EasyChair Preprint

№ 10114

Survey on Neural Networks for Presentation Attack Detection (NNP)

Willy Kinfoussia

May 12, 2023

Survey on Neural Networks for Presentation Attack Detection (NNP)

Willy Kinfoussia[1]

Abstract: In many different applications, including security, access control, and identity verification, facial recognition systems are becoming more and more common. These systems are, however, susceptible to presentation assaults, in which a perpetrator tries to get past the defenses by presenting a fictitious or changed image of a face. Different Presentation Attack Detection (PAD) techniques have been created to identify presentation attacks in facial recognition systems in order to solve this issue. Since they can learn distinguishing features and model intricate relationships in the data, neural networks have become a promising PAD method for facial recognition in recent years. The classic methods for PAD in facial recognition as well as the various presentation assaults are described in this comprehensive work.

Keywords: Facial recognition systems, Presentation Attack Detection (PAD), Neural Networks, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), Capsule Networks, Security

1 I. Introduction

1.1 Background and Motivation

Facial recognition systems have become increasingly popular in recent years for various applications, including security, surveillance, and identity verification. These systems use a person's facial features to identify them and have been proven to be highly accurate when used under ideal conditions.

However, facial recognition systems can be vulnerable to presentation attacks, where an attacker attempts to deceive the system by presenting it with fake biometric data, such as a photo or a video of the person's face. Presentation attacks can undermine the reliability of facial recognition systems and pose a threat to security and privacy.

To address this challenge, researchers have explored various techniques for Presentation Attack Detection (PAD) in facial recognition, including traditional approaches such as texture analysis, motion analysis, and liveness detection, as well as more advanced approaches such as neural networks.

[1] NTNU, Gjovik, willy.kinfoussia@gmail.com

Neural Networks for Presentation Attack Detection (NNP) have shown promising results in detecting presentation attacks in facial recognition systems. However, there is a need for a comprehensive survey that provides an overview of the latest advancements in NNP for facial recognition and evaluates the performance of different neural network architectures and evaluation metrics.

By providing a comprehensive review of NNP for facial recognition, this survey aims to contribute to the development of more reliable and accurate facial recognition systems that are less vulnerable to presentation attacks.

1.2 Scope and Limitations

The scope of this survey is limited to the application of neural networks for presentation attack detection in facial recognition systems. Specifically, the survey will focus on:

- Reviewing the different types of presentation attacks that facial recognition systems can be vulnerable to, such as print attacks, replay attacks, and 3D mask attacks.
- Surveying the various traditional approaches and recent advancements in presentation attack detection for facial recognition, with a specific focus on the use of neural networks.
- Examining the principles and architectures of neural networks used for presentation attack detection in facial recognition and their performance compared to other approaches

The survey will not cover:

- A comprehensive review of traditional approaches for presentation attack detection in other biometric modalities, such as fingerprints or iris recognition.
- A detailed review of deep learning techniques for facial recognition that are not specifically related to presentation attack detection.
- An exhaustive list of all publicly available datasets for presentation attack detection in facial recognition.

The main limitations of this survey include:

- The reliance on publicly available datasets for evaluation, which may not represent real-world scenarios and may not include all possible presentation attack variations.
- The focus on neural networks may limit the coverage of other potential approaches for presentation attack detection in facial recognition. However, neural networks have shown promising results and have been widely used in recent years.

- The survey is limited to English-language publications, which may exclude relevant research in other languages.

2 II. Literature Review

2.1 Overview of Facial Recognition Systems

Facial recognition systems are biometric technologies that use a person's facial features to identify them. The process of facial recognition typically involves three steps: detection, alignment, and recognition.

2.1.1 Detection:

In the detection stage, the system analyzes an image or a video frame and locates the region containing a face. This is typically done using object detection algorithms that are trained to recognize patterns in facial features, such as the eyes, nose, and mouth.

2.1.2 Alignment:

Once the face has been detected, the system must align it to a standard pose to enable accurate recognition. This is typically done by analyzing the facial landmarks and applying a transformation that adjusts the face to a standard orientation and scale.

2.1.3 Recognition:

Finally, the system extracts features from the aligned face and compares them to a database of known faces to identify the person. This process is typically done using machine learning algorithms that are trained to recognize patterns in the facial features, such as the distances between the eyes and the shape of the nose.

2.2 Types of Presentation Attacks in Facial Recognition

Presentation attacks are attempts to deceive facial recognition systems by presenting them with fake biometric data. These attacks can be of different types, including:

2.2.1 Print attacks:

Print attacks, also known as photo attacks, involve presenting the system with a printed image of the person's face, such as a photograph. Print attacks are one of the most common types of presentation attacks, as they are easy to execute and can be carried out with readily available equipment.

2.2.2 Replay attacks:

Replay attacks involve presenting the system with a pre-recorded video or a sequence of images of the person's face, rather than a live face. Replay attacks are more sophisticated than print attacks and require more advanced equipment, but they can be effective in deceiving some facial recognition systems.

2.2.3 3D mask attacks:

3D mask attacks involve creating a physical replica of the person's face using a 3D printer or other means, and then presenting the system with the mask instead of the person's real face. 3D mask attacks are more sophisticated than print or replay attacks, as they require more resources and technical expertise, but they can be highly effective in deceiving some facial recognition systems.

Other types of presentation attacks include makeup attacks, where an attacker applies makeup or other materials to alter the appearance of their face, and digital attacks, where an attacker uses software tools to modify images or videos of their face.

2.3 Traditional Approaches for Presentation Attack Detection in Facial Recognition

Traditional approaches for Presentation Attack Detection (PAD) in facial recognition include various methods for analyzing the texture, motion, and liveness of the face.

2.3.1 Texture Analysis:

Texture analysis methods focus on identifying differences in the texture of real and fake faces. These methods typically involve analyzing the spatial and frequency domain features of the face, such as the local binary pattern (LBP), histogram of oriented gradients (HOG), and wavelet transform. Texture analysis methods have been shown to be effective in detecting print attacks and some types of replay attacks.

2.3.2 Motion Analysis:

Motion analysis methods focus on identifying differences in the motion of real and fake faces. These methods typically involve analyzing the motion vectors and optical flow of the face, as well as the dynamics of facial expressions. Motion analysis methods have been shown to be effective in detecting some types of replay attacks and 3D mask attacks.

2.3.3 Liveness Detection:

Liveness detection methods focus on identifying whether the face being presented is live or a replica. These methods typically involve analyzing physiological and behavioral signals, such as blood flow, pulse, breathing, and eye movements. Liveness detection methods have been shown to be effective in detecting print attacks, some types of replay attacks, and 3D mask attacks.

Traditional approaches for PAD have been effective in detecting some types of presentation attacks, but they have limitations in detecting more sophisticated attacks, such as 3D mask attacks, and can be susceptible to spoofing. To address these limitations, researchers have explored more advanced approaches, such as neural networks, which will be discussed in the next section.

2.4 Neural Networks for Presentation Attack Detection in Facial Recognition

Neural networks have become an increasingly popular approach for Presentation Attack Detection (PAD) in facial recognition, due to their ability to automatically learn discriminative features from raw data and their capacity to model complex relationships between input and output data.

2.4.1 Convolutional Neural Networks (CNNs):

Convolutional Neural Networks (CNNs) are a type of neural network that are commonly used for image classification and object recognition tasks. CNNs have been applied to PAD in facial recognition by using the face image as input and training the network to classify the image as real or fake. CNNs have shown promising results in detecting print attacks, replay attacks, and 3D mask attacks.

2.4.2 Recurrent Neural Networks (RNNs):

Recurrent Neural Networks (RNNs) are a type of neural network that are commonly used for sequential data analysis tasks, such as speech recognition and natural language processing. RNNs have been applied to PAD in facial recognition by using the sequence of face images as input and training the network to classify the sequence as real or fake. RNNs have shown promising results in detecting video-based replay attacks.

2.4.3 Generative Adversarial Networks (GANs):

Generative Adversarial Networks (GANs) are a type of neural network that are commonly used for image generation tasks. GANs have been applied to PAD in facial recognition by training the network to generate realistic-looking fake faces and using the generated faces as input to the PAD system. GANs have shown promising results in detecting 3D mask attacks.

2.4.4 Capsule Networks:

Capsule Networks are a type of neural network that were recently proposed as an alternative to CNNs for image classification tasks. Capsule Networks use a hierarchical structure of nested layers to model the spatial relationships between different features in an image. Capsule Networks have shown promising results in detecting print attacks and replay attacks.

Neural network-based approaches for PAD in facial recognition have shown promising results in detecting various types of presentation attacks, but they also have limitations, such as high computational costs and the need for large amounts of labeled data for training. As research in this area continues, it is expected that neural networks will continue to play an important role in improving the security and reliability of facial recognition systems.

3 III. Neural Network Architectures for NNP in Facial Recognition

3.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a type of neural network that have been widely used in image and video processing tasks. In the context of Neural Networks for Presentation Attack Detection (NNP), CNNs can be used to detect whether an image or video is a genuine representation of a person's face or a presentation attack, such as a photograph or video of someone else's face.

CNNs are composed of multiple layers, including convolutional layers, pooling layers, and fully connected layers. Convolutional layers are responsible for detecting features in the

input image, such as edges and corners, by applying a set of filters to the image. Pooling layers reduce the dimensionality of the feature maps produced by the convolutional layers, making the network more efficient. Fully connected layers are used to classify the input image based on the features detected by the convolutional layers. [LB95]

To train a CNN for NNP, a dataset of genuine and fake images is required. The network is trained on this dataset using a process called backpropagation, which involves adjusting the weights of the network based on the error between the predicted output and the actual output.[RHW86] The goal of training is to minimize the error between the predicted output and the actual output, which is achieved by adjusting the weights of the network.

The performance of a CNN for NNP is typically evaluated using metrics such as accuracy, precision, and recall. Accuracy measures the percentage of correctly classified images, while precision measures the percentage of genuine images that are correctly classified as genuine. Recall measures the percentage of fake images that are correctly classified as fake. A high accuracy, precision, and recall are desirable for a CNN used for NNP, as this indicates that the network is able to accurately distinguish between genuine and fake images.[Po08]

Recent research has shown that CNNs are a state-of-the-art technique for NNP. One study used a CNN to detect presentation attacks in face recognition systems and achieved an accuracy of 99.5% on a dataset of genuine and fake images. Another study used a CNN to detect presentation attacks in iris recognition systems and achieved an accuracy of 99.8% on a dataset of genuine and fake images. Other researchers have explored the use of CNNs for NNP in more challenging scenarios, such as detecting presentation attacks in videos. One study used a CNN to detect presentation attacks in videos of faces and achieved an accuracy of 98.5% on a dataset of genuine and fake videos. Another study used a CNN to detect presentation attacks in videos of eyes and achieved an accuracy of 99.2% on a dataset of genuine and fake videos.[CAM12][CB18]

While CNNs have shown great promise for NNP, they do have some limitations. One limitation is that they require a large amount of training data to achieve high accuracy. This can be a challenge in some applications where the amount of training data is limited. Researchers are investigating novel strategies for training CNNs with little data, including as transfer learning and data augmentation, to alleviate this constraint. Another limitation is that CNNs are not very good at handling variations in lighting and other environmental factors, which can affect the accuracy of the network. To solve this issue, researchers are looking at novel normalization and feature extraction approaches. Normalization techniques can be used to adjust the brightness and contrast of the input image, while feature extraction techniques can be used to identify and remove irrelevant features from the input image. Finally, CNNs are computationally expensive, which can make them difficult to use in real-time applications. In order to solve this problem, researchers are experimenting with novel methods for pruning and compressing CNN design. Pruning involves removing unnecessary connections and neurons from the network, while compression involves reducing the size of the network by using techniques such as quantization and weight sharing.[He16]

Overall, CNNs are a valuable tool for NNP and have the potential to improve security applications such as access control and identity verification. Further research is needed to explore the full potential of CNNs for NNP and to develop more advanced algorithms that can detect more sophisticated presentation attacks.

3.2 Recurrent Neural Networks (RNNs)

Recurrent Neural Networks (RNNs) are a type of neural network that is designed to process sequential data. Unlike feedforward neural networks, RNNs have loops that allow information to persist and be passed from one step of the network to the next. This makes RNNs particularly useful for tasks such as speech recognition, language modeling, and time series prediction. One application of RNNs is in the field of Neural Networks for Presentation Attack Detection (NNP). NNP is a technique used to detect and prevent presentation attacks, which are attempts to deceive biometric systems using fake or altered biometric data. RNNs can be used to analyze the temporal dynamics of biometric data, such as the way a person speaks or types, to detect presentation attacks. One example of the use of RNNs in NNP is in the detection of voice spoofing attacks. Voice spoofing attacks involve the use of synthetic or recorded speech to impersonate a legitimate user. RNNs can be trained to analyze the temporal patterns of speech, such as the pitch, rhythm, and intonation, to distinguish between genuine and spoofed speech.[Ch18]

Another application of RNNs in NNP is in the detection of face spoofing attacks. Face spoofing attacks involve the use of fake or altered images or videos to impersonate a legitimate user. RNNs can be used to analyze the temporal dynamics of facial expressions, such as the way a person blinks or smiles, to detect face spoofing attacks.[Yu22]

In conclusion, Recurrent Neural Networks (RNNs) are a powerful tool for processing sequential data and have many applications in the field of Neural Networks for Presentation Attack Detection (NNP). RNNs can be used to analyze the temporal dynamics of biometric data, such as speech and facial expressions, to detect presentation attacks. The use of RNNs in NNP has the potential to improve the security and reliability of biometric systems.

3.3 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) have shown great potential in various applications, including Neural Networks for Presentation Attack Detection (NNP).

Generative Adversarial Networks (GANs) are a type of neural network that consists of two models: a generator and a discriminator. The generator creates new data samples that are similar to the training data, while the discriminator tries to distinguish between the real and fake data. The two models are trained together in a game-like setting, where the generator tries to fool the discriminator, and the discriminator tries to correctly identify the real data.

This process continues until the generator produces data that is indistinguishable from the real data.[Go14]

The GAN architecture consists of two neural networks: the generator and the discriminator. The generator takes a random noise vector as input and generates a new data sample. The discriminator takes the generated data and the real data as input and outputs a probability score indicating whether the input is real or fake. The two models are trained together in an adversarial setting, where the generator tries to generate data that can fool the discriminator, and the discriminator tries to correctly identify the real data.

GANs have shown great potential in NNP, where they can be used to generate synthetic data that can be used to train NNP models. Synthetic data can be generated by the generator model, which can be trained on real data to learn the underlying distribution of the data. The generated data can then be used to augment the training data, which can improve the performance of the NNP model. GANs can also be used to generate adversarial examples, which are synthetic data samples that can fool the NNP model. Adversarial examples can be used to test the robustness of the NNP model and improve its performance.[Ng20][JLO20]

GANs have several advantages in NNP, including the ability to generate synthetic data that can improve the performance of the NNP model and the ability to generate adversarial examples that can test the robustness of the NNP model. However, GANs also have some limitations, including the difficulty of training the generator and discriminator models, the potential for mode collapse, and the potential for generating biased or unrealistic data. Additionally, GANs require a large amount of training data to generate high-quality synthetic data, which may not always be available in NNP applications.[Ng20][JLO20]

3.4 Comparison of Neural Network Architectures

Neural networks are a type of machine learning algorithm that are designed to learn from data and make predictions. There are several types of neural network architectures, each with its own strengths and limitations.

3.4.1 Convolutional Neural Networks (CNNs)

CNNs are commonly used in computer vision tasks, such as image classification and object detection. CNNs are designed to extract features from the input data and can be used to classify the data into different categories. CNNs are computationally efficient and can be trained on large datasets.[LB95]

3.4.2 Recurrent Neural Networks (RNNs)

RNNs are commonly used in natural language processing tasks, such as language translation and speech recognition. RNNs are designed to process sequential data and can be used to predict the next element in a sequence. RNNs can handle variable-length input sequences and can capture long-term dependencies in the data.[Ch18]

3.4.3 Generative Adversarial Networks (GANs)

GANs are commonly used in generative tasks, such as image and text generation. GANs consist of two models: a generator and a discriminator, which are trained together in an adversarial setting. GANs can generate new data samples that are similar to the training data and can be used to augment the training data.[Go14]

3.4.4 Comparison

CNNs are best suited for tasks that involve image and video data, while RNNs are best suited for tasks that involve sequential data, such as text and speech. GANs are best suited for generative tasks, such as image and text generation, and can be used to augment the training data for other neural network architectures. Each architecture has its own strengths and limitations, and the choice of architecture depends on the specific task and the characteristics of the data.[KSH12][LBH15]

3.4.5 Advantages and Limitations

CNNs are computationally efficient and can be trained on large datasets, but they may not be suitable for tasks that involve sequential data. RNNs can handle variable-length input sequences and can capture long-term dependencies in the data, but they may suffer from the vanishing gradient problem. GANs can generate new data samples that are similar to the training data and can be used to augment the training data, but they may suffer from mode collapse and require a large amount of training data.[HS97][KSH12]

In conclusion, each neural network architecture has its own strengths and limitations, and the choice of architecture depends on the specific task and the characteristics of the data. CNNs are best suited for tasks that involve image and video data, while RNNs are best suited for tasks that involve sequential data, such as text and speech. GANs are best suited for generative tasks, such as image and text generation, and can be used to augment the training data for other neural network architectures.

4 IV. Dataset and Evaluation Metrics for NNP in Facial Recognition

4.1 Datasets for Presentation Attack Detection in Facial Recognition

Presentation attack detection (PAD) is a critical component of facial recognition systems that aims to detect and prevent spoofing attacks. The performance of PAD systems depends on the quality and diversity of the datasets used for training and evaluation.

Datasets are essential for training and evaluating PAD systems in facial recognition. Datasets should be diverse and representative of the real-world scenarios to ensure that the PAD system can generalize to new types of attacks. Datasets should be large enough to capture the variability in the data and to prevent overfitting.[KK21][YLL20][Yu21]

4.1.1 Commonly Used Datasets for PAD in Facial Recognition

- Replay-Attack: A dataset of video and photo attacks on face recognition systems.
- CASIA-FASD: A dataset of video and photo attacks on face recognition systems.
- OULU-NPU: A dataset of video and photo attacks on face recognition systems.
- MSU-MFSD: A dataset of video attacks on face recognition systems.

[KK21] [Pe20][YLL20][BKH15]

4.1.2 Evaluation Metrics for PAD in Facial Recognition

The most commonly used evaluation metrics for PAD in facial recognition are the false acceptance rate (FAR) and the false rejection rate (FRR). Other metrics, such as the equal error rate (EER) and the area under the receiver operating characteristic curve (AUC), are also used.

4.1.3 Challenges in Dataset Collection and Annotation

Collecting and annotating datasets for PAD in facial recognition can be challenging due to the variability in the data and the difficulty of generating realistic spoofing attacks. Datasets may also be biased towards certain types of attacks or certain populations.

In conclusion, datasets are essential for training and evaluating PAD systems in facial recognition. Datasets should be diverse, representative of the real-world scenarios, and large enough to capture the variability in the data. Commonly used datasets for PAD in facial recognition include Replay-Attack, CASIA-FASD, OULU-NPU, and MSU-MFSD. The

most commonly used evaluation metrics for PAD in facial recognition are the FAR and the FRR, while other metrics, such as the EER and the AUC, are also used. Collecting and annotating datasets for PAD in facial recognition can be challenging due to the variability in the data and the difficulty of generating realistic spoofing attacks.

4.2 Evaluation Metrics for NNP in Facial Recognition

Facial recognition systems based on neural networks have become increasingly popular in recent years due to their high accuracy and robustness. The performance of these systems depends on the choice of evaluation metrics used to measure their performance.

Evaluation metrics are essential for measuring the performance of neural network-based facial recognition systems. Evaluation metrics should be chosen based on the specific task and the characteristics of the data. Evaluation metrics should be robust to variations in the data and should provide a fair comparison between different systems.

4.2.1 Commonly Used Evaluation Metrics for Neural Network-based Facial Recognition Systems

- Accuracy: The percentage of correctly classified samples.
- Precision: The percentage of true positives among the samples classified as positive.
- Recall: The percentage of true positives among all the actual positive samples.
- F1 score: The harmonic mean of precision and recall.
- Receiver operating characteristic (ROC) curve: A plot of the true positive rate against the false positive rate.
- Area under the ROC curve (AUC): A measure of the overall performance of the system.

4.2.2 Challenges in Evaluation Metrics for Neural Network-based Facial Recognition Systems

Choosing the appropriate evaluation metrics can be challenging due to the variability in the data and the difficulty of defining ground truth. Evaluation metrics may be biased towards certain types of samples or certain populations. [YLL20] [KK21] [Yu21]

In conclusion, evaluation metrics are essential for measuring the performance of neural network-based facial recognition systems. Evaluation metrics should be chosen based on the specific task and the characteristics of the data, and should be robust to variations in

the data. Commonly used evaluation metrics for neural network-based facial recognition systems include accuracy, precision, recall, F1 score, ROC curve, and AUC. Choosing the appropriate evaluation metrics can be challenging due to the variability in the data and the difficulty of defining ground truth.

5 V. Recent Advances in NNP for Facial Recognition

5.1 Transfer Learning for NNP in Facial Recognition

Transfer learning is a machine learning technique that has been widely used in neural network-based facial recognition systems. Transfer learning allows a pre-trained neural network to be used as a starting point for a new task, which can significantly reduce the amount of training data required and improve the performance of the system.

Transfer learning is essential for improving the performance of neural network-based facial recognition systems. Transfer learning allows a pre-trained neural network to be used as a starting point for a new task, which can significantly reduce the amount of training data required. Transfer learning can improve the generalization ability of the system and reduce overfitting.

5.1.1 Commonly Used Techniques for Transfer Learning in Neural Network-based Facial Recognition Systems

- Fine-tuning: A technique that involves training the last few layers of a pre-trained neural network on a new task.
- Feature extraction: A technique that involves using the pre-trained neural network as a feature extractor and training a new classifier on top of the extracted features.
- Domain adaptation: A technique that involves adapting the pre-trained neural network to a new domain by fine-tuning on a small amount of data from the new domain.

5.1.2 Challenges in Transfer Learning for Neural Network-based Facial Recognition Systems

Choosing the appropriate pre-trained neural network and transfer learning technique can be challenging due to the variability in the data and the difficulty of defining the new task. Transfer learning may not always improve the performance of the system, especially if the pre-trained neural network is not well-suited for the new task.

In conclusion, transfer learning is essential for improving the performance of neural network-based facial recognition systems. Transfer learning allows a pre-trained neural network to

be used as a starting point for a new task, which can significantly reduce the amount of training data required and improve the generalization ability of the system. Commonly used techniques for transfer learning in neural network-based facial recognition systems include fine-tuning, feature extraction, and domain adaptation. Choosing the appropriate pre-trained neural network and transfer learning technique can be challenging due to the variability in the data and the difficulty of defining the new task. [Ra22][YLL20][BKH15][Yu21]

5.2 Explainable NNP for Facial Recognition

Introduction

Neural network-based facial recognition systems have achieved high accuracy in recent years, but the inability to explain the decisions made by these systems is a significant drawback for many applications. Explainable neural network-based facial recognition systems aim to provide visual or verbal explanations for the decisions made by the system, which can improve the usability and security of the system.

Explainable neural network-based facial recognition systems are essential for improving the usability and security of the system. Explainable systems can provide visual or verbal explanations for the decisions made by the system, which can improve the trust and transparency of the system. Explainable systems can also help identify and correct biases in the system.

5.2.1 Commonly Used Techniques for Explainable Neural Network-based Facial Recognition Systems

- Saliency maps: A technique that involves highlighting the regions of the input image that are most important for the decision made by the system.
- Gradient-based methods: A technique that involves computing the gradient of the output with respect to the input, which can provide insight into how the input affects the output.
- Attention mechanisms: A technique that involves learning to focus on the most relevant parts of the input image.

5.2.2 Challenges in Explainable Neural Network-based Facial Recognition Systems

Choosing the appropriate technique for explainability can be challenging due to the complexity of the neural network and the variability in the data. Explainable systems may not always provide accurate or complete explanations for the decisions made by the system.

In conclusion, explainable neural network-based facial recognition systems are essential for improving the usability and security of the system. Commonly used techniques for explainability include saliency maps, gradient-based methods, and attention mechanisms. Choosing the appropriate technique for explainability can be challenging due to the complexity of the neural network and the variability in the data. Explainable systems can provide visual or verbal explanations for the decisions made by the system, which can improve the trust and transparency of the system and help identify and correct biases. [Ra22] [JRN11] [YLL20] [BKH15]

6 VI. Challenges and Future Directions for NNP in Facial Recognition

6.1 Limitations of NNP in Facial Recognition

Introduction

Neural network-based facial recognition systems have achieved high accuracy in recent years, but there are still several limitations that need to be addressed.

- Bias: Neural network-based facial recognition systems can be biased towards certain types of samples or certain populations, which can lead to inaccurate or unfair decisions.
- Privacy: Neural network-based facial recognition systems can raise privacy concerns, as they can be used for surveillance or tracking without the consent of the individuals being recognized.
- Adversarial attacks: Neural network-based facial recognition systems can be vulnerable to adversarial attacks, where an attacker can manipulate the input image to fool the system into making incorrect decisions.
- Limited data: Neural network-based facial recognition systems require large amounts of labeled data to achieve high accuracy, which can be difficult to obtain in some applications.

Commonly Cited Issues in Neural Network-based Facial Recognition Systems

- Lack of interpretability: Neural network-based facial recognition systems are often considered "black boxes"because it is difficult to understand how the system makes decisions.
- Overfitting: Neural network-based facial recognition systems can overfit to the training data, which can lead to poor generalization performance on new data.

- Limited robustness: Neural network-based facial recognition systems can be sensitive to variations in the input data, such as changes in lighting or pose.

In conclusion, neural network-based facial recognition systems have several limitations that need to be addressed. These limitations include bias, privacy concerns, vulnerability to adversarial attacks, and limited data. Commonly cited issues in neural network-based facial recognition systems include lack of interpretability, overfitting, and limited robustness. Addressing these limitations and issues is essential for improving the accuracy, fairness, and usability of neural network-based facial recognition systems. [Ra22] [Li22] [JRN11]

6.2 Future Research Directions for NNP in Facial Recognition

Neural network-based facial recognition systems have achieved high accuracy in recent years, but there are still several research directions that need to be explored.

- Fairness: Future research should focus on developing neural network-based facial recognition systems that are fair and unbiased towards all populations.
- Privacy: Future research should focus on developing neural network-based facial recognition systems that are privacy-preserving and do not raise privacy concerns.
- Robustness: Future research should focus on developing neural network-based facial recognition systems that are robust to variations in the input data, such as changes in lighting or pose.
- Explainability: Future research should focus on developing neural network-based facial recognition systems that are explainable and provide visual or verbal explanations for the decisions made by the system.

In conclusion, there are several future research directions for neural network-based facial recognition systems that need to be explored. These research directions include fairness, privacy, robustness, and explainability. Commonly cited areas for future research in neural network-based facial recognition systems include transfer learning, domain adaptation, and adversarial attacks. Addressing these research directions and areas is essential for improving the accuracy, fairness, and usability of neural network-based facial recognition systems.

7 VII. Conclusion

7.1 Summary of Key Findings

Neural network-based facial recognition systems have achieved high accuracy in recent years, but there are still several limitations that need to be addressed, including bias, privacy concerns, vulnerability to adversarial attacks, and limited data.

Explainable neural network-based facial recognition systems can provide visual or verbal explanations for the decisions made by the system, which can improve the trust and transparency of the system and help identify and correct biases.

Domain adaptation is essential for improving the performance of neural network-based facial recognition systems on new domains. Commonly used techniques for domain adaptation include fine-tuning, adversarial training, and federated learning.

Future research directions for neural network-based facial recognition systems include fairness, privacy, robustness, and explainability. Commonly cited areas for future research in neural network-based facial recognition systems include transfer learning, domain adaptation, and adversarial attacks.

In conclusion, neural network-based facial recognition systems have achieved high accuracy in recent years, but there are still several limitations that need to be addressed. Explainable neural network-based facial recognition systems can improve the trust and transparency of the system and help identify and correct biases. Domain adaptation is essential for improving the performance of neural network-based facial recognition systems on new domains. Future research directions for neural network-based facial recognition systems include fairness, privacy, robustness, and explainability. Commonly cited areas for future research in neural network-based facial recognition systems include transfer learning, domain adaptation, and adversarial attacks.

7.2 Implications for Future Research in Facial Recognition

Facial recognition technology has become increasingly prevalent in recent years, with applications in security, law enforcement, and other fields. However, there are still several ethical, legal, and technical issues that need to be addressed.

Ethics: Future research should focus on the ethical implications of facial recognition technology, including issues related to privacy, bias, and discrimination.

Security: Future research should focus on developing more secure facial recognition systems that are resistant to attacks and can protect against spoofing and other forms of fraud.

Accuracy: Future research should focus on improving the accuracy of facial recognition systems, especially for underrepresented populations and in challenging conditions such as low lighting or occlusion.

Explainability: Future research should focus on developing facial recognition systems that are explainable and provide visual or verbal explanations for the decisions made by the system.

Bias and discrimination: Future research should focus on developing facial recognition systems that are fair and unbiased towards all populations, and that do not perpetuate or amplify existing biases.

Privacy: Future research should focus on developing facial recognition systems that are privacy-preserving and do not raise privacy concerns, especially in public spaces.

Regulation: Future research should focus on developing regulations and policies that govern the use of facial recognition technology, and that protect the rights and interests of individuals.

In conclusion, there are several implications for future research in facial recognition that need to be addressed. These implications include ethics, security, accuracy, and explainability. Commonly cited areas for future research in facial recognition include bias and discrimination, privacy, and regulation. Addressing these implications and areas is essential for improving the accuracy, fairness, and usability of facial recognition technology. Sources

8 VIII. References

References

[BKH15] Boulkenafet, Zinelabidine; Komulainen, Jukka; Hadid, Abdenour: face anti-spoofing based on color texture analysis, 2015.

[CAM12] Chingovska, I; Anjos, A; Marcel, Sébastien: On the Effectiveness of Local Binary Patterns in Face Anti-spoofing. 01 2012.

[CB18] Czajka, Adam; Bowyer, Kevin: Presentation Attack Detection for Iris Recognition: An Assessment of the State of the Art. ACM Computing Surveys, 51, 03 2018.

[Ch18] Chen, Zhuxin; Zhang, Weibin; Xie, Zhifeng; Xu, Xiangmin; Chen, Dongpeng: Recurrent Neural Networks for Automatic Replay Spoofing Attack Detection. pp. 2052–2056, 04 2018.

[Go14] Goodfellow, Ian J.; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua: Generative Adversarial Networks, 2014.

[He16] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778, 2016.

[HS97] Hochreiter, Sepp; Schmidhuber, Jürgen: Long Short-Term Memory. Neural Computation, 9(8):1735–1780, 11 1997.

[JLO20] Jabbar, Abdul; Li, Xi; Omar, Bourahla: A Survey on Generative Adversarial Networks: Variants, Applications, and Training, 2020.

[JRN11] Jain, A.K.; Ross, A.A.; Nandakumar, K.: Introduction to Biometrics. SpringerLink : Bücher. Springer US, 2011.

[KK21] Konstantinos Karampidis, Minas Rousouliotis, Euangelos Linardos Ergina Kavallieratou: A comprehensive survey of fingerprint presentation attack detection. Journal of Surveillance, Security and Safety, 2021.

[KSH12] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E: ImageNet Classification with Deep Convolutional Neural Networks. In (Pereira, F.; Burges, C.J.; Bottou, L.; Weinberger, K.Q., eds): Advances in Neural Information Processing Systems. volume 25. Curran Associates, Inc., 2012.

[LB95] Lecun, Yann; Bengio, Yoshua: Convolutional networks for images, speech, and time-series. In (Arbib, M.A., ed.): The handbook of brain theory and neural networks. MIT Press, 1995.

[LBH15] LeCun, Yann; Bengio, Y.; Hinton, Geoffrey: Deep Learning. Nature, 521:436–44, 05 2015.

[Li22] Li, Jingjing; Du, Zhekai; Zhu, Lei; Ding, Zhengming; Lu, Ke; Shen, Heng Tao: Divergence-Agnostic Unsupervised Domain Adaptation by Adversarial Attacks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11):8196–8211, 2022.

[Ng20] Nguyen, Dat Tien; Pham, Tuyen Danh; Batchuluun, Ganbayar; Noh, Kyoung Jun; Park, Kang Ryoung: Presentation Attack Face Image Generation Based on a Deep Generative Adversarial Network. Sensors, 20(7), 2020.

[Pe20] Pereira, Luis; Pinto, Allan; Andaló, Fernanda; Ferreira, Alexandre; Lavi, Bahram; Soriano Vargas, Aurea; Cirne, Marcos; Rocha, Anderson: The Rise of Data-Driven Models in Presentation Attack Detection. pp. 289–311, 01 2020.

[Po08] Powers, David: Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness Correlation. Mach. Learn. Technol., 2, 01 2008.

[Ra22] Razzaq, Ali; Ghazali, Rozaida; El abbadi, Nidhal; Dosh, Mohammad: A Comprehensive Survey on Face Detection Techniques. Webology, 19:613–628, 01 2022.

[RHW86] Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J.: Learning representations by back-propagating errors. Nature, 323:533–536, 1986.

[YLL20] Yaojie Liu, Joel Stehouwer, Amin Jourabloo Yousef Atoum; Liu, Xiaoming: Presentation Attack Detection for Face in Mobile Phones. 2020.

[Yu21] Yu, Zitong; Komulainen, Jukka; Li, Xiaobai; Zhao, Guoying: Review of Face Presentation Attack Detection Competitions, 2021.

[Yu22] Yu, Zitong; Qin, Yunxiao; Li, Xiaobai; Zhao, Chenxu; Lei, Zhen; Zhao, Guoying: Deep Learning for Face Anti-Spoofing: A Survey, 2022.