Srinath S , 2025, 13:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

A Novel High-Performance Face Anti-Spoofing Detection Method

Srinath S, Dr. Krithika. D. R
Department Of Computer Application-PG
VISTAS, Chennai.

Abstract: Existing face anti-spoofing models using deep learning for multi-modality data suffer from low generalization when using a variety of presentation attacks such as 2D printing and high-precision 3D face masks. One of the main reasons is that the non-linearity of multi-spectral information used to preserve the intrinsic attributes between a real and a fake face is not well extracted. To address this issue, we propose a multi-ability data-based two-stage cascade framework for face anti-spoofing. The proposed framework has two advantages. Firstly, we design a two-stage cascade architecture to selectively fuse low-level and high-level features from different modalities to improve feature representation. Secondly, we use multi-modality data to construct a distance-free spectral on RGB and infrared (IR) to augment the non-linearity of data. The presented data fusion strategy is different from popular fusion approaches, since it can strengthen the discrimination ability of network models on physical attribute features rather than identity structure features under certain constraints. In addition, a multi-scale patch-based weighted fine-tuning strategy is designed to learn each specific local face region. Experimental results show that.

Keywords:: Face anti-spoofing, Deep learning, Multi-modality data, Presentation attacks.

I. INTRODUCTION

Human face is one of the most salient and stable biometrics, it often relies on various kinds of interactive AI systems and has been widely used in many crowd gathering and sensitive areas [1-3] such as attendance registration, security surveillance, etc. Despite successful applications in many types of face authentication scenarios, most existing face recognition systems are easily spoofed by presentation attacks (PAs) ranging from a 2D printing attack to a vivid 3D mask attack [4, 5]. For example, with the help of silicone or latex masks, users easily portray another identity or obfuscate their identity for entertainment purposes. However, such masks have been treated as criminal tools to deceive automatic face recognition systems. Therefore, it is important to distinguish a real face and a fake face for face recognition and authentication systems. In general, a robust face recognition system can cope with variants of face states, such as face partial occlusion, the change of face expression, etc. On the contrary, variant face presentations should be strictly restricted on face anti-spoofing tasks, and an entire frontal face presentation is required. More importantly, an advanced face anti-spoofing model needs to show strong discriminability on intra-dataset with prior defined face knowledge and performs well on inter dataset with unknown faces.

Most popular face anti-spoofing methods extract generalized liveness features under binary supervision. In practice, a well-discriminative feature map is composed of structure clues, texture clues, depth clues, material clues, etc. There are many structure and texture distinctions between original and recaptured images. Here we present

two obvious distinctions, the first is light reflection. Typically, fake face materials are much flatter and smoother than real faces, and easily cause specular reflection, especially under active infrared light spectral. Secondly, Moire´s Pattern image is formed due to superimposing of the gratings and can be extracted by traditional feature descriptors such as Local Binary Pattern (LBP), Histograms of Oriented Gradients (HOG), Difference of Gaussian (DOG), Speeded Robust Features (SURF), etc. In addition, face anti-spoofing models are also improved by considering the influence of image blur, distortion, noise, etc.

II. LITERATURE SURVEY

Title: CASIA-SURF: A Large-Scale Multi-Modal Benchmark for Face Anti-Spoofing
Author: Shifeng Zhang, Ajian Liu
Year: 2020
Description: The paper introduces CASIA-SURF, a large-scale, multi-modal dataset designed to advance face anti-spoofing research. Unlike existing benchmarks with limited subjects and modalities, CASIA-SURF includes 1,000 subjects and 21,000 videos across 3 modalities (RGB, Depth, and IR). The dataset offers comprehensive evaluation metrics, diverse protocols, and subsets for training, validation, and testing. Additionally, the paper presents a multi-modal multi-scale fusion method that re-weights features to prioritize more informative channels and suppress less useful ones. Extensive experiments on the dataset demonstrate its significance and generalization capabilities for improving face anti-spoofing models.
Title: Cross-ethnicity Face Anti-spoofing Recognition Challenge: A review
Author: Ajian Liu, Xuan Li,
Year: 2020
Description: The paper addresses ethnic bias in face anti-spoofing by introducing the CASIA-SURF cross-ethnicity face anti-spoofing (CeFA) dataset, the largest of its kind. This dataset includes 1,607 subjects, covering three ethnicities and three modalities (RGB, Depth, and Infrared), with both 2D and 3D attack types. It also includes explicit ethnic labels, making it a valuable resource for studying ethnic bias in anti-spoofing systems. To promote research on mitigating this bias, the paper details the Chalearn Face Anti-spoofing Attack Detection Challenge, which featured both single-modal (RGB) and multi-modal (RGB, Depth, Infrared) tracks. The challenge attracted 340 teams, with 11 teams in the single-modal track and 8 teams in the multi-modal track submitting final solutions. The paper presents the challenge's design, evaluation protocol, results, and analysis of top-performing solutions, while also suggesting future research directions to further address ethnic bias in face anti-spoofing.
Aim & Objective:
This work aims to develop a robust face anti-spoofing model that improves the generalization and performance of deep learning-based systems when handling a variety of presentation attacks, such as 2D printing and high-precision 3D face masks.

III. SYSTEM ANALYSIS

Proposed System: The paper proposes a two-stage cascade framework for face anti-spoofing that leverages multi-modality data (RGB and infrared). The framework has two main advantages: it selectively fuses low-level and high-level features from different modalities to improve feature representation, and it constructs a distance-free spectral representation to enhance data non-linearity. This fusion strategy focuses on physical attributes rather than identity features, improving the model's discriminative power. Additionally, a multi-scale patch-based fine-tuning strategy is used to learn specific local face regions. Experimental results show that the proposed framework outperforms state-of-the-art methods, especially in handling multi-material mask spoofing.
Advantage

- An advanced face anti-spoofing model needs to show strong discriminability on intra-datasets with prior defined face knowledge and perform well on inter-datasets with unknown faces.
- Spectral signature between real and fake faces provides additional spectral-spatial information that helps improve face anti-spoofing.

Existing system: Face anti-spoofing is an important task in full-stack face applications including face detection, verification, and recognition. Previous approaches build models on datasets that do not simulate real-world data well. Existing models may rely on auxiliary information, which prevents these anti-spoofing solutions from generalizing well in practice.

Although ethnic bias has been verified to severely affect the performance of face recognition systems, it remains an open research problem in face anti-spoofing.

Disadvantage

- Based on the uniqueness of a person's biological attributes or traits. Face recognition is a challenging field in the framework.
- Biometric systems are vulnerable to attacks made by persons showing photos, videos, or masks to spoof the real identity.

IV. MODULES

i. System Architecture
ii. Depth-Image Pre-processing and Multimodality Fusion
    a) Depth Image Processing (Depth Face Normalization)
    b) Multispectral Image Processing (Reflectance Analysis)
    c) Data Augmentation
iii. Design of Network
iv. Multiscale Patch and Weighted Fine-Tuning

V. ALGORITHM

Deep Neural Network

What are Deep Neural Networks?

Deep learning is a machine learning technique used to build artificial intelligence (AI) systems. It is based on the idea of artificial neural networks (ANN), designed to perform complex analysis of large amounts of data by passing it through multiple layers of neurons.

There is a wide variety of deep neural networks (DNN). Deep Neural Networks (DNN or DDNN) are the type most commonly used to identify patterns in images and video. DDNNs have evolved from traditional artificial neural networks, using a three-dimensional neural pattern inspired by the visual cortex of animals.

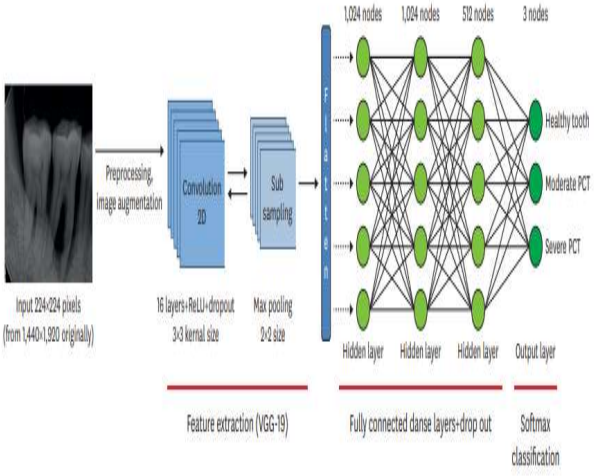
Deep Neural Networks are mainly focused on applications like object detection, image classification, and recommendation systems, and are also sometimes used for natural language processing.

The architecture of the deep DNN algorithm

DNN is a type of machine learning that is used in various fields, especially in image and sound recognition. Deep DNNs imitate the connectivity patterns of neurons in the animal visual cortex. DNNs consist of 1 or more convolutional layers, a pooling layer, and a fully connected layer. Every convolutional layer responds to stimuli only in a restricted region of the visual field known as the receptive field. This structure is distinguished from conventional image classification algorithms and other deep learning algorithms since DNN can learn the type of filter that is hand-crafted in conventional algorithms.

This study used 16 convolutional layers and 3 fully connected dense layers; this network is illustrated in Figure 1. Each convolutional layer was designed with a kernel size of 3×3 pixels, the same padding, and a rectified linear unit activation function.

The maximum pooling layers were designed with strides of 2×2 pixels. After extracting the feature quantities of images using convolutional layers, we used the maximum pooling layers to reduce the

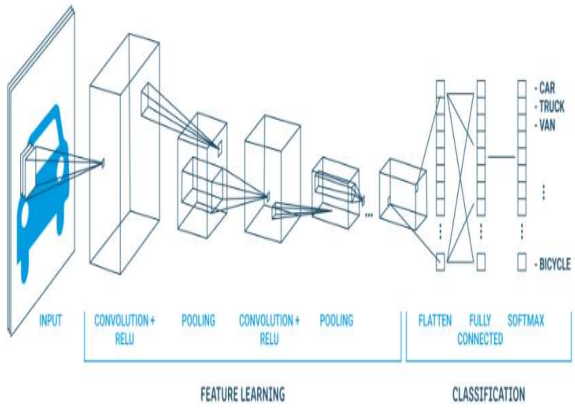


Deep Neural Networks Explained

The strength of DDNNs is in their layering. A DDNN uses a three-dimensional neural network to process the Red, Green, and Blue elements of the image at the same time. This considerably reduces the number of artificial neurons required to process an image, compared to traditional feed-forward neural networks.

Deep Neural Networks receive images as an input and use them to train a classifier. The network employs a special mathematical operation called a "convolution" instead of matrix multiplication.

The architecture of a convolutional network typically consists of four types of layers: convolution, pooling, activation, and fully connected.



Convolutional Layer:

Applies a convolution filter to the image to detect features of the image. Here is how this process works:

- **A convolution**—takes a set of weights and multiplies them with inputs from the neural network.
- **Kernels or filters**—during the multiplication process, a kernel (applied for 2D arrays of weights) or a filter (applied for 3D structures) passes over an image multiple times. To cover the entire image, the filter is applied from right to left and from top to bottom.
- **Dot or scalar product**—a mathematical process performed during the convolution. Each filter multiplies the weights with different input values. The total inputs are summed, providing a unique value for each filter position.

3_0	3_1	2_2	1	0
0_2	0_2	1_0	3	1
3_0	1_1	2_2	2	3
2	0	0	2	2
2	0	0	0	1

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

ReLU Activation Layer:

The convolution maps are passed through a nonlinear activation layer, such as Rectified Linear Unit (ReLu), which replaces negative numbers of the filtered images with zeros.

Pooling Layer

The pooling layers gradually reduce the size of the image, keeping only the most important information. For example, for each group of 4 pixels, the pixel having the maximum value is retained (this is called max pooling), or only the average is retained (average pooling).

Pooling layers help control overfitting by reducing the number of calculations and parameters in the network.

After several iterations of convolution and pooling layers (in some Deep Neural Network architectures this may happen thousands of times), at the end of the network, there is a traditional multi-layer perceptron or "fully connected" neural network.

Fully Connected Layer

In many DNN architectures, there are multiple fully connected layers, with activation and pooling layers in between them. Fully connected layers receive an input vector containing the flattened pixels of the image, which have been filtered, corrected, and reduced by convolution and pooling layers. The softmax function is applied at the end to the outputs of the fully connected layers, giving the probability of a class the image belongs to – for example, is it a car, a boat, or an airplane?

Related content: read our guide to deep learning for computer vision.

What are the Types of Deep Neural Networks?

Below are five Deep Neural Network architectures commonly used to perform object detection and image classification.

R-DNN

Region-based Convolutional Neural Network (R-DNN), is a network capable of accurately extracting objects to be identified in the image. However, it is very slow in the scanning phase and the identification of regions.

The poor performance of this architecture is due to its use of the selective search algorithm, which extracts approximately 2000 regions of the starting image. Afterward, it executes N DNNs on top of each region, whose outputs are fed to a support vector machine (SVM) to classify the region.

Fast R-DNN

Fast R-DNN is a simplified R-DNN architecture, which can also identify regions of interest in an image but runs a lot faster. It improves performance by extracting features before it identifies regions of interest. It uses only one DNN for the entire image, instead of 2000 DNN networks on each superimposed region. Instead of the SVM which is computationally intensive, a softmax function returns the identification probability. The downside is that Fast R-DNN has lower accuracy than R-DNN in terms of recognition of the bounding boxes of objects in the image.

GoogleNet (2014)

GoogleNet, also called Inception v1, is a large-scale DNN architecture that won the ImageNet Challenge in 2014. It achieved an error rate of less than 7%, close to the level of human performance. The architecture consists of a 22-layer deep DNN based on small convolutions, called "inceptions", batch normalization, and other techniques to decrease the number of parameters from tens of millions in previous architectures to four million.

VGG Net (2014)

A Deep Neural Network architecture with 16 convolutional layers. It uses 3x3 convolutions and is trained on 4 GPUs for more than two weeks to achieve its performance. The downside of VGG Net is that, unlike Google Net, it has 138 million parameters, making it difficult to run in the inference stage.

Res Net (2015)

The Residual Neural Network (Res Net) is a DNN with up to 152 layers. Res Net uses "gated units", to skip some convolutional layers. Like Google Net, it uses heavy batch normalization. Res Net uses an innovative design that lets it run many more convolutional layers without increasing complexity. It participated in the ImageNet Challenge 2015, achieving an impressive error rate of 3.57% while beating human-level performance on the trained dataset.

Business Applications of Convolutional Neural Networks

Image Classification

Deep Neural Networks are the state-of-the-art mechanism for classifying images. For example, they are used to:

- **Tag images**—an image tag is a word or combination of words that describes an image and makes it easier to find. Google, Facebook, and Amazon use this technology. Labeling includes identifying objects and even analyzing the sentiment of the image.
- **Visual search**—matching the input image with an available database. Visual search analyzes the image and searches for an existing image with the identified information. For example, Google search uses this technique to find different sizes or colors of the same product.
- **Recommendation engines**—using DNN image recognition to provide product recommendations, for example on websites like Amazon. The engine analyzes user preferences and returns products whose images match previous products they viewed or bought, for example, a red dress or red shoes with red lipstick.

Medical Image Analysis

DNN classification on medical images is more accurate than the human eye and can detect abnormalities in X-ray or MRI images. Such systems can analyze sequences of images (for example, tests taken over a long period of time) and identify subtle differences that human analysts might miss.

This also makes it possible to perform predictive analysis.

Classification models for medical images are trained on large public health databases. The resulting models can be used on patient test results, to identify medical conditions and automatically generate a prognosis.

Optical Character Recognition

Optical character recognition (OCR) is used to identify symbols such as text or numbers in images. Traditionally OCR was performed using statistical or early machine learning techniques, but today many OCR engines use Deep Neural Networks.

OCR powered by DNNs can be used to improve search within rich media content, and identify text in written documents, even those with poor quality or hard-to-recognize handwriting. This is especially important in the banking and insurance industries. Another application of deep learning OCR is for automated signature recognition.

Deep Neural Networks with Run: AI

Run: AI automates resource management and orchestration for machine learning infrastructure. With Run: AI, you can automatically run as many compute-intensive experiments as needed.

Here are some of the capabilities you gain when using Run: AI:

- **Advanced visibility**—create an efficient pipeline of resource sharing by pooling GPU compute resources.
- **No more bottlenecks**—you can set up guaranteed quotas of GPU resources, to avoid bottlenecks and optimize billing.
- **A higher level of control**—Run: AI enables you to dynamically change resource allocation, ensuring each job gets the resources it needs at any given time.

Run: AI simplifies machine learning infrastructure pipelines, helping data scientists accelerate their productivity and the quality of their models.

Learn more about the Run: AI GPU virtualization platform.

DNN Advantage

The main advantage of DNN compared to its predecessors is that it automatically detects the important features without any human supervision. For example, given many pictures of cats and dogs, it learns distinctive features for each class by itself. DNN is also computationally efficient.

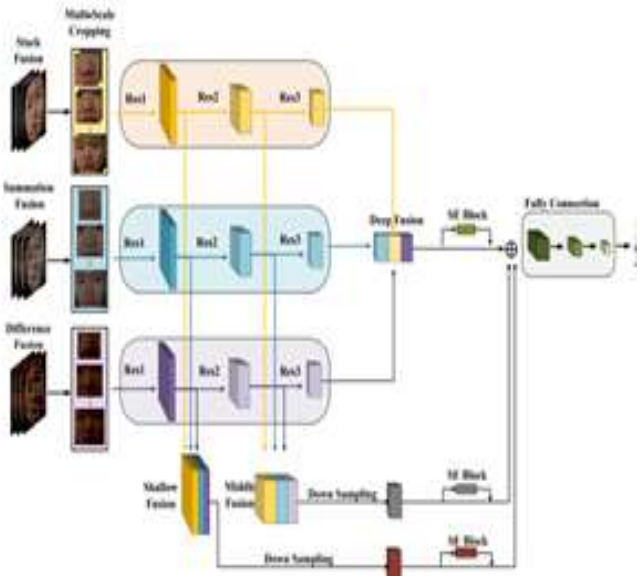
DNN Disadvantage

Some of the disadvantages of DNNs: include the fact that a lot of training data is needed for the DNN to be effective and that they fail to encode the position and orientation of objects. They fail to encode the position and orientation of objects. They have a hard time classifying images with different positions.

VI. SYSTEM DESIGN

I.

Architecture Diagram



VII. CONCLUSION

In this work, we have studied the task of face anti-spoofing for preventing both 2D and 3D face attacks under several identification verification scenarios. For this task, we have developed a two-stage cascade framework to extract both face reflectance features and multi-level face texture features by considering data non-linearity fusion strategy and network skip-connection architecture. Experimental results show that the proposed anti-spoofing framework can prevent a diversity of face-attacking forms such as dim light, realistic face camouflage, static or motion patterns, etc. Furthermore, the proposed model shows strong generalization ability on presentation attacks since

it fuses features from coarse to fine network levels and utilizes the non-linearity of multi-modality information.

VIII. FUTURE SCOPE

For future works, we will establish a more pervasive face spoofing dataset to analyze the generalization ability of the proposed framework. Moreover, the proposed cascade strategy can also be extended to other tasks of biometric modality attack detection, such as the print attack in the iris and palm.

REFERENCES

[1] I. Chingovska, A. R. d. Anjos, and S. Marcel, “Biometrics evaluation under spoofing attacks,” IEEE Trans. Inf. Forensics Secur., vol. 9, no. 12, pp. 2264-2276, Dec. 2014.

[2] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, “Biometrics systems under spoofing attack: An evaluation methodology and lessons learned,” IEEE Signal Process. Mag., vol. 32, no. 5, pp. 20-30, Sept. 2015.

[3] X. Zhu, S. Li, X. Zhang, H. Li, and A. C. Kot, “Detection of spoofing medium contours for face anti-spoofing,” IEEE Trans. Circuits Syst. Video Technol.,2019,doi:10.1109/TCSVT.2019.2949868.

[4] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar, “Detecting silicone mask-based presentation attack via deep dictionary learning,” IEEE Trans. Inf. Forensics Secur., vol. 12, no. 7, pp. 1713-1723, Jul. 2017.

[5] Z. Boulkenafet, J. Komulainen, and A. Hadid, “Face anti-spoofing using speeded-up robust features and fisher vector encoding,” IEEE Signal Process. Lett., vol 24, no. 2, pp. 141-145, Feb. 2017.

[6] J. Galbally, and S. Marcel, "Face anti-spoofing based on general image quality assessment," in Proc. 22nd Int. Conf. Pattern Recognit. (ICPR), Stockholm, 2014, pp. 1173-1178.

[7] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," IEEE Trans. Inf. Forensics Secur., vol. 10, no. 4, pp. 746-761, Apr. 2015.

[8] P. P. K. Cha, W. Liu, D. Chen, D. S. Yeung, F. Zhang, X. Wang, and Chien-Chang Hsu, "Face liveness detection using a flash against 2D spoofing attack," IEEE Trans. Inf. Forensics Secur., vol. 13, no. 2, pp. 521-534, Feb. 2018.

[9] A. Zaliha Abd Aziz, H. Wei, and J. Ferryman, "Face anti-spoofing countermeasure: Efficient 2D materials classification using polarization imaging," in Proc. Int. Workshop Biometr. Forensics (IWBF), Coventry, 2017, pp. 1-6.

[10] J. Yang, Z. Lei, S. Liao, and S. Z. Li, “Face liveness detection with component dependent descriptor,” in Proc. Int. Conf. Biometr. (ICB), Madrid, 2013, pp. 1-6.