https://doi.org/10.2352/ISSN.2470-1173.2018.07.MWSF-319
© 2018, Society for Imaging Science and Technology

Domain Adaptation in Steganalysis for the Spatial Domain

Li Lin[+]*, Jennifer Newman*[+]*, Stephanie Reinders*[+]*, Yong Guan*[†]*, and Min Wu*[⋆]
[+]*Department of Mathematics,* [†]*Department of ECPE, Iowa State University, Ames, Iowa, USA,* {*llin, jlnewman, srein, guan*}*@iastate.edu,* [⋆]*Department of ECE, University of Maryland, College Park, MD, USA, minwu@umd.edu*

Abstract

A scenario of domain adaptation (DA) in machine learning occurs when training and test data are drawn from some population with different distributions. In steganalysis, this scenario can arise when images used for training and testing come from different cameras, especially in blind detection. Although there has been some work in this area, it is still not clear that one can design a feasible detection scheme for all devices from one camera model. In this research, Spatial Rich Models (SRM) and ensemble classifiers have been applied for feature extraction and classification, respectively. After carefully collecting images from several camera models from mobile phones, with at least two devices for each model, we identify two measurable factors that affect detection: ISO speed and exposure time. This allows us to adapt the classifier from one device to a different one of the same model, even when images from the two devices are significantly different in visual appearance, by choosing specific training data. Our experiments show that a well-trained stego detector based on data from one source shows more adaptability to new target data if the training images have similar distributions of ISO speed and exposure time as the target images.

Motivation

Digital image steganalysis is the analysis of image data to discover if hidden content is contained within the image. To classify an image as cover (no hidden content, or innocent) or stego (with hidden content), many machine learning (ML) feature selection methods and classification algorithms, such as [1, 2, 3, 4], have been developed. The evaluation of such ML detectors are typically based on the empirical errors from the experiments on a given database, in which the training samples and the test samples are assumed to have the same sources. However, the rapid development in digital mobile devices brings new challenges to the steganalysis community. Individuals are more likely to take pictures using their mobile phones, and there are many different kinds of mobile phones and camera apps. Moreover, there are many stego apps that allow users to embed secret messages into images conveniently. Creating a stego image is much easier than ten years ago. Therefore, it is very necessary to extend stego detection experiments from several fixed cameras to a broader range of image sources.

To explore the possibility of identifying a stego image from (practically) unlimited sources, it is unlikely to maintain the assumption that the target device is still included in the training database. Some previous work, such as [5], defines this as the cover-source mismatch problem, and shows that in the worst case, no matter what adjustments are made to the ML algorithm, no improvement is seen without including the source data. Note that the paper [5] only discusses cases where the target camera models are not contained in the training database. It is still not clear that with knowledge of the camera model, one can successfully detect stego images from other cameras in this same model by training on one device from that particular camera model. We note here that, even after limiting to one single camera model, this problem is not trivial, since the images from the target devices are very likely to have different properties than the images in the training database. Applying a well-trained classifier directly to new data may bring unacceptably high errors. This is a typical domain adaptation problem, which aims at transferring the knowledge from a source domain to a different test domain.

Numerous progress has been made in domain adaptation and transfer learning through the past years. The formal definition of domain adaptation and its relationship to transfer learning are well explained in [6, 7, 8]. Domain adaptation has been widely applied in many fields, including speech recognition and face recognition [9, 10]. If we introduce the terms of domain adaptation into the context of steganalysis, then images collected by one device and labeled as " cover " or " stego " in the training database form a sample from one *source domain*, and unlabeled images collected by another, distinct device for testing is a sample from the *target domain*. In fact, readers will learn from our experiments described below that, if the data collection for two such devices produces data that are independent from each other, with regards to certain factors, that a classifier that successfully separates covers from stegos in the source domain, may fail to classify the stego images for the target device, even under the condition that the target camera is the same model as the training source. Thus, a very natural question is, what kind of factors can be used to describe the similarity between the source domain and the target domain.

Many factors, including the complexity of the scene contents, the saturation level of images, and the noise level of images, affect the empirical error rate of detecting stego images. Some previous work, such as [11, 12, 13], has already shown that noise levels affect the performance of an image classifier. To study if noise is also a factor which can separate the target domain from the source, we look for measurable variables that can represent the noise level of a image, rather than computed noise values. Since all computations of noise values in an image have their shortcomings, we chose to focus on having an indirect measure of the noise in an image that is part of the camera system. ISO speed and exposure time are the first two parameters we start with, as they have strong correlation to the signal-to-noise ratio (SNR) of images in auto-exposure mode [14].

In short, the goal of this paper is to explore domain adaptation problems in a more practical steganalysis setting. We outline our work in the following steps: 1. Under what conditions can domain adaptation be introduced into a more practical steganalysis framework? 2. How can we detect stego images from many var-

ious cameras, by training with image data from limited sources? 3. How well do the factors of ISO and exposure time work in creating an adaptive classifier that works on unseen devices from the same model? To answer the above questions, we construct our own dataset and design a series of well-controlled experiments using our data.

Preparation of experiments

The particular choice of image dataset plays a crucial role in any research field involving image processing and machine learning. To run experiments for domain adaptation in steganalysis, image data from various sources with rich and varied properties are very much in demand. In addition, to analyze certain factors, we require that images are collected through a series of well-controlled procedures. There are several image datasets utilized by the image forensics communities. These include BOSSbase [15], constructed for steganalysis; RAISE [16], designed for image forgery; and the Dresden Image Database [17], created for digital image forensics. Unfortunately, none of above image databases is designed for the study of domain adaptation problems. Therefore, we create a database expressly for our purposes, which we describe next.

Devices

With the development of the mobile Internet, the built-in cameras in cell phones are very commonly used to take photos on a daily basis. Mobile phones are preferred to individual digital still cameras for many reasons, including lighter weight and readiness to capture pictures; improved quality of mobile phone cameras is also leading the general public away from the use of digital still cameras. Therefore, unlike other benchmarking databases for image forensics, in which images were collected solely by digital still cameras, we choose mobile phones as the devices for our image data collection. The widespread use of mobile phones and the cameras associated with them not only allows us to collect images from a variety of sources, but also gives us a great opportunity to test the performance of stego detection algorithms on image data collected by a large number of devices.

The initial data collection for our database, called *StegoDB*, utilized six iPhones purchased for our lab representing three different phone models: two devices of the iPhone 6s model, two devices of the iPhone 6sPlus model, and two devices of the iPhone7 model. Table 1 lists some technical specifications of the camera system for these devices. In order to run experimental tests on different devices using the same camera model, we acquired two devices from the same model, and label the individual devices as indicated in the first column in Table 1.

Data Collection

After iOS 10 was announced, software engineers were able to develop third-party camera apps that allow a user to shoot photos in a manual mode and save them in raw formats. After investigating a few such apps, we chose the app "ProCam" [18] to collect data for our experiments. This app allows more convenient selection of ISO setting and exposure time, and enables us to save the raw image in .dng or .tiff formats. With ProCam installed, all six lab iPhones were assigned to different student photographers to take photos. To better control the variability of lighting conditions, we required all images to be taken indoors. More than 20 student photographers were recruited, with each photographer checking out a single device at a time to collect his or her indoor photos. A photographer was required to take photos using a specific procedure. The procedure stated that a set of 10 individual photographs were to be acquired while the camera was handpointed to a specific scene of the photographer's choice, using 10 different exposure settings: one at the camera's auto-exposure setting, and the remaining nine with all combinations of three ISO speeds: 100, 200, 1000, and three exposure times: 1/10 seconds, 1/50 seconds, and 1/200 seconds. The scene remained fixed for the 10 photos of such a set. All original images were saved in .tiff format by the app. In Figure 1, we display a sample of a set of 10 images. A minimum number of 150 sets of 10 images were collected for each iPhone, resulting in at least 1500 original photographs over the required range of exposure settings for each device.

Table 1. Camera Specification.

Device	Rear Camera	MegaPixel	Image Size	Aperture	Image Stabilization
iPhone6s-1 iPhone6s-2	Sony Exmor RS IMX315	12	4032×3024	f/2.2	Digital
iPhone6sPlus-1 iPhone6sPlus-2	Sony Exmor RS IMX315	12	4032×3024	f/2.2	Digital and Optical
iPhone7-1 iPhone7-2	2nd-generation Sony Exmor RS	12	4032×3024	f/1.8	Digital and Optical

Figure 1. *A set of 10 sample images.*

In this manner, a total of more than 10,000 original images with the native image dimensions (roughly 4000 × 3000 pixels) were collected, representing 1000 different indoor scenes with the six iPhones. To increase our sample size of cover images, we extracted five subimages of dimension 512 × 512 from each original color image, and saved them as grayscale images, as shown in Figure 2. This produced a minimum of 750 images of size 512 × 512 that we designated as cover images for each phone for each of the 10 exposure settings. Some devices had more images available, providing an excess of 2000 cover images.

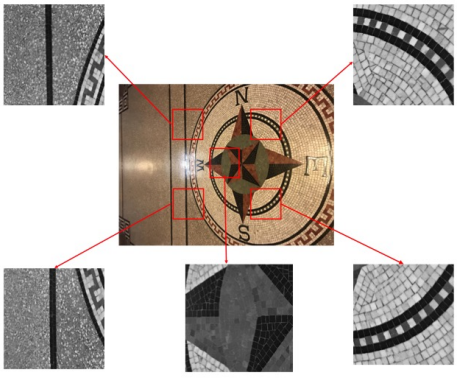

Figure 2. *Extracting five grayscale subimages from a color image.*

A sample size of two iPhone devices of the same model is not sufficient to discuss the performance of a well-trained steganalysis classifier applied to other devices from the same camera model. Thus, we performed another experiment to collect images from a large number of different devices of the same model. Selecting the iPhone7, we recruited more than 50 iPhone7 devices from volunteers (52 to be precise), and from each of the 52 devices, collected 40 original raw images under the 10 exposure settings, using the same procedure described above for our lab iPhones. With more than 2000 original images and 10,000 smaller cover images from 50 different iPhone7 devices, we are able to study the performance of a steganalysis classifier trained by images from a single device and tested on data with more than 50 different origins.

Creation of Stego Images

For our experiments, we implemented three spatial domain embedding algorithms, using the associated code available on Dr. Jessica Fridrich's website: "Wavelet Obtained Weights"(WOW) [19], "Spatial version of the Universal Wavelet Relative Distortion"(S-UNIWARD) [20], and "Minimizing the power of the most Powerful Detector"(MiPOD)[21]. Stego images were generated from the cover images using these algorithms. We fix the embedding payload rate at 0.1 bits per pixel (bpp), and leave all other parameters at the default settings, since the goal of this paper is not concerned with the security of the embedding algorithms, but simply to compare the relative effect of our experiments using state-of-the-art embedding algorithms.

Steganalysis Method

Many machine learning algorithms have proven to be the workhorse of steganalysis. Classic ML requires feature extraction, an ML algorithm, and large amounts of data. In the following experiments, the Spatial Rich Model (SRM) [2], with 34,641 features is used as the feature set. The Fisher Linear Discriminant (FLD) ensemble classifier [4] is implemented for the classification of stego images. We chose this ML procedure because this combination has been identified by the steganography community as being one of the top-ranked ML steganalysis algorithms. The performance of classifiers is evaluated by the average error on test data. That is, if we let P_{MD} denote the percentage of misdetections and P_{FA} the percentage of false alarms, then for a dataset constituting 50% cover images and 50% stego images, the average error rate P_E for the detection is defined as

$$P_E = \frac{1}{2}(P_{MD} + P_{FA}). \tag{1}$$

Discussion of Experimental Design

A discussion of the statistics involved in our experiments is informative. The design of our experiments was approached to produce meaningful results. We attempted to fix as many of the variables as reasonably possible and then vary only one or two, to observe the results and how they varied when the factors were varied. Thus, for each phone, budget limitations kept our purchase to two phones of one model. We chose to use the same camera app on all phones, and fix all settings in the app for each photo taken, excepting the two factors of ISO and exposure time that were varied. The operating systems of the phones and the app version were kept the same during the initial photo collection, although they were later updated due to security reasons. Unfortunately, the 50 phones from volunteers had a variety of different operating systems which did not match our labs' iPhones data. The choice of ISO and exposure settings were selected carefully to cover as broad a range as possible, yet still produce mostly visibly identifiable photos for one scene, thus emulating most photographs that might be taken by people (see Figure 1 for an example). The number 700 for the number of images was chosen as a compromise between the number of samples needed to produce statistically significant results (more is typically better) and the computation time required to process each image, including feature extraction, and the associated classifiers. The feature extraction was the most computationally expensive. Some initial experiments using several thousands of images did produce results with lower detection error, but, with limited resources for computation purposes, we decided ultimately to use 700 images across most of our experiments. We decided that at least 30 volunteer phones would suffice, but more was better and ultimately we collected from 52 iPhone7 devices.

Experiments on Auto-Exposure Images using Six Lab iPhones

We start our first experiments on image data collected in auto-exposure mode for the six lab iPhone devices. For each device, the original auto-exposure photos consist of more than 150 different scenes. For each embedding method, cropping each original photo into 5 smaller gray images generates 750 cover-stego pairs as the sample size (for the distribution of images) for each device. SRM and ensemble classifiers are applied for feature extraction and classification, respectively. For each device, 700 cover-stego pairs are randomly selected first, and are used to generate a stego classifier for that particular device. Before applying a classier generated by images from one device to the other five image datasets with different sources, we record the ten-fold

Table 2. Average Cross-Device Detection Error Rate, based on Image Data taken in Auto-Exposure Mode using Embedding Method MiPOD.

Training:	iPhone6s-1	iPhone6s-2	iPhone6sPlus-1	iPhone6sPlus-2	iPhone7-1	iPhone7-2
Target:						
iPhone6s-1	**14.8%(CV)**	**17.5%**	25.3%	20.1%	37.4%	42.8%
iPhone6s-2	**14.9%**	**13.7%(CV)**	33.1%	28.0%	31.3%	41.1%
iPhone6sPlus-1	23.0%	23.2%	**13.9%(CV)**	**12.1%**	47.3%	42.7%
iPhone6sPlus-2	22.0%	23.4%	**15.1%**	**11.6%(CV)**	46.7%	42.5%
iPhone7-1	46.3%	43.1%	36.4%	31.2%	**22.2%(CV)**	**39.3%**
iPhone7-2	48.3%	41.7%	40.6%	36.9%	**44.3%**	**21.8%(CV)**

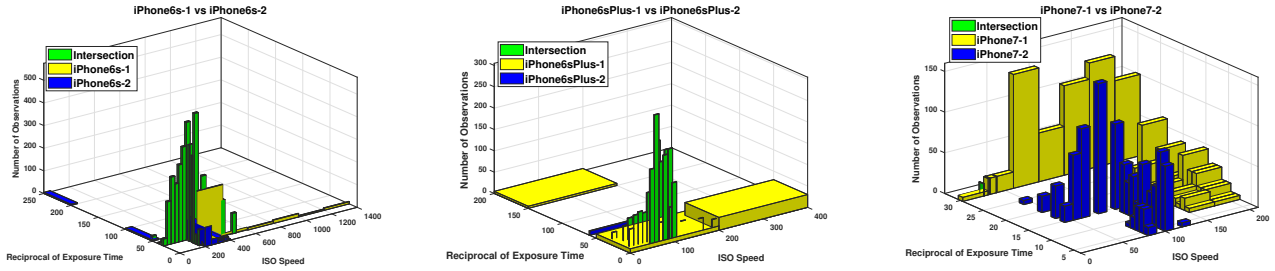

Figure 3. *2D Histogram of ISO and exposure time for original photos taken in auto-exposure mode.*

cross-validation (CV) error of classifying the stego images from the same device, and view it as a baseline of detecting stego images from the target device. The result of cross-device testing on the embedding algorithm MiPOD is provided in Table 2. We omit the error table of cross-device experiment for embedding algorithms WOW and S-UNIWARD, since they have almost the same trend as presented in Table 2.

As we can see from Table 2, for every targeted dataset in the experiment, the lowest error occurs when the training images are drawn from the same phone models as used for the test data. This is not a surprise, as mentioned earlier, some previous work on the cover-source model mismatch problem reveals that if the target data are collected by camera models that are not included in the training database, a high error rate occurs [5]. The relatively low error rates of 20%-30% in the cross-camera-model experiments on iPhone6s and iPhone6sPlus (Table 2) is consistent with the fact that the rear cameras on the iPhone6s and the iPhone6sPlus are similar (Table 1).

If we focus on the case when the training data and testing data are from the same camera models, the results are even more interesting. As Table 2 shows, applying a well-trained classifier based on data from one iPhone6s to the image data collected by the other iPhone6s, produces an error rate almost as low as the respective CV errors, and a similar case holds for the iPhone6sPlus phones. However, this is not the case for iPhone7 devices, where the cross-device errors are significantly higher than the CV error. Considering the good quality control of iPhones, we are more inclined to believe that there are some other factors related to the image noise that affects the results of the cross-testing.

The connection between the noise and the exposure settings is a well-known phenomenon in photography (see, for example, [22]). Thus, we analyze the meta data of our original images (which, although can be easily changed, we know are authenticated as we took the photos), and then display our discovery in Figure 3, in which the two-variable histogram of ISO speed and exposure time is plotted as a 3D graph for every phone model. In Figure 3, image data from the two iPhone6s' share most of the settings of ISO speed and exposure time with each other, and a similar case holds for the plot for the two iPhone6sPlus'. However, in the case of the iPhone7 devices, the settings of auto-exposure images from iPhone7-1 and iPhone7-2 barely intercept. Since the auto-exposure program usually adjusts setting parameters based on the lighting condition, it is fair to view the images collected by the device iPhone7-1 as having properties quite a bit different from the images taken by the device iPhone7-2. One possible explanation is that the student photographers for iPhone7-1 have very different hobbies in scene contents or light conditions than the student photographers for iPhone7-2. That is the reason why we need to introduce the domain adaption method in steganalysis, and even when we limit the source of data to a single camera model, such as iPhone7, a well-trained classifier is not always adaptive.

A Domain Adaptation Solution to the Cross-Device-Test Problem

In the previous experiment, a well-trained classifier based on data from one iPhone7 fails to classify the target data collected by the other iPhone7. In the language of domain adaptation, that implies the image data from these two iPhone7 devices are from two separate domains (of distribution). Figure 3 provides two parameters that separate these two image sets. One natural idea is to view these two parameters, ISO speed and exposure time, as factors that may represent the distribution domain of image data. So, our second experiment is to redo the cross-device experiment on images that have the same ISO speed and exposure time.

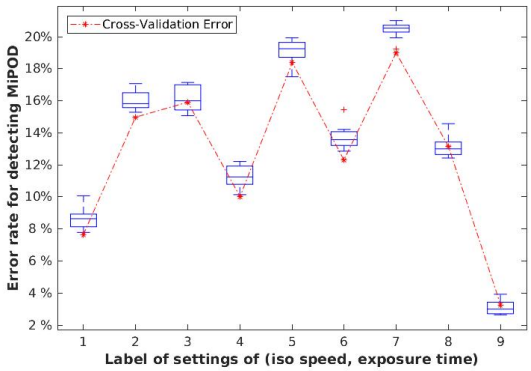

Figure 4. *Boxplots of the error rates of the adaptive classifiers tested on iPhone7-2 data v.s. the CV errors of ML classifiers on iPhone7-2 (red stars), where adaptive classifiers are trained by image data from iPhone7-1 (MiPOD).*

To that end, we first partition the images from all six iPhones into nine subsets such that all images in the same set have the same ISO speed and exposure time. Then for each subset and the same embedding method, we randomly select 700 cover-stego pairs of images taken by iPhone7-1, build a stego-detection classifier, and then test it on (a different set of) 700 pairs of images from iPhone7-2. After performing this random experiment twenty times, the errors for MiPOD at 10% embedding rate are plotted in Figure 4, in which the CV errors (red stars) are plotted as the

Table 3. Labels of 9 Sets of Different Exposure Settings.

ISO, Exposure Time:	100, 1/200 s	100, 1/50 s	100, 1/10 s	200, 1/200 s	200, 1/50 s	200, 1/10 s	1000, 1/200 s	1000, 1/50 s	1000, 1/10 s
Label:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

baseline against which to evaluate the performance of our adaptive classifiers that are plotted as boxplots. Table 3 shows the labels identifying the nine different exposure settings, which we use to efficiently label the x-axis for Figure 4. The boxplots for the other two embedding algorithms can be found in Figure 8. We use the term *adaptive classifier* here to mean that the classifier is trained on one set of data and tested on a set of data from a different device or model of phone.

In Figure 4, when we fix the ISO speed and exposure times, even the highest error rate generated by the adaptive classifier tested on the iPhone7-2 data is still below 25% on average. Note also that the boxplots of the errors made by the adaptive classifiers follow the same trend as the CV errors across all nine exposure settings. Moreover, for each setting, if we compare the value of the CV error from the iPhone7-1 data to the range of the boxplots of errors made by the adaptive classification, we discover that the adaptive classifiers work almost as well as the ML classifier this time, as compared to the data in Table 2. We omit the case when the roles of the training source and test source are exchanged between our two iPhone7 devices, since the result is quite similar. To give further experimental support to the results we see with our iPhone7 devices, we run the same experiments on the pairs of iPhone6s and iPhone6sPlus models. These results are provided in Figure 9 and Figure 10. We remark that for those interested in the relation between image noise and classification accuracy, the red stars representing the CV errors in Figure 9 and Figure 10 are also worth further study, but is not included in this paper.

Training an adaptive classifier on data from one camera device and one set of exposure settings, then testing on data from a different device but the same camera model and same exposure settings, can be viewed as a very restricted experiment, since it can be argued that this scenario does not emulate practical situations, or even that these experiments are not interesting and have obvious results. However, we contend that investigations of this sort are necessary to explore the effect of specific data that is used for training and testing steganalysis classifiers, which, in time, may lead to better understanding of the entire steganography embedding and detecting process. Access to creating this data allowed us to pursue these experiments, whose results are not necessarily so predictable.

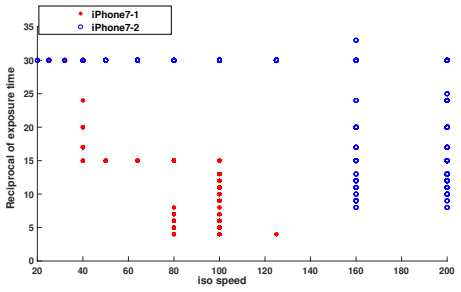

Figure 5. *Scatter plot of exposure time v.s. ISO, for images collected at auto-exposure settings by iPhone7-1 and iPhone7-2.*

Since meta or EXIF information is easily manipulated, and is not a reliable source of model information for a digital image, we have the problem of the *unknown domain* for an unknown image. In the context of domain adaptation, one method to solve this problem for the unknown domain is to develop a combination of weighted classifiers based on the current knowledge. We refer readers to the paper [7], for more rigorous details. Another way to solve this problem is to draw the training data from a large number of different sources such that the test domain is included in the training domain.

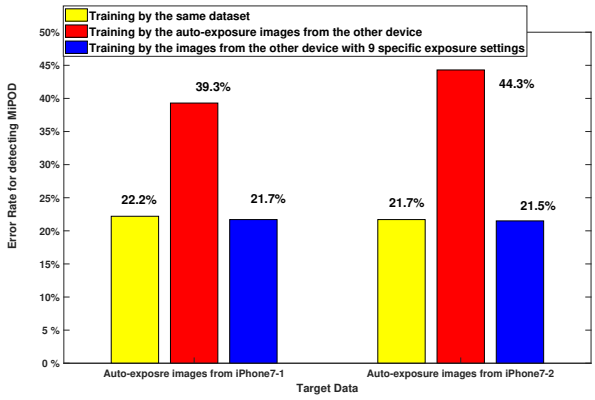

Figure 6. *Performance of the adaptive classifiers, trained on three different datasets, and tested on auto-exposure images from the two iPhone7 devices.*

For the last problem, the auto-exposure images taken by our two iPhone7 devices make a very good target source to try our idea, although with only two phones. To view the distribution of ISO and exposure time more clearly, we provide a scatter plot of these two variables in Figure 5. As we can see from Figure 5 and Figure 3, the auto-exposure images collected by iPhone7 devices have most of their ISO speed values less than 200, and their exposure times vary from 1/35 second to 1/5 second. Comparing them to the range of the fixed exposure settings in Table 3, we predict that we might see a fair performance of an adaptive classifier built by training from combinations of images from these nine subsets and then testing on the auto-exposure image data. We speculate this because the range of (ISO, exposure time) pairs for the auto-exposure images falls within the range of the (ISO, exposure time) pairs for the training data. Thus, we randomly select 700 pairs of images from the union of the nine subsets of fixed-exposure settings from one iPhone7, build a stego classifier for embedding algorithm MiPOD as before, and then test it on the auto-exposure images from the other iPhone7. This time the CV error is generated by training on the auto-exposure images for each target device. The result is presented in Figure 6. The CV error in Figure 6 is indicated by the yellow bars (and is also given in Table 2). In Figure 6, the new adaptive classifiers trained by

images with nine different exposure parameters have fairly low error rate when detecting the stego images from the other source, and are comparable to the CV error. Noticing the fact that only nine exposure settings in the (ISO, 1/exposure time) plane have been used to develop this classifier, we predict that an even better adaptive steganalysis classifier can be built by adding more training data from additional and different (ISO, exposure time) pairs, which is part of our future work.

Experiments on Image Data with Rich Origins

All results and conclusions in previous sections involve only two devices from the same camera model. In this section, we select one camera model, which is the rear camera in iPhone7, and recruit as many iPhone7 devices as possible from volunteers during a time period of several months to conduct our next experiments. The data collection is described above, in which a total number of 2000 original photos and 10,000 cover images from 52 different iPhone7 devices are collected and processed. After cleaning the data, we chose the images from one of our lab devices, the iPhone7-1, to generate nine adaptive classifiers based on a random sample of 700 cover-stego pairs of its images for each of the nine different exposure settings, and then test each adaptive classifier on 700 pairs of images from the 52 devices. The first target subset of images is a random sample of 700 pairs of auto-exposure images from the 52 iPhone7 devices. By training on 700 labeled pairs from iPhone7-1 with nine different settings, we build the classifier based on the data from just one source. The result is present in Table 4. Although there is a gap between the first two error rates in Table 4, we still believe that the performance is fair enough, especially considering that the test data are from 52 different devices and only nine exposure settings of training data have been involved in the experiment.

Another explanation for the gap between the CV error and the prediction error is that the data collected for both our lab devices iPhone7-1 and iPhone7-2 was completed 10 months before the images were collected for the 52 iPhone7 devices, and therefore the iOS versions are different, which may cause some changes in the camera APIs. In addition, the training data from iPhone7-1 are images taken from more than 200 scenes in different buildings, but the images taken by the 52 iPhone7 devices are collected in two different rooms with relatively fixed scenes.

We also tested the performance of the adaptive classifiers for each fixed (ISO, exposure time) combination in the dataset for the 52 sources. We trained nine distinct classifiers using 700 pairs of image data from the iPhone7-1 device, and tested each classifier on corresponding (ISO, exposure time) data from the 52 iPhone7 devices. We also tested an additional 700 pairs of iPhone7-1 image data, and plot this next to the error rate for the 52 iPhone7 devices. The results are summarized as bar plots of the average

Table 4. Experimental Results of Detecting MiPOD on Auto-Exposure Images from 52 iPhone7's (sample size =700)

Training Source	Test Error
Auto-Exposure Images from 52 iPhone7 devices	**26.0% (CV)**
Images with 9 exposure settings from iPhone7-1	**32%**
Auto-Exposure Images from iPhone7-1	**41%**

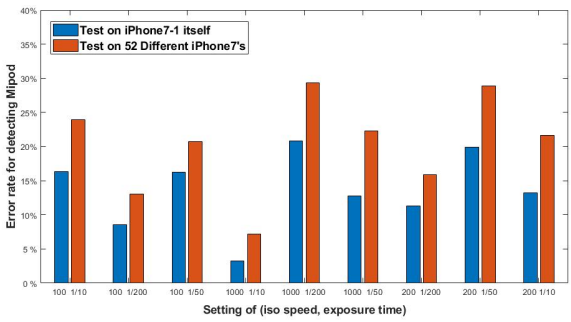

Figure 7. *Performance of the steganalysis classifiers trained on 700 pairs of images from iPhone7-1, and tested on images taken by 52 iPhone7 devices.*

error rates for 52 iPhone7 devices in Figure 7. In Figure 7, it is quite obvious that the average error rates for 52 iPhone7 devices are noticeably higher than the errors testing on iPhone7-1 data itself. But we have to point out that the test errors on iPhone7-1 data are not the CV errors for the target (52 phones) datasets. We computed the CV errors for the 52 phones, and found they are very small due to the way we collected those images. Therefore, considering the fact that greatest error for the (ISO, exposure time) setting of (1000, 1/200) is around 30% for detecting MiPOD with 10% spatial embedding for 52 targeted devices, one can not deny that our proposed adaptive method has a decent performance. However, by analyzing the meta data of the 52 iPhone7 devices, we noticed that the iOS versions for all phone are not identical. This may also play an important role in contributing to the errors, and to show a complete result with a fair sample size, we leave this topic to a future experiment.

Conclusions and Future Work

In this paper, the experiments on auto-exposure images from a pair of iPhone7 motivated us to explore domain adaptation to steganalysis. Two main exposure parameters, ISO speed and exposure time, which are well known for their relationship to the noise of images, have been taken into account as factors for building adaptive classifiers. Our experimental results include a test using 50 iPhone7 devices, and show that a well-trained stego detector, trained using data from one device, has the ability to classify fairly competently on unseen target data from the same model (but different devices) if the training images exhibit similar distributions of ISO speed and exposure time as the target images. One way to view the domain adaptation process here is that by changing the sampling procedure of the data, that is, changing which population the data was sampled from and consequently used to train the ML algorithm, a classifier is produced that is more representative of the population of data that will be tested on.

We remark that all results we show here are limited to the case when the training device and target device are from the same camera model. For a target device whose model is different from the training camera model, our preliminary experiments show the performance of the adaptive classifiers can be very terrible in some special cases, even when the target images share the exact same ISO and exposure time parameters as the training images. Considering the fact that the camera model identification prob-

lem is still at the heart of research in image forensics and camera model identification could be used as a first step in a steganography detection procedure using our approach, the assumption of having some knowledge of the target device is not very strong. One direction of our future work is to apply the knowledge of camera models to build a suitable stego image classifier for the target data mixed with unknown camera models.

We implement only three spatial domain embedding algorithms and apply the classical SRM for feature extraction. Thus, another future work is to study the domain adaptation problem for other steganography methods, especially for those embedding algorithms working in the frequency domain.

Acknowledgments

This work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement #70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia.

References

[1] Y. Miche, B. Roue, A. Lendasse, and P. Bas, "A feature selection methodology for steganalysis," in *International Workshop on Multimedia Content Representation, Classification and Security*. Springer, 2006, pp. 49–56.

[2] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.

[3] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," in *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*, Dec 2014, pp. 48–53.

[4] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, 2012.

[5] J. Kodovskỳ, V. Sedighi, and J. J. Fridrich, "Study of cover source mismatch in steganalysis and ways to mitigate its impact." in *Media Watermarking, Security, and Forensics*, 2014, p. 90280J.

[6] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct 2010.

[7] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine learning*, vol. 79, no. 1, pp. 151–175, 2010.

[8] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, p. 9, 2016.

[9] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital signal processing*, vol. 10, no. 1-3, pp. 19–41, 2000.

[10] S. Shekhar, V. M. Patel, H. V. Nguyen, and R. Chellappa, "Generalized domain-adaptive dictionaries," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 361–368.

[11] J. Fridrich, "Sensor defects in digital image forensic," in *Digital Image Forensics*. Springer, 2013, pp. 179–218.

[12] V. Sedighi, J. Fridrich, and R. Cogranne, "Toss that BOSSbase, Alice!" *Electronic Imaging*, vol. 2016, no. 8, pp. 1–9, 2016.

[13] H. Gou, A. Swaminathan, and M. Wu, "Noise features for image tampering detection and steganalysis," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 6. IEEE, 2007, pp. VI–97.

[14] International Standard Organization, "Photography – digital still cameras – determination of exposure index, iso speed ratings, standard output sensitivity, and recommended exposure index," ISO 12232:2006, 2006.

[15] P. Bas, T. Filler, and T. Pevnỳ, "Break Our Steganographic System: The Ins and Outs of Organizing BOSS," in *Information Hiding*. Springer, 2011, pp. 59–70.

[16] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "RAISE: a raw images dataset for digital image forensics," in *Proceedings of the 6th ACM Multimedia Systems Conference*. ACM, 2015, pp. 219–224.

[17] T. Gloe and R. Böhme, "The 'Dresden Image Database' for Benchmarking Digital Image Forensics," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, ser. SAC '10. New York, NY, USA: ACM, 2010, pp. 1584–1590.

[18] S. Azzam, "ProCam," https://itunes.apple.com/us/app/procam-5/id730712409?mt=8.

[19] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *2012 IEEE International Workshop on Information Forensics and Security (WIFS)*, Dec 2012, pp. 234–239.

[20] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, p. 1, 2014.

[21] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, 2016.

[22] B. London, J. Stone, and J. Upton, *Photography*. Pearson, 2017.

Author Biography

Li Lin received his B.S. degree in Mathematics from Capital Normal University, Beijing, China. He is currently pursuing the Ph.D degree in Applied Mathematics at Iowa State University, Ames, Iowa. His research interests include statistical image forensics, steganalysis, and statistical learning.

Dr. Jennifer Newman received her BA in Physics from Mount Holyoke College and her PhD in Mathematics from the University of Gainesville, FL. She is an Associate Professor of Mathematics at Iowa State University in the Department of Mathematics, her research focusing on image processing, stochastic modeling, steganalysis and image forensics. She is a member of SIAM and IS&T.

Stephanie Reinders received her BA in Journalism and Asian Languages and Literatures from the University of Minnesota (2005). After working for several non-profit organizations as an administrative assistant, she returned to school to earn a graduate degree in mathematics. She received a post-baccalaureate

certificate in Mathematics from Smith College (2013) and currently is pursuing a PhD in Applied Mathematics and Computer Engineering at Iowa State University.

Dr. Min Wu is a Distinguished Scholar-Teacher at the University of Maryland, College Park in the ECE Department. She graduated from Tsinghua University in Beijing, China, and holds a Ph.D. degree in electrical engineering from Princeton University. She leads the Media and Security Team (MAST), with main research interests on information security and forensics and multimedia signal processing. She is an IEEE Fellow for contributions to multimedia security and forensics.

Dr. Yong Guan is an Associate Professor of Electrical and Computer Engineering, the Associate Director for Research of Information Assurance Center, and the cyber forensics coordinator for NIST-CSAFE at Iowa State University. He received his Ph.D. degree in Computer Science from Texas A&M University. Supported by NSF, NIST, IARPA, ARO and Boeing, his research focuses on security and privacy issues, including digital forensics, network security, and privacy-enhancing technologies for the Internet.

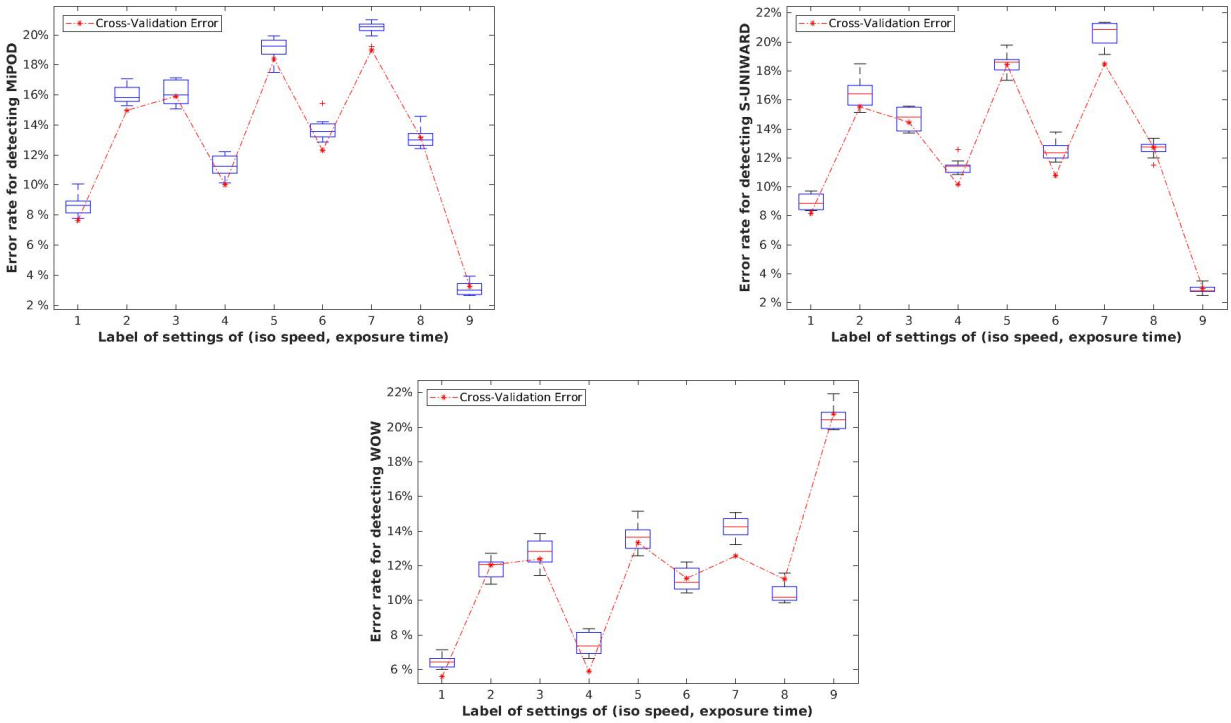

Figure 8. *Boxplots of the error rates of the adaptive classifiers tested on iPhone7-2 data v.s. the CV errors of ML classifiers on iPhone7-2 (red stars), where the adaptive classifiers are trained by image data from iPhone7-1 (S-Uniward and WOW).*

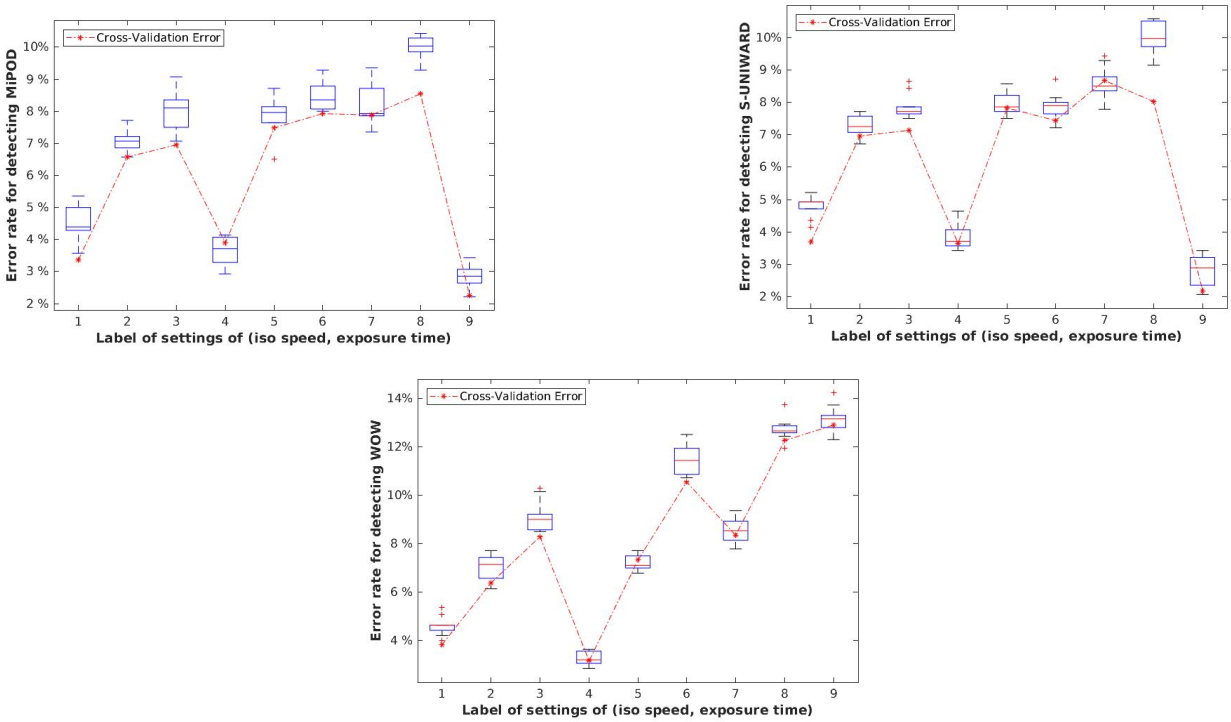

Figure 9. *Boxplots of the error rates of the adaptive classifiers tested on iPhone6s-2 v.s. the CV errors of ML classifiers on iPhone6s-2 (red stars), for three embedding algorithms, where the adaptive classifiers are trained by image data from iPhone6s-1 (MiPOD, S-Uniward and WOW).*

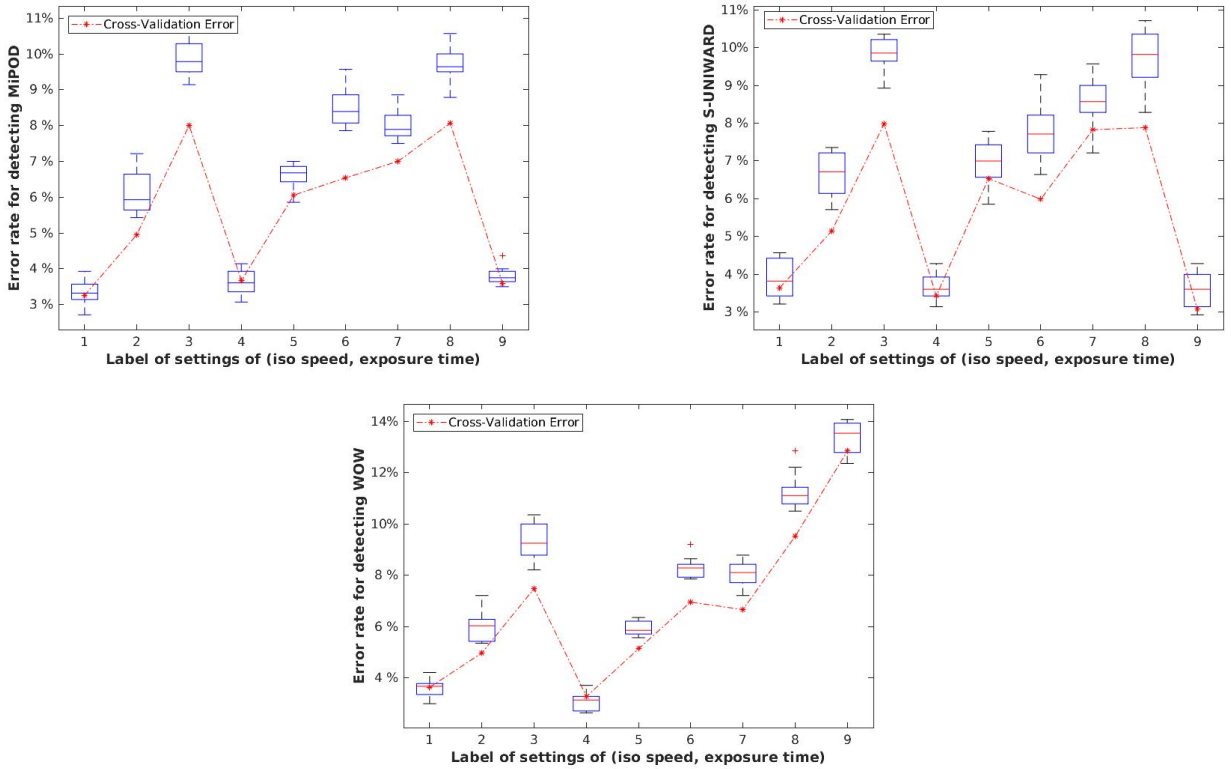

Figure 10. *Boxplots of the error rates of the adaptive classifiers tested on iPhone6sPlus-1 v.s. CV errors made by ML classifiers on iPhone6sPlus-1 (red stars), for three embedding methods, where the adaptive classifiers are trained by image data from iPhone6sPlus-2 (MiPOD, S-Uniward and WOW).*