CSEC-472 (Authentication) Literature Review

Improving Facial Recognition Anti-Spoofing: Theoretical and Experimental Adversarial Domain Adaptation Implementations

Wallert, Rick[1], Forward, Karch[1], Follmer, Timothy[1], and Schanzenbacher, Ryan[1]

[1]Department of Computing Security, B. Thomas Golisano College of Computing and Information Science, Rochester Institute of Technology

May 8, 2023

Abstract

With the growing prevalence of facial recognition technology, the need for secure and robust systems has become increasingly important. One major threat to these systems is spoofing attacks, where an individual can deceive the system into accepting a false identity. This literature review aims to explore the potential of Adversarial Domain Adaptation (ADA) to improve facial recognition anti-spoofing systems. A total of 81 papers were reviewed to extract themes regarding ADA in the context of facial recognition anti-spoofing. The three themes identified are "What is ADA?," "Design Types," and "Data Usage." After thorough investigation, it was found that ADA can be utilized to improve the robustness of facial recognition anti-spoofing systems through its dynamic nature. At its core, ADA draws additional testing data from source datasets to target datasets and manipulates the testing data to fit the purpose of its working environment. When working with facial recognition anti-spoofing, one of the strongest ways to improve accuracy and efficiency is with increased testing data, which is why ADA is so effective. Future work in the field of ADA with respect to facial recognition anti-spoofing should allow for the development and integration of advanced technologies into preexisting systems and systems yet to be created.

Keywords

Adversarial Domain Adaptation, domain adaptation, facial recognition, facial recognition anti-spoofing.

1 Introduction

Within recent years facial recognition has become widely used for a variety of purposes. From unlocking personal devices to government security access, this particular form of biometric authentication is seemingly everywhere. If these systems are spoofed, the range of security consequences can be staggering. This interested our team and caused us to focus on the field of facial recognition anti-spoofing. We wanted to see what different systems existed, along with how they functioned. Throughout our research, we stumbled across varying techniques such as depth-detection or infrared sensing. One common theme emerged with all anti-spoofing techniques, though: they all require testing data. Upon further inspection, the vast majority of techniques use their own specialized datasets, and the quality of anti-spoofing seems to increase along with the amount of testing data available [1]. So, this moved our interests elsewhere. Instead of asking the question "what facial recognition anti-spoofing techniques exist," we instead ask "how can you improve both the quality and quantity of testing data?" This is how we discovered the field of Adversarial Domain Adaptation (hereinafter referred to as ADA).

This literature will answer the following questions:

- What is ADA?
- How has it been utilized?
- Does it improve facial recognition anti-spoofing systems?
- Where can ADA be used in the future?

We answer these questions through the use of a literature review. One theme that will be commonly seen is that there is an inadequate amount of testing data being shared between facial recognition anti-spoofing algorithms [2]. As will be proven later on in this paper, the problem with ADA is not a technical one; Adversarial Domain Adaptation already exists, it's just not widely adopted yet [1]. Through conducting a literature review on the current state of ADA, we have answered the questions posed above.

As prior mentioned, facial recognition anti-spoofing is a piece of critical technology in modern society. As biometric systems - specifically facial recognition ones - increase in use and popularity, the attacks targeting them will become stronger accordingly [3]. It is through the use of ADA that we can improve the anti-spoofing systems that are not only already in use, but ones that have yet to be created. Within this paper, the terms "system," "technique," and "tool" will be repeatedly used. The terms "system" and "technique" are synonymous, and both refer to algorithms - specifically facial recognition anti-spoofing algorithms, both ones that exist and have not been created yet. The term "tool" refers to any technology that aids a system without being developed enough to innately be a system. At its core, ADA is a tool - not a system. ADA's use can improve systems as a whole, which is shown throughout our findings.

Section 3, Research Method, clearly documents our data collection techniques. However, it is important to briefly mention our methods here. We combed through 81 different sources to fully understand the field of ADA. These sources canvassed various different fields such as how facial recognition systems and their anti-spoofing algorithms work, what domain adaptation is, how it can be made adversarial, and key techniques used within the prior mentioned subjects. Our research also included literature reviews documenting specific applications of the aforementioned topics. By reviewing publications from well-known journals and conferences, we confidently collected accurate data and effectively informed ourselves on this area of expertise.

Finally, the potential implications of our findings must be discussed. As seen through this review, it is clear that the use of ADA within facial recognition anti-spoofing algorithms is positive. Its use can improve the effectiveness of any technology that requires an increased amount of testing data. So, as more algorithms begin to use ADA, anti-spoofing fields should increase in proficiency at an accelerated rate.

2 Related Work

While research into ADA isn't entirely new, the scope of ADA's facial recognition anti-spoofing application is. This is a direct effect from the lack of information, experimentation, and data into the extent of how anti-spoofing works in tandem with ADA. Over the course of our research, we found articles which seek to unlock further understanding of improving anti-spoofing algorithms using ADA.

Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling: An article which, at present time, is newly published with information regarding the implementation of Convolutional Neural Networks (CNNs) into ADA with the hints of combining CNNs into other technologies [4]. With adequate research and experimentation, anti-spoofing could be one of these technologies, allowing for the creation of datasets which are lacking within the research community.

ExprADA Adversarial domain adaptation for facial expression analysis: Completes a course of experimentation within ADA, specifically adding the implementation of Convolutional Neural Networks (CNNs) which increase the performance of ADA compared to Generative Neural Networks (GANs). There is an important conclusion made within the article suggesting the implementation of ADA with other technologies. This would improve facial quality within images [5]. Further experimentation with relevant data with this article's approach combined with anti-spoofing technology could unlock even further understanding of how ADA is applicable with anti-spoofing.

SoFA: Source-Data-Free Feature Alignment for Unsupervised Domain Adaptation: An article detailing an application of Unsupervised Domain Adaptation (UDA). This publication discusses how UDA, a different form of ADA, has been proven to be an effective approach to solve the problem of domain shift by leveraging both data from the scenario that the model was trained on (source) and the new scenario (target) [6]. The SoFA method works to solve the issue of in-applicability of most UDA methods in the absence of source data by only using the trained source data and unlabeled target data. Although UDA is not the same as ADA, both are forms of domain adaptation. The SoFA technique proves that domain adaptation can improve facial recognition systems even if it is not adversarial.

Gradually Vanishing Bridge for Adversarial Domain Adaptation: An extensive 2020 study

on using ADA with the Gradually Vanishing Bridge (GVB) mechanism. This study discussed how the use of ADA with GVB instead of the use of UDA reduced both the overall transfer difficulty of data and the influence of residual domain-specific characteristics in domain-invariant representations [7]. This specific piece of work did not focus on facial recognition specifically, but instead focused on ADA as a whole, proving that its use can be effective even in fields other than facial recognition anti-spoofing algorithm improvement.

3 Research Method

In order to fully understand the concepts discussed within this paper, we performed an in depth review on varying sources regarding facial recognition, facial recognition anti-spoofing, domain adaptation, and Adversarial Domain Adaptation. These sources fell under two main categories of research:

1. Theoretical Research
2. Experimental Research

The Theoretical Research papers were literature reviews. These papers go into detail on assets of the fields prior mentioned, and they provide insight into the depth that Adversarial Domain Adaptation has within the anti-spoofing facial recognition field. The Experimental Research papers all provide data on individual methods being put to the test. By reading these papers, our team learned about the application side of ADA and the practical results it provides. It's through this precise delineation of sources that we educated ourselves on this particular field.

3.1 Procedures

In order to conduct our research, our team followed a general protocol. This began by traveling to one of three separate well know databases. These were:

- Google Scholar
- IEEE Xplore
- Science Direct

Upon visiting one of these databases, we searched for different keywords. These initially began with searching for the phrase "adversarial learning and domain adaption in facial recognition," but as our research continued we chose to search for terms we saw within multiple different sources. Each source we found had to be within this set of criteria, otherwise, we deemed it unfit for our work:

- Published within the past five years, with an emphasis on publications made within the past three.
- The conference or journal the source was published in needed to be ranked by the Scimago Journal and Country Rank (SJR), yet no limitations were placed on where they fell in the rankings.
- Had to be cited by at least 25 people, with minimal exceptions made.

3.2 Limitations and Risks

When it comes to finding sources on any topic, there are always going to be issues with the quality of publications. We were not exempt from this issue. When researching ADA, we often came across outdated data. Since the field is constantly growing, we had to discount older publications, as they could potentially hold wrong information. Additionally, there was a surprising lack of detailed analysis on the field of ADA. Most sources that we found were pieces of Experimental Research which, although informational, did not always fit our research needs. On top of this, some publications did not have the best H-Index rankings on SJR. However, the difficulties we faced finding informational sources led us to place a lower emphasis on the importance of having high H-Indexes. Finally, as research was conducted, we came across duplicate sources. While this pushed us to consult the different databases listed above, it did make it increasingly difficult to find quality research. Despite all of these limitations, however, we believe that our method of finding research did compromise the integrity of the information we were provided.

4 Findings

Throughout our research into the topic of ADA within facial recognition anti-spoofing, there were several similarities within varying sources. These themes bring out key understandings of how ADA functions, how it has been applied, and how it can be applied.

4.1 What is ADA?

As prior stated, ADA stands for "Adversarial Domain Adaptation." ADA has two key components:

1. It uses Domain Adaptation
2. It is Adversarial

Domain Adaptation is the process of adapting a machine learning model trained on one domain, often referred to as the source domain, to perform well on another domain, often referred to as the target domain. This improves the distribution of data for the target domain. In the context of deep visual Domain Adaptation, this means adapting a deep neural network that has been trained on images from one domain to work effectively on images from another domain. Its use can fix issues where images have differences in lighting, color, texture, or other visual characteristics. Ultimately, domain adaptation minimizes the performance gap between source and target domains to improve the overall generalization ability of the model. [8]

Adversarial Domain Adaptation is a type of unsupervised domain adaptation that seeks to align the feature distributions of the source and target domains by training a domain discriminator to distinguish between the two domains. At the same time, it also trains the feature extractor to produce features that are domain-invariant or indistinguishable by the discriminator. Essentially, the idea of ADA is to create a feature representation that is robust to differences between the source and target domains, allowing a classifier to perform well on the target domain without the need for labeled target data. [9] [7] [1]

4.2 Design Types

ADA is versatile, and there are various different forms of it. One of the many design types encountered in our research was the Single Sample Per Person Domain Adaptation Network (SSPP-DAN) type. SSPP-DAN is a novel deep domain adaptation network for face recognition that is specifically designed to address the problem of Single Sample Per Person (SSPP). SSPP-DAN consists of two main components:

- A feature extraction network
- A domain adaptation network

The feature extraction network is trained on a source domain with a large number of labeled images, and it is used to extract deep features from those input images. The domain adaptation network is then trained on a target domain with a small number of labeled images, and it is used to adapt the feature representation learned from the source domain to the target domain. In order to encourage the feature distributions of the source and target domains to be similar, a Maximum Mean Discrepancy (MMD) loss function is used. The main innovation of SSPP-DAN is its ability to effectively adapt the feature representation learned from the source domain to the target domain with very limited labeled data. [10]

Another commonly used ADA type are those that use Convolutional Neural Networks (CNNs). CNNs are a type of artificial neural network that is commonly used in computer vision tasks such as image recognition, object detection, and image segmentation. CNNs are designed to automatically learn hierarchical feature representations from the input image by applying a series of convolutional, pooling, and activation functions. The convolutional layers consist of a set of filters that slide over the input image, convolving it with the filter and producing a feature map. The pooling layers then downsample the feature maps by taking the maximum or average value of a local region. The activation functions add non-linearity to the network and help to capture complex patterns in the data. [11] [12] [13] [14] [2]

Similarly to CNNs, there also exists Generative Adversarial Networks (GANs). GANs are a type of deep learning model that can generate realistic images by learning the underlying distribution of a given dataset. GANs consist of two parts:

- A generator
- A discriminator

The generator takes a random noise vector as input and produces a synthetic image as output. The discriminator is a binary classifier that takes an image as input and outputs a probability score indicating whether the image is real or fake. During training, the generator and discriminator play a two-player minimax game where the generator tries to generate realistic images that can fool the discriminator. At the same time, the discriminator tries to correctly classify the real and synthetic images. As the generator improves, it becomes increasingly difficult for the discriminator to distinguish between real and synthetic images, resulting in high-quality synthetic images. The use of GANs can, overall, improve face recognition tasks such as data augmentation, domain adaptation, and feature learning. [2] [11] [15]

Another type of network used often with ADA networks are Deep Neural Networks (DNNs). DNNs are a type of neural network that consists of multiple layers of neurons, allowing them to learn complex representations of data. They have been widely used in many machine learning tasks, including image recognition, natural language processing, and speech recognition. In image recognition, DNNs have even been shown to

outperform traditional machine learning methods and have become the dominant approach to machine learning. [16] [5] [13] [17]

These different ADA techniques have seen intensive application. One specific facial recognition ADA system is referenced within §2-Related Work. This specific system is ExprADA, and it is a CNN-based ADA system. By using a CNN model instead of a GAN model, the creators of ExprADA yield higher expression accuracy percentages. By extrapolating data directly from their study, it can be seen that ExprADA is, on average, 2.43% more accurate than GAN counterparts. Although this might sound like a small percentage, these systems are being used on thousands of images. For a situation where 3000 faces are being scanned, an increase of 2.43% accuracy can properly identify roughly 73 additional people. For technology that can have potential security implications, this is a vital increase in accuracy. [5]

4.3 Data Usage

When using ADA, there are various predetermined datasets for facial recognition testing purposes. These datasets contain different facial images to test the efficiency of ADA when paired with anti-spoofing. One highly regarded dataset in the research community is CASIA-WebFace. This dataset contains 494,414 facial images collected from across the Internet. These images are comprised of 10,575 real identities, and they are also under fair conditions with ambient lighting for testing purposes. [12]

The CASIA-SURF dataset is also relevant within the research community for anti-spoofing purposes. The CASIA-SURF dataset is similar to CASIA-WebFace, where it consists of 492,522 images. These images have three different modalities: RGB, Depth, and IR [18]. The dataset also contains various facial images across three different ethnicity groups with various levels, and is intended to prevent spoofing attacks such as:

- Print attacks,
- Replay attacks,
- and 3D mask attacks

A print attack is one wherein an attacker uses a printed photo of someone's face. A replay attack is one where a looped video or image is played of the victim's face with the goal of making it seem more natural compared to holding someone's photo. A 3D mask attack aims to use a 3D printed mask created from an image of a victim to deceive advanced anti-spoofing filters. Using this dataset allows ADA models paired with anti-spoofing to test the efficiency and accuracy of the model. [19]

5 Conclusions

As discussed within this paper, ADA can help improve the robustness of facial recognition systems against spoofing attacks through its dynamic nature. By implementing Domain Adaptation, ADA is able to draw in additional data for target domains in order to improve the availability of testing data. Additionally, through its adversarial use, ADA can be modified to manipulate the testing data for the particular purpose of the system it is being applied to. This approach has shown promising results in several studies including improving the accuracy of facial recognition anti-spoofing systems on cross-dataset evaluations. The use of ADA with facial recognition anti-spoofing can lead to more reliable and secure facial recognition systems which are increasingly important in various domains (such as security and privacy).

Implications

Future work in the field of ADA with respect to facial recognition anti-spoofing would allow for the development and integration of advanced anti-spoofing technologies into preexisting systems. This implementation would increase both the accuracy and security of varying systems. Future work must also take into consideration the challenges of developing these technologies by creating "spoofable" architectures to implement the ADA models to better understand how anti-spoofing systems are defeated. [3]

References

[1] Wang G, Han H, Shan S, Chen X. Improving Cross-database Face Presentation Attack Detection via Adversarial Domain Adaptation. In: 2019 International Conference on Biometrics (ICB); 2019. p. 1–8.

[2] Guo G, Zhang N. A survey on deep learning based face recognition. Computer Vision and Image Understanding. 2019;189:102805. Available from: https://www.sciencedirect.com/science/article/pii/S1077314219301183.

[3] Galbally J, Marcel S, Fierrez J. Biometric Antispoofing Methods: A Survey in Face Recognition. IEEE Access. 2014;2:1530–1552.

[4] Hossain S, Umer S, Rout RK, Tanveer M. Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling. Applied Soft Computing. 2023 Feb;134:109997. Available from: https://doi.org/10.1016/j.asoc.2023.109997.

[5] Bozorgtabar B, Mahapatra D, Thiran JP. ExprADA: Adversarial domain adaptation for facial expression analysis. Pattern Recognition. 2020;100:107111. Available from: https://www.sciencedirect.com/science/article/pii/S0031320319304121.

[6] Yeh HW, Yang B, Yuen PC, Harada T. SoFA: Source-Data-Free Feature Alignment for Unsupervised Domain Adaptation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2021. p. 474–483.

[7] Cui S, Wang S, Zhuo J, Su C, Huang Q, Tian Q. Gradually vanishing bridge for adversarial domain adaptation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 12455–12464.

[8] Wang M, Deng W. Deep visual domain adaptation: A survey. Neurocomputing. 2018;312:135–153. Available from: https://www.sciencedirect.com/science/article/pii/S0925231218306684.

[9] Shu R, Bui HH, Narui H, Ermon S. A DIRT-T Approach to Unsupervised Domain Adaptation. arXiv; 2018. Available from: https://arxiv.org/abs/1802.08735.

[10] Hong S, Im W, Ryu J, Yang HS. SSPP-DAN: Deep domain adaptation network for face recognition with single sample per person. In: 2017 IEEE International Conference on Image Processing (ICIP); 2017. p. 825–829.

[11] Wang X, Wang X, Ni Y. Unsupervised Domain Adaptation for Facial Expression Recognition Using Generative Adversarial Networks. Computational Intelligence and Neuroscience. 2018 Jul;2018:1–10. Available from: https://doi.org/10.1155/2018/7208794.

[12] Wang M, Deng W. Deep face recognition with clustering based domain adaptation. Neurocomputing. 2020 Jun;393:1–14. Available from: https://doi.org/10.1016/j.neucom.2020.02.005.

[13] Luo Z, Hu J, Deng W, Shen H. Deep unsupervised domain adaptation for face recognition. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE; 2018. p. 453–457.

[14] Aggarwal A, Kumar M. Image surface texture analysis and classification using deep learning. Multimedia Tools and Applications. 2021;80:1289–1309.

[15] Ali W, Tian W, Din SU, Iradukunda D, Khan AA. Classical and modern face recognition approaches: a complete review. Multimedia tools and applications. 2021;80:4825–4880.

[16] Zhong L, Fang Z, Liu F, Yuan B, Zhang G, Lu J. Bridging the Theoretical Bound and Deep Algorithms for Open Set Domain Adaptation. IEEE Transactions on Neural Networks and Learning Systems. 2021:1–15.

[17] He G, Liu X, Fan F, You J. Classification-Aware Semi-Supervised Domain Adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops; 2020. .

[18] Zhang S, Wang X, Liu A, Zhao C, Wan J, Escalera S, et al.. A Dataset and Benchmark for Large-scale Multi-modal Face Anti-spoofing; 2019.

[19] Li A, Tan Z, Li X, Wan J, Escalera S, Guo G, et al.. CASIA-SURF CeFA: A Benchmark for Multi-modal Cross-ethnicity Face Anti-spoofing; 2020.