Machine Learning in Cyber Security:

Developing an Algorithm to Detect Signature Authenticity

Word Count: 4185

Abstract

This paper studies whether machine learning, be used to develop an algorithm that can autonomously detect the authenticity of human signatures as a means to decrease fraud. In this current day, over $21.84B are lost to credit card fraud globally. 37% of this statistic comes from the usage of counterfeit cards. This makes it increasingly important to study new ways to counter credit-card fraud. To test the applicability of machine learning, a dataset containing genuine and counterfeit signatures for three test subjects were used to train 3 convolutional neural networks. With such methodology, it is possible to emulate how similar algorithms can be implemented in the real world to prevent credit-card fraud. Ultimately, all the of the models trained had a 100% accuracy. Apart from certain limitations, this proves that using convolutional neural networks as a means to detect signature authenticity is an effective way of countering credit/debit card fraud.

Keywords: Machine Learning, signature, cybersecurity, signature detection, convolutional, neural network

Introduction

Especially with the advent of the digital age, the usage of credit cards in both Realtime and online payments has been increasing. With the abundance of computers and digital information, there also grows the risk of cyber fraud occurring. One of the most significant cases of this fraud is in credit card transactions. This payment method since its first introduction about half a decade ago has proven to be an important part of daily life, with over 5 billion credit cards being in circulation at this day (Steel & Gonzalez).

In the year 2015 alone, over 21.84 Billion dollars were lost to credit card fraud globally. That number has been constantly growing up to this date, with it being projected to be $32.91

Billion in by 2020. Within the United States, $8.45 billion was lost to credit card fraud (Holmes). Due to credit/ debit card transaction primarily taking place in the form of digital addresses, it becomes especially easy to steal them. Credit card fraud can be separated into two major categories: card-present (CP) and card-not-present (CNP). Card present refers to the usage of credit cards by physically using the card (ie. at a store). This category includes counterfeit cards. CNP refers to the usage of credit cards by their number, such as in online transactions. Statistics for CP fraud in the United States has been constantly growing until October 2015, when EMV smart chip authentication was implemented. This regression is also valid for global statistics, however such cannot be pin-pointed to a specific date as different countries implemented EMV at different times. According to the Federal Trade Commission, FTC, credit card fraud is the largest form of identity theft. The largest form of credit card fraud is counterfeit card production, at 37% at 2014 per USA Statistics (Bennet).

Literature Review

Application of Machine Learning

Machine learning is a method of data analysis based on the notion of machines being able to learn from certain predictions from patterns found in data. It is one of the fastest growing fields in technology, which can be seen by the fact that machine learning patents have grown since 2013 with a 34% compound annual growth rate, making it a field with huge potential (Columbus). Specifically, within the field of cyber security and namely credit card security, machine learning can be applied to have significant benefits. One way this can be done is by evaluating customer transaction patterns. One study uses a hybrid approach into this idea by combining various neural network models, both supervised and unsupervised. The networks

were based off 6 key attributes used for input: cardholder number, card number, merchant category code, amount, date, and time. The 6 networks used included: decision tree (DT), random forest (RF), Bayesian network (BN), Naïve Bayes (NB), support vector machine (SVM), and K*models. This theoretically eliminates the limitations of each by combining the result of the models based on a voting criterion (Kültür & Çağlayan). This research studies a way to detect fraudulent transactions by deriving an algorithm that studies the transactional patterns of users in both CP and CNP Transactions. A positive of this type of algorithm and data is that regardless of the type of security measures present in a transaction, there always is a customer transaction pattern to study off of, meaning that there will never be a shortage of data. A limitation, however, is that there are a huge amount of variables in the buying pattern of a person, so being able to make a meaningful algorithm out of them can and will be challenging.

On the other hand, another study rather than using machine learning, uses more basic statistics as a means to detect fraudulent transactions. This is done by hard-coding various patterns that are generally from fraudulent transactions. For example, a $10,000 transaction may not be regarded as fraud however multiple of them being done in short succession likely will be fraudulent (Bolton & Hand). By compiling various statistical use cases from patterns of fraudulent transactions, an effective filter that stops fraudulent cases can be derived. This case can be relatively easy to implement, however, it is not very scalable since its usability is limited to the patterns that were previously thought of/ discovered and later implemented.

When it comes to detecting the performance for any model, an often overlooked but critical aspect is deciding the criterion to measure by. What may be optimal criteria in a linear decision surface (one level) can be orthogonal to another. With the usage of neural networks,

there are two types of training: supervised and unsupervised. Supervised training involves training the algorithm via both positives and negatives. For example, if one is developing an algorithm for detecting fraud and the data they use involves both fraudulent and genuine examples, that would be supervised. Unsupervised learning is when a machine is trained with data as it is fed and generally involves only positives, so with the previous example, the algorithm would only be trained by genuine examples. Another way to use the two types of training is to combine them where the input for the different models will be the same, and the models will run in parallel.

Utilization of Machine Learning

The unique thing about developing computer algorithms that manipulate signatures of people is that signatures as simple as they do seem due to their similarity to normal image recognition, they are quite divorced from such. Unlike other security features used with payment cards, such as a pin for example, a signature is much more spontaneous. Something like a pin is a simple one dimensional number. A pin with 4 digits returns to just over 10,000 combinations, however, a signature which is two dimensional, results in millions of unique combinations. This ultimately means that a signature offers orders of magnitude better security and hence fraud prevention. Picture recognition works primarily by studying prior knowledge of a given object, or in simpler terms, by recognizing commonalities within an example dataset of images with and without the object being classified (Karpathy). This creates a gap in image recognition with current approaches because they adopt various handcrafted features (based on the prior knowledge from the training dataset). The limitation within this is that these models easily get

invalidated when the scene structure is different. For example, as the angle or location of an object change, it can lead to the developed algorithm being invalid (Lei et. Al.).

Another limitation within the usage of machine learning in this field has to do with Unbalanced class sizes. False-positives and false-negatives are common performance measures when it comes to supervised machine learning. One issue arises, however, when there are a disproportionate number of positives and negatives, such as it is in fraud detection. This makes the measure of false-negatives and false-positives rather inaccurate. For example, if an algorithm that can correctly identify 99% of fraudulent cases and 99% of legitimate cases may sound very accurate on paper, with a false-negative and false-positive rate of only 1%. However, if the fraud cases consist only 0.1% of all cases, then 91% of the cases recognized as fraud are legitimate, which is much less attracting in terms of accuracy. (Hand et al.) One thing to keep in mind on this matter is that the monetary loss of misclassified fraudulent cases are much higher than mis-classified genuine cases.

When it comes to signature classification, namely detecting the authenticity of a signature, there can be a lot of variation even between genuine signatures due to various reasons such as the location, environment and the device the signature was taken in. A viable way to counter such is to use a Naïve Bayes classifier first hand to separate the signature from background static. One study uses such along with layers of restricted Boltzmann machines that automatically extract features from data at the process of modeling the data's distribution (Lei et. al). Although this study primarily focuses on image recognition, it directly applies to the usage of machine learning within signatures. Signatures are not just like any other image as each signature

is a variable entity, in that there can still be a significant variance between them, even amongst signatures of the same one owner.

This paper intends to study the applicability of using machine learning in order to decrease fraud in credit/ debit card transactions. Such will be done by developing a machine learning algorithm that can evaluate the authenticity of signatures. Namely, these will be the signatures done on payment-pads used throughout commercial places, however, in this study, signatures will be directly collected in paper, as the collection method for the signatures do not result in any change on the applicability of an algorithm to calculate the authenticity of a signature. The primary method used to train the neural network will be via supervised training. Supervised training involves prior datasets to use for training, which in this case involves records of fraudulent and legitimate transactions (Bolton & Hand).

With the usage of recurrent neural networks, handwriting recognition models can be written, however in this specific case, a model will be designed towards not only understanding signatures but also to detect the difference of counterfeit signatures from those done by the owner of the signature. There are numerous examples of machine learning algorithms being used to identify fraud as previously mentioned in this paper, however, there is not any documentation regarding the usage of machine learning being used to detect fraud by classifying signatures—namely whether they are authentic or not. Moreover, the usage of signatures has not been combined with today's computing power to automatically detect fraud. This can provide beneficial results as every person's handwriting is highly variable and is mostly unique, and by deriving an algorithm for each person's handwriting, via their signature, a model that is able to differentiate between the genuine signature of a person and a counterfeit signature can be

effectively derived. With there being over 21.84 Billion dollars being lost to credit card fraud annually, with the statistic increasing, it becomes increasingly important to study new ways to detect fraud. On top of this, for future studies, detection of counterfeit signatures could not only be used to detect payment fraud in credit cards, but it also can be applied to other fields such as documents and checks to recognize whether those may be fraudulent. Ultimately, there is no study regarding the usage of machine learning with signatures and that is what this paper intends to address.

Methods

The primary goal for this research is to evaluate the potential applicability of machine learning within the field of cybersecurity as a means to prevent credit card fraud. In order to do so, a machine learning algorithm will be developed to differentiate between authentic and counterfeit signatures.

Acquiring and Manipulating Data

The first portion of the study was getting access to the required data. In this case, the data to be collected is signatures of people, and their simulated counterfeit signatures of the aforementioned people done by others. the required data simulated the genuine and the counterfeit signatures received by a credit card company in the real world. By using the data, this current study aims at detecting whether a given signature may be originating from the stated customer (ie. Whether the signature is counterfeit). A primary source used was the ICHHR 2010 Signature Verification Competition from the German Research Center for Artificial Intelligence. The dataset includes 3 subjects for genuine signatures and then counterfeit signatures for the

people. Three subjects are an effective number because it allows for an algorithm to be tested with a variety of signature types and also, it can allow for testing the effectiveness of an algorithm while using a varying quantity of signatures. Precisely, for person 1,2 and 3, there were 250, 100, and 100 total (both counterfeit and genuine) signatures respectively. All signatures were collected on paper, and then scanned into a .jpg image. This is done to stimulate how a signature would be received from a real-world case where a person would sign a credit-card scanner (or an equivalent). Once a signature is acquired into a digital format, it makes very little difference as to whether the signature was made first on paper or digitally. In order to derive the machine learning algorithm to be used with the signatures, the programming language Python, was used due to the vast amount of research and libraries for working with machine learning with the language. The reason for creating an algorithm/ model to detect the authenticity of signatures is to be able to simulate how an algorithm could be created in a realistic case to detect the authenticity of signatures as a means to decrease credit/ debit card fraud, which directly answers the goal of this research. For example, by having an algorithm such as the current one being tested, it could be used in tandem with pre-existing infrastructure and technology such as credit card readers that take signatures to automatically detect the authenticity of a signature.

With the images of signatures collected, they were split into three categories: the training set, the validation set and the test set. The training set is what the model will see and base its parameters out of. The Validation set is what the model uses to test itself while training the data. Lastly, the test set is completely separate from the model while training and is only used at the end to measure the performance of the model. By splitting the data into these 3 separate groups,

the model can be prevented from over-focusing, where the model only gets used to the training data and is not effective in labeling data separate from the initial data used for training.

With the training set, each image will be converted to a black and white image. Since the authenticity of a signature does not have to do with the color they were taken/written in, simplifying each signature enables the neural network to be lighter and more efficient. To do such, the 8-bit red, green, and blue value for each pixel was taken and then averaged out. This provides a single grayscale value from 0 to 255 which then converted to a binary black or white value by dividing by 255, yielding 0-1, and then rounding it to the nearest integer (0 or 1). An algorithm written with the Open CV library was used to automatically size every signature into the same 512x1024 plane. This prevents the neural network that will later be used from detecting a signature as different than another just because one signature may be of different size than another.

Developing the Neural Network

Next, a convolutional neural network (CNN) was made with the TensorFlow library for Python. CNN's can detect images with multiple layers of layers of abstraction much better than a conventional neural network can (Ishitsuka). This makes it a much more viable option to use CNN's in image recognition or in this case, for signature recognition. Also, by applying neural networks to the field of cyber security and credit-card fraud, a novel way for decreasing fraud can be found, which can further justify this research's goal. Each pixel value was inputted into the neural network, making a total of 2^20input nodes. The first two layers of the neural network was a max-pooling layer. It took the max value of each 2x2 square in each 1024x1024 image to simplify each picture into a smaller one. This ultimately results in a 256x256 image. By doing

so, the neural network algorithm can become much more efficient to train with minimal sacrifice on final accuracy. Then, the three fully connected convolutional layers take the output of the previous layer to recognize a certain attribute in the signature. The attribute that each convolutional layer caries will vary by the signature that is being detected.

Analysis of the Result

Lastly, a fully connected neural network takes the output of the final convolutional layer to result in a percentage value, stating how close the signature being studied for authenticity is to the trial signatures. This percentage value is used to evaluate how genuine the signature is to the CNN. A threshold value will then be set depending on the regression of the output percentages and whether they are genuine. Due to artificial techniques for evaluating the authenticity of signatures being relatively new, there is no previous data or system to compare this with. Ultimately, the accuracy of this system will define its applicability in the real world. Since such system is will not necessarily be replacing a preexisting system, it can be used in parallel with another fraud detection method. This means that even if the accuracy of the system is not very high (this metric will be defined later), the system as a whole may still provide a benefit.

Results

After a model was created and trained for all three of the unique signature sets, the precision and recall values were looked into in detail. The precision value signifies the percentage of genuine signatures that were correctly identified. A higher precision value will signify a lower number of false positives, which are genuine signatures that get labeled as being counterfeit. On the other hand, the recall value measures the exact opposite of the precision

value. It evaluates the percentage of counterfeit signatures that were labeled by the model correctly as being counterfeit. A higher recall value means that there were a smaller amount of false negatives, which are counterfeit signatures being incorrectly detected by the model as being genuine. Ultimately, the goal of a model is to have both the precision and recall values as being 100%.

Starting with the first model, there were a total of 246 images, with a 158 of them being counterfeit and 88 being genuine. After creating a model via the process mentioned in the methods, the predicted labels and true labels of all of the images from the test set were collected. For the models for person 2 and 3, there were a total of 100 signatures for each person. Person 2 had 58 genuine signatures and 42 simulated counterfeits. Person 3 had 71 counterfeit and 29 genuine signatures. A separate model was created for each person and again, the precision and recall values were collected. The process for creating a model for each person is identical. The only difference is that during training, each model is separately fitted to per the data received to specialize to the signatures of a person. The reason for having a separate model is to be able to fit a model to the individual, unique aspects of a person's signature.

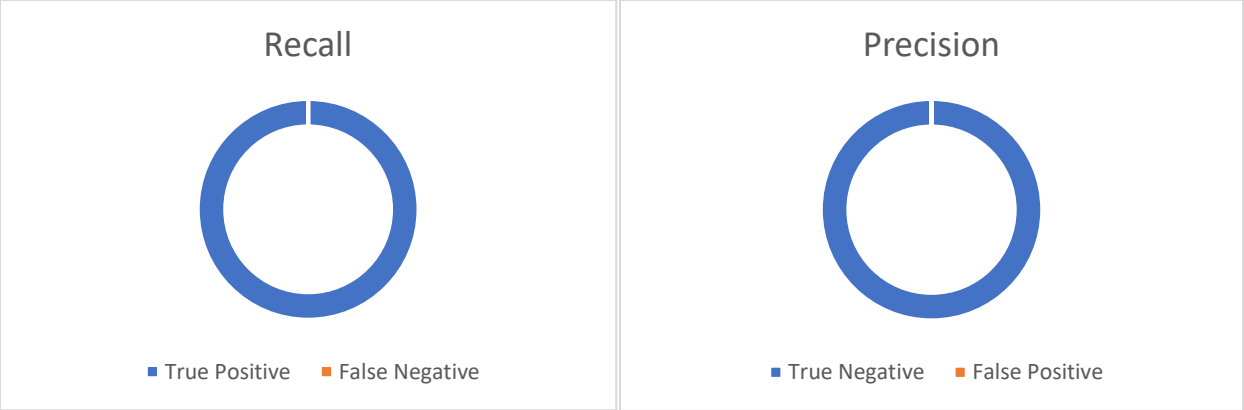

Figure 1a (left) and 1b (right). Both signify the results of the model based on the signatures of person 1 and its respective forgeries.

Figure 2a (left) and 2b (right). Each signifies the results of the model based on person

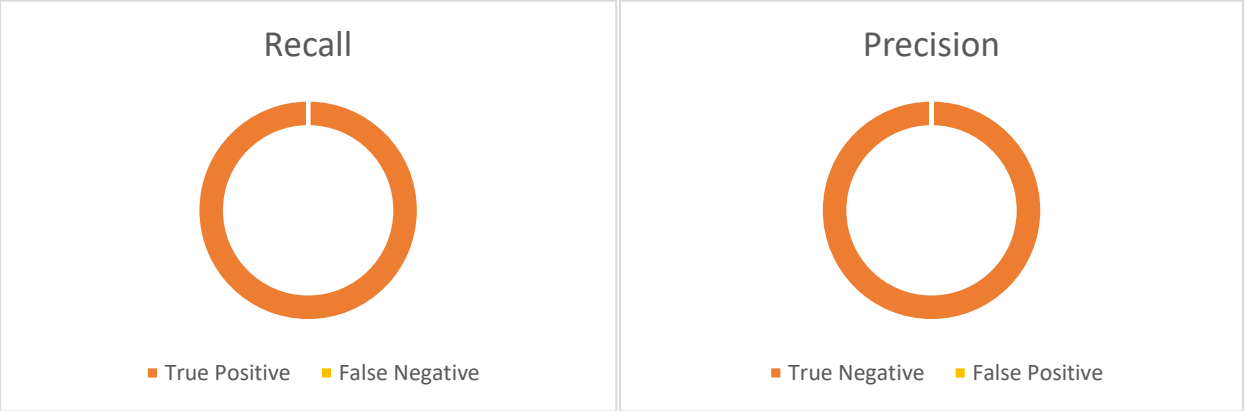

Figure 3a (left) and 3b (right). Both signify the results of the model based on the respective signature of person 3.

(1, 2, 3)a depict how the genuine signatures of person one were classified by the model. True negatives represent correctly identified genuine signatures while false positives show genuine signatures being incorrectly classified as being counterfeit. (1, 2, 3)b depicts how the forged signatures of person one were classified. True positives are the forgeries being correctly identified and false negatives depict that forgeries that were incorrectly seen as being genuine by the model.

As can be seen in the charts above, all three of the models were able to perfectly classify the signatures based on the respective person. The rate of false negatives that occurred were 0 for all three of the models. This means that the amount of times the model received a counterfeit signature and calculated it as being genuine is exactly zero. Such result is incredibly important in a real-world case because having a larger number means having counterfeit signatures that signify a fraudulent activity to be let off the radar. Ultimately, this exactly what this research paper intends to counter—fraudulent activity such as credit/debit card fraud. On top of this, the rate of false positives being detected is also exactly zero for all three of the algorithms. In a real world case, a false positive would mean a genuine transaction being classified as fraudulent activity. A high value of this, similar to false negatives, would lead to an ineffective algorithm. All in all, since both the negatives (genuine signatures) and positives (counterfeit signatures) were all correctly identified, a machine learning model designed to identify the authenticity of a signature is a perfectly viable option in the real world for preventing credit/ debit card fraud.

		Predicted label	
		Counterfeit/ Forged	Genuine
Actual Label	Genuine	False Positive: 0%	True Negative: 100%
	Counterfeit/ Forged	True Positive: 100%	False Negative: 0%

As shown above, the confusion matrix depicts the data for all of the algorithms. Since the three models had identical results, one set of data represent all of them in this instant.

Discussion

Conclusion

In this study, there were three separate models trained with a unique set of signatures. All three of them had perfect results with a 100% precision and recall rate. This proves that a convolutional neural network model for detecting the authenticity of signatures is a very viable and effective choice due to both its efficiency and accuracy. In a more realistic perspective, this technology can be applied in the real world to limit fraud in credit card and debit card transactions. Since this technology is not a replacement of another preexisting system, it can be used parallel to other means of detecting fraud. Nevertheless, the application of this research will increase the security of credit-card transactions as it realistically will have little to no negative effect on security. Due to the proposed solution in this research being

Implications

The usage and research of machine learning and more specifically of artificial neural networks have been on the rise in usage primarily during the past decade. Although there have been major studies on using such in the field of cyber security, there is a gap in the usage of using these techniques for detecting the authenticity of signatures. By studying such, a way to

limit fraud, namely in the field of credit card transactions, has been found. Although it is impossible to ever truly get rid of fraud, by being able to approach it from various avenues, it becomes much easier to limit it. The unique thing about signatures is that the technology and infrastructure to be able to collect them already exists, however, it is not considered to be an effective means of countering fraud, and as a matter of fact is even fading out. In the year 2018, both VISA, MasterCard, Discover, and American Express have all stated plans about removing the requirement of signatures (Egan). The usage of signatures for countering fraud is increasingly being seen as ineffective due to there not being any effective means to detect a signatures authenticity in a (monetarily) cheap, accurate, and timely fashion. The application of a convolutional neural network model to have an automated process to detect the authenticity of signatures extremely accurately and affordably directly counters this issue. Ultimately, studying this subject both directly fills a gap in research and puts a new form of countering credit/ debit card fraud onto the limelight.

Another application of this study could also be used to counter other forms of fraud. There are numerous cases of signatures being used for verification of things apart from just credit card transactions (ex. Checks, forms, documents, bills, etc.). This study focused on the fraudulent transactions from payment cards as a means to both focus the scope of this research and because of the fact that the usage of signatures in credit card transaction is both digital and preexistent, meaning that this technology would be relatively easier to implement in the field.

Limitations

One major thing to put attention on is the data used for training convolutional neural network models. Since every node inside a neural network is based solely off of the data used to train it, the accuracy or lack thereof in a model is directly correlates to the data used to train it.

Such in this research with the data being signatures, since a person's signature can vary significantly according to the environment it takes place in. Therefore, it is crucial to have a wide range of data (ie. signatures) depicting these various environmental variations so that at the end, the model will be effectively trained to work in different environments and cases.

If there is limited variation in the data used to train a model, the algorithm trained may on paper have a very high precision and recall rate, however, realistically speaking, it may be ineffective since it is not made to work with the various cases and environments it is supposed to. It is impossible to gauge such without testing a model within the real world, and this is a major limitation to this research. Although the results to all three of the models are extremely optimistic, it is very hard to predict as to whether the models still will be as applicable in a real usage case.

Furthermore, another limitation that arises from limited data used for training is over-fitting. This occurs when the model becomes too accustomed to the data it sees in training. This again partially results from a limited amount of data, however, there are some ways to counter it. This study did so by splitting up the data into training, validation, and test sets as mentioned in the methodology. Although this can do an effective job in preventing over-fitting with most cases, it still will be inadequate if there is not enough data to begin with. This may have been the case with this research, however, it is impossible to give a definitive answer without experimenting with larger sets of data.

Works Cited

Bennet, Michael. "11 Types Of the Most Common Credit Card Fraud | Consumer Protect.com." Consumer Protect, ConsumerProtect, 8 Sept. 2015, www.consumerprotect.com/11-types-of-credit-card-fraud/.

Bolton, Richard J., and David J. Hand. "Statistical Fraud Detection: A Review." *Statistical Science*, vol. 17, no. 3, 2002, pp. 235–249. *JSTOR*, JSTOR, www.jstor.org/stable/3182781.

Columbus, Louis. "Roundup Of Machine Learning Forecasts And Market Estimates, 2018." Forbes, Forbes Magazine, 19 Feb. 2018, www.forbes.com/sites/louiscolumbus/2018/02/18/roundup-of-machine-learning-forecasts-and-market-estimates-2018/#2e70e7872225.

Egan, John. "No Signatures Required: Mastercard, Discover, AmEx and Visa Ditch Them." CreditCards.com, Creditcards.com, 8 Aug. 2018, www.creditcards.com/credit-card-news/signatures-soon-may-not-be-required.php.

Hand, D. J., et al. "Performance Criteria for Plastic Card Fraud Detection Tools." *The Journal of the Operational Research Society*, vol. 59, no. 7, 2008, pp. 956–962. *JSTOR*, JSTOR, www.jstor.org/stable/20202156.

Holmes, Tamara. "Credit Card Fraud and ID Theft Statistics." NASDAQ.com, CreditCards.com, 16 Sept. 2015, www.nasdaq.com/article/credit-card-fraud-and-id-theft-statistics-cm520388.

Ishitsuka, Kazuya, et al. “Object Detection in Ground-Penetrating Radar Images Using a Deep Convolutional Neural Network and Image Set Preparation by Migration.” International Journal of Geophysics, Nov. 2018, pp. 1–8. EBSCOhost, doi:10.1155/2018/9365184.

Karpathy, Andrej. “Connecting Images and Natural Language.” Stanford , 2016.

Kültür, Yiğit, and Mehmet Ufuk Çağlayan. “Hybrid Approaches for Detecting Credit Card Fraud.” Expert Systems, vol. 34, no. 2, Apr. 2017, p. n/a-N.PAG. EBSCOhost, doi:10.1111/exsy.12191.

Lei, Jun, et al. “Convolutional Restricted Boltzmann Machines Learning for Robust Visual Tracking.” *Neural Computing & Applications*, vol. 25, no. 6, Nov. 2014, pp. 1383–1391. *EBSCOhost*, doi:10.1007/s00521-014-1625-x.

Marcus Liwicki, Elisa van den Heuvel, Bryan Found, Muhammad Imran Malik. "Forensic Signature Verification Competition 4NSigComp2010 – Detection of Simulated and Disguised Signatures", Proc. 12th Int. Conference on Frontiers in Handwriting Recognition, 2010

Steele, Jason, and Jamie Gonzalez. “Credit Card Fraud and ID Theft Statistics.” CreditCards.com, Creditcards.com, 24 Oct. 2017, www.creditcards.com/credit-card-news/credit-card-security-id-theft-fraud-statistics-1276.php.