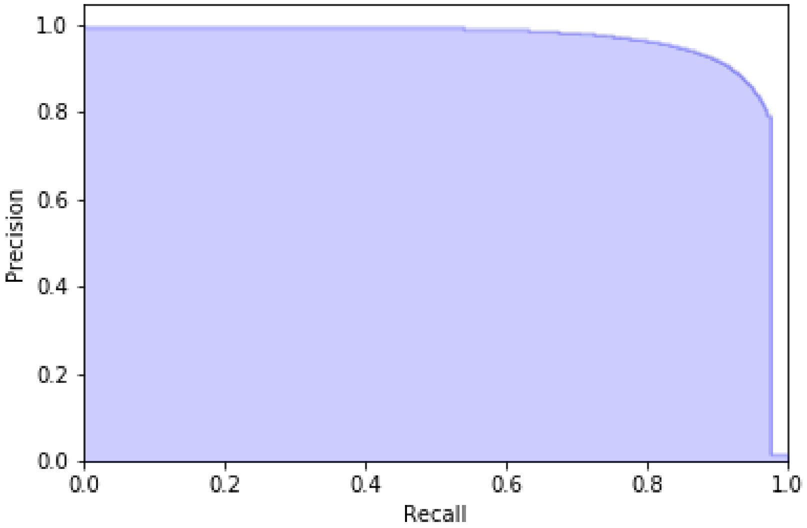


(b) ROC curves for forgery localization depicting comparison between BCE and L_1 loss for $\mathcal{L}_R$

Fig. 4.13. ROC curves for detection and localization of spliced forgeries



(a) PR curve for forgery detection using BCE loss for $\mathcal{L}_R$



(b) PR curve for forgery localization using BCE loss for $\mathcal{L}_R$

Fig. 4.14. PR curves for detection and localization of spliced forgeries

5. CONCLUSION

5.1 Summary

In this thesis, we propose a forensic image analysis method based on a cGAN for splicing detection and localization in satellite images. The proposed technique exploits a data driven approach. Thus, it learns how to distinguish forged regions from pristine ones directly from the available training data. Results show that the developed methodology accomplishes both tampering detection and localization with incredibly high accuracy on the used dataset. Moreover, it is interesting to notice how the proposed solution generalizes to forgeries of different sizes than those seen during training.

5.2 Future Work

While the results of this experiment look very promising, more work needs to be done to improve the robustness of the system. A first step would be to investigate how our method performs on datasets containing images coming from different satellites, since the experiments mentioned in this thesis utilize images all captured from the same satellite. The next step would be to extend the method to different types of forgeries, such as copy-move and inpainting forgeries. A third improvement would be to separate the training and testing datasets based on the object spliced into the images. For example, the training data would only contain plane objects while the testing data would only contain cloud objects. In creating this dataset, it would be good to include pristine images containing planes and clouds to verify that the method only flags spliced objects as forgeries. A final suggestion for future work is to apply this method to different types of satellite imagery, including Synthetic Aperture

Radar (SAR) and Multispectral Imagery (MSI). It would be interesting to see how the technique performs in all of these scenarios and to further test our method's generalization capabilities.

5.3 Contributions of this Thesis

In this thesis, we developed a Conditional Generative Adversarial Network (cGAN) to detect and localize spliced forgeries in satellite images. We demonstrated high performance of the method on the used dataset.

5.4 Publication Resulting from the Thesis

- **E. R. Bartusiak**, S. K. Yarlagadda, D. Gera, F. M. Zhu, P. Bestagini, S. Tubaro, E. J. Delp. Splicing Detection And Localization In Satellite Imagery Using Conditional GANs. *IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR)*, March 2019. San Jose, CA.

REFERENCES

REFERENCES

[1] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein, "Vision of the unseen: Current trends and challenges in digital image and video forensics," *ACM Computing Surveys*, vol. 43, pp. 1–42, Oct. 2011.

[2] A. Piva, "An overview on image forensics," *ISRN Signal Processing*, vol. 2013, p. 22, Nov. 2013.

[3] M. C. Stamm, Min Wu, and K. J. R. Liu, "Information forensics: An overview of the first decade," *IEEE Access*, vol. 1, pp. 167–200, May 2013.

[4] "Conspiracy files: Who shot down MH17?" BBC News *http://www.bbc.com/news/magazine-35706048*, (Accessed on 12/01/2018).

[5] "Satellite images show clearly that russia faked its MH17 report," Mashable *http://mashable.com/2015/05/31/russia-fake-mh17-report*, (Accessed on 12/01/2018).

[6] "15 free satellite imagery data sources," GIS Geography *http://gisgeography.com/free-satellite-imagery-data-list*, (Accessed on 12/01/2018).

[7] M. Barni, A. Costanzo, and L. Sabatini, "Identification of cut&paste tampering by means of double-JPEG detection and image segmentation," *Proceedings of the IEEE International Symposium on Circuits and Systems*, May 2010, Paris, France.

[8] M. Kirchner and T. Gloe, "Forensic Camera Model Identification," in *Handbook of Digital Forensics of Multimedia Data and Devices*. John Wiley & Sons, Ltd, 2015.

[9] D. Cozzolino, G. Poggi, and L. Verdoliva, "Splicebuster: A new blind image splicing detector," *Proceedings of the IEEE International Workshop on Information Forensics and Security*, Nov. 2015, Rome, Italy.

[10] L. Bondi, S. Lameri, D. Güera, P. Bestagini, E. J. Delp, and S. Tubaro, "Tampering detection and localization through clustering of camera-based CNN features," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1855–1864, July 2017, Honolulu, HI. [Online]. Available: dx.doi.org/10.1109/CVPRW.2017.232

[11] A. T. S. Ho, X. Zhu, and W. M. Woon, "A semi-fragile pinned sine transform watermarking system for content authentication of satellite images," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Jan. 2005, Seoul, Korea.

[12] S. K. Yarlagadda, D. Güera, P. Bestagini, F. M. Zhu, S. Tubaro, and E. J. Delp, “Satellite image forgery detection and localization using GAN and one-class classifier,” *arXiv:1802.04881*, Jan. 2018.

[13] L. Ali, T. Kasetkasem, F. G. Khan, T. Chanwimaluang, and H. Nakahara, “Identification of inpainted satellite images using evalutionary artificial neural network (EANN) and k-nearest neighbor (KNN) algorithm,” *Proceedings of the IEEE International Conference of Information and Communication Technology for Embedded Systems*, May 2017, Chonburi, Thailand.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” *Advances in Neural Information Processing Systems*, pp. 2672–2680, Dec. 2014, Montréal, Canada.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* Cambridge, MA: MIT Press, 2016.

[16] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5967–5976, Jul. 2017, Honolulu, HI. [Online]. Available: https://doi.org/10.1109/CVPR.2017.632

[17] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Oct. 2015, Munich, Germany. [Online]. Available: https://doi.org/10.1007/978-3-319-24574-4_28

[18] “Landsat on AWS,” Amazon Web Services Inc. *https://aws.amazon.com/public-datasets/landsat/*, (Accessed on 12/01/2018).

[19] “Landsat science,” National Aeronautics and Space Administration *https://landsat.gsfc.nasa.gov/*, (Accessed on 12/01/2018).

[20] “NASA,” National Aeronautics and Space Administration *https://www.nasa.gov/*, (Accessed on 12/01/2018).

[21] “Usgs.gov — science for a changing world,” U.S. Geological Survey *https://www.usgs.gov/*, (Accessed on 12/01/2018).

[22] E. R. Bartusiak, S. K. Yarlagadda, D. Güera, F. Zhu, P. Bestagini, S. Tubaro, and E. J. Delp, “Splicing detection and localization in satellite imagery using conditional gans,” *IEEE International Conference on Multimedia Information Processing and Retrieval*, March 2019, San Jose, CA.

VITA

VITA

As a proud graduate with both a Bachelor's degree and a Master's degree in Electrical Engineering from Purdue University, Emily is excited to continue contributing to technology's advancement and serving humanity through innovation.

During her collegiate career, she focused on signal processing, image processing, and machine learning through her coursework, Master's thesis, three summer internships with Qualcomm, and two summer internships with Motorola Solutions. Emily is also a member of the IEEE and the Eta Kappa Nu Honor Electrical and Computer Engineering Honor Society. Emily is a recipient of the the Eta Kappa Nu Outstanding New Member Award for her unparalleled initiative and the Purdue College of Engineering Magoon Award for excellence in teaching.

Emily has been consistently praised as persevering and willing to overcome complex challenges while exhibiting good engineering judgement and a positive attitude. In situations where the challenges do not play into her strengths, Emily's managers and advisors have commended her on her determination and ability to pick up new skills as required.

PUBLICATION

PUBLICATION

E. R. Bartusiak, S. K. Yarlagadda, D. Gera, F. M. Zhu, P. Bestagini, S. Tubaro, E. J. Delp. Splicing Detection And Localization In Satellite Imagery Using Conditional GANs. *IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR)*, March 2019. San Jose, CA.