# LIVENESS DETECTION VÀ ỨNG DỤNG TRONG BÀI TOÁN NHẬN DIỆN KHUÔN MẶT

## LIVENESS DETECTION AND ITS APPLICATION IN FACE RECOGNITION PROBLEM

**Đỗ Mạnh Quang[1,*], Vũ Việt Thắng[2], Đăng Quỳnh Nga[2]**

DOI: https://doi.org/10.57001/huih5804.2023.169

**TÓM TẮT**

Một trong những phương pháp sinh trắc học được sử dụng rộng rãi nhất là nhận dạng khuôn mặt. Nhận dạng khuôn mặt được sử dụng trong nhiều lĩnh vực. Ứng dụng phổ biết trong lĩnh vực này là xác thực khuôn mặt trên thiết bị di động. Trong khi số lượng người dùng thiết bị di động tăng lên hàng năm, nhu cầu về bảo mật di động cũng ngày càng tăng. Trên thực tế hệ thống vẫn tồn tại lỗ hổng bảo mật đó là khi người dùng không sử dụng khuôn mặt thật mà sử dụng các hình ảnh chụp khuôn mặt hoặc hình ảnh khuôn mặt trong video. Lợi dụng điều này những kẻ gian lận có thể sử dụng mặt nạ của một người được ủy quyền để đánh lừa camera nhận dạng thành người thật. Liveness Detection là một chủ đề nghiên cứu quan trọng giúp phát hiện khuôn mặt giả mạo. Phương pháp được đề xuất trong bài báo này là một mô hình học sâu (Deep learning). Thực nghiệm cho thấy mô hình chúng tôi đề suất có độ chính xác tương đối tốt trên tập dữ liệu thực nghiệm [1].

***Từ khóa:*** *Học sâu, mạng tích chập, Liveness Detection.*

**ABSTRACT**

One of the most widely used biometric methods is facial recognition. Facial recognition is used in many fields. One of the popular applications in this field is facial authentication on mobile devices. While the number of mobile device users increases every year, the demand for mobile security is also increasing. However, in fact, the system still has a security hole that is when users do not use real faces but use face images or face images in videos. Taking advantage of this, fraudsters can use an authorized person's mask to trick the camera into looking real. Liveness Detection is an important research topic for fake face detection. The method proposed in this paper is a deep learning model. Experiments show that our proposed model has relatively good accuracy on the experimental data set [1].

***Keywords:*** *Deep learning, convolutional neural networks (CNN), Liveness Detection.*

[1]R&D Team, Công ty Cổ phần giao hàng tiết kiệm

[2]Khoa Công nghệ thông tin, Trường Đại học Công nghiệp Hà Nội

[*]Email: quangdm2@ghtk.co

Ngày nhận bài: 25/02/2023

Ngày nhận bài sửa sau phản biện: 30/6/2023

Ngày chấp nhận đăng: 15/10/2023

### 1. GIỚI THIỆU

Nhu cầu của người tiêu dùng đối với công nghệ xác thực khuôn mặt và sinh trắc học đang tăng lên, thị trường nhận dạng khuôn mặt trị giá gần 4 tỷ USD vào năm 2020 và dự kiến sẽ chỉ tăng trong thập kỷ tới [2]. Trong thế giới thực, các chuyên gia phòng chống gian lận chống giả mạo bằng cách sử dụng tính năng phát hiện giả mạo. Còn được gọi là "chống giả mạo" hoặc "kiểm tra độ giả mạo", tính năng phát hiện giả mạo sinh trắc học mô tả một loạt các kỹ thuật được người xác thực sử dụng để đảm bảo rằng công nghệ của họ đang đọc nguồn sinh trắc học thực sự.

Hiện nay có ba hình thức kiểm tra giả mạo chính: thụ động, chủ động và kết hợp. Quá trình kiểm tra giả mạo diễn ra mà không cần người dùng nhập dữ liệu, chẳng hạn như hệ thống nhận dạng khuôn mặt của điện thoại sẽ quét khuôn mặt của người dùng cũng như các chuyển động tự nhiên như chớp mắt để xác minh tính xác thực. Những kiểm tra này thường được coi là một giải pháp kiểm tra giả mạo dễ dàng. Thử thách và phản hồi là một ví dụ về kiểm tra giả mạo. Kiểm tra thử thách và phản hồi sẽ yêu cầu người dùng trả lời các lời nhắc như chớp mắt, di chuyển đầu, mỉm cười,... Ý tưởng này đánh bại các biểu diễn sai như ảnh 2D hoặc phát lại video bằng cách yêu cầu người dùng chứng minh rằng họ là người thật.

Trong bài báo này, chúng tôi đề xuất mô hình mạng tích chập để phát hiện mức độ giả mạo của mỗi video selfie/chụp mặt có độ dài từ 1 - 5 giây trên bộ dữ liệu thực nghiệm. Phần 2 của bài báo sẽ giới thiệu các nghiên cứu liên quan, phần 3 sẽ trình bày về mô hình đề xuất, phần 4 sẽ là kết quả của mô hình.

### 2. CÁC NGHIÊN CỨU LIÊN QUAN

Chống giả mạo khuôn mặt là vấn đề gây nhiều khó khăn cho các nhà khoa học khi nghiên cứu và xây dựng các hệ thống nhận diện khuôn mặt. Người dùng thay vì sử dụng khuôn mặt thật của mình để hệ thống nhận diện thì sử dụng ảnh chụp trên điện thoại hoặc video quay trên điện thoại hay thậm chí là ảnh in 2D, 3D của người khác để cho máy xác nhận. Chúng tôi đã nghiên cứu và phân tích một vài ưu nhược điểm của các nghiên cứu trước đó như bảng 1.

Như các phân tích ở trên, các phương pháp hiện có chưa giải quyết được các vấn đề bao gồm cần nhiều dữ liệu huấn luyện, chưa giải quyết được bài toán trong điều kiện môi trường khác nhau, chưa phát hiện được giả mạo trong video.

Do đó trong nghiên cứu này, chúng tôi đề xuất một mô hình mạng tích chập phát hiện giả mạo.

Bảng 1. Ưu và nhược điểm của các mô hình chống giả mạo

| Phương pháp | Ưu điểm | Nhược điểm |
|---|---|---|
| Eye blink check [3] | Dễ triển khai, chi phí thấp và có thể triển khai trên mobile một cách dễ dàng. Người dùng cũng khá thuận tiện vì không cần thực hiện thêm thao tác gì để nhận diện. | Không phát hiện được video giả mạo. Dễ bị vượt qua bằng cách đục 2 lỗ trên tấm ảnh, một ai đó chỉ cần đưa 2 mắt vào 2 lỗ trên tấm ảnh và đặt trước camera hệ thống. Cần số lượng lớn ảnh vì có người chớp mắt nhiều, chớp ít không thể biết trước được. |
| Challenge response [4] | Chi phí triển khai thấp, không cần mua thêm thiết bị. Triển khai đơn giản. Có thể triển khai trên thiết bị di động hiệu quả. Kiểm tra được cả ảnh và video. | Người dùng sẽ có trải nghiệm không tốt do phải thực hiện khá nhiều thao tác trước khi được nhận diện. |
| LPB + SVM [5] | Khá dễ triển khai, không cần thiết bị, có thể làm trên mobile. | Khó áp dụng trong nhiều điều kiện khác nhau về ánh sáng, môi trường, hướng khuôn mặt. |
| CNN [6] | Dễ triển khai, không cần thiết bị, có thể làm trên mobile. Nhận diện tốt hơn LBP vì CNN trích chọn tốt các đặc trưng và bỏ qua được nhiễu. | Cần khá nhiều dữ liệu để huấn luyện. Trong nghiên cứu này còn tồn tại các trường hợp quá khớp với khuôn mặt đã huấn luyện thì phân biệt thật hoặc giả mạo khá ổn nhưng đổi khuôn mặt khác (da đen hơn, nhăn nheo hơn,...) thì ngay lập tức mô hình nhận sai. |

## 3. MÔ HÌNH ĐỀ XUẤT

### 3.1. Dữ liệu thực nghiệm

Để huấn luyện chúng tôi sử dụng ảnh được trích xuất từ dữ liệu ban đầu được cung cấp bởi ban tổ chức cuộc thi Zalo AI [1]. Bộ Zalo AI do ban tổ chức cung cấp bao gồm 1168 video selfie/chụp mặt có độ dài từ 1 - 5 giây đã được gán nhãn. Kết quả chung cuộc sẽ được tính với bộ dữ liệu riêng tư bao gồm 1253 video không có gán nhãn. Ngoài ra khi huấn luyện mô và kiểm tra chúng tôi trích xuất 5 khung hình trên mỗi video. Mỗi video selfie/chụp mặt có độ dài từ 1-5 giây nên chúng tôi sẽ trích xuất 5 khung hình theo các mốc thời gian của video lần lượt là 0,15; 0,25; 0,5; 0,75; 0,95. Mỗi khung hình sẽ có kích thước là chiều cao là 300 và chiều rộng là 100.

$$\text{frame}_i = \text{totalTime} \times \text{paramaterTime}_i \quad (1)$$

Trong đó:

- $\text{Frame}_i$ là khung hình thứ i của video
- TotalTime là tổng thời gian của video
- $\text{ParamaterTime}_i$ là hệ số thời gian lấy khung hình

Chi tiết các bộ dữ liệu được mô tả như bảng 2.

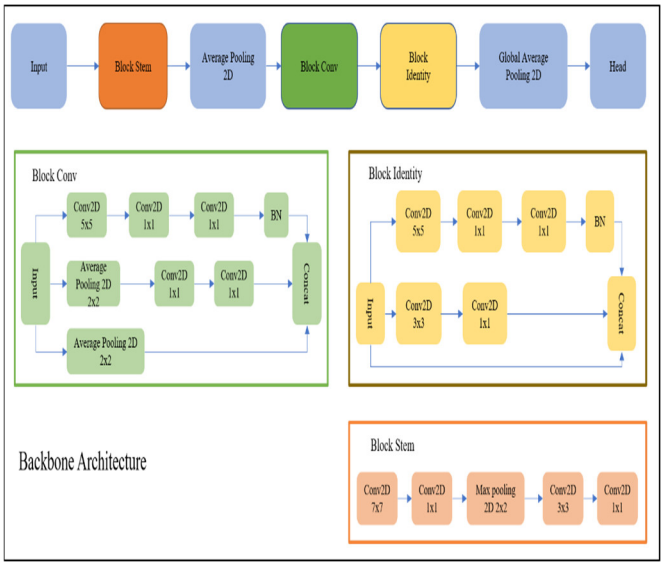
Bảng 2. Mô tả dữ liệu thực nghiệm

| | Dữ liệu huấn luyện | Dữ liệu đánh giá |
|---|---|---|
| Dữ liệu nhãn "real" | 2990 | 6265 |
| Dữ liệu nhãn "fake" | 2850 | |

### 3.2. Mô hình đề xuất

Mô hình đề xuất của chúng tôi bao gồm 3 khối chính:

- Khối Stem: Được sử dụng giảm kích thước dữ liệu.
- Khối Conv: Sử dụng khung 5x5, 2x2 và 1x1 để tăng khả năng học của mô hình.
- Khối Identity: Khối kết nối để bảo đảm mô hình không những không giảm hiệu suất mà có thể làm hiệu suất mô hình tăng lên.




Hình 1. Mô hình đề suất cho bài toán

Bảng 3. Chi tiết kiến trúc mô hình đề xuất với các tham số tối ưu

| Kiến trúc đề xuất | | | | Block Stem | | | |
|---|---|---|---|---|---|---|---|
| Layer | Kernel | Strides | Output shape | Layer | Kernel | Strides | Output shape |
| Input | | | 300x100x3 | Input | | 1x1 | 300x100x3 |
| Block Stem | | | 150x50x128 | Conv2D | 7x7 | 1x1 | 300x100x256 |
| Average Pooling 2D | 2x2 | | 75x25x128 | Conv2D | 1x1 | 1x1 | 300x100x256 |
| Block Conv | | | 38x13x160 | Max Pooling 2D | 2x2 | 1x1 | 150x50x256 |
| Block Identity | | | 38x13x168 | Conv2D | 3x3 | 1x1 | 150x50x128 |
| Global Average Pooling 2D | | | 168 | Conv2D | 1x1 | 1x1 | 150x50x128 |
| Head | | | Classification | Output | | | 150x50x128 |

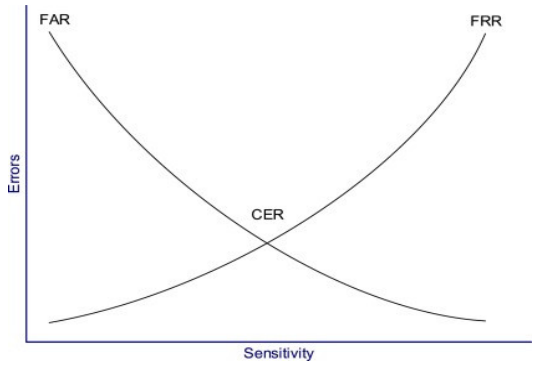
### 3.3. Đánh giá hiệu quả mô hình đề xuất

Để đánh giá hiệu suất mô hình, chúng tôi sử dụng Equal Error Rate (EER hay có một tên gọi khác là CER) (công thức 4). Trong đó, giá trị EER càng nhỏ thì hiệu suất mô hình càng cao. Để tính toán giá trị của EER, chúng tôi sử dụng tham số False Acceptance Rate (FAR) (công thức 2) và False Rejection Rate (FRR) (công thức 3). FAR và FRR thay đổi tùy theo độ nhạy (Sensitivity). Vị trí tại đó FAR và FRR bằng nhau được gọi là Tỷ lệ lỗi EER hoặc tỷ lệ lỗi chéo (CER).

$$\mathrm{FAR} = \frac{\text{False accept}}{\text{Fake ground truth}} \quad (2)$$

$$\mathrm{FRR} = \frac{\text{False reject}}{\text{Real ground truth}} \quad (3)$$

$$\mathrm{EER} = \frac{\mathrm{FPR}+\mathrm{FNR}}{2} \quad (4)$$




Hình 2. Mô tả Equal Error Rate (EER) [1]

Trong đó:

- False accept là tỉ lệ người dùng trái phép có thể truy cập vào hệ thống như những người dùng hợp lệ (tỉ lệ chấp nhận sai sót).
- Fake ground truth là số khuôn mặt giả.
- False reject là tỉ lệ người dùng trái phép bị từ chối truy cập vào hệ thống như những người dùng không hợp lệ (tỉ lệ từ chối sai sót)
- Real ground truth là số khuôn mặt thật.

$$\mathrm{S}(x) = \frac{1}{1+\mathrm{e}^{-x}} \quad (5)$$

Trong quá trình huấn luyện chúng tôi sử dụng hàm tối ưu Adam [7] với tỷ lệ học là 0,0001, hàm mất mát BinaryCrossentropy (công thức 5), cùng với các tham số tăng cường dữ liệu ảnh để giảm việc quá khớp trong quá trình huấn luyện như tăng giảm kích thước, thu phóng ảnh, hướng góc cắt, lật ngẫu nhiên. Các đặc trưng sẽ được đưa vào hàm kích hoạt sigmoid (công thức 4) để đánh giá giả mạo.

$$\mathrm{Loss}_{\mathrm{BinaryCrossentropy}} = -\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{y_i} \times \log(\mathrm{y_{pred}}) + (1-\mathrm{y_i}) \times \log(1-\mathrm{y_{pred}}) \quad (6)$$

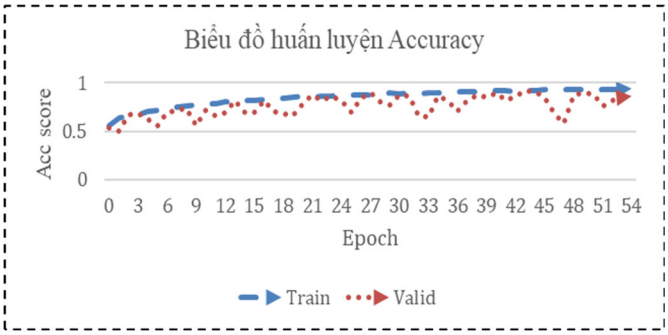
Bảng 4. Mô tả chi tiết khối conv với các tham số tối ưu

| Block Conv | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Layer | Kernel | Strides | Output Shape | Layer | Kernel | Strides | Output Shape | Layer | Kernel | Strides | Output Shape |
| Input | | | 75x25x128 | Input | | | 75x25x128 | Input | | | 75x25x128 |
| Conv2D | 5x5 | 2x2 | 38x13x32 | Average Pooling 2D | 2x2 | 2x2 | 38x13x128 | Average Pooling 2D | 2x2 | 2x2 | 38x13x128 |
| Conv2D | 1x1 | 1x1 | 38x13x32 | Conv2D | 1x1 | 1x1 | 38x13x32 | | | | |
| Conv2D | 1x1 | 1x1 | 38x13x16 | Conv2D | 1x1 | 1x1 | 38x13x16 | | | | |
| BN | | | 38x13x16 | | | | | | | | |
| Concat | | | | | | | 38x13x160 | | | | |

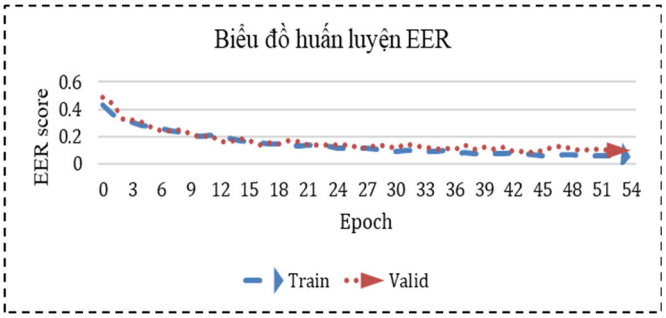
Bảng 5. Mô tả chi tiết khối Identity với các tham số tối ưu

| Block Identity | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Layer | Kernel | Strides | Output Shape | Layer | Kernel | Strides | Output Shape | Layer | Output Shape |
| Input | | | 38x13x160 | Input | | | 38x13x160 | Input | 38x13x160 |
| Conv2D | 5x5 | 1x1 | 38x13x8 | Conv2D | 3x3 | 1x1 | 38x13x8 | | |
| Conv2D | 1x1 | 1x1 | 38x13x8 | Conv2D | 1x1 | 1x1 | 38x13x4 | | |
| Conv2D | 1x1 | 1x1 | 38x13x4 | | | | | | |
| BN | | | | | | | | | |
| Concat | | | | | | | 38x13x168 | | |

### 3.4. Kết quả thực nghiệm




Hình 3. Biểu đồ huấn luyện Accuracy




Hình 4. Biểu đồ huấn luyện EER

Kết quả nhận được sau khi thực nghiệm mô hình là EER = 0,18382. Đây cũng là kết quả nhận được khi chúng tôi tham gia cuộc thi do ban tổ chức Zalo thực hiện. Chúng tôi nhận được vị trí thứ 26/249 của bảng xếp hạng chung kết [8]. Thông qua kết quả trên cho thấy mô hình đề xuất có tính thực tế và hiệu quả cao.

## 4. KẾT LUẬN

Trong nghiên cứu này, chúng tôi đã tự đề xuất được một mô hình nhận diện khuôn mặt có tính hiệu quả cao, tuy nhiên mô hình này cần được cải thiện để nhận diện được tốt hơn. So với các các kết quả khác trong cuộc thi Zalo AI Challenge 2022 thì phần lớn các đội đạt thứ hạng cao hơn đều sử dụng mô hình Pre-train. Do đó chúng tôi sẽ kết hợp mô hình của mình với các mô hình Pre-train hoặc tăng cường thêm dữ liệu huấn luyện nhằm cải thiện hiệu suất mô hình trong thời gian tới. Toàn bộ mã nguồn nghiên cứu có ở Github [9].

### TÀI LIỆU THAM KHẢO

[1]. Zalo AI Challenge 2022 (https://challenge.zalo.ai)

[2]. https://thanhnien.vn/cham-cong-bang-ung-dung-ai-nhan-dien-khuon-mat-xu-huong-moi-nang-tam-trai-nghiem-lam-viec-so-1851531918.htm.

[3]. Soukupova T., Cech J., 2016. *Eye-Blink Detection Using Facial Landmarks*. 21st Computer Vision Winter Workshop, Rimske Toplice, Slovenia.

[4]. W. Xu, J. Tian, Y. Cao, S. Wang, 2020. *Challenge-Response Authentication Using In-Air Handwriting Style Verification*. in IEEE Transactions on Dependable and Secure Computing, vol. 17, no. 1, pp. 51-64, DOI: 10.1109/TDSC.2017.2752164.

[5]. Fernandes S.L., Bala G.J., 2016. *Developing a Novel Technique for Face Liveness Detection*. Procedia Computer Science, 78, 241-247.

[6]. A. A. Mohamed, M. M. Nagah, M. G. Abdelmonem, M. Y. Ahmed, M. El-Sahhar, F. H. Ismail, 2021. *Face Liveness Detection Using a sequential CNN technique*. IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), NV, USA, pp. 1483-1488, doi: 10.1109/CCWC51732.2021.9376030.

[7]. Kingma D. P., Ba J. 2014. *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980.

[8]. https://challenge.zalo.ai/portal/liveness-detection/final-leaderboard

[9]. https://github.com/DoManhQuang/liveness-detection

### AUTHORS INFORMATION

**Do Manh Quang[1], Vu Viet Thang[2], Dang Quynh Nga[2]**

[1]R&D Team, Giao Hang Tiet Kiem Joint Stock Company, Vietnam

[2]Faculty of Information Technology, Hanoi University of Industry, Vietnam