The Fake vs Real Goods Problem: Microscopy and Machine Learning to the Rescue

Ashlesh Sharma
Entrupy Inc
New York, NY
ashlesh@entrupy.com

Vidyuth Srinivasan
Entrupy Inc
New York, NY
vidyuth@entrupy.com

Vishal Kanchan
Entrupy Inc
New York, NY
vishal@entrupy.com

Lakshminarayanan Subramanian
Entrupy Inc and New York University
New York, NY
lakshmi@entrupy.com

ABSTRACT

Counterfeiting of physical goods is a global problem amounting to nearly 7% of world trade. While there have been a variety of overt technologies like holograms and specialized barcodes and covert technologies like taggants and PUFs, these solutions have had a limited impact on the counterfeit market due to a variety of factors - clonability, cost or adoption barriers. In this paper, we introduce a new mechanism that uses machine learning algorithms on microscopic images of physical objects to distinguish between genuine and counterfeit versions of the same product. The underlying principle of our system stems from the idea that microscopic characteristics in a genuine product or a class of products (corresponding to the same larger product line), exhibit inherent similarities that can be used to distinguish these products from their corresponding counterfeit versions. A key building block for our system is a wide-angle microscopy device compatible with a mobile device that enables a user to easily capture the microscopic image of a large area of a physical object. Based on the captured microscopic images, we show that using machine learning algorithms (ConvNets and bag of words), one can generate a highly accurate classification engine for separating the genuine versions of a product from the counterfeit ones; this property also holds for "super-fake" counterfeits observed in the marketplace that are not easily discernible from the human eye. We describe the design of an end-to-end physical authentication system leveraging mobile devices, portable hardware and a cloud-based object verification ecosystem. We evaluate our system using a large dataset of 3 million images across various objects and materials such as fabrics, leather, pills, electronics, toys and shoes. The classification accuracy is more than 98% and we show how our system works with a cellphone to verify the authenticity of everyday objects.

KDD '17, , August 13-17, 2017, Halifax, NS, Canada

ACM ISBN 978-1-4503-4887-4/17/08…$15.00
https://doi.org/10.1145/3097983.3098186

CCS CONCEPTS

• **Security and privacy → Biometrics**; • **Computing methodologies → Biometrics**; *Neural networks*;

KEYWORDS

physical authentication, convolutional neural networks, microscopy, computer vision

ACM Reference format:
Ashlesh Sharma, Vidyuth Srinivasan, Vishal Kanchan, and Lakshminarayanan Subramanian. 2017. The Fake vs Real Goods Problem: Microscopy and Machine Learning to the Rescue. In *Proceedings of KDD '17, August 13-17, 2017, Halifax, NS, Canada,* 9 pages.
https://doi.org/10.1145/3097983.3098186

1 INTRODUCTION

Counterfeit goods represent a massive worldwide problem with nearly every high valued physical object or product being directly impacted by this issue. Anecdotal evidence and business reports point to counterfeit trade representing a significant 7% of world trade today [4]. It is also a known fact that the profits garnered by counterfeiters across a variety of markets has been one of the important funding sources for several illegal and potentially harmful activities around the world [3].

Fighting the battle against counterfeit goods has been a never-ending one. A wide-range of overt and covert technological solutions have been proposed to address the counterfeit detection problem [23–25, 29, 38, 41]. Overt technologies such as different versions of holograms [9], barcodes and RFID supply integrity solutions [12] physically add a tag to a physical object and each of these solutions have their own set of security limitations where the tag can be removed, forged, potentially duplicated or replaced on other physical goods. Covert technologies such as taggants [10] and Physical Unclonable Functions (PUFs) [6] offer stronger authenticity guarantees but these solutions are often expensive or even hard to adopt; across many markets for high valued objects, manufacturers may place objections to adding covert solutions to the physical objects especially for luxury, fashion or art items.

We use the term *authenticating a physical object* to refer to the task of identifying whether a physical object is genuine or counterfeit. In this paper, we introduce a novel and robust method to

authenticate physical goods in the microscopic regime using machine learning techniques. The key idea of our solution is based on the premise that objects manufactured using prescribed or standardized methods tend to have similar visual characteristics at a microscopic level compared to those that are manufactured in non-prescribed methods, which are typically counterfeits. Using these characteristics, distinct groups of objects can be classified and differentiated as authentic or not authentic.

Given a physical object of a particular product line, we are focused on developing a non-intrusive solution to easily distinguish authentic versions of the product produced by the original manufacturer and fake versions of the product produced by counterfeiters. The basic building block of our solution is to extract *microscopic image* of a physical object and use them to differentiate genuine versions of a product from counterfeit ones. To build a highly accurate classification engine, our solution takes a supervised approach where we train a machine learning algorithm against a known set of genuine and counterfeit products of a particular product line. Essentially, the machine learning algorithm is trained on a high-dimensional feature space on the classification task of learning the differences in the microscopic characteristics of genuine and counterfeit versions of the same product.

Capturing high quality microscopic images in a consistent fashion, especially at large magnifications (100-300x) is a potentially complicated and arduous task for a user. At high magnifications, conventional microscopes also have a limited field of view. To address these limitations, we designed a low-cost, high quality portable microscope that can capture a wide-angle, high resolution microscopic image of a physical object. This is a key building block for our solution since it enables a single image to capture microscopic features over a larger surface area of a physical object.

Given a few genuine and counterfeit objects of a particular product line, we use the portable microscopic hardware to generate a large labeled training set of microscopic images gathered from different regions within each object. We trained two different classes of supervised machine learning algorithms with a different set of features for the classification task. In the first algorithm, we use an SVM based classification engine using bag of visual words by extracting features based on histogram of oriented gradients, performing quantization and using a SVM/kNN classifier. To improve upon the classification accuracy, the second algorithm trained a deep (multi-layered) convolutional neural network (CNN) architecture to classify fine-grained features, and then used region based CNN to select regions in an image to improve classification accuracy.

We evaluate our technique using a large dataset of 3 million microscopic images (counting data augmentation) extracted from our microscopic imaging device. The various types of materials we classify include 20 types of leather, 120 types of fabrics, 10 types of paper, 10 types of plastic, authentic and fake NFL jerseys and 2 types of Viagra pills. Using the SVM based algorithm, we achieve a classification accuracy of more than 95% for authenticating based on a single microscopic image of a new object and the CNN based algorithm offers an accuracy of 98%. We note that in practice to authenticate a physical object, one would test with several microscopic images gathered from random positions within the physical object, which should substantially increase the overall classification accuracy of a given object. We demonstrate the practical value of our system by showing how it can be applied to authenticate luxury handbags and we are working towards deploying our system to be used by luxury resale vendors.

2 RELATED WORK

In the object authentication space, the related work can be categorized into two sets. i) Object authentication using over and covert technology, ii) Image classification using machine learning.

Object authentication using overt and covert technologies: There are several overt technology that are used in authentication of physical products. Taggants such as holograms, barcodes, specially manufactured paper [24, 25, 38], various types of ink [23–25, 29, 38, 41], special taggants such as ProofTag [10] and Cryptoglyps [1]. There are few companies that operate in the microscope based forensic analysis sector such as IQStructures [7] that use microscopes in forensic analysis to identify specific embedded structures. Recently, World Health Organization released a report on anti-counterfeiting technology to protect medicines [16]. Microtaggants is one of the solutions which can be detected by using microscopes to identify a specific class of pills. Nano-printing is also used in identifying classes of medicines in packaging.

In covert techniques there is Laser Surface Authentication from Ingenia Technology [6, 18], fiber fingerprinting by Smith*et. al.* [22, 36], PaperSpeckle by Sharma *et al.* [34], print signatures work by Zhu *et.al.* [42] and modeling 3D fiber structure of paper by Clarkson *et. al.* [20].

Our technique differs from the above in three significant ways. i) In the overt/covert techniques mentioned, they need to apply their technique at the source of creation or manufacturing of the product. Whereas in our case, we don't need to be present at the source of manufacturing of the product. Unlike overt technologies such as inks, barcodes, holograms, micro-structures etc., we do not need to embed any substance within the product or object. Our technique is non-invasive and does not modify the object in any way. ii) There is no need to tag every single item. We can classify original and duplicate based on the microscopic variations. iii) Current overt/covert authentication techniques cannot authenticate objects there were not tagged earlier. In our case, since we use machine learning techniques, we can authenticate new instances of the object. iv) Most of techniques such as nano-printing, micro-taggants are expensive to embed onto the product. Plus their detection based on specialized, expensive microscopic handheld devices which is a problem in consumer/enterprise adoption. Our device and cloud based authentication solution that works with a mobile phone is low cost and simple to use (as described in Section 6).

Image classification using machine learning: Supervised, semi-supervised and unsupervised learning techniques are used in large scale classification of images. Bag-of-words [19, 31] and Convolutional neural networks (CNNs) [27, 30, 32] are two important techniques in large scale image classification. Recently, deep CNNs such as Googlenet [37], VGGnet [35], Residual Nets [26] have given state-of-the-art performance in Imagenet [21] classification and localization tasks. Fine grain visual classification is an active area of research where fine grained visual classification has been

applied to datasets ranging from birds [17] to aircrafts [33]. Our approach differs from the above in three ways: i) Feature extraction and training to identify microscopic variations, ii) classifying microscopic images of objects based on the mid-level and fine-grained features, iii) using a combination of techniques (bag of words, deep convolutional nets) and microscopic imaging hardware in order to authenticate objects.

3 THREAT MODEL

The game of counterfeit goods production has been alive for centuries. Our goal in this paper is to design a non-intrusive counterfeit detection system that significantly makes it harder for the counterfeiter to design non-authentic goods that resemble the original. The key intuition of our system design is to extract features of a physical object at a granularity at least an order smaller the precision of manufacturing used by both the original manufacturer and the counterfeiter. Specifically, we aim to examine physical objects at a microscopic granularity with a magnification of $100 - 300x$ with a precision of 5 microns. This precision typically represents an order of magnitude or more compared with the mechanical precision of manufacturing of most physical goods. For a certain class of electronic goods including chips and circuits that rely on nano-fabrication techniques, our counterfeit detection solution may not be applicable. The first assumption in our threat model is that the precision in manufacturing of a physical good that we aim to authenticate should be at least 1-2 orders of magnitude lower than the precision of microscopic imaging of the physical surface of the good. One area we are specifically interested is luxury and fashion goods. Given that many high end goods are tailor-made, the precision assumption holds for most of these objects.

Secondly, we empirically show based on a wide range of objects that objects manufactured using prescribed or standardized methods tend to have similar visual characteristics at a microscopic level. This is the key property we exploit in our system design. These visual characteristics often represent random patterns or contours that appear over a very small area in a microscopic image. Here, we assume that given that the limitation in the precision of manufacturing of a counterfeiter, it becomes a very expensive proposition for the counterfeiter to replicate the same visual patterns in his manufacturing. In theory, a counterfeiter with a large amount of resources and access to very high precision manufacturing machinery can generate a counterfeit object with similar microscopic visual characteristics as the original object. We assume that common-case counterfeiters who aim to make a quick buck do not have the resources to launch such an attack on our system.

Finally, our system aims to deal with different levels of counterfeit sophistication ranging from fakes to super-fakes. The most common case counterfeiter assumes no direct access to the supply chain of the original product line and simply aims to replicate/reverse engineer the manufacturing process. However, in the case of luxury handbags, a sophisticated counterfeiter can get access to the original fabric from the tannery but only changes the final production process. For our counterfeit detection system, we assume that while even portions of a physical object may potentially be from the original supply chain, there are at least a few well defined regions of a physical object that are modified by the counterfeiter using a different process from that of the original manufacturer. For example, an LVMH super-fake luxury bag may use the same canvas as the original but the counterfeiter may use a different process for printing the logo on the bag. In the extreme case, if the counterfeiter infiltrates the final supply chain of a product and gets access to originals to be sold in the counterfeit market (due to a rogue employee), our solution is not designed to detect such counterfeits.

4 WIDE-FOV MICROSCOPIC IMAGING

A key building block of our system for authenticating physical objects is the use of microscopic images. Capturing high quality microscopic texture images in a consistent fashion, especially at large magnifications (100-300x) is a potentially complicated and arduous task for a user. At high magnifications, conventional microscopes also have a limited field of view. In this section, we provide a brief outline of our wide-area microscopic imaging device and how it helps in capturing consistent microscopic images. Figure 1 shows the handheld device. The key aspects of our microscopic imaging hardware are: wide field of view, high magnification and high resolution.

4.1 Need for Microscopy

In the counterfeiting industry, most of the counterfeits are manufactured or built without paying attention to the microscopic details of the object. Even if microscopic details are observed, manufacturing objects at a micron or nano-level precision is both hard and expensive. This destroys the economies of scale in which the counterfeiting industry thrives. Hence we use microscopic images to analyze the physical objects.

Microscopic images are different from ordinary macroscopic images or photos due to a variety of factors, i) complex multiple scattering in the medium produces artifacts not present in the media –speckles, shadows and inter-reflections ii) images vary based on the reflective or diffuse nature of the medium, iii) repeatedly obtaining or registering the same image is difficult due to its microscopic size. While there are off-the-shelf microscopes [15], the images are inconsistent and the build quality is low. If the input image to system is inconsistent and of low quality, then microscopic image analysis also will be of low quality. Due to address this problem, we have designed our own handheld microscope device.

4.2 Special Features of our Device

Due to space constraints, we summarize the salient features of our device which is currently deployed and sold as part of our authentication service.

(1) **Wide angle Field-of-View(FOV)**: A traditional microscope that provides a magnification of 200-300x has a limited field of view of 2mm x 2mm. In contrast, our device can provide a FOV of 1cm x 1cm under 200-300x magnification with very low distortion and no barrel effects where the corner parts of the microscopic image can often be blurred in traditional microscopes.
(2) **Diffraction-limited**: The device is designed such that the for the field of view the imaging system is diffraction limited.

Figure 1: Image of the device

(3) **High resolution, low distortion**: The device is designed to almost match the theoretical capability of the lens configuration. This gives us high resolution in the order of 7 microns or less. The resolution can be further improved by varying the focal length to achieve higher resolution. The field curve and distortion are minimal (within 1%).
(4) **Asymmetrical Double Gauss architecture**: The lens architecture reduces optical aberrations and distortions (such as barrel effect, astigmatism, pin cushion effect, spherical distortion, chromatic aberration and field curvature)
(5) **Portability and ease of use**: The device is portable (size: 50mmx50mmx100mm) and can be used by anyone with little or no help. This is in stark contrast to microscopes that require precision stand and table to extract images. Our device can be simply placed on top of the object without any stand or table and the image can be extracted using a cellphone or computer. Since the focus is fixed the object's surface will always be in focus. Due to this feature the user is easily able to register and verify the authenticity of objects.
(6) **Uniform illumination**: We have used a light guide based diffusion technique to provide uniform illumination to the field of view. This helps us in providing consisting results across devices.

5 ALGORITHM

We use two types of techniques to identify the authenticity of a physical object. One is the bag of features or bag of visual words classifier, which comprises of a traditional feature detector, quantization scheme and an SVM classifier to categorize images. Another is the Convolutional Networks framework which uses convolution and pooling layers with gradient descent to classify images. In this section we describe these two types of classification techniques and how we use them in the context of classifying authentic versus fake objects.

5.1 Bag of visual words

We use a five stage process in classifying microscopic images to verify authenticity. i) Extract features of microscopic images using a patch, corner or blob based feature descriptor, ii) quantize the descriptors such that the nearest neighbors fall into the same or nearby region (bag), which form the visual words iii) histogram the visual words based on the descriptors of the microscopic images, iv) use a kernel map and linear SVM to classify the images (as either authentic or fake) v) during the testing phase, a new microscopic image is classified using the same procedure to verify if the image is authentic or not. Some of the the hyperparameters that need to optimized to achieve high level accuracy are: the level of quantization, feature extraction parameters such as number of features, location, feature size and number of visual words.

Feature extraction. Once the image is captured using the microscope imaging hardware, it is split into chunks of smaller images for processing. Splitting an image into smaller chunks is important for multiple reasons: i) the field of view of the microscopic imaging hardware is large (compared to off-the-self digital microscopes) around 12mm x 10mm. We need to look at microscopic variations at the 5μm-10μm range, so the images have to be split such that we are able to process these minor variations. ii) splitting the image into smaller chunks helps in preserving the minor variations in the visual vocabulary. Since during the quantization process minor variations of the image tend to get lost.

Each image chunk or patch is then processed using a dense feature descriptor such as DSIFT or DAISY to extract the descriptors. Dense feature descriptors are important because we would want to capture the microscopic variations across the entire image rather than at specific interest points as in traditional object detection. Since both DSIFT and DAISY do not process the image at multiple scales, we scale the image at different levels and feed it into DSIFT/DAISY system. Features are extracted at various step sizes ranging from 4 to 12 pixels and descriptors are generated at every location. The step size affects the sensitivity of the classifier to classify fine-grained features. If the step size is small 4 pixels, then more number of features are generated which capture more fine-grained details in an image. The DSIFT descriptor is a 128 dimensional vector and while the DAISY descriptor varies based on the rings and histograms per ring, we have used a 2-ring 6-histogram per ring, 104 dimensional DAISY descriptor.

Quantization. The descriptors are clustered using k-means clustering based on the number of visual words. The number of visual words which are essentially the number of clusters is used to control the granularity required in forming the visual vocabulary. For example, in hierarchical image classification, at a higher level with inter-object classification the vocabulary can be small; while in fine-grained image classification such as in this case, the vocabulary needs to be large in order to accommodate the different microscopic variations. Hence, we don't use a fixed number of visual words, but a range so that we capture the diversity in microscopic variations. This means running k-means clustering for a range of clusters instead of a fixed sized cluster. The k-means cluster centers now form the visual vocabulary or codebook that is essential in finding whether a reference image as enough words to classify it as authentic (or fake).

Histogram of visual words and classification. The next step in the algorithm is to compute the histogram of visual words in the image chunk. The keypoint descriptors are mapped to the cluster centers (or visual words) and a histogram is formed based on the frequency of the visual words.

Once we have the visual words of training images, we use SVM to train the system. We use three types of SVMs: i) linear SVM, ii) non-linear RBF kernel SVM, and iii) a χ^2-linear SVM.

While linear SVM is faster to train, the non-linear and the χ^2-linear SVM provide superior classification results when classifying

large number of categories. We train the images using one vs. all classification, but this approach becomes unscalable as the training set increases (number of categories increase). Another approach is the one vs. one approach where the pairs of categories are classified. We employ both the approaches and see that both provide comparable performance under different scenarios.

We also apply the k-nearest neighbor(k-NN) classifier to the histogram. While k-NN classifier works well for small datasets, for large datasets that the memory requirements for near neighbor search becomes unwieldy. The bag of visual words results for both SVM and k-NN based techniques are presented in the evaluation section.

5.2 Convolutional Neural Networks

Feature extraction in object recognition tasks using bag of visual words method involves identifying distinguishing features. Hand crafted feature extraction using DSIFT, DAISY and other techniques are often used. If the image statistics is already known then hand-crafting features would be ideal. But if the image statistics are unknown then hand-crafting features would be a problem since it is unclear what would be the set of distinguishing features –features that help to classify the image. Both fine-grained and macro features in an image might be lost because the hand crafted feature might fail to identify them as regions or points of interest. To avoid this issue in classifying microscopic images, we use Convolutional Neural Networks (CNN) which is fast becoming a commodity in image classification.

We describe two types of convolutional neural networks (CNN) architecture that achieves a high level of accuracy across the datasets of various microscopic images of materials. Our CNN is based on the architecture of Krizhevsky *et al.* [30] which won the ILSVRC 2012 Imagenet classification challenge. The first architecture is an 8-layer network of convolution, pooling and fully-connected layers that essentially fine-tunes the Krizhevsky et al. architecture to suit the microscopic image dataset. In the second architecture we remove one of the fully connected layers, but reduce the filter size and stride in the first convolution layer in order to aid the classification of fine-grained features.

CNN1: First Architecture. The first network architecture consists of 3 convolution layers along with 3 max-pooling layers and ReLU (Rectified Linear Unit), followed by 2 independent convolution layers (which do not have max-pooling layers) and 3 fully connected layers in the final section. The final classifier is a softmax function which gives the score or probabilities across all the classes.

The architecture is presented in Figure 2. The input RGB (3 channel) image is down sampled to 256x256x3 and is then center cropped to 227x227x3 before entering the network. In the first convolution layer the input image is convolved with 96 different filters with a kernel size of 11 and stride 4 in both x and y directions. The output 110x110x96 feature map is processed using ReLU, max-pooled with kernel size 3, stride 2 and is normalized using local response normalization to get 55x55x96 feature map. Similar operations are done on the feature maps in subsequent layers. In layer 2, the feature maps are convolved, processed using ReLU, max-pooled and normalized to obtain a feature map of size 13x13x256. The next two layers (layers 3,4) are convolution layers with ReLU but no max-pooling and normalization. The output feature map size is 13x13x384. Layer 5 consists of convolution, ReLU, max-pooling and normalization operations to obtain a feature map of size 6x6x256. The next two layers (layers 6,7) are fully connected which outputs a 4096 dimensional vector. The final layer is C-way softmax function that outputs the probabilities across C classes. In our case, we use the pre-trained weights of the Krizhevsky model, but train the final layer which is the input to the softmax function. This "fine-tuning" of the pre-trained model works achieves good accuracy in classifying microscopic images of different materials (with different granularity).

Figure 3 shows the CNN1 pipeline in action classifying two images. One is a microscopic image of the outer fabric of an authentic Louis Vuitton Monogram bag and another is a microscopic image of the outer fabric of a counterfeit Louis Vuitton Monogram bag. To the naked eye, it is hard to distinguish between the authentic and fake images, as both the images look almost the same. But CNN1 successfully distinguishes/classifies the images into authentic and fake classes. The convolution layer 1 shows the first 36 filters (out of the 96) for each image and convolution layer 3 shows the 384 filters of each image. While both images look similar there are minor differences. In the fully connected layer 6 (fc6), the 4096 dimensional vector of each image is different. Similarly in the fully connected layer 7 (fc7), the 4096 vectors corresponding to each image is different (we can now distinguish the two vectors and thereby the images). After fc7, the softmax function takes the 4096 vector as input and outputs the scores/probabilities for each class.

CNN2: Second Architecture. We use various types of convolution kernels on the candidate image to generate a feature maps of different sizes, as part of the convolution layers. These convolution capture diverse set of distortions possible on the microscopic image. Since the image is subjected to variations and distortions from image capture and tampering of the object's surface, we apply convolutions to the image, to make the network robust against such distortions.

In the second architecture, we reduce the filter size and stride in the first convolution layer. Instead of kernel size of 11, we use 8 and instead of stride 4 we use stride 2. This change increases the number of parameters hence we have to train with a much smaller batch size. We reduce the training batch size from 250 images to 50 images. This type of technique of reducing the filter size and decreasing the stride is done by Zeiler et al. [40] to increase the recognition/classification of fine grained features. The only change in the second architecture compared to the first architecture is the reduction in the filter and stride sizes in the first convolution layer. Since the first layer is different, we cannot use the pre-trained weights of the Krizhevsky model. We train the entire network from scratch using new sets of weight initialization, biases, learning rates and batch sizes. Due to the depth of the network it is prone to over fitting, so we used data augmentation to artificially increase the number of images in the dataset. We used label-preserving data augmentation techniques such as translation, shifts, horizontal and vertical flips, random cropping of 227x227 regions (from the original 256x256) and rotations. These augmentation techniques increased the dataset by 50x. Also, we used random dropouts in the final two layers to regularize and reduce over fitting.

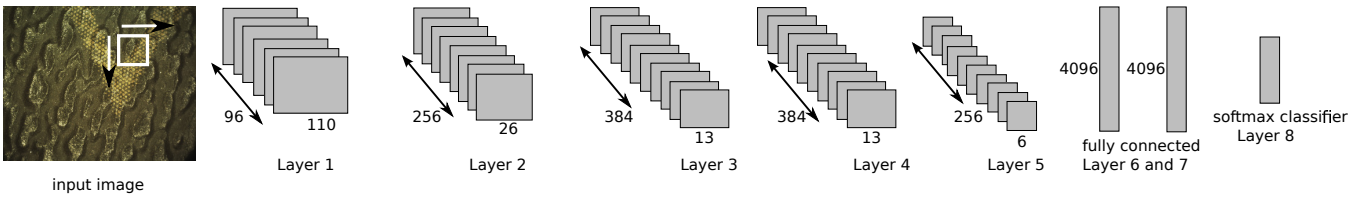

Figure 2: Classification and authentication of physical objects from microscopic images using 8-layer convolutional neural network

6 EVALUATION

We evaluate our system on 1.2 million microscopic images spread across the following objects and materials.

(1) Leather: We capture 30,000 microscopic images from 20 types of leather. The leather samples are obtained from Restoration Hardware, Abea Leather in New York and Tanneries Haas from France, which supplies leather to most of the top leather brands in the world [13].
(2) Fabric: We extract 6,000 images from 120 types of fabric. The fabric samples are obtained from the Fabric Science kit [5].
(3) Plastic: We extract 2000 images from 15 types of plastic surfaces.
(4) Paper: We extract 2000 images from 10 types of paper. The paper samples are from Neenah Paper [8].
(5) Jersey: We extract 500 images from 2 authentic NFL jerseys purchased from NFL store; and 2 fake NFL jerseys obtained from street hawkers.
(6) Pills: We extract 200 images from few Viagra pills to show the variation and classification results.

6.1 Methodology

Each object/material dataset is randomly split into three sets: training set, validation set, test set, such that training set contains 70% images, validation set contains 20%, and the test set contains 10% of the images. The algorithm runs on the training set and the validation accuracy is tested on the validation set. Once the learning cycle (training, validation) is completed (either by early stopping, or until the max iteration is reached), the algorithm is run on the test set to determine the test set accuracy. In results in Table 1 refer to the 10-fold cross validation accuracy on the test set. (The dataset is split into training, validation, testing set 10 times and the accuracy is determined each time. 10-fold cross validation accuracy is the average test accuracy across 10 trails)

From the bag of visual words perspective, we apply four types of classification methods. i) DSIFT for dense feature extraction, k-means for quantization, and SVM for final classification, ii) DAISY for dense feature extraction, k-means for quantization and SVM for final classification. For the rest, we use k-NN instead of SVM in the final step. For DSIFT we use VLFeat library [39] and for DAISY descriptor we use the implementation in skimage [11]. The results for each of these cases is presented in Table 1.

From a CNN perspective, we apply three types of methods. i) CNN1 which is a fine-tuning of the Krizhevsky model, ii) CNN2, in which the first convolution layer and last layers are changed for fine-grained classification, iii) R-CNN, which is a region based CNN that selects specific regions within an image and then uses CNN for classification. The results for each of the techniques are presented in Table 1.

For CNN, in order to avoid overfitting and get good test accuracy, we increase the size of the dataset artificially by generating label-preserving distortions such as 4 rotations, flips in each rotation, 12 translations (wrap side and up) and cropping the 256x256 input image into 30 randomly cropped 227x227 regions. This increases the dataset size by 50x to 3 million images. (Note that this data augmentation is performed once the dataset is split into training, validation and test sets. Else we would be validating/testing for different distortions of the same training images)

The training parameters for CNN are as follows. For CNN1, the learning rate is 0.001, step size is 20000, weight decay is 0.0005, momentum is 0.9 and batch size of 50. For CNN2, the learning rate is 0.0001 and the step size is 200000. Since we are training CNN2 from scratch the learning rate is significantly lower and the step size is higher than CNN1. Caffe [28] was used to train the CNNs on a NVidia K520 GPU with 1536 CUDA cores and 4GB of video memory.

6.2 Results

Leather. The test accuracy across 30,000 leather samples is given in Table 1. (After data augmentation, the size of the dataset increases to 1.5 million images). For the bag of visual words model, the average test accuracy after 10-fold cross validation is about 93.8%. k-NN based method tends to perform lower than the SVM based method and DSIFT performs slightly better than the DAISY descriptor. Perhaps if the descriptor size in DAISY is increased we would be able to get higher accuracy rates. For the CNNs, the average test accuracy is 98.1%. The last layer is a 20-way softmax classifier to classify 20 types of leather.

Fabric. The average test accuracy for the bag of words model is 92%. One of the reasons for the decrease in accuracy rate compared to leather samples is due to increase in the class size. The test accuracy for CNNs is 98.3%. The data augmentation and dropout techniques increase the accuracy rates when compared to the bag of visual words model. Due to data augmentation the dataset increases to 300,000 images.

Plastic. This is a 10-way classification across 10 different types of plastic materials. The average test accuracy for bag of words model is 92.5%. For CNNs, the average test accuracy is 95.3%.

Paper. The average test accuracy for paper across 2000 images and 10 types of paper is, 94.3% for the bag of words model and

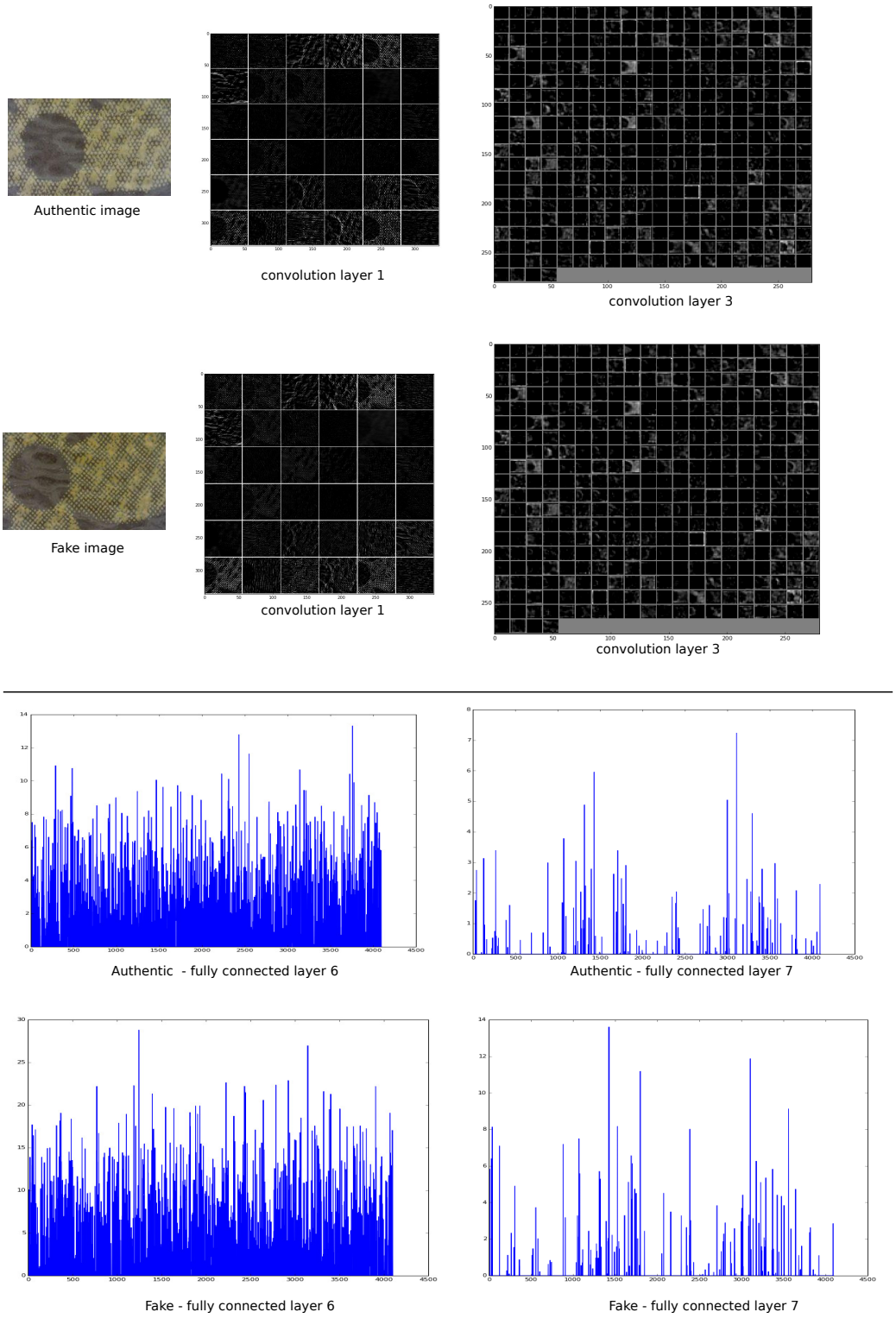

Figure 3: Visualizations of the convolutional layers of an authentic image and a fake image. The authentic image is of an authentic Louis Vuitton Monogram bag and fake image is of a fake Louis Vuitton Monogram bag. Convolution layer 1 shows the first 36 filters (out of the 96 filters); convolution layer 3 shows the 384 filters; fully connected layer 6 shows the differences in the 4096 vector of each image; similarly, fully connected layer 7 shows the differences in the 4096 dimensional vector of the authentic and the fake image.

Method	Leather	Fabric	Plastic	Paper	Jersey	Pills
DSIFT + k-means + SVM	97.8	95.3	93.6	96.2	98.5	98.6
DSIFT + k-means + k-NN	91.2	90.2	94.2	93.1	95.3	98.6
DAISY + k-means + SVM	95.6	94.3	94.2	96.5	96.2	96.5
DAISY + k-means + k-NN	90.7	88.2	95.2	90.3	97.5	96.4
CNN1	97.6	98.7	98.1	97.4	98.3	98.7
CNN2	98.6	98.0	92.5	92.8	99.3	98.4

Table 1: 10-fold cross validation test accuracy (%) for bag of visual words and convolutional neural networks based authentication system

95.1% for the CNNs. The results of both bag of words and CNNs are comparable with respect to classification of paper samples.

Jersey. With NFL jerseys we perform binary classification. Given an input image we determine if the image is authentic or fake. The average test accuracy for bag of words model is 94% and CNNs is 98.8%. CNN2 is able to capture the fine-grained details in some of the images, which gives it a superior performance compared to the rest of the methods.

Pills. In this dataset (which consists of Viagra pills), we do not have fake Viagra pills. So we perform binary classification of classifying two different types of Viagra pills. The average test accuracy for bag of words model is 96.8% and for CNNs it is 98.5%.

7 DEPLOYMENT EARLY EXPERIENCES

We have developed and deployed a mobile-cloud based authentication system where our microscopy hardware captures and transmits microscopic images over WiFi to a mobile client. The mobile client application interacts with a cloud based authentication service and uploads the microscopic data from a physical object to the cloud and the cloud service runs the machine learning algorithms to determine whether an object is authentic or fake. We have built an iOS app that interacts with the device and the server, for the authentication phase. The iOS app is currently available in the Apple AppStore; the authentication service in the application can only be initiated in conjunction to our microscopic hardware. During the authentication phase the steps are as follows. i) the user opens the mobile app, places the device on the object, ii) the device streams live video of the microscopic surface of the object via Wifi onto the app, iii) the user captures the image (or multiple images) using the app and uploads it to the server, iv) in a few seconds the server responds with a message saying the object the was either "Authentic" or "Fake".

Training phase and Data Collection: In the training phase, we have worked with data collectors in several supply chain points to extract a large database of microscopic images from different products or classes of products to form a training set. These images are trained and tested to generate a model that is ready for authentication (details are presented in Section 5.2). In the case of authenticating luxury handbags, we acquire bags of one particular brand, say Louis Vuitton (LV) and extract lot of microscopic images using our device. Every region of the bag is scanned and the images are uploaded, processed and trained in the back end server. This procedure is done for both authentic and counterfeit versions of the physical product. We have worked with several suppliers and human authenticators in the supply chain to obtain labeled information of authentic and fake versions of a physical object. Once trained, cross validated and tested, the model is ready for the authentication phase. During the training phase, we repeat the process of holding out different sets of input data for internal testing the consistency of the input dataset to identify any potential mislabeled data; any such mislabeled data is re-examined by a human authenticator and potentially removed if determined to be spurious. In addition, as a first pre-filtering step, all data entered in the system is also manually verified and cleaned by a human authenticator. In essence, we run the training dataset for several internal tests before executing any machine learning algorithms on the data.

Rollout for Luxury Handbags: We have initially rolled out our authentication service for luxury handbags. In the case of authenticating luxury handbags, the user uploads multiple images from different regions of the bag to check for authenticity. Here, we have trained independent machine learning algorithms on a per region basis and can achieve high end-to-end authentication accuracy with negligible false positive rates. The service is currently in use by several reputed luxury resale stores and the initial feedback has been highly positive. The system and the app have been designed to be highly user friendly to promote easy adoption. Based on tests conducted by customers, our system is able to also easily identify "superfake" bags which may tend to use the same material on some regions. Even if the fake good has a significant characteristic difference across one region, our current system is able to detect such fakes with high accuracy.

8 CONCLUSION

This paper describes the design, implementation and evaluation of a practical and non-intrusive system for authenticating physical objects and classifying genuine and counterfeit goods with high accuracy. The key idea of our system is to use a supervised machine learning algorithm to learn the microscopic characteristics of genuine physical objects corresponding to a product line and differentiate them against the microscopic characteristics of counterfeit versions of the same product line. We describe the design of a new wide-angle image capture device that can capture high quality, high resolution microscopic images of a relatively large area at $100 - 300x$ magnification. We train our system using a collection of microscopic images gathered from pre-labeled versions of genuine and counterfeit goods of a particular product line and test it on

any new object of that product line. Our SVM based supervised algorithm provides an accuracy of 95% for authenticating based on a single microscopic image and the convolutional neural network provides an enhanced accuracy of 98% per image. Assuming each image capture of the same physical object is relatively independent, when the system is tested on several microscopic images of the same physical object, our system provides a strong guarantee to differentiate authentic goods from counterfeit ones. We have tested out system on a variety of different physical objects and surfaces and we believe our system is quite generic in scope; we aim to apply it to authenticate a wide-spectrum of real-world physical goods. We are currently beta-testing our solution at scale with a luxury resale vendor for the specific vertical of luxury handbags and fashion goods. While our solution may not be perfect to eradicate the problem of counterfeit goods, we hope that our solution when adopted at scale does make it significantly harder for the counterfeiter to defeat this approach.

REFERENCES

[1] Alpvision. http://www.alpvision.com/cryptoglyph-covert-marking.html.

[2] Ambarella Video Security and Surveillance IP-Camera Solutions. http://www.ambarella.com/products/security-ip-cameras.

[3] Counterfeit goods are linked to terror groups - Business - International Herald Tribune. http://www.nytimes.com/2007/02/12/business/worldbusiness/12iht-fake.4569452.html.

[4] Counterfeiting Intelligence Bureau. http://www.iccwbo.org/products-and-services/fighting-commercial-crime/counterfeiting-intelligence-bureau/.

[5] Fabric Science kit. http://www.amazon.com/J-J-Pizzutos-Fabric-Science-Swatch/dp/1609013581.

[6] Ingenia technology ltd. http://www.ingeniatechnology.com/.

[7] IQStructures. http://www.iqstructures.com/.

[8] Neenah Paper. http://www.neenahpaper.com/.

[9] NovaVision. http://www.novavisioninc.com/.

[10] Prooftag. http://www.prooftag.net/.

[11] skimage: Image processing in Python. http://scikit-image.org/.

[12] Sproxil. http://sproxil.com/.

[13] Tanneries Haas - cuir de veau - BareniaÂő, novocalfÂő. http://www.tanneries-haas.com/.

[14] TI DM368. http://www.ti.com/product/tms320dm368.

[15] Veho Discovery Deluxe 004. http://www.veho-uk.com/main/shop_detail.aspx?article=40.

[16] WHO Anti-counterfeiting technologies for protection of medicines. http://www.who.int/impact/events/IMPACT-ACTechnologiesv3LIS.pdf.

[17] Berg, T., Liu, J., Lee, S. W., Alexander, M. L., Jacobs, D. W., and Belhumeur, P. N. Birdsnap: Large-scale fine-grained visual categorization of birds. In *Proc. Conf. Computer Vision and Pattern Recognition (CVPR)* (June 2014).

[18] Buchanan, J. D. R., Cowburn, R. P., Jausovec, A.-V., Petit, D., Seem, P., Xiong, G., Atkinson, D., Fenton, K., Allwood, D. A., and Bryan, M. T. Forgery: 'Fingerprinting' documents and packaging. *Nature 436* (July 2005), 475.

[19] Caputo, B., and Jie, L. A performance evaluation of exact and approximate match kernels for object recognition. *Electronic Letters on Computer Vision and Image Analysis 8*, 3 (2009), 15–26.

[20] Clarkson, W., Weyrich, T., Finkelstein, A., Heninger, N., Halderman, A., and Felten, E. Fingerprinting blank paper using commodity scanners. In *IEEE Security and Privacy* (2009).

[21] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09* (2009).

[22] E. Metois, P. Yarin, N. S., and Smith, J. R. Fiberfingerprint identification. In *Third Workshop on Automatic Identification* (2002).

[23] et. al, B. J. Surface treated security paper and method and device for producing surface treated security paper, US Patent Number 5,193,854, 1993.

[24] et al, E. B. G. Coatings and ink designs for negotiable instruments, us patent number 6,155,604, 2000.

[25] Greene, E. B. Negotiable instrument, us patent number 4,634,148, 1987.

[26] He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385* (2015).

[27] Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. What is the best multi-stage architecture for object recognition? In *Computer Vision, 2009 IEEE 12th International Conference on* (2009), IEEE, pp. 2146–2153.

[28] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093* (2014).

[29] Kimura, and Yoshihiro. Woven security label, us patent number 6,068,895, 2000.

[30] Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *NIPS* (2012), vol. 1, p. 4.

[31] Lazebnik, S., Schmid, C., and Ponce, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (2006), vol. 2, IEEE, pp. 2169–2178.

[32] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*, 11 (1998), 2278–2324.

[33] Maji, S., Kannala, J., Rahtu, E., Blaschko, M., and Vedaldi, A. Fine-grained visual classification of aircraft. Tech. rep., 2013.

[34] Sharma, A., Subramanian, L., and Brewer, E. A. Paperspeckle: microscopic fingerprinting of paper. In *ACM Conference on Computer and Communications Security* (2011), pp. 99–110.

[35] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2014).

[36] Smith, J. R., and Sutherland, A. V. Microstructure based indicia. In *Proceedings of the Second Workshop on Automatic Identification Advanced Technologies* (New York, NY, USA, 1999), ACM, pp. 79–83.

[37] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions. In *CVPR 2015* (2015).

[38] van Renesse, R. L. *Optical Document Security, Second Edition.* Artech House, Inc, Norwood, MA, 1998.

[39] Vedaldi, A., and Fulkerson, B. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

[40] Zeiler, M. D., and Fergus, R. Visualizing and understanding convolutional networks. *CoRR abs/1311.2901* (2013).

[41] Zeira, Eitan; Ellett, D. Verification methods employing thermally–imageable substrates, us patent number 6107244, August 2000.

[42] Zhu, B., Wu, J., and Kankanhalli, M. S. Print signatures for document authentication. In *ACM CCS '03* (New York, NY, USA, 2003), pp. 145–154.