




# FacialCueNet: unmasking deception - an interpretable model for criminal interrogation using facial expressions

Borum Nam<sup>1</sup> · Joo Young Kim<sup>2</sup> · Beomjun Bark<sup>2</sup> · Yeongmyeong Kim<sup>2</sup> · Jiyeon Kim<sup>2</sup> · Soon Won So<sup>2</sup> · Hyung Youn Choi<sup>3,4</sup> · In Young Kim<sup>2</sup> 

Accepted: 13 August 2023 / Published online: 7 September 2023  
© The Author(s) 2023

## Abstract

Polygraphs are used in criminal interrogations to detect deception. However, polygraphs can be difficult to administer under circumstances that prevent the use of biosensors. To address the shortcomings of the biosensors, deception-detection technology without biosensors is needed. We propose a deception-detection method, FacialCueNet, which is a multi-modal network that utilizes both facial images and facial cues based on deep-learning technology. FacialCueNet incorporates facial cues that indicate deception, such as action-unit frequency, symmetry, gaze pattern, and micro-expressions extracted from videos. Additionally, the spatial-temporal attention module, based on convolutional neural network and convolutional long short-term memory, is applied to FacialCueNet to provide interpretable information from interrogations. Because our goal was developing an algorithm applicable to criminal interrogations, we trained and evaluated FacialCueNet using the DDCIT dataset, which was collected using a data acquisition protocol similar to those used in actual investigations. To compare deception-detection performance with state-of-the-art works, a public dataset was also used. As a result, the mean deception-detection F1 score using the DDCIT dataset was 81.22%, with an accuracy of 70.79%, recall of 0.9476, and precision of 0.7107. When evaluating against the public database, our method demonstrated an evaluation accuracy of 88.45% and achieved an AUC of 0.9541, indicating a improvement of 1.25% compared to the previous results. We also present interpretive results of deception detection by analyzing the influence of spatial and temporal factors. These results show that FacialCueNet has the potential to detect deception using only facial videos. By providing interpretation of predictions, our system could be useful tool for criminal interrogation.

**Keywords** Attention network · Deception detection · Deep learning · Investigation · Facial expression

## 1 Introduction

The phenomenon of lying is the subject of various fields of psychological research [1]. Based on psychological evidence of lying, there are various studies on deception in the field of applied technology. Deception-detection involves linguistic, behavioral, and physiological domains [2, 3]. Traditionally, deception-detection techniques measure physiological changes [4]. To detect physiological changes during lying, a polygraph [5] is most commonly used. Because polygraphs used in criminal interrogations measure various

biosignals, they cannot be used if it is difficult to attach a sensor to the body for physical or legal reasons. Attachment of biosensors can make a suspect's psychological state unstable and cause difficulties in deception detection [6]. In addition, as the suspect's body is restrained, consent of a prosecutor is required. Because of this, suspects sometimes refuse to take deception-detection tests [7]. To address these shortcomings, and with the advent of voice recognition, imaging technology, and deep-learning analysis, deception-detection technologies using voice and video are also being developed [8–15]. However, in actual criminal interrogations, most biosignal analysis is conducted using traditional deception detectors [5], as deep learning-based deception-detection technology is difficult to apply to interrogations.

### Lack of interpretation

In deception-detection technologies, the accuracy of the algorithm is of utmost importance. However, little research has been conducted on what kind of changes occur when

Borum Nam and Joo Young Kim are both contributed equally to this work.

✉ In Young Kim  
iykim@hanyang.ac.kr

Extended author information available on the last page of the article

an individual lies in a video recording, and there are cases where such the appearance of changes varies according to the method of data acquisition [7, 16]. This is a major obstacle in the application of deception-detection technology. Although there is evidence that polygraph performance based on physiological signals reaches 98% [17], the data produced is only used as a reference in court and is not recognized as evidence [18]. This is because the accuracy of deception-detection results cannot be guaranteed [19]. In deception detection based on audio and video records, there are several physiological hypotheses involving cognitive dissonance [20, 21]. However, it is necessary to present the changes caused by deception more concretely and quantitatively using observable indicators.

### Nonconformity of the data acquisition protocol

Most deception-detection studies are conducted by collecting data from forensic video footage, fake news, fabricated statement experiments, and games played between participants [22–24]. However, when applied to investigations, the degree of accuracy in the development stage of such algorithms may become meaningless due to variables that differ from those of experimental environments. Concealed information tests (CITs) are used by investigative agencies to detect deception [25]. The theoretical basis of CIT is habituation and orienting response. To use a CIT, the investigator must ask questions by applying the relevant stimulus and mixing it with irrelevant stimuli in the same category. The answers are limited to “yes” or “no.” To prove the reliability of deception-detection research requires protocols that based on real-world test methods.

Among various non-contact modalities for deep learning-based deception detection, we developed a method that can utilize facial expressions in images. Humans express emotions through their faces, and there are several universal facial expressions across various cultures for happiness, sadness, anger, fear, and disgust [26]. Even when lying, facial expressions change according to emotional states [24, 27]. Changes in facial expression at this time are related primarily to anxiety, and facial clues about lies are largely classified into two types based on studies of the effect of anxiety on facial expressions [28–30]. Even in those who try to hide their expressions, brief micro-expressions cannot be hidden because emotional expressions are generated not by the motor cortex but subcortical impulses [31]. Deep-learning technology can recognize patterns that are difficult for humans to recognize and make decisions based on them.

Because our main goal is to develop a deception-detection algorithm that can be used in criminal interrogations by utilizing facial cues and deep-learning, we proposed a video-based deep-learning network for deception detection that can effectively use instantaneous changes in facial expression and provide spatial and temporal interpretation of prediction results. In addition, we collected a database using a protocol

based on investigative methods to train and evaluate proposed deception-detection model.

## 2 Related work

Deception-detection studies using facial expressions focus primarily on facial cues from data collected in the laboratory and an automated deception-detection algorithm based on data collected in the laboratory or within a public dataset.

### 2.1 Facial cues and dataset

When lying, lip pressing caused by emotional anxiety [30], frequent swallowing [32], unnatural duration of an expression, slips of the tongue [33], asymmetry in the face [34], fewer facial movements [35], a higher rate of blinking during anxiety [7], a lower rate of blinking during an experience of cognitive complexity [16], eye fixation accompanied by a reduced blinking rate [36], micro-expressions, and various other responses are often present. To quantify this, facial motion can be expressed as action units (AUs) using a facial action coding system [37]. As revealed by previous studies, several AUs can distinguish between deception and truthful responses [38–42]. For example, AU15 (lip corner depressed) is rarer during expressions of truth than deception; AU17 (chin raised) is rarer in truth than deception; AU20 (lip stretched) is less common when lying; AU25 (lips parted) is rarer in truth than deception; and AU45 (blinking) is less frequent when lying. Similarly, symmetry of the face can be an important deception cue when using the automated face-landmark detection method [42]. In addition, previous studies have confirmed that micro-expressions, in which real emotions are visible for brief periods (less than 0.5 s) due to involuntary emotional responses, also appear more frequently during lying [24, 43]. Other studies indicate that gaze is an important feature in deception detection [8, 44].

Methods adopted by research groups to induce lies in the laboratory to identify various lie cues on the face or to detect lies using facial expressions include role-playing and mock crimes, which requires a person to assume a specific role before the experiment [23]; a memorizing a script for a specific question [24]; and a method of generating a sudden situation and making up a story [22]. In addition, deception data obtained from television programs and court statements have been analyzed to identify cues that appear on the face when lying [42, 45]. However, because these deception-inducement experiments differ from deception-detection techniques used in the interrogation of actual criminals, it is difficult to apply them to criminal interrogations. In the case of data obtained from the internet, such as television programs and court statements, deception and truth can be mixed in a single statement, resulting in

unclear labels, and the quality of the video may be too poor to analyze facial expressions (e.g., if the face is covered by subtitles, the recording resolution is too low, or an individual is not facing the camera). In these cases, it is difficult to use such data to train a neural network that can be applied to criminal interrogation. In this study, an algorithm using various facial cues was developed by collecting experimental data based on a real polygraph questioning technique. The ultimate goal was an automated, non-contact deception-detection algorithm that can be applied to the interrogation of real criminals.

## 2.2 Deep learning-based techniques

Various deep-learning structures have been used to create automated deception-detection systems. For example, because the responses are measured as time series data, a classifier structure using a long short-term memory (LSTM) [46] recurrent neural network (RNN) can be applied to analysis of gaze features [47]. One study [23] utilized a dynamic graph-embedding model based on face-to-face interactions to detect deception. To develop an algorithm that can be used for criminal interrogations, which is the target of this study, it is necessary to assess facial expressions in response to specific questions rather than interactions among individuals.

In many deception-detection studies, video-based algorithms developed as classifiers based on a convolutional neural network (CNN) show remarkable performance [10, 11]. Because deception-detection algorithms have the task of “classifying” videos with human facial expressions [48, 49], several attempts have been made to develop classifiers using a CNN and RNN in combination for more efficient video classification [9, 22]. However, these models are more suitable for classifying videos that show distinct characteristics by label (e.g., distinguishing between volleyball and swimming). Also, the number of public datasets for deception detection is not large, and performance cannot be guaranteed when models suitable for large-scale datasets are applied to deception detection. To tackle these problems, we developed a deception-detection model by designing a network structure that can utilize facial cues revealed in many studies, as well as a video-classification model. One study added interpretability to a classification model using attention beyond simply distinguishing deceptions from truth with a deep-learning network to reveal deception cues in a case study [9]. For a deception-detection algorithm to be used at a crime interrogation scene and provide helpful information to investigators, interpretation of the prediction results would be required. In this study, we developed a deception-detection model based on the latest video-recognition model and embedded a spatial-temporal attention module for classification interpretation [50].

## 3 Method

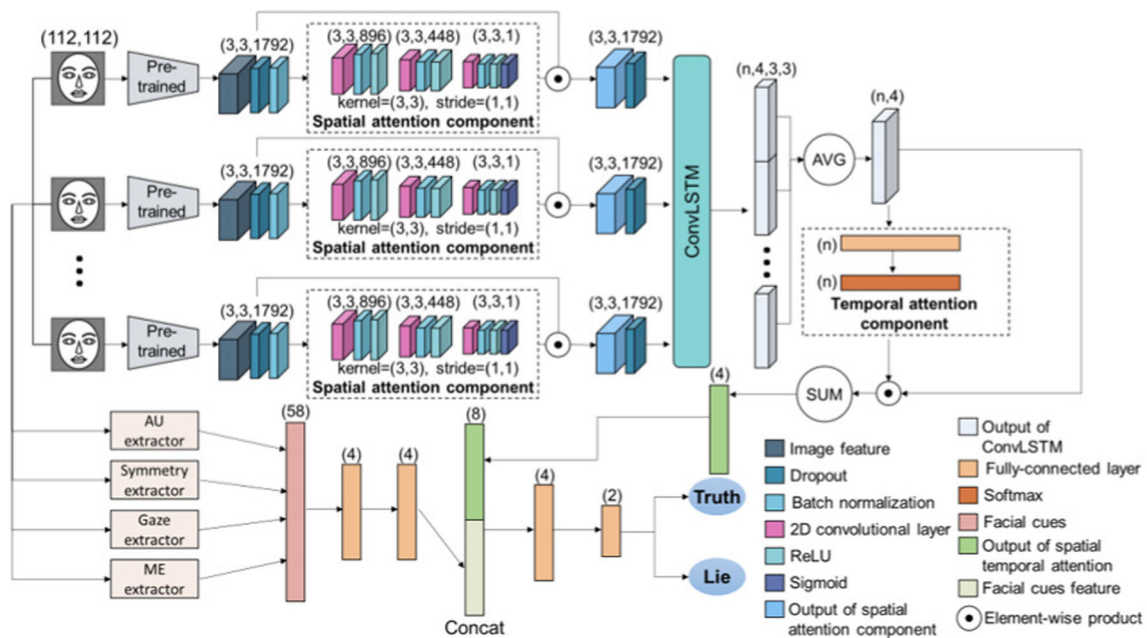
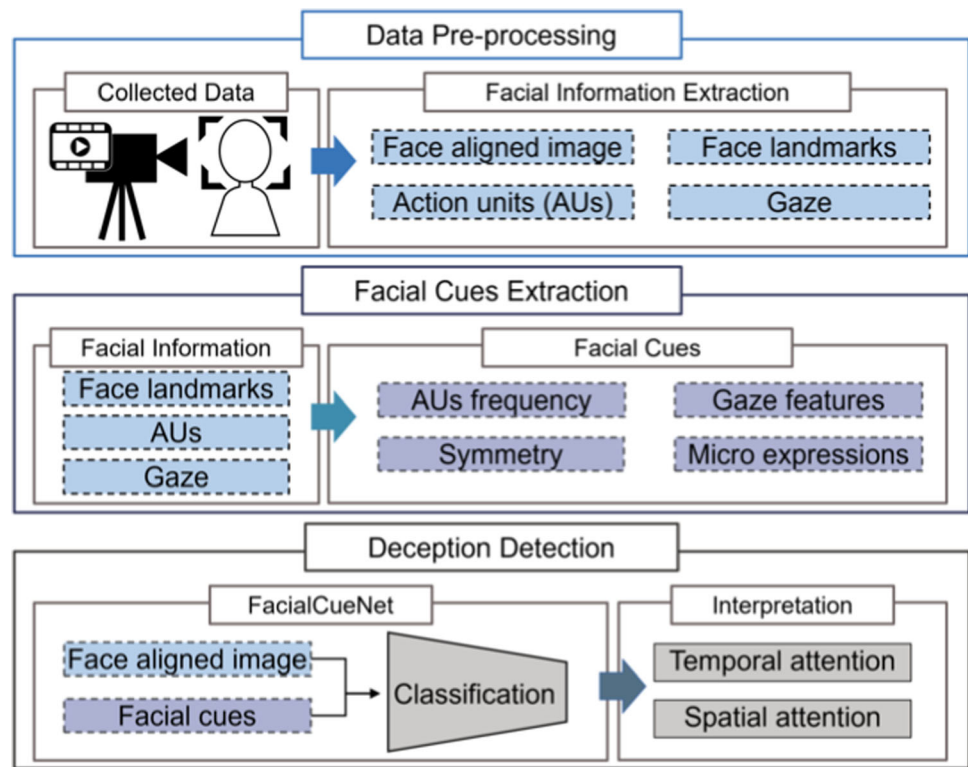
To develop an automated deception-detection model using facial deception cues that can be applied to real-world criminal interrogations, an effective model structure should be used for video classification, various facial cues should be utilized, and an investigator should provide model interpretability, which is a description of the prediction. In this study, we tried to achieve this goal by developing FacialCueNet. As shown in Fig. 1, the first step was to collect video data based on criminal interrogation and then pre-process the data. Second, important cues appearing on the face during lying were extracted, and these cues and preprocessed face images were used to train FacialCueNet. Finally, the results were interpreted after classification.

### 3.1 Spatial-temporal attention network

A video-recognition model was applied to the deception-detection model using facial expressions. A video-action-recognition model [50] including a spatio-temporal attention mechanism that shows excellent performance among video-recognition models and has spatial and temporal interpretability, was used as the basic structure of FacialCueNet. As can be seen in Fig. 2, because videos contain time series image data, we employed a convolutional LSTM (ConvLSTM) [51] to use time series data as input and inserted a temporal attention component using the output of the convLSTM. Several CNN layers were used in the spatial attention component using frames obtained from the video as input. The spatial attention component was designed for the CNN to learn the importance mask  $M_i$ , which was calculated to be the spatially significant representation of the image feature  $X_i$  of the  $i$ -th frame from the video, and the output  $\tilde{X}_i$  of the spatial attention component was an element-wise multiplication product, where  $\tilde{X}_i = X_i \odot M_i$ . The values of  $i$  were from 1 to  $n$ , where  $n$  is the number of frames. The spatial attention module had three 2D convolutional layers with the number of channels [896, 448, 1] and a kernel size of 3, stride 1. The importance mask  $M_i$  spanned 0 to 1 and attenuated certain regions of the feature map based on the model’s estimated importance. The output of the ConvLSTM in the spatial-temporal attention network was  $\bar{H} = \frac{1}{n} \sum_{i=1}^n H_i$  for a time length  $n$ , which was expressed by calculating the average of the hidden states in the ConvLSTM. Finally, the temporal attention component learned the importance weight of each frame from a video. The importance weight at each time step  $t$  obtained from the temporal attention mechanism can be defined as follows, where  $1 \leq i \leq n$ :

$$w_i = \frac{\exp(\Phi(H_i))}{\sum_{i=1}^n (\exp(\Phi(H_i)))} \quad (1)$$

**Fig. 1** Deception detection process using FacialCueNet



**Fig. 2** FacialCueNet framework. The input images are face aligned (112,112) images. ‘Pretrained’ represents FaceNet, which is Inception-ResNetV1 trained with the VGG-face2 dataset. The size of each layer’s output is written on the top of each layer. The ConvLSTM has 1 layer with kernel size (3,3) and 4 hidden units. “n” is the time length of the

input video. “AVG” and “SUM” represent the average of the output of the ConvLSTM, which was performed over multiple axes (the first and second) and the summation, which was performed along the second axis, respectively



where  $\Phi$  is a feed-forward neural network and  $H_i$  is a ConvLSTM hidden state at time  $i$ . Two fully connected layers were used for temporal attention module with a dimensionality of  $n$ . The final output of the spatial-temporal attention network was  $\tilde{H} = w_i \times H_i$ , which is used as the input of the fully connected layer, the final classification layer. Regarding the detailed structure of the spatial attention component, the optimal model structure in the previous study [50] was used. The loss function  $L$  for this model was defined as follows to learn reasonable spatio-temporal importance and to increase classification accuracy:

$$L = L_{CE} + \lambda_{TV} L_{TV} + \lambda_{contrast} L_{contrast} + \lambda_{unimodal} L_{unimodal} \quad (2)$$

$L_{CE}$  is the cross-entropy loss for classification.  $L_{TV}$  is the total variation regularization [52] for spatial smoothness of the importance mask.  $L_{contrast}$  is the contrast regularization of the learnable attention mask, and  $L_{unimodal}$  is the unimodality regularizer [50] that supports the unimodality of temporal attention, biasing against unimportant temporal weights. In this study,  $L_{unimodal}$  was modified for our temporal attention component as follows:

$$L_{unimodal} = \frac{\sum_{i=2}^{n-1} \sqrt{\max\{0, w_{i-1} w_{i+1} - w_i^2\}}}{n-2} \quad (3)$$

where  $\lambda_{TV}$ ,  $\lambda_{contrast}$ , and  $\lambda_{unimodal}$  are the weights for the corresponding regularizers. The algorithm of spatial temporal attention network is presented below.

---

**Algorithm 1** Spatial temporal attention network.

---

**Input:** Image features in a video  $[X_1 \dots X_n]$

- 1:  $i \leftarrow 1$
- 2: **while**  $i \neq n$  **do**
- 3:  $\tilde{X}_i \leftarrow X_i \odot M_i$  // masked image features (spatial attention)
- 4:  $\tilde{H} \leftarrow \text{ConvLSTM}(\tilde{X}_i)$
- 5:  $i \leftarrow i + 1$
- 6: **end while**
- 7:  $\tilde{H} = \frac{1}{n} \sum_{i=1}^n H_i$
- 8:  $\Phi$ : feed-forward neural network
- 9:  $j \leftarrow 1$
- 10: **while**  $j \neq n$  **do**
- 11:  $w_i \leftarrow \frac{\exp(\Phi(H_i))}{\sum_{i=1}^n \exp(\Phi(H_i))}$
- 12:  $j \leftarrow j + 1$
- 13: **end while**

**Output:**  $Y = \tilde{H} \odot w_i$  // weighted output (temporal attention)

---

### 3.2 Facial cue extraction

To improve the classification ability of the video-based deception-detection model, we used deception cues presented in previous studies as hints for the classification

model. Based on the various psychological grounds presented in the introduction section, AUs [38–42], symmetry on the left and right sides of the face [42], the presence or absence of micro-expressions [24, 43], and gaze features [8, 44] show a significant difference between deception and truth. Using OpenFace [53], which is used in various facial expression analysis, to extract the modality of each facial cue, AUs and face landmarks corresponding to the images were obtained. The face recognition accuracy of OpenFace using the LFW dataset was 0.9292 [53].

#### 3.2.1 Action unit frequency

First, in constructing facial cues, AUs, which have been identified as significant factors when distinguishing lies from truth, [38–42] were extracted. The AUs were extracted using OpenFace, and pre-processing was performed by calculating whether they occurred as binary values. Because the number of specific AU occurrences in both truth and deception responses in previous studies differed, the frequency of frames in which AU15 (lip corner depressor), AU17 (chin raiser), AU20 (lip stretched), AU25 (lips part), and AU45 (blink) appeared were calculated as follows:

$$freqAU = \frac{|AU_{frames}|}{n} \quad (4)$$

where  $n$  represents the number of total frames in an input video and  $|AU_{frames}|$  represents the number of frames in which a specific action unit appears. When the total number of videos in the dataset is  $K$ , set  $AUs_k = \{freqAU15_k, freqAU17_k, freqAU20_k, freqAU25_k, freqAU45_k\}$  is extracted from each input video  $V_k$  for  $1 \leq k \leq K$ .

#### 3.2.2 Facial symmetry

The symmetry of facial movement was extracted using the detection method of the changes in the Euclidean distances of face landmarks presented in a previous study [54] that use the left and right symmetry of the face as one of the facial cues. As shown in Fig. 3, Euclidean distances were obtained for the face landmarks of the left/right eyebrow (index 20/25) and the face landmarks of the left/right eye (index 40/43), and the correlation of the distances was calculated. Where  $1 \leq i \leq n$ , each left distance (ld) and right distance (rd) were calculated in the  $i$ -th frame, in the form of two time-series signals. Cross-correlation of the mean-removed sequences [55, 56] was used to calculate the correlation between these two signals as follows:

$$\phi_{LR}(\mu) = E\{(L - \mu_L)(R - \mu_R)^*\} \quad (5)$$

where  $\phi_{LR}$  represents the cross correlation of the ld signal (L) and rd signal (R),  $E$  is the expected value operator,  $\mu_L$  and  $\mu_R$  are the means of each signal, and  $(R - \mu_R)^*$  represents the complex conjugate of  $(R - \mu_R)$ . A symmetry value  $Sym_k = \{Corr_k\}$  was extracted from each input video  $V_k$ , where  $Corr_k$  represents the cross-correlation of mean-removed L and R.

### 3.2.3 Gaze pattern

Gaze features were extracted to utilize the gaze pattern as a facial cue to be used as an input feature. For the  $i$ -th frame, the gaze pattern has x,y, and z values for each left and right eye, and it is expressed as six signals for  $1 \leq i \leq n$ . Gaze features were extracted using a method used in a previous study [47] from the gaze pattern expressed by 6 signals. The extracted gaze features are the mean, standard deviation, skewness, kurtosis, minimum value, and maximum value of each gaze signal on the left and right. Therefore, the following set consisting of a total of 36 gaze features was used as one of the facial cues:

$$gaze_k = \{L_xmean_k, L_xstd_k, L_xskew_k, L_xkur_k, L_xmin_k, L_xmax_k, \\ L_ymean_k, L_ystd_k, L_yskew_k, L_ykur_k, L_ymin_k, L_ymax_k, \\ L_zmean_k, L_zstd_k, L_zskew_k, L_zkur_k, L_zmin_k, L_zmax_k, \\ R_xmean_k, R_xstd_k, R_xskew_k, R_xkur_k, R_xmin_k, R_xmax_k, \\ R_ymean_k, R_ystd_k, R_yskew_k, R_ykur_k, R_ymin_k, R_ymax_k, \\ R_zmean_k, R_zstd_k, R_zskew_k, R_zkur_k, R_zmin_k, R_zmax_k\} \quad (6)$$

where L and R represent the left eye and right eye, std means standard deviation, skew means skewness, kur means kurtosis, and min and max mean minimum value and maximum value, respectively.

### 3.2.4 Micro-expression

The last facial cue was a micro-expression extracted from the video. Micro-expressions were detected using the action unit obtained from OpenFace. Among the 18 types of action units that can be obtained with OpenFace, the occurrence of micro-expressions was counted using the duration of 17 types of action units excluding AU45 (blinking). Because the latency of a micro-expression is up to 0.5 s and considering the frames per second (fps) of video  $V_k$ , a micro-expression was counted when the expression time of each action unit was less than  $0.5 \times \text{fps}$ . When the AU expression latency was only a single frame, it was considered an error in AU detection and excluded from the count. Therefore, the number of

micro-expression occurrences for each AU extracted from the input video  $V_k$  was obtained as follows:

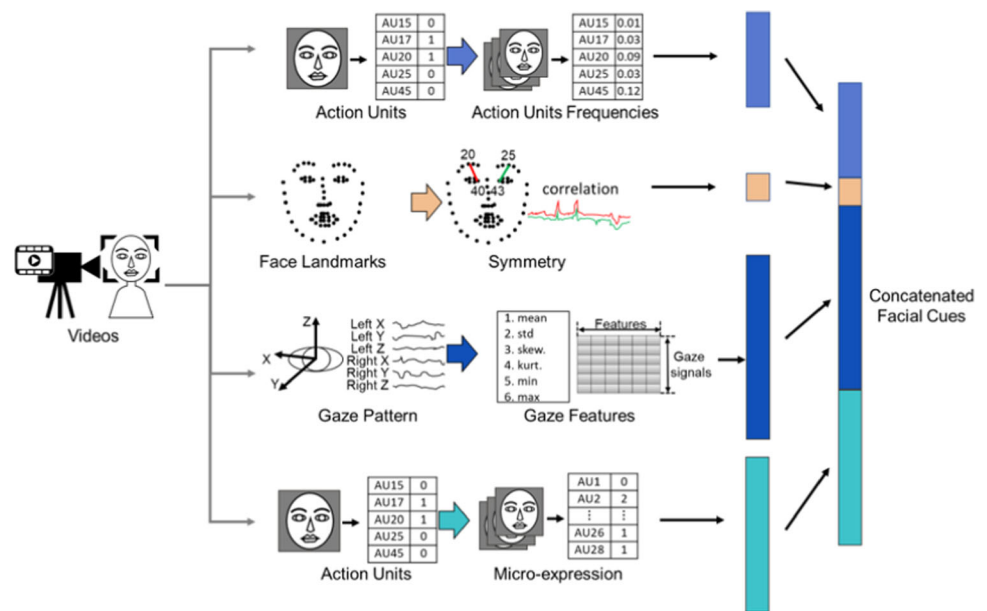
$$ME_k = \{\text{count AU1}_k, \text{count AU2}_k, \text{count AU4}_k, \text{count AU5}_k, \\ \text{count AU6}_k, \text{count AU7}_k, \text{count AU9}_k, \text{count AU10}_k, \\ \text{count AU12}_k, \text{count AU14}_k, \text{count AU15}_k, \text{count AU17}_k, \\ \text{count AU20}_k, \text{count AU23}_k, \text{count AU25}_k, \text{count AU26}_k, \\ \text{count AU28}_k\} \quad (7)$$

## 3.3 FacialCueNet

The FacialCueNet architecture was developed to use video input and simultaneously provide psychological cues to the deception detection model. The overall structure of FacialCueNet was optimized empirically. FacialCueNet has a multi-modal network structure using a spatial-temporal attention network described in the Spatial-temporal Attention Network section and facial cues described in the Facial Cue Extraction section were used for input. (Figure 2) Image feature  $X_i$  of the  $i$ -th frame was extracted using an appropriate pre-trained model that learned face information from  $n$  frames of video  $V_k$ , and used as an input to the spatial-temporal attention network. The pre-trained model used for image-feature extraction was FaceNet, which is Inception-ResNetV1 [57] trained on a VGG-face2 dataset [58]. In a previous study that used facial information extracted from the pre-trained InceptionResNetV1 as an input feature [59], the Inception 4e block was adopted from the entire model structure. The input feature shape was (3,3,1792). To prevent FacialCueNet from overfitting, we added dropout(p=0.5) and batch normalization layer before the spatial attention component. As a result, an output  $f(V_k)$  was calculated for the input video  $V_k$  using a pre-trained network with spatial-temporal attention network  $f$ . At the same time, we added two fully connected randomly initialized layers for  $g$ , with input dimensionalities of 58 (the first fully connected layer in Fig. 2), representing the size of concatenated facial cues) and 4 (the second fully connected layer in Fig. 2).  $g$  used  $facialcues_k$  as input, where  $facialcues_k = AU_s_k \cup Sym_k \cup gaze_k \cup ME_k$  represents concatenated facial cues combining  $AU_s_k$ ,  $Sym_k$ ,  $gaze_k$ , and  $ME_k$  extracted from  $n$  video frames (see the Facial Cue Extraction section). The input to the final classification layer  $H$  was the feature concatenating the output of the spatial-temporal attention network and the output of the fully connected layer using facial cues in Fig. 3. The structure  $H$  included randomly initialized fully connected layers with output dimensionalities of 4 and 2. FacialCueNet  $F$  can therefore be expressed as:

$$F(V_k) = H(f(V_k) \oplus g(V_k)) \quad (8)$$

**Fig. 3** Facial cue extraction framework



## 4 Experiments

The general deception-detection performance of FacialCueNet was validated using a public database, the process of collecting a dataset suitable for FacialCueNet to be applied to criminal interrogation was presented, and the performance of FacialCueNet on the collected dataset was validated. In addition, the usefulness of FacialCueNet was validated by checking model interpretability to provide useful information to investigators in actual criminal interrogations.

### 4.1 Dataset

Two datasets were used to validate the performance of FacialCueNet. First, to evaluate the general-purpose deception detection performance of FacialCueNet, we used “Real-life Trial Dataset” [45], which is the most widely used dataset in previous studies of deception detection. Second, we used the “Deception detection using the concealed information test” (DDCIT) dataset, which was collected using a deception detection test technique used by professional investigators in an environment similar to criminal interrogations.

#### 4.1.1 Real-life trial dataset

The general performance of FacialCueNet was evaluated using the Real-life Trial Dataset [45] used in previous studies on deception detection. This study was reviewed and approved by the institutional review board of Hanyang University (HYU-2019-01-006-4), and the requirement for informed consent was waived. The database consisted of 121

video clips from the “The Innocence Project” website, court-room trial videos, police interrogations, and statements for deception detection. In the video, the testimony is in the form of freely given responses to questions, and the label of the video was determined according to the verdict. The videos in the Real-life Trial Dataset were obtained from the internet. In cases in which a person was not seen in the video, a subtitle passed over the face, or the scene was changed, the videos were pruned from the dataset. Therefore, as in a previous study [60], 104 videos were selected, including 50 truthful videos and 54 deceptive videos, out of the 121 videos for FacialCueNet.

#### 4.1.2 DDCIT(deception detection using the concealed information test) dataset

We collected a deception-detection dataset using the CIT, an actual polygraphic technique used for interrogation, with advice provided by a professional deception-detection investigation team to develop a deception-detection model that can be applied to criminal interrogations. The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Hanyang University (HYU-2019-01-006-4). The experiment was conducted by recruiting healthy adult males and females. Before taking the deception test, the subject picked one of three types of gift cards (for a cinema, coffee shop, or drug store) placed on a table in an empty laboratory and then hid the gift card anywhere in the laboratory, including in their clothes. In addition, several steps were taken to satisfy the structured circumstances and proper pretesting conditions with the help of an investigator who specializes in deception

detection [61–64]. First, a detailed explanation of the purpose and procedure of the experiment was provided to the subject. Next, the subject filled out a questionnaire containing their name, date of birth, lie-detection-test experience, medical history, history of drug use, current medical history, physiological status, and medication history before the test. Finally, we interviewed the subjects about the prepared questionnaire to build rapport, which fosters trust in the subjects [6, 65, 66]. In the experiment, the galvanic skin response (GSR) signal from a finger was obtained along with the video using existing polygraph methods. Deception detection, referred to as stimulation testing, was performed, in which the subject was asked to draw a card labeled with the numbers 3, 4, or 5, and all answers were “no” (e.g., Q: Is the selected card number 3? A: No) By observing the GSR signal, the subject was informed of the accuracy of conventional deception detection so that a clearer deception response could appear before starting this experiment. After this process, the experimenter showed the image of gift card and asked the subjects questions such as “Is the gift card that you hidden?” for each type of gift card. Approximately 1–2 seconds after the question was asked by the experimenter, the subject gave an answer. At this time, the subject had to answer “no” to the three questions, and the next question was asked 10 seconds after the answer. Accordingly, two truth samples and one deception sample were obtained. The question order was shuffled, and a total of four truth samples and two deception samples were obtained from one subject during a total of two sessions. If the experimenter could not guess the hidden card after seeing the subject’s reaction, the subject’s compensation was doubled to give the subject an incentive to continue participating usefully in the experiment. The experimenter checked the hidden card and noted whether the subject’s answer on the video was truthful or deceptive. From 105 subjects, 315 samples were obtained for each session, including 210 truths and 105 deceptions. A total of 630 samples were obtained from the experiment. A summary of the DDCIT dataset is shown in Table 1.

## 4.2 Model validation

To create a deception-detection model that can be used for criminal interrogations, it was necessary to validate the objective performance of FacialCueNet for deception detection using facial video as a preliminary step. The generality of the model was checked using the Real-life Trial Dataset for objective performance validation of the developed FacialCueNet.

### 4.2.1 Data pre-processing

Facial cues that change according to time series data, such as AU frequency, facial symmetry, gaze features, and micro-expressions, could not be extracted using frames from raw video because the videos in the Real-life Trial Dataset do not all have the frame rate. Therefore, frames were sampled so that all videos have the same frame rate. Because micro-expressions last for less than 0.5 s [67], facial cues were extracted from frames sampled for 15 fps, which is sufficient to cover them. Face-aligned images were extracted from frames sampled at 1 fps as in the pre-research [9] for deception detection using images because general facial expression changes are unconscious biopsychosocial reactions caused by emotions and typically last for less than 4 s [68]. To compensate for the unequal number of frames in the videos, we adjusted the length of the other videos to the longest video, which was 79 s. If the length of the sampled video was shorter than 79 s, the video was zero-padded using a blank image.

### 4.2.2 Model settings

FacialCueNet was trained using the Real-life Trial Dataset, and its performance was evaluated using a 10-fold cross-validation method, following the approach of previous studies on deception detection, with videos from the same Real-life Trial Dataset. Regarding the hyperparameters, a

**Table 1** Summary of the DDCIT dataset

Total number of subjects	105		
Total number of video clips	630		
Number of truth and deception videos	Truth	Deception	
	420	210	
Frames per video clip	150		
Language	Korean		
Race	Asian		
Gender	Male	Female	
	51	54	
Ages	20s	30s	40s
	96	8	1



batch size of 12, an initial learning rate of 0.0005 with a decrease rate of 0.99 per epoch, and a dropout rate of 0.5 were used. The convolutional LSTM in FacialCueNet had one layer, the kernel size was (3,3), and there were two hidden units in the LSTM. To increase the learning efficiency, we employed L2 regularization [69], and the lambda for L2 regularization was 0.00001. The values for  $\lambda_{TV}$ ,  $\lambda_{contrast}$ , and  $\lambda_{unimodal}$  were 0.000001, 0.000001, and 1, respectively. The dropout ratio was 0.5.

### 4.3 Deception detection for criminal interrogation

FacialCueNet was optimized in a direction that can be applied to actual criminal interrogation using the DDCIT dataset. As an input of FacialCueNet, the DDCIT dataset was pre-processed. The overall model structure used was the same as the structure used for the Real-life Trial Dataset in the Model Validation section, but the optimal learning parameters for the DDCIT dataset were determined empirically.

#### 4.3.1 Data pre-processing

Because the experimenter annotated the question start and end times during the experiment, all videos were divided based on this annotation time. The parts of the videos after the subjects answered were used because the subjects' answers to the question were all 'no,' and there was the same change in facial movement in all subjects during the answer. In previous studies, facial expression after lying was important for deception detection. The subjects' answers, which appeared about 1 s after the experimenter's questions, were detected by voice activity detection [70]. For efficient data-handling of the prediction model, we cropped videos 5 s after the end of speech. Because the frame rate of the collected videos was 30 fps, an input sample with 150 time steps was generated.

#### 4.3.2 Model settings

The FacialCueNet architecture for the DDCIT dataset was constructed based on the architecture for the Real-life Trial Dataset. A polygraph session, a deception-detection method used in criminal interrogation, includes a pre-test in which a question with a known answer is asked before the main test. For practical use, it was necessary that the pre-test data obtained first were used as the training set, and the main test data obtained later were used as the test set for the deception-detection model. As FacialCueNet is being applied to criminal interrogation, the training and testing protocols were imitated. The 315 samples obtained from the first session of the DDCIT dataset were used for the training set, and the 315 samples obtained from the second session were used for the test set. Seven-fold cross-validation was used to determine the parameters used for training using the

train set. The hyperparameters from the seven-fold cross-validation were a batch size of 9, an initial learning rate of 0.004 with a decrease rate of 0.99 per epoch, and a dropout rate of 0.4. The convolutional LSTM in FacialCueNet had 1 layer, a (3,3) kernel size, and four hidden units. To increase the efficiency of training, we used L2 regularization [69], and the lambda for L2 regularization was 0.00001. The values for  $\lambda_{TV}$ ,  $\lambda_{contrast}$ , and  $\lambda_{unimodal}$  were  $10^{-11}$ ,  $10^{-12}$ , and 1, respectively. The dropout ratio was 0.4. These hyperparameters were used for training the whole Session 1 samples and testing FacialCueNet performance using the whole Session 2 samples. Because the DDCIT dataset is a class-imbalanced dataset with a truth:deception ratio of 2:1, The  $L_{CE}$ , cross-entropy loss in (3), was replaced with the weighted cross-entropy loss:

$$-\frac{1}{M} \sum_{m=1}^M \omega r_m \log(p_m) \quad (9)$$

where  $M$  is the number of samples in the training set,  $r_m$  is the target label for sample  $m$ ,  $p_m$  is the  $m$ -th score vector of the output, and  $\omega$  is the weight matrix. Using a weight matrix of [0.3, 0.7], we weighted the loss on the truth label as 0.3 and the loss on the deception label as 0.7, which was determined empirically.

#### 4.3.3 Model interpretability

The interpretability of the model trained on the DDCIT dataset was evaluated using the spatial and temporal attention modules of FacialCueNet. Spatial attention was expressed as the value of the importance mask  $M$ , and temporal attention was expressed as the importance weight at the  $i$ -th frame  $w_i$ . The  $w_i$  of the frame was normalized from 0 to 1 for each video to express intuitive spatial and temporal attention values. The importance mask obtained in each frame was visualized by multiplying the importance mask  $M$  and  $w_i$ .

## 5 Results

We present the experimental results using the FacialCueNet structure and parameters required for learning. The Real-life Trial Dataset was used to evaluate the versatility of FacialCueNet, and the DDCIT dataset was used to optimize FacialCueNet so that it could be used for criminal interrogations, which was the goal of this study. Based on the DDCIT dataset, performance was measured using a combination of facial cues as input for FacialCueNet. In addition, we checked which facial cues were actually considered important in predicting deception using the attention module embedded in FacialCueNet.

**Table 2** Comparison results of the presented approach with baseline facial deception detection models

Method	Accuracy (%)	AUC <sup>1</sup>
Wu et al. [60]	—	0.8456
Karimi et al. [9]	75.00	—
Ding et al. [11]	84.33	0.8411
Şen et al. [12]	80.79	0.9416
Avola et al. [13]	92.01	0.9337
Ours	88.45	0.9541

‘—’ means ‘not provided’

<sup>1</sup> Area under the receiver operating characteristic curve

### 5.1 Model reliability for deception detection

The generality FacialCueNet for deception detection was obtained using the Real-life Trial Dataset. With the above-mentioned settings, the accuracy and area under the ROC curve of a 10-fold cross-validation were 88.45% and 0.954, respectively (Table 2). Additionally, the sensitivity (recall), specificity, and precision were 0.9062, 0.9064, and 0.8917, respectively. When comparing the results of deception-detection methods in previous research using only images and features extracted from faces with a frame sampling rate of 15 or higher and using 104 videos excluding images with no faces in the Real-life Trial Dataset, FacialCueNet showed comparable performance (Table 2).

### 5.2 FacialCueNet performance

The performance of FacialCueNet was evaluated using the DDCIT dataset. To determine the hyperparameters for training, we conducted seven-fold cross-validation using samples collected in Session 1 of the DDCIT dataset. As a result, an accuracy of 71.75%, an F1 score of 0.8125, a recall of 0.9238, and a precision rate of 0.7286 were obtained. The performance of FacialCueNet on the test set is shown in Table 3. This was similar to the seven-fold cross-validation result with the training set. As a result of FacialCueNet optimization using various hyperparameters for training, we found that the batch size and the number of the convolutional

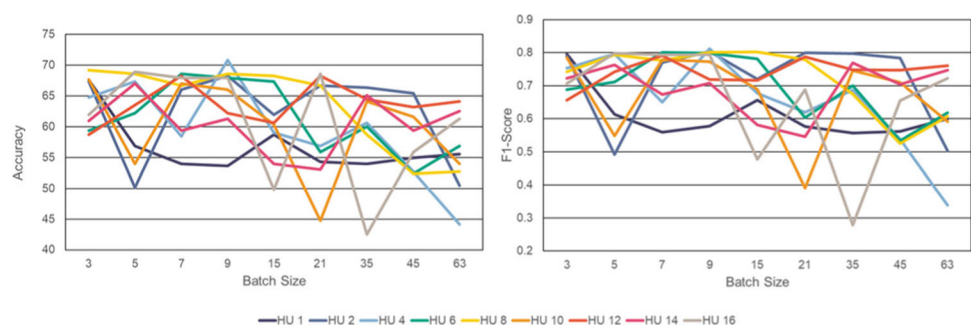
**Table 3** FacialCueNet performance on the DDCIT dataset

Accuracy (%)	F1 Score	Recall	Precision
70.79	0.8122	0.9476	0.7107

LSTM hidden units had a significant effect on performance. This was tested by comparing the results of the test set. FacialCueNet was trained using the entire train set by changing the batch size and number of hidden units. A comparison of the performances with the test set (Fig. 4) did not reveal a clear trend, but the best performance was obtained with batch sizes of 7 and 9 and when two, four, six, or eight hidden units were used. In addition, when the batch size was larger than 15 the overall performance deteriorated regardless of the number of hidden units, and the number of hidden units was greatly affected by the value of the batch size. Because FacialCueNet utilizes various facial cues, we conducted an ablation test on facial cues to confirm the effectiveness of the facial cues used. As shown in Table 4, among the cases where the combination of facial features extracted from the video frame and each facial cue was used as an input to FacialCueNet, the best performance was achieved when all four facial cues were used: AU, symmetry, gaze feature, and micro-expression. To develop a deception-detection model with interpretability, an attention module was added to FacialCueNet. Because the attention module has the advantage of improved performance, we also presented the ablation test results for the attention module (Table 5). In FacialCueNet, the spatio-temporal attention module had an effect on performance, and temporal attention had a greater effect on performance compared with spatial attention.

### 5.3 Deception detection cue

Using the attention module in the FacialCueNet, we present interpretation results of deception detection. The results are shown in Fig. 5, in which three examples of the detected deception cues mentioned in introduction section are shown. The differences between deception and truth were seen in the videos at the time when deception cues appeared. In the

**Fig. 4** Experimental results of FacialCueNet with different parameters. HU represents the number of hidden units in the LSTM

**Table 4** FacialCueNet accuracy, F1 score, recall, and precision results with facial cue combination on the DDCIT dataset

Input	Accuracy (%)	F1 score	Recall	Precision
Video	64.44	0.7617	0.8524	0.6885
Video+AU	65.71	0.7533	0.8714	0.6932
Video+sym	64.45	0.8000	0.8143	0.7008
Video+gaze	65.08	0.7718	0.8857	0.6838
Video+ME	66.03	0.7914	0.9429	0.6758
Video+AU+sym	66.67	0.7799	0.8857	0.6966
Video+AU+gaze	63.49	0.7392	0.7762	0.7057
Video+AU+ME	65.40	0.7625	0.8333	0.7028
Video+sym+gaze	64.13	0.7601	0.9333	0.6667
Video+sym+ME	64.44	0.7714	0.9000	0.6750
Video+gaze+ME	61.90	0.7087	0.6952	<b>0.7228</b>
Video+AU+sym+gaze	66.67	0.7771	0.8714	0.7011
Video+AU+sym+ME	67.30	0.7928	0.9381	0.6864
Video+AU+gaze+ME	66.67	0.7826	0.9000	0.6923
Video+sym+gaze+ME	67.62	0.7820	0.8714	0.7093
Video+AU+sym+gaze+ME	<b>70.79</b>	<b>0.8122</b>	<b>0.9476</b>	0.7107

Video, AU, sym, ME represent extracted facial features from video, action unit, facial symmetry, gaze features, and micro expression, respectively

case of the first subject, the blinking rate decreased in the deception case compared with the truthful case, and the spatial attention value was large in the eye area. In the second subject, lip pressing occurred in the case of deception, and the spatial attention value was larger in the lower part of the face. Finally, in the third subject, eye fixation occurred in the deception case, and many eye movements were observed in the truth video, relatively.

## 6 Discussion

In this study, FacialCueNet, a deception-detection model using various cues that appear on the face during deception, was developed. For FacialCueNet to be applied to an actual criminal interrogation, FacialCueNet was trained on data collected from an actual application environment. In addition,

beyond lie detection, we were able to analyze time zones and facial regions that are important for deception detection. FacialCueNet, a deception-detection model developed with a focus on criminal interrogations, showed sufficient performance to be used for deception-detection tasks using a public database (Table 2). As a result of training FacialCueNet with the DDCIT dataset, it achieved a high F1 score and recall compared with accuracy (Table 3). In the actual investigation and interrogation process, it is important to detect deception, but it is also important not to judge truth as deception in order to not persecute an innocent victim. In this regard, FacialCueNet has a potential application in criminal interrogation. As shown in Table 4, the basic structure of FacialCueNet uses facial features, action units, facial symmetry, gaze, and micro-expressions in video images. If all five features were not used, the performance was generally similar, but when four or more features were used, the performance was relatively good. All four facial cues, along with image features, contributed to the detection of deception. We used various parameters to optimize FacialCueNet. As shown in Fig. 4, as the batch size increased, the performance decreased. Although performance was poor when the number of hidden units was too small, performance also deteriorated when the number of hidden units was larger than the optimum. When we tried to optimize FacialCueNet, the loss of the model did not decrease when the model architecture was complicated. We assumed that this can be attributed to the simplicity of the data. The dataset used for deception detection was a relatively simple video with content that changed only slightly, unlike the datasets (e.g., playing basketball, or swimming) with marked differences

**Table 5** Spatio-temporal attention module ablation test in FacialCueNet using the DDCIT dataset

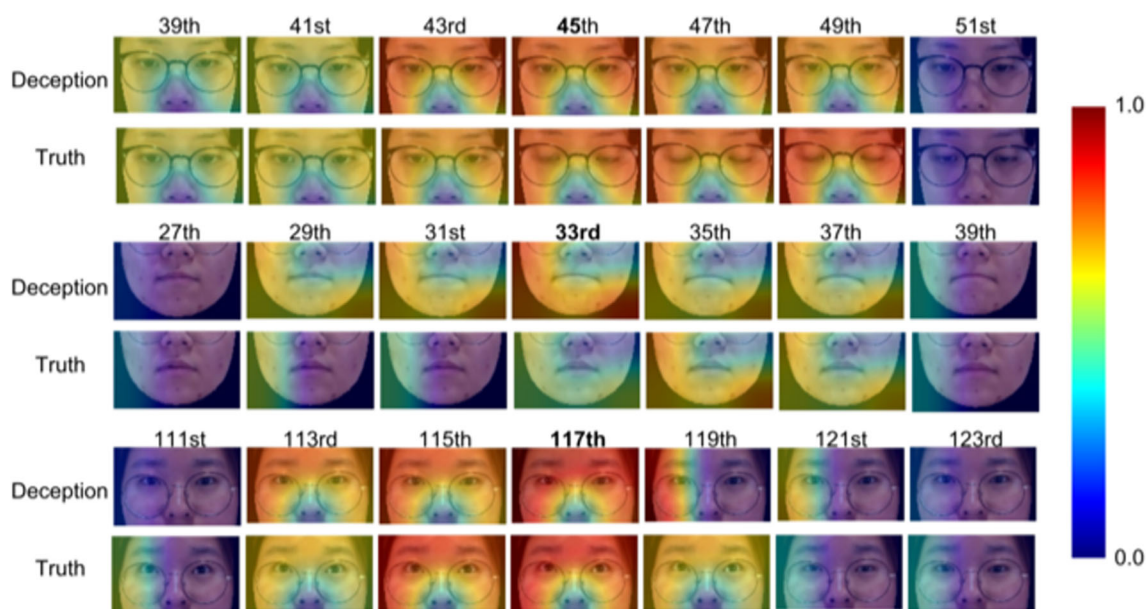
	Accuracy (%)	F1 Score	Recall	Precision
w/o Att. <sup>1</sup>	52.06	0.5546	0.4476	0.7287
Only S. Att. <sup>2</sup>	56.83	0.6201	0.5286	0.7500
Only T. Att. <sup>3</sup>	68.89	0.7967	0.9143	0.7058
S.T. Att. <sup>4</sup>	70.79	0.8122	0.9476	0.7107

<sup>1</sup>FacialCueNet without spatio-temporal attention module

<sup>2</sup>FacialCueNet only with the spatial attention module

<sup>3</sup>FacialCueNet only with the temporal attention module

<sup>4</sup>FacialCueNet with the spatio-temporal attention module

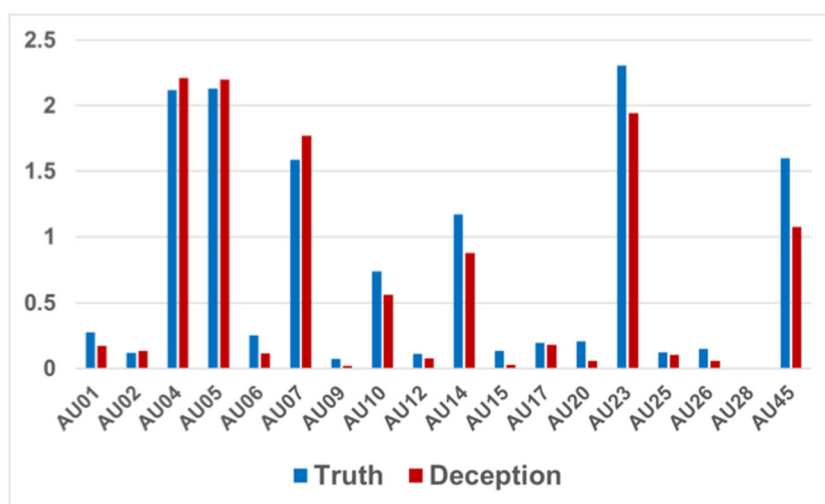


**Fig. 5** Spatial-temporal attention visualization. Each pair of lines shows the deception and truth aspects of the subject. The number above the picture represents the frame order in each video. The attention maps correspond to the right color bar

by class used for general video classification. Therefore, we concluded that this tendency was because complex models can make training difficult when using simple data as FacialCueNet tried to increase interpretability when used in an actual criminal interrogation by using the spatial-temporal attention module as a model structure. In this paper, facial cues appearing during deception reported in various previous studies using attention modules were found, and among them, blinking rate reduction, lip pressing, and eyes fixation were presented during deception (Fig. 5). When we analyzed the spatial attention of FacialCueNet for all samples in the DDCIT dataset, there were relatively larger spatial attention values on the upper part of the face close to the eyes compared with the lower parts of the face closer to the mouth. In

addition, we analyzed video frames with normalized temporal attention weights greater than 0.9. The presence of AUs was counted in video frames with normalized temporal attention weights greater than 0.9, and the average value for each AU was calculated for all videos (Fig. 6). As a result, the top three differences between deception and truth were AU45 (blink), AU23 (lip tightener), AU14 (dimpler). AU45 appeared less often in deception videos. This shows that the blinking rate decreases when lying, as reported in previous studies. AU23 and AU14 also appeared less often during deception. This suggests that there are fewer small movements at the bottom of the face than at the top of the face during deception in criminal interrogation situations. Although FacialCueNet performed relatively well, there is

**Fig. 6** Average of action unit presence in video frames with normalized temporal attention weights of 0.9 or higher





room for improvement. First, because the DDCIT dataset was collected from Koreans, it will be necessary to recruit and experiment with subjects of diverse nationalities to obtain a more generalizable model. Second, when the interpretation method is applied to actual criminal interrogations, a technique that can quantify the value of spatial and temporal attention will be required. Thirdly, an ethical anonymization method is required to render the extracted facial features unidentifiable. By utilizing data anonymization techniques, such as data quantization, which ensures that the data cannot be restored to its original image form, we can enhance security and improve the ethical feasibility of its application in deception detection during criminal interrogations. In addition, if various non-contact modalities, such as voice and infrared recording, are used, superior performance can be expected.

## 7 Conclusion

In this study, we present FacialCueNet, a non-contact deception-detection deep-learning model that utilizes facial expressions to aid in criminal investigations. The performance evaluation of FacialCueNet was conducted on public and additional datasets resembling real investigation conditions, demonstrating its potential for practical application in criminal investigations. The inclusion of the attention module in FacialCueNet enhances its utility by providing valuable insights into facial expression changes associated with deception. This research contributes to the advancement of deception detection methods, highlighting FacialCueNet as a reliable tool for improving the efficiency and accuracy of criminal investigations. Our future plans involve recruiting subjects of diverse nationalities to obtain a more generalizable model, developing a technique to quantify spatial and temporal attention in criminal interrogations, and exploring the use of various non-contact modalities for improved performance. These endeavors will enhance the applicability and effectiveness of FacialCueNet in real-world criminal investigation scenarios.

**Funding** This work was partly supported by (1) the Bio and Medical Technology Development Program of the NRF funded by the Korean government, MSIP(NRF-2017M3A9E1064781) (2) the Bio and Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No.NRF-2021M3E5D2A01022397)

**Data Availability** (1) Real-life Trial Dataset

The data is available under the terms of GNU General Public License. (<https://web.eecs.umich.edu/mihalcea/downloads.html>)

(2) DDCIT

Data sharing is not applicable to this study, as the collected data had to be discarded after being stored for 3 years from the end of the study (Institutional Review Board of the Hanyang University, HYU-2019-01-006-4).

**Code Availability** The models used in the paper are available at the following links

(1) FacialCueNet

<https://github.com/BorumNam/FacialCueNet>

(2) FaceNet

<https://github.com/timesler/facenet-pytorch>

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

**Ethics approval** For the use of the Real-life Trial Dataset dataset, this study was reviewed and approved by the institutional review board of Hanyang University (HYU-2019-01-006-4), and the requirement for informed consent was waived. Also, we collected dataset according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Hanyang University (HYU-2019-01-006-4).

**Informed consent** Informed consent was obtained from all individual participants included in the study.

**Consent for publication** The manuscript has not been sent to any other journal for publication.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References


1. Serota KB, Levine TR (2015) A few prolific liars: Variation in the prevalence of lying. *J Lang Soc Psychol* 34(2):138–157
2. Abd SH, Hashim IA, Jalal ASA (2021) Automated deception detection systems, a review. *Iraqi J Sci* 70–80
3. Speth J, Vance N, Czajka A, Bowyer KW, Wright D, Flynn P (2021) Deception detection and remote physiological monitoring: A dataset and baseline experimental results. In: 2021 IEEE International Joint Conference on Biometrics (IJCB). IEEE, pp 1–8
4. Matsumoto D, Hwang HS, Skinner L, Frank M (2011) Evaluating truthfulness and detecting deception. *FBI L. Enforcement Bull.* 80:1
5. Kircher JC, Raskin DC (2019) Polygraph techniques: History, controversies, and prospects. *Psychol Soc Pol* 295–308
6. Lewis JA, Cuppari M (2009) The polygraph: The truth lies within. *The Journal of Psychiatry & Law* 37(1):85–92
7. Porter S, Brinke L (2010) The truth about lies: What works in detecting high-stakes deception? *Legal Criminol Psychol* 15(1):57–75
8. Borza D, Itu R, Danescu R (2018) In the eye of the deceiver: Analyzing eye movements as a cue to deception. *J Imaging* 4(10):120

9. Karimi H, Tang J, Li Y (2018) Toward end-to-end deception detection in videos. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE, pp 1278–1283
10. Karnati M, Seal A, Yazidi A, Krejcar O (2021) Lienet: A deep convolution neural network framework for detecting deception. *IEEE Trans Cogn Dev Syst* 14(3):971–984
11. Ding M, Zhao A, Lu Z, Xiang T, Wen J-R Face-focused cross-stream network for deception detection in videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 7802–7811
12. Şen MU, Perez-Rosas V, Yanikoglu B, Abouelenien M, Burzo M, Mihalcea R (2020) Multimodal deception detection using real-life trial data. *IEEE Trans Affect Comput* 13(1):306–319
13. Avola D, Cascio M, Cinque L, Fagioli A, Foresti GL (2021) Lietome: An ensemble approach for deception detection from facial cues. *Int J Neural Syst* 31(02):2050068
14. Fernandes SV, Ullah MS (2021) Use of machine learning for deception detection from spectral and cepstral features of speech signals. *IEEE Access* 9:78925–78935
15. Sanaullah M, Gopalan K (2013) Deception detection in speech using bark band and perceptually significant energy features. In: 2013 IEEE 56th International Mid-west Symposium on Circuits and Systems (MWSCAS). IEEE, pp 1212–1215
16. Vrij A (2004) Why professionals fail to catch liars and how they can improve. *Legal Criminol Psychol* 9(2):159–181
17. Ansley N (1997) The Validity and Reliability of Polygraph Testing. Forensic Research Incorporated, Maryland
18. Honts CR, Perry MV (1992) Polygraph admissibility: Changes and challenges. *Law and Human Behavior* 16(3):357
19. Faigman DL, Fienberg SE, Stern PC (2003) The limits of the polygraph. *Issues Sci Technol* 20(1):40–46
20. Festinger L, Carlsmith JM (1959) Cognitive consequences of forced compliance. *J Abnorm Soc Psychol* 58(2):203
21. Cash CS (2014) Think about your heart: The effects of cognitive dissonance on cardiovascular functioning. PhD thesis, The Ohio State University
22. Ngo LM, Wang W, Mandira B, Karaoglu S, Bouma H, Dibeklioglu H, Gevers T (2021) Identity unbiased deception detection by 2d-to-3d face reconstruction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp 145–154
23. Kumar S, Bai C, Subrahmanian V, Leskovec J (2021) Deception detection in group video conversations using dynamic interaction networks. *Proceedings of the International AAAI Conference on Web and Social Media* 15:339–350
24. Yap MH, Rajoub B, Ugail H, Zwiggelaar R (2011) Visual cues of facial behaviour in deception detection. In: 2011 IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE). IEEE, pp 294–299
25. Verschuere B, Ben-Shakhar G, Meijer E (2011) Memory Detection: Theory and Application of the Concealed Information Test. Cambridge University Press, Cambridge
26. Matsumoto D, Ekman P (2008) Facial expression analysis. *Scholarpedia* 3(5):4237
27. Feldman RS, Jenkins L, Popoola O (1979) Detection of deception in adults and children via facial expressions. *Child Dev* 350–355
28. Dow MG (1985) Peer validation and idiographic analysis of social skill deficits. *Behavior Therapy* 16(1):76–86
29. Fydrich T, Chambless DL, Perry KJ, Buergener F, Beazley MB (1998) Behavioral assessment of social performance: A rating system for social phobia. *Behaviour Research Therapy* 36(10):995–1010
30. Egloff B, Schmukle SC (2002) Predictive validity of an implicit association test for assessing anxiety. *J Pers Soc Psychol* 83(6):1441
31. Yan W-J, Wu Q, Liang J, Chen Y-H, Fu X (2013) How fast are the leaked facial expressions: The duration of micro-expressions. *J Nonverbal Behav* 37:217–230
32. Bartlett MS, Hager JC, Ekman P, Sejnowski TJ (1999) Measuring facial expressions by computer image analysis. *Psychophysiol* 36(2):253–263
33. Ekman P (2009) *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*, revised. WW Norton & Company, New York
34. Graber RB (1981) Ekman: The face of man: Expressions of universal emotions in a new guinea village. *Stud Vis Commun* 7(1):83–85
35. Yap MH, Ugail H, Zwiggelaar R (2013) A database for facial behavioural analysis. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). IEEE, pp 1–6
36. Peth J, Kim JS, Gamer M (2013) Fixations and eye-blinks allow for detecting concealed crime related memories. *Int J Psychophysiol* 88(1):96–103
37. Ekman P, Friesen WV (1978) Facial action coding system. *Environmental Psychology & Nonverbal Behavior*
38. Mondal A, Mukhopadhyay P, Basu N, Bandyopadhyay SK, Chatterjee T (2016) Quantitative analysis of euclidean distance to complement qualitative analysis of facial expression during deception. *Ind Psychiatry J* 25(1):78
39. Cutrow RJ, Parks A, Lucas N, Thomas K (1972) The objective use of multiple physiological indices in the detection of deception. *Psychophysiol* 9(6):578–588
40. Marchak FM (2013) Detecting false intent using eye blink measures. *Front Psychol* 4:736
41. Rauch EH (2015) Cues to deception: eye blinking
42. Shen X, Fan G, Niu C, Chen Z (2021) Catching a liar through facial expression of fear. *Front Psychol* 12:675097
43. Ekman P (2009) Lie catching and microexpressions. *The Philosophy of Deception* 1(2):5
44. Wang JT-Y, Spezio M, Camerer CF (2010) Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *Am Econ Rev* 100(3):984–1007
45. Pérez-Rosas V, Abouelenien M, Mihalcea R, Burzo M (2015) Deception detection using real-life trial data. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, pp 59–66
46. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Computation* 9(8):1735–1780
47. Gallardo-Antolín A, Montero JM (2021) Detecting deception from gaze and speech using a multimodal attention lstm-based framework. *Appl Sci* 11(14):6393
48. Baygin M, Tuncer I, Dogan S, Barua PD, Tuncer T, Cheong KH, Acharya UR (2023) Automated facial expression recognition using exemplar hybrid deep feature generation technique. *Soft Comput* 1–17
49. Tuncer T, Dogan S, Subasi A (2023) Automated facial expression recognition using novel textural transformation. *J Ambient Intell Humaniz Comput* 1–15
50. Meng L, Zhao B, Chang B, Huang G, Sun W, Tung F, Sigal L (2019) Interpretable spatio-temporal attention for video action recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp 0–0
51. Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K, Woo W-c (2015) Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Adv Neural Inf Process Syst* 28
52. Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60(1–4):259–268

53. Amos B, Ludwiczuk B, Satyanarayanan M et al (2016) Openface: A general-purpose face recognition library with mobile applications. *CMU School Comput Sci* 6(2):20
54. Beh KX, Goh KM (2019) Micro-expression spotting using facial landmarks. In: 2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA). IEEE, pp 192–197
55. Orfanidis SJ (1988) Optimum Signal Processing: an Introduction. Macmillan publishing company, New York
56. Larsen J (2009) Correlation functions and power spectra. Section for Cognitive Systems, Informatics and Mathematical Modelling
57. Szegedy C, Ioffe S, Vanhoucke V, Alemi A (2017) Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 31
58. Cao Q, Shen L, Xie W, Parkhi OM, Zisserman A (2018) Vggface2: A dataset for recognising faces across pose and age. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE, pp 67–74
59. Schoneveld L, Othmani A, Abdelkawy H (2021) Leveraging recent advances in deep learning for audio-visual emotion recognition. *Pattern Recognit Lett* 146:1–7
60. Wu Z, Singh B, Davis L, Subrahmanian V (2018) Deception detection in videos. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 32
61. Bradley M, Cullen M (1993) Polygraph lie detection on real events in a laboratory setting. *Perceptual and Motor Skills* 76(3):1051–1058
62. Capps MH (1991) Predictive value of the sacrifice relevant. *Polygraph* 20(1):1–6
63. Ginton A, Daie N, Elaad E, Ben-Shakhar G (1982) A method for evaluating the use of the polygraph in a real-life situation. *J Appl Psychol* 67(2):131
64. Patrick CJ, Iacono WG (1989) Psychopathy, threat, and polygraph test accuracy. *J Appl Psychol* 74(2):347
65. Pérez-Campos Mayoral E, Langer AA (2019) Devil's advocate tactic to generate rapport. *Polygraph* 52(5)
66. Nelson R (2015) Scientific basis for polygraph testing. *Polygraph* 44(1):28–61
67. Hurley CM, Anker AE, Frank MG, Matsumoto D, Hwang HC (2014) Background factors predicting accuracy and improvement in micro expression recognition. *Motivation and Emotion* 38:700–714
68. Svetieva E, Frank MG (2016) Empathy, emotion dysregulation, and enhanced microexpression recognition ability. *Motivation and Emotion* 40:309–320
69. Khalil M, McGough AS, Pourmirza Z, Pazhoohesh M, Walker S (2022) Machine learning, deep learning and statistical analysis for forecasting building energy consumption-a systematic review. *Eng Appl Art Intell* 115:105287
70. Sohn J, Kim NS, Sung W (1999) A statistical model-based voice activity detection. *IEEE Signal Process Lett* 6(1):1–3

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Borum Nam<sup>1</sup> · Joo Young Kim<sup>2</sup> · Beomjun Bark<sup>2</sup> · Yeongmyeong Kim<sup>2</sup> · Jiyeon Kim<sup>2</sup> · Soon Won So<sup>2</sup> · Hyung Youn Choi<sup>3,4</sup> · In Young Kim<sup>2</sup> 

Borum Nam  
boruem0125@hanyang.ac.kr

Joo Young Kim  
jykim1026@hanyang.ac.kr

Beomjun Bark  
pbj7629@hanyang.ac.kr

Yeongmyeong Kim  
ymkim1202@lguplus.co.kr

Jiyeon Kim  
jiyeon0630@gmail.com

Soon Won So  
swso0530@outlook.kr

Hyung Youn Choi  
yelly301@naver.com

<sup>1</sup> Department of Electronic Engineering, Hanyang University, 222 Wangsimni-ro, Seoul 04763, Republic of Korea

<sup>2</sup> Department of Biomedical Engineering, Hanyang University, 222 Wangsimni-ro, Seoul 04763, Republic of Korea

<sup>3</sup> Department of Police and Criminal Justice, Dongguk University, 30, Pildong-ro 1-gil, Seoul 04620, Republic of Korea

<sup>4</sup> Department of Forensic Investigation Management, Seoul Metropolitan Police Agency, 18-8, Jahamun-ro 19-gil, Seoul 03035, Republic of Korea