# A Practical Tool for Phishing Detecting

Fujiao Ji
University of Tennessee, Knoxville
Knoxville, Tennessee
fji1@vols.utk.edu

Mengjun Wang
University of Tennessee, Knoxville
Knoxville, Tennessee
mwang43@vols.utk.edu

Tianhao Wu
University of Tennessee, Knoxville
Knoxville, Tennessee
twu21@vols.utk.edu

Qian Liu
University of Tennessee, Knoxville
Knoxville, Tennessee
qliu30@vols.utk.edu

## Abstract

In the realm of cybersecurity, the detection of phishing websites—fraudulent sites mimicking legitimate ones to steal sensitive user information—remains a paramount challenge. This study introduces an advanced phishing detection framework that enhances accuracy and fairness in identifying such threats. By integrating Vision Transformer (ViT) technology for logo analysis with traditional domain verification methods, our approach provides a robust mechanism for evaluating website authenticity. Unlike conventional systems that rely predominantly on domain information, our framework employs a dual-check strategy. First, it analyzes the visual components, specifically the logos, using ViT to identify potential discrepancies from legitimate counterparts. Second, it examines the domain information for anomalies. This combined analysis ensures a comprehensive assessment, capable of detecting sophisticated phishing attempts that might otherwise evade single-faceted detection methods. The framework's effectiveness is validated through extensive testing on diverse datasets, demonstrating its capability to significantly reduce false positives and enhance the detection of genuine phishing activities with high precision.

***Keywords:*** Phishing Detection, Vision Transformer, Logo Analysis, Domain Verification, Cybersecurity, Artificial Intelligence, Image Recognition

**ACM Reference Format:**
Fujiao Ji, Mengjun Wang, Tianhao Wu, and Qian Liu. 2024. A Practical Tool for Phishing Detecting. In . ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 Introduction

The expansion of the Internet has brought with it a surge in digital transactions and communications, creating fertile ground for cybercriminals to exploit unsuspecting users. Among the various forms of cyber threats, phishing attacks are particularly insidious, cleverly disguised to mislead users into providing personal and financial information to seemingly trustworthy entities. These attacks not only lead to direct financial loss but also compromise the integrity of personal data, posing long-term security risks. As such, the need for effective and reliable phishing detection mechanisms is more pressing than ever.

The crux of the challenge in detecting phishing websites lies in their increasing sophistication. Modern phishing schemes are expertly crafted, with attackers employing advanced techniques to replicate the visual and functional aspects of legitimate sites. Traditional detection methods, primarily based on domain name analysis, are often inadequate. They fail to capture the nuanced visual manipulations that can make a phishing site appear authentic, such as the strategic use of logos and branding elements. Furthermore, reliance solely on domain analysis can lead to false positives, unjustly flagging legitimate sites that happen to have domain names similar to those of known phishing sites.

This backdrop sets the stage for the development of our phishing detection framework, which innovatively combines the analysis of both visual and textual website components. By leveraging the Vision Transformer (ViT) for image recognition, our approach brings a novel perspective to phishing detection. The ViT model, adapted from advancements in natural language processing, treats images as sequences of patches. It analyzes these patches to capture complex patterns and dependencies, offering a granular insight into the authenticity of website logos. This level of detail allows for the detection of subtle discrepancies that are typical of counterfeit logos used on phishing sites.

Simultaneously, our framework employs a sophisticated domain verification process that goes beyond mere name matching. It analyzes the structural and syntactical features of domain names, comparing them against an extensive database of domain characteristics associated with phishing and legitimate sites alike. This dual-layered approach ensures a comprehensive evaluation of potential threats, significantly enhancing the detection accuracy while reducing the likelihood of false alarms.

*Conference'17, July 2017, Washington, DC, USA*
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

In the subsequent sections, this paper will detail the methodology used to integrate these technologies into a cohesive phishing detection system. We will discuss the data collection processes, model training, and the algorithmic strategies employed to fuse visual and textual analyses effectively. Our findings demonstrate the robustness of this integrated approach, showcasing its superior performance in various test scenarios and its potential to revolutionize the field of cybersecurity by setting new benchmarks for phishing detection.

## 2 Background and Related Work

Phishing detection is paramount in the domain of cybersecurity, as phishing attacks have become increasingly complex and frequent. These attacks pose significant risks as they aim to deceive individuals and organizations into revealing confidential and sensitive information. The necessity to counteract these threats has led to the development of various detection techniques.

Conventional methods for detecting phishing primarily rely on heuristic-based strategies and blacklisting. Heuristic approaches scrutinize URLs and webpage content to spot suspicious patterns and red flags typical of phishing sites, including specific keywords, URL discrepancies, or concealed redirects. Alternatively, blacklisting entails compiling and updating a list of known phishing URLs to block access to them. Although this method effectively neutralizes recognized threats, the need for frequent updates makes it challenging to combat the swiftly emerging new phishing sites, diminishing its utility against immediate threats.

Regarding list-based detection[5], browsers such as Microsoft Edge and Google Chrome employ these methods. These include whitelisting and blacklisting; the former permits access only to URLs deemed safe, while the latter blocks URLs identified as harmful or deceptive. The major drawback here is that even minor changes in a URL can elude these list-based systems, necessitating continuous updates to handle new phishing threats effectively.

Researchers like Jain and Gupta[3] have explored how heuristic methods fare in identifying phishing sites. This approach relies on a range of attributes that distinguish phishing sites from legitimate ones, drawing data from URLs, textual content, DNS records, digital certificates, and web traffic patterns. The effectiveness of these methods hinges on the chosen features, training samples, and classification algorithms. One notable benefit of heuristic techniques is their potential to detect zero-hour phishing attacks. However, they are also susceptible to high false positive rates, which can lead to legitimate sites being incorrectly flagged as malicious, causing user inconvenience and potential overblocking.

In recent years, the use of machine learning (ML) models for phishing detection has gained significant traction. These models are trained using a diverse array of features extracted from websites, such as URL structures, HTML content, and patterns of network traffic, to assess whether a site is engaged in phishing. Advanced methods like deep learning have notably enhanced the precision of these predictions. However, ML-based detection systems encounter several challenges. They require ongoing training to adapt to new and evolving phishing tactics, ensuring their relevance. Moreover, sophisticated attackers frequently utilize techniques to circumvent detection, including the obfuscation of malicious content or the simulation of legitimate websites, which can diminish the effectiveness of ML models.

Machine learning has emerged as a prevalent method for detecting phishing websites, as outlined by Sindhu and colleagues in 2020[6]. Key features such as URL information, website structure, and JavaScript attributes are gathered to characterize phishing URLs and associated websites. Based on these attributes, datasets are compiled, and machine learning classifiers are then trained to identify phishing site[8]. This approach is particularly effective when applied to large datasets characterized by high velocity, variety, volume, value, and veracity. According to Alkawaz and others in 2021[1], machine learning-based classifiers have achieved accuracy rates exceeding 99%, making them one of the most effective methods currently available for phishing detection.

Overall, while current phishing detection techniques have made strides in identifying and mitigating threats, they also exhibit inherent limitations that can be exploited by attackers. This ongoing challenge underscores the need for continuous innovation and adaptation in cybersecurity strategies to keep up with the evolving tactics of cybercriminals.

## 3 Model Structure

In this section, we developed practical phishing detection tools utilizing either Vision Transformer (ViT) or Involution technologies. For detailed descriptions of the process, please refer to Figure 1. Specifically, we maintained a reference list to compare the legality of content information across different brands. For testing samples, we initially employed Selenium to extract content information, focusing particularly on screenshots. Subsequently, using these screenshots, we applied the Faster R-CNN algorithm to extract the logo component. After extracting the logo, we compared it with our maintained reference list to verify the target brand. Finally, we compared the target brand domain with the testing URL domain to determine the authenticity of the content.

It's important to note that in our approach, testing samples are initially presumed to be benign by default. If a sample's similarity score falls below the established threshold, we continue to consider it benign. This decision is based on the practical reality of the internet, where billions of websites exist. Most detection methods maintain a blacklist rather than a whitelist because setting phishing as the default assumption would severely restrict user experience. Such an approach allows for a more user-friendly interaction with

the vast majority of legitimate sites, while still safeguarding against malicious ones.

### 3.1 Logo Extraction

Logo extraction refers to taking screenshots as inputs and generating a set of bounding boxes to annotate the position and size of the logo on the image with confidence scores. In this scenario, we find Faster R-CNN (Region-based Convolutional Neural Network) is an influential model designed for object detection tasks and suitable for our task. It contains four key components: Convolutional Layers, Region Proposal Network (RPN), ROI Pooling, and Classification and Bounding Box Regression. Following [4], we finetuned the Faster RCNN on benign30k data based on Detectron2 [2].

### 3.2 Target Match

The key point in the target matching process is to determine the similarity between the maintained reference list and the logos extracted from testing samples. To achieve this, we compute the cosine similarity between the logo images. Our reference list includes various logo variants and legitimate domains for each brand. We compile and regularly update this list based on the premise that phishing attackers often create counterfeit webpages that closely mimic legitimate ones to deceive users. By maintaining a diverse collection of logo variants for each brand, we can achieve more precise and flexible logo matching. Additionally, a single brand may be associated with multiple legitimate domains. For example, both 'drive.google.com' and 'google.co.uk' are affiliated with Google. Recognizing such relationships helps us minimize false positives by accurately identifying legitimate domains associated with the brand logos.

Before calculating the similarity between the logos being tested and those in our reference list, it is essential to extract features from each logo. To accomplish this, we employed two distinct methods: Vision Transformer (ViT) and Involution. We trained each model on a classification task using the Logo2K+ dataset, as described by Wang et al. [7]. Then, we extracted logo features from the penultimate layer, just before the classification layer. To establish the appropriate thresholds for detecting phishing attempts, we utilized data from the study by [4]. Using these thresholds, we then applied the trained models to effectively identify phishing examples.

**3.2.1 ViT.** The Vision Transformer (ViT) applies the transformer architecture, traditionally used in natural language processing (NLP), to computer vision tasks. It processes images by splitting an image into fixed-size patches, which are then flattened and processed through a sequence of transformer blocks that use self-attention mechanisms to understand the relationships between different patches. Key features of ViT include patch-based processing, positional embeddings, and self-attention. ViT has shown that transformers can be highly effective in image recognition tasks, often outperforming traditional convolutional neural networks (CNNs), prompting its integration into our tool.

**3.2.2 Involution.** The Involution model is proposed as an alternative to the traditional convolution operation in deep learning, focusing on improving adaptability and efficiency with spatial data. Unlike convolution, which aggregates input features through a fixed kernel, involution generates dynamic kernels based on the input features themselves. This allows for more flexibility as the kernel adapts to the specific features of the input at different spatial positions. It includes dynamic kernels and spatial-specific processing. It provides a promising direction for designing more efficient and flexible neural networks, particularly in tasks requiring detailed spatial awareness and adaptability. Motivated by these advantages, we incorporated Involution into our tool.

Both ViT and Involution represent significant improvements in the way deep learning models process spatial and visual information, each offering unique advantages and potential applications in various fields of computer vision. Therefore, we choose these two methods to help us get the features of logos.

## 4 Results

Table. 1 presents the performance evaluation of the Vision Transformer (ViT) and Involution models on a dataset comprising 1,941 websites, including 501 benign and 1,440 phishing websites associated with three major brands: eBay, Google, and Facebook. This dataset was sourced from Archive.org, capturing variations over the past seven years.

### 4.1 Overall Performance

The Involution model demonstrated a superior overall detection rate of 93.35% compared to the ViT model, which had a rate of 38.79%. This stark contrast in performance underscores the effectiveness of Involution's dynamic, adaptive kernel technology over the fixed, patch-based approach of ViT.

### 4.2 Brand-Specific Performance

- **eBay**: Involution achieved a near-perfect detection rate of 99.23%, while ViT detected 80.51% of phishing instances. eBay's consistent logo design over the years likely contributed to the higher detection rates.
- **Google**: ViT exhibited a detection rate of only 28.75%, significantly lower than Involution's 87.28%. The frequent modifications to Google's logo for various holidays and events introduce a high degree of variability in logo appearances, which likely affected the ground truth collection and subsequently the model performance.

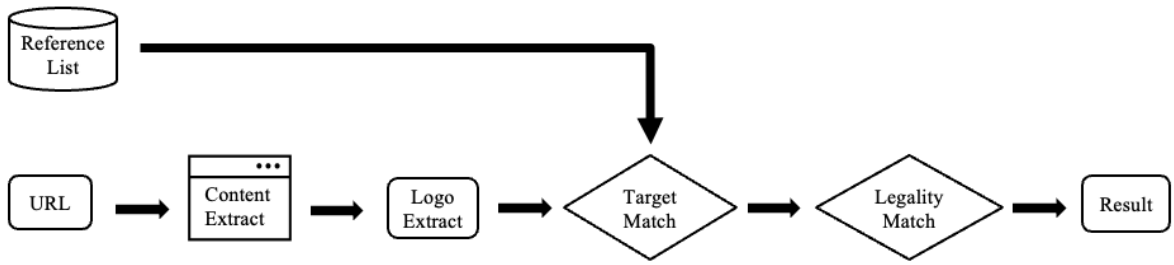


**Figure 1.** Phishing Detection Process

- **Facebook**: Involution performed robustly with a detection rate of 93.44%, in stark contrast to ViT's 28.15%. Although Facebook's logo undergoes fewer variations compared to Google's, the subtle complexities in changes can still pose significant challenges for models dependent on static visual cues like ViT.

**Table 1.** Detection rate (DR) for the ViT and Involution on collected datasets

| Brand | Total | ViT DR | Involution DR |
|---|---|---|---|
| Ebay | 390 | 0.805 | 0.992 |
| Google | 393 | 0.288 | 0.873 |
| Facebook | 1158 | 0.282 | 0.934 |
| Total | 1941 | 0.388 | 0.934 |

## 5 Discussion

The results from Table 1 reveal significant disparities in the performance of the Vision Transformer (ViT) and Involution models across different brands, which can be primarily attributed to the inherent characteristics of the models and the variability in logo designs:

- **Impact of Logo Variability**: The frequent logo changes by Google, often tied to various global and cultural events, introduce a high degree of visual variability. This aspect significantly challenges models like ViT, which rely on stable, static visual cues. The performance degradation in Google's case highlights the limitations of applying fixed-pattern recognition techniques to dynamic visual contexts. Involution, with its adaptive kernels, demonstrates resilience by adjusting its processing dynamically, which is crucial for handling real-world variability in phishing detection scenarios.
- **Model Architecture Differences**: The contrasting architectures of ViT and Involution underpin their performance differences. While ViT uses a sequence of image patches processed through transformer mechanisms, which excel in environments with minimal visual change, Involution adapts its kernels based on the input feature map, offering a more flexible and context-aware approach. This adaptability is especially advantageous in distinguishing subtle manipulations in phishing attempts that may escape more rigid models.
- **Dataset Diversity and Accuracy**: The robustness of phishing detection models also hinges on the diversity and accuracy of the dataset used for training. The variability in detection rates, particularly for ViT, suggests potential gaps in the dataset's representation of real-world variations. Enhancing dataset diversity to include a wider range of logo transformations could help in developing more generalizable models that maintain high accuracy across different operational environments.

## 6 Conclusion & Future Directions

The study elucidates the significant role of model architecture and adaptability in detecting phishing websites, especially in the face of varying logo designs. Involution's superior performance across the dataset underscores the importance of using adaptive models for cybersecurity applications where visual content can vary dramatically. However, the varied performance of ViT highlights the need for careful consideration of model capabilities and limitations in relation to the specific characteristics of the task at hand.

Building on the findings of this research, several avenues can be explored to enhance phishing detection capabilities:

1. **Advanced Data Curation Techniques**: Developing advanced techniques for data curation that more accurately reflect the dynamic nature of logos used in phishing attacks. This could involve using machine learning to predict and simulate logo variations for training purposes.
2. **Hybrid and Ensemble Models**: Investigating hybrid and ensemble approaches that combine the strengths of different model architectures, such as integrating ViT's global processing capabilities with Involution's local adaptability. This could potentially lead to a robust model that leverages the best features of both architectures.

3. **Real-Time Detection Systems**: Designing real-time phishing detection systems that integrate these models with live data streams for immediate website verification. This would require optimizing the models for speed and efficiency without compromising accuracy.
4. **Cross-Domain Application and Testing**: Expanding the application and testing of these models to include a broader range of industries and digital contexts. This would help in understanding the models' effectiveness across different sectors and adapting them to diverse cybersecurity needs.
5. **Continuous Model Training and Updating**: Implementing continuous training frameworks to keep the models updated with the latest phishing tactics and logo variations. This could involve creating a feedback loop where new phishing attempts are continually used to refine and adapt the model.
6. **User-Centric Design Considerations**: Developing user-friendly interfaces and warnings that can assist in educating end-users about potential phishing threats, thereby integrating human-centered design principles into cybersecurity measures.

These future directions aim to expand the scope and efficacy of AI-driven phishing detection tools, ensuring they remain effective against the continually evolving tactics of cyber threats.

## 7 Citations and Bibliographies

## References

[1] Mohammed Hazim Alkawaz, Stephanie Joanne Steven, Asif Iqbal Hajamydeen, and Rusyaizila Ramli. 2021. A Comprehensive Survey on Identification and Analysis of Phishing Website based on Machine Learning Methods. In *2021 IEEE 11th IEEE Symposium on Computer Applications Industrial Electronics (ISCAIE)*. 82–87.

[2] Detectron2:online [n. d.]. GitHub - facebookresearch/detectron2: Detectron2 is a platform for object detection, segmentation and other visual recognition tasks. https://github.com/facebookresearch/detectron2. (Accessed on 05/11/2024).

[3] Ankit Kumar Jain B. B. Gupta. 2019. Two-level authentication approach to protect from phishing attacks in real time. *Journal of Ambient Intelligence and Humanized Computing* (2019). https://doi.org/10.1007/s12652-017-0616-z

[4] Yun Lin, Ruofan Liu, Dinil Mon Divakaran, Jun Yang Ng, Qing Zhou Chan, Yiwen Lu, Yuxuan Si, Fan Zhang, and Jin Song Dong. 2021. Phishpedia: A Hybrid Deep Learning Based Approach to Visually Identify Phishing Webpages. In *30th USENIX Security Symposium (USENIX Security 21)*.

[5] Asadullah Safi and Satwinder Singh. 2023. A systematic literature review on phishing website detection techniques. *Journal of King Saud University - Computer and Information Sciences* 35, 2 (2023), 590–611. https://doi.org/10.1016/j.jksuci.2023.01.004

[6] Smita Sindhu, Sunil Parameshwar Patil, Arya Sreevalsan, Faiz Rahman, and Ms. Saritha A. N. 2020. Phishing Detection using Random Forest, SVM and Neural Network with Backpropagation. In *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*. 391–394.

[7] Jing Wang, Weiqing Min, Sujuan Hou, Shengnan Ma, Yuanjie Zheng, Haishuai Wang, and Shuqiang Jiang. 2020. Logo-2K+: A Large-Scale Logo Dataset for Scalable Logo Classification. In *AAAI Conference on Artificial Intelligence. Accepted.*

[8] Erzhou Zhu, Yinyin Ju, Zhile Chen, Feng Liu, and Xianyong Fang. 2020. DTOF-ANN: An Artificial Neural Network phishing detection model based on Decision Tree and Optimal Features. *Applied Soft Computing* 95 (2020), 106505. https://doi.org/10.1016/j.asoc.2020.106505

## A Work Distribution

- **FuJiao Ji**: training model design, model training, result analysis
- **TianHao Wu**: phishing Detector framework design, training model design, data collection
- **MengJun Wang**: data pre-prossessing Model design, data collection, result analysis
- **Qian Liu**: data pre-prossessing Model design, relate work study, data collection