Facial Expression Based Imagination Index and a Transfer Learning Approach to Detect Deception

Md Kamrul Hasan[1], Wasifur Rahman[1], Luke Gerstner[2], Taylan Sen[1],
Sangwu Lee[1], Kurtis Glenn Haut[1], Mohammed (Ehsan) Hoque[1]
{1-Department of Computer Science, 2-Goergen Institute for Data Science},University of Rochester, USA
{mhasan8,echowdh2}@cs.rochester.edu, lgerstn3@u.rochester.edu, tsen@cs.rochester.edu
{slee232,khaut}@u.rochester.edu, mehoque@cs.rochester.edu

Fig. 1: **The temporal progression of facial expressions in *Imagining* vs. *Remembering*. Can you guess which sequence corresponds to each? (Answers are given in the Acknowledgement Sec. and the explanation is given in Sec.VIII)**

***Abstract*—In this paper, we introduce a framework to automatically distinguish between facial expression sequences associated with imagining vs. remembering while answering a question. Our experiment includes a baseline and relevant questioning technique in the context of deception with 220 participants (20 hours long). Baseline questioning includes participants being separately asked to remember and imagine an arbitrary experience. During the relevant questioning, participants were prompted to either lie or tell the truth about a certain task. We trained a neural network model on the baseline data and achieved an accuracy of 60% on classifying imagining vs. remembering, whereas human performance for this task is 51%. Relevant questioning included a set of questions, each of which became an independent response segment. Using a transfer learning approach, we use the pretrained model from the baseline to obtain an imagination probability score for each relevant response segment. We define this individual probability per response as the *Imagination Index*. We apply the imagination indices as a feature vector to classify the whole relevant section as truth vs. bluff with an accuracy of 70%, significantly outperforming the human performance of 52%.**

***Index Terms*—transfer learning, non-verbal behavior, deception, facial expression**

I. Introduction

Many argue that facial expressions have universal meaning [1]. However, how one uses them given a particular context depends on many factors, making it very difficult to model computationally. As a result, the task of developing a one-size-fits-all model to classify nonverbal behavior is a grand challenge. Most prior work in the domain of facial expression analysis have relied on collecting a specific dataset and applying a model for the given task [2]–[4], further limiting the utility of the model in other contexts. Recently, there has been a growing trend to apply transfer learning – a way to transfer the knowledge learned in a data-rich context, to a data-scarce, but relevant context – to make a model more versatile. This approach has shown great potential in image classification, action recognition, text classification and spam filtering [5]–[7].

Studying non-verbal behaviors during human communication is an active research area in affective computing. Recent success has been demonstrated in transferring non-verbal signatures across domains [8], [9]. Due to extensive variations of facial expressions coupled with nuanced temporal dynamics, modeling complicated behaviours is an extremely challenging task. Recurrent Neural network based architectures like LSTM [10] have shown great promise in modeling complex temporal pattern and therefore, can be helpful in studying non-verbal signatures in human communication. Unfortunately, harnessing these advantages of deep learning architectures require a large amount of data, and in most cases, data is scarce in human behavioral studies. Transfer learning can help alleviate the problem by transferring knowledge from a data-rich context to a data-scarce context.

Our proposed transfer learning framework can alleviate these problems by gaining insights about facial expressions from one context and then using it in a relevant task. To illustrate the potential that transfer learning can have in affective computing, we have chosen to apply our model to the domain of deception detection. The nature of deception can vary widely based on specific context, but some non-verbal signatures can be generalizable across different contexts. Moreover, collecting large quality dataset of deception is very challenging and time consuming [11].

This research was supported in part by grant W911NF-15-1-0542 and W911NF-19-1-0029 with the US Defense Advanced Research Projects Agency (DARPA) and the Army Research Office (ARO).

978-1-7281-3888-6/19/$31.00 ©2019 IEEE

The interrogation game dataset used in our experiment involves two phases: baseline and relevant questioning [11]. In baseline questioning, participants are asked to alternatively remember and imagine about an arbitrary experience (Table.I). During the relevant questioning, participants are prompted to either bluff or tell the truth about a certain task (Table.II). Having baseline knowledge about a person's non-verbal facial expressions can help us detect deception during the relevant phase. Besides, it is much easier to collect a lot of data by asking baseline questions and train a neural network model on those data than it is to collect data samples on interpersonal deception. These baseline questions are not necessarily tied to deception and knowledge gained from them should be applicable in a broad range of contexts.

We trained a Neural Network architecture called *Baseline Knowledge Model* that can classify imagining vs.remembering given the facial expressions expressed while answering a baseline question. This model achieved a 60% accuracy (human performance is 51%). Fig.1 illustrates this classification task and its complicated temporal nature. As shown in Fig.1, even though participants tended to look away and frown during both remembering and imagining, the combined changes over time is needed to distinguish the two. Then we applied transfer learning by using this pre-trained model as a feature extractor for the target classification task: deception detection in the relevant phase. For each answer during relevant phase, the pre-trained model outputs a probability metric that we define as the *Imagination Index*. This index indicates the level of imagining vs. remembering involved for a particular answer. Our experiments indicate that imagination indices are a valuable feature in classifying the relevant phase as truth vs. bluff. Our model achieves an accuracy of 70%, significantly outperforming the human performance of 52% in the deception classification task. To the best of our knowledge, this is the first attempt of using a transfer learning approach based on non-verbal features in deceptive behavior analysis.

In summary, our contributions are:

- The development of the novel facial expression based *Imagination Index* to distinguish the level of imagining vs. remembering using a neural network.
- The introduction of a transfer learning-based framework for applying the *Imagination Index* for deception detection in an interrogation-based dyadic communication game.

II. Background

Facial expression analysis remains an active area of research in affective computing. Ekman developed the concept of micro-expressions and introduced the Facial Action Unit Coding System to identify the movements of a set of visually discernible facial muscles [12]. These facial Action Units are related to various emotions [13], [14]. Researchers have utilized micro-expressions in human affect analysis like sentiment [15], and various communication behavior analysis like deception [16]–[18], speed dating [2] and depression [19]. Affect analysis and recognition has been conducted in dyadic conversations as well [20]–[22].

With the advancement of deep learning architectures, people achieved significant performance on various human affective behavior recognition like sentiment [15], [23], [24] and depression [25], [26]. Recently, transfer learning approaches have become successful in various domains such as image classification [27], [28], human activity classification [29], and text classification [30]. It has the potential to transfer general facial expressions from one context to study another relevant context [8]. This kind of approach can be very useful to study human behavior like deception where data is very scarce.

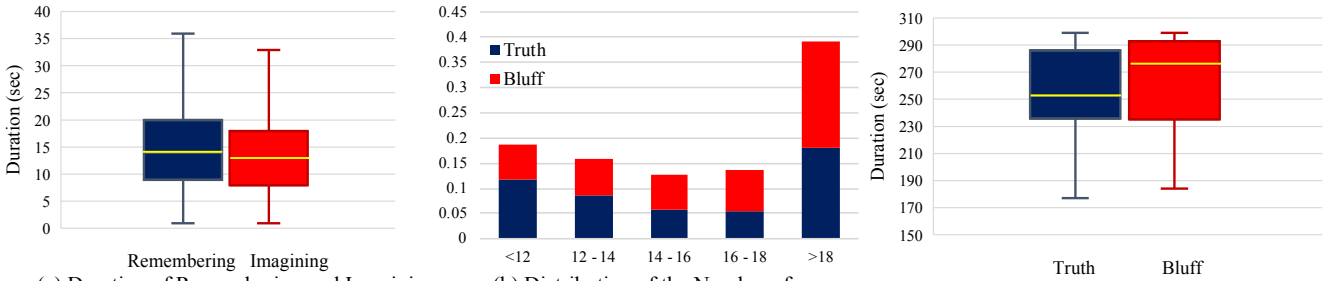
Researchers have been studying micro-expressions as indicators of deception [16]–[18]. Other non verbal behavior cues such as head movements [31], pupil dilation [32], eye blinking rates [33] and particular hand gestures [34] are also the indicators of deceptive behavior. Machine learning approaches to detect deception from video [35], [36] and deep learning based approaches to detect fake news [37], [38] and spam [39] have been studied too. To the best of our knowledge, there has been no work that involve transfer learning approach to study the deceptive behavior based on facial expression.

III. Dataset

In this paper, we use an openly available interrogation game dataset (N=220, #Truth=110, #Bluff=110) collected via Automatic Dyadic Data Recorder Framework [11]. Total duration of the videos in the dataset is 20 hours. Crowd sourced participants are paired to play an interrogation game where one participant acts as interrogator and another as witness. At the beginning of the game, the witness is shown an evidence (image) and randomly instructed by the system to tell either Truth or Bluff about the image. The interrogator is guided by the framework to ask a set of questions to the witness and participants take their turns to continue the conversation. At the end of the game interrogator has to determine whether witness is truthful or bluffing about the image. Participants are motivated by the bonus money to follow their role. The interrogator gets bonus if he correctly identifies the role of the witness and the witness gets bonus if the interrogator believes him. Each interrogation game comprises of two main phases: baseline and relevant.

A. Baseline Phase

During the baseline phase interrogator is prompted with a set questions by the system which are not related to the evidence image. The system instructs the witness to be honest during the Baseline Questions regardless of his assigned role of either telling the truth or bluff about the image. These questions are designed to evoke non-verbal expressions associated with mental states like slight confusion, analytic thought, remembrance and imagination. Table. I shows some example of questions related to remembrance and imagination. We have extracted 1126 video clips involving only imagining and remembering mental states (#imagining=563,#remembering=563) from the baseline phase of 107



(a) Duration of Remembering and Imagining Questions During the Baseline Phase

(b) Distribution of the Number of Questions During the Relevant Phase

(c) Duration of the Relevant Phase

Fig. 2: Overview of the Deception Dataset. (Best viewed in zoomed and color)

TABLE I: 'Remembering' and 'Imagining' questions used During the Baseline Phase

Remembering Questions
Describe what your childhood home looked like.
Describe what clothes you wore yesterday look like.
Describe what your elementary school looked like for 15 seconds.
Describe what burning your tongue feels like.
Please describe the characteristics of how thunder sound.
Imagining Questions
Imagine you are in a kingdom in the bottom of the sea, describe what it looks like
Imagine that you are in a castle in the clouds, describe what it looks like.
Describe the characteristics of the sounds a half-frog-half-deer would make.
Describe the characteristics of the sounds two planets colliding.
What would it feel like to be in a pool of ketchup like to you?

interrogation games. Utilizing these clips, we designed a neural network model that can distinguish between imagining vs. remembering based on non-verbal facial expression features. Fig. 2.a shows the distribution of *Remembering* and *Imagining* answers' duration in seconds.

B. Relevant Phase

Several questions related to the evidence image is prompted to the interrogator by the ADDR system (Table II). At the end of the relevant phase, the system asks the interrogator whether he/she believes that the witness is telling the truth or bluffing regarding the image. As a result, the total number of relevant phase data is 220 (#Truth=110, #Bluff=110). The distribution of number of questions and average duration of the relevant phase is shown in Fig.2.b and Fig.2.c.

C. Extracted Features

The interrogation game video is recorded at 15 frame/sec. We extract the non-verbal features like facial expressions and emotions only. Verbal features like audio and language can have task dependent bias. Non-verbal features are more universal [1], [40] and is more appropriate for transfer learning approach.

TABLE II: 'Relevant questions' used in the relevant phase

What was your image?
Could you give me some more details about the image?
If there were something to count in the image, what would it be and what would be the count?
Were there any other objects in the image?
What were the colors in the image?
Please tell me about the background in the image.
Where do you think the photograph was taken?
Were parts of the object in the image man-made?

1) Facial Expression Features: OpenFace behavioral analysis tool [41] is used to analyze the facial expression of witness. Facial Action Unit (AU) features are extracted based on the Facial Action Coding System (FACS) [12]. In addition, we also extracted eye gaze directions and head pose features [41].

2) Affect Features: We have also extracted all the affective features like joy, fear, disgust, sadness, anger, surprise, contempt, valence, engagement, smile and more form the facial expression analysis tool called affectiva [42].

IV. Research Questions

Our research focused on the following questions:

1) Can we model temporal facial expression features to distinguish whether someone is remembering directly from memory or imagining something new?
2) Can we transfer information from the imagining vs. remembering network to detect deception during relevant phase?

V. Methods

Every interrogation game data consists of two phases: a Baseline phase (B) and a Relevant phase (R). From Baseline phase, we take N question-answering sessions; each of them is labeled as either Imagining (Im) or Remembering (Re). We can denote the Baseline phase (B) as $B = (B_1, B_2, \dots B_N)$.

During relevant phase the interrogator asks a set of M questions to the witness. The Relevant phase consists of M question-answering session and the entire phase is labeled as either Truth (Tr) or Bluff (Bl). We denote Relevant phase as $R = (R_1, R_2, \dots R_M)$; where R_i is i^{th} question-answering session.

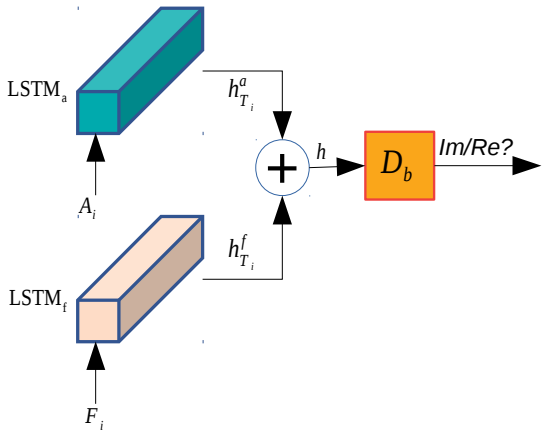


Fig. 3: Baseline Knowledge Model. (Best viewed in zoomed and color)

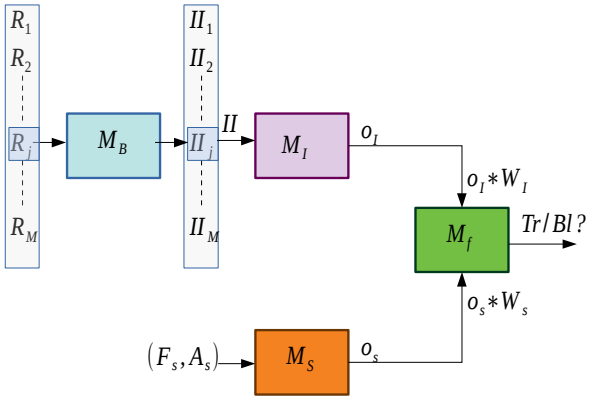


Fig. 4: Weighted Fusion Model. (Best viewed in zoomed and color)

Each baseline question-answering session (B_i) uses facial expression features (F_i) and affect features (A_i) and hence, we can define $B_i = (F_i, A_i)$. Each B_i session consists of T_i segments - each one second long. We can express F_i as as $F_i = (f_1, f_2, \dots f_{T_i})$ where f_j – facial expression features averaged across jth segment – is a d_f dimensional vector. Similarly, $A_i = (a_1, a_2, \dots a_T)$ where a_j is a d_a dimensional vector. Similarly, a relevant question-answering session (R_j) will also have facial expression features(F_j) and affect features(A_j) and therefore, we can define $R_j = (F_j, A_j)$.

To build *summary* features, we average all the facial expression features and affect features on the entire length of R. For facial expression features, we get $F_s = (\hat{f}_1, \hat{f}_2, \dots \hat{f}_{d_f})$ where $\hat{f}_i$ is the average value of the ith facial expression feature over R. We represent the affect features as $A_s = (\hat{a}_1, \hat{a}_2, \dots \hat{a}_{d_a})$ where $\hat{a}_i$ is the average value of the ith affect feature over R.

Our framework consists of:

- **Baseline Knowledge Model** that determines whether the label of a baseline clip B_i is Imagining (Im) or Remembering (Re)
- **Pre-trained Feature Extraction** that extracts *Imagination Indices* from the relevant phase question-answers. Pre-trained *Baseline Knowledge Model* is used to extract these features to use in deception detection task
- **Deception Detection Models** that predicts whether the label of the Relevant phase (R) is Truth (Tr) or Bluff (Bl).

A. Baseline Knowledge Model

We model each baseline question-answering session (B_i) using two LSTMs [10] : LSTM$_f$ for modeling F_i and LSTM$_a$ for modeling A_i. Given $F_i = (f_1, f_2, \dots f_{T_i})$, LSTM$_f$ creates a sequence of hidden states $(h^f_1, h^f_2, \dots h^f_{T_i})$ where h^f_i encodes the information in the first i elements of F_i. We use the last hidden state $h^f_{T_i}$ as the encoding of F_i. Similarly, LSTM$_a$ creates a sequence of hidden states $(h^a_1, h^a_2, \dots h^a_{T_i})$ and we use the last hidden state $h^a_{T_i}$ as the encoding of A_i. We concatenate both of them to form a new vector $h = (h^f_T, h^a_T)$. We apply an affine transformation $D_b : |h| \mapsto 1$ on h to produce $y^i_B = D_b(h)$ – the probability that the label of B_i is Imagining (Im).

B. Pre-trained Feature Extraction

We define the model trained in V-A as M_B. Although the model was trained to classify Imagining (Im) vs. Remembering (Re) on the baseline phase of our data, we can use it as a feature extractor for the relevant phase. Given a relevant question-answering session $R_j = (F_j, A_j)$, the model outputs $II_j = M_B(R_j)$ which is the probability of whether the witness is imagining that answer (instead of remembering it). We define this probability metric as ***Imagination Index***

Given a relevant phase $R = (R_1, R_2, \dots R_M)$ of M question-answering sessions, we extract imagination indices feature vector $II = (II_1, II_2, \dots II_M)$ using the pre-trained model. Since R is labeled as Truth (Tr) or Bluff (Bl); either all of the answers in R are Tr or all of them are Bl.

C. Deception Detection Models

1) Weighted Fusion Model (WFM): This model will have two sub-models: M_I to model II and M_S to model (F_s, A_s).

M_I is made of two linear transformations with a ReLU (ReLU$(v) = max(v, 0)$) activation function [43] in between. M_I maps II to a real number o_I. M_S has the identical structure as M_I and it maps (F_s, A_s) to a single real number o_s. We combine (o_I, o_s) through another affine transformation M_f to get final prediction.

$$o_I = M_I(II)$$
$$o_s = M_S(F_s, A_s)$$
$$\hat{y} = M_f(W_I * o_I + W_s * o_s)$$

W_I and W_s are two learnable weights assigned to the output of M_I and M_S respectively.

2) Support Vector Machine: We trained a Support Vector Machine(SVM) [44] using a Radial Basis Function(RBF) kernel [45] on the imagination indices (II) and summary(F_s, A_s) features of the Relevant session(R).

VI. Experiments

In the experiments of this paper, our goal is to show that pre-trained feature *Imagination Index* contains useful information about imagining and remembering pattern.

First, we use the Baseline Knowledge Model(Section. V-A) to distinguish between imagining and remembering. We used 1126 video clips – 563 imagining answers and 563 remembering answers – from the baseline phase of 107 interrogation games. We divide the train, development and test folds in such a way that these folds share no speaker among them - hence standard folds are speaker independent [46].

Second, we use the best Baseline Knowledge Model as pre-trained feature extractor for the deception detection task during relevant phase. For this experiment, the relevant phase of 220 interrogation games are used (#Truth=110, #Bluff=110). We use Weighted Fusion Model (WFM) described in Section V-C1 for the task. Both the summary of facial expression and affective features over whole relevant phase and *Imagination Indices* are used to train this model. For hyper-parameter tuning we again divide this dataset into the train, development and test folds.

Aside from the proposed WFM, the following variants are also studied:

WFM (S): This variant of WFM uses summary of facial expression and affective features over whole relevant phase only. The score of this model demonstrates how much information the summary features contain.

WFM (P): This variant of WFM only uses *Imagination Indices* extracted from the Baseline Knowledge Model only. We aim to understand how much information is transferred from Baseline Knowledge Model model to relevant phase.

We also use a Support Vector Machine (SVM) (V-C2) classifier. Both the summary and *Imagination Indices* are used to train this model. Similarly, we also use two variation of SVM model, **SVM (S)**: uses summary of facial expression and affective features only, **SVM (P)**: uses *Imagination Indices* extracted from the Baseline Knowledge Model only.

VII. Results

First, we are going to present the result of the experiments done on imagining vs. remembering detection from the answer of witnesses. Then, we present the results of the experiments involving transfer learning and deception detection during relevant phase.

A. Exploratory Study about Imagining vs Remembering

The task of detecting imagining vs. remembering from nob-verbal features is a very challenging task. From Table. I, we can see that imagining a story involves remembering from past experiences too. Therefore, imagining can evoke similar facial expressions like remembering. Even human performance for this task is around 51%. We will use the human performance as baseline for this experiment. Our baseline knowledge model achieves 60% accuracy and 60% F1-score which outperforms the human baseline.

TABLE III: binary accuracy for different variants of WFM, SVM outlined in Section VI. 'S' denotes the training scenario where only summary facial expression and affective features are used.'P' denotes the scenario where pre-trained imagination indices are used only. Others use both features

Models	Accuracy	F1 Score
WFM (S)	0.68	0.68
WFM (P)	0.70	0.68
WFM	**0.70**	**0.70**
SVM (S)	0.52	0.55
SVM (P)	0.57	0.63
SVM	0.61	0.59
Human Performance	0.52	

B. Experiment Results of Deception Detection

The results of these experiments are presented in Table III. At the end of each game, the interrogator determine whether witness is truthful or bluffing from the image. From those decisions, we calculate the human performance for this task to be 52%. All of our models outperform the human baseline using non-verbal facial expressions and affective features only. On the contrary, human interrogators use all the modalities – text, audio and vision – to take their decisions.

Models (WFM(P) and SVM(P)) trained with Imagination Indices feature alone can outperform the models (WFM(S) and SVM(S)) that use summary features only. Even the Weighted Fusion Model based on the Imagination Index feature alone (WFM (P)) achieved same accuracy as the best model (WFM) that uses all features. Our results suggest that the pre-trained features extracted from the Baseline Knowledge Model contains useful information. The best performance, in terms of accuracy and f1 score, was obtained from the WFM model that uses both summary and pretrained features. From this best performing model, we observe the weights in last layer to be $W_I = 0.73$ and $W_S = 0.48$. It is evident that the model put more attention on the Imagination Index furthering underscoring the importance of our pre-trained features.

VIII. Discussion

The result of our experiments clearly demonstrate that the pre-trained Imagination Index feature is useful for the deception detection task. This follows human intuition, because when an individual is being deceptive, they are often imagining a hypothetical situation that did not occur. In contrast, when an individual is telling the truth, they are often remembering an event from their memory. The *Imagination Index* is not strictly

Human performance for detecting imagining vs. remembering is calculated by averaging the performance of three annotators over a shuffled set of 100 imagining and 100 remembering cases. The annotators are given the same input as the machine learning models. They watched the video without any sound.

TABLE IV: statistical comparison results of facial expression and affect features between imagining and remembering answers, mean values are normalized

Features	t-value	p-value	Imagining Mean	Remembering Mean
Action Unit 45	-2.795	0.005	0.0697	0.079
Action Unit 12	1.996	0.046	0.412	0.370
Lip Press	-1.978	0.048	0.049	0.059
Eye Closure	-2.445	0.015	0.11	0.131
Lid Tighten	2.639	0.008	0.045	0.032
Dimpler	-2.208	0.027	0.042	0.053
Wink	1.977	0.048	0.103	0.087

limited to this kind of dyadic deceptive communication and could be applied to other scenarios, such as airport security screening, real life trials etc. It could be applied to other affective computing analyses as well. For example, it would be beneficial to learn whether successful speed daters, public speakers, sales people, and psychiatric counselors tend to have a higher or lower *Imagination Index*.

The importance of temporal sequences of facial expressions and head pose over time is shown in Fig.1. Sequences 1 and 4 depict the facial expressions over time while an individual is remembering an experience. Sequences 2 and 3 depict when an individual is imagining an experience. An inspection of the individual frames reveals that both remembering and imagining involve breaking eye contact and an engaged head pose, as well as some expression of a lip corner depressor (i.e. AU 15). However, looking at the individual frames, it may difficult to clearly distinguish between imagining and remembering. Indeed, when we analyze the static frame frequencies of each of the individual facial expression, head pose, and eye gaze features gathered, there was no clear distinction between imagining and remembering. It is only when the full sequence is examined with our temporal model that we gain the ability to distinguish between imagining and remembering. Though our baseline knowledge model achieved 60% accuracy, it is still better than the human performance (51%). In future work, we will attempt to explain the temporal pattern that the model has learned in comprehensible terms.

Sequential network architecture like RNN, LSTM could have been more appropriate for modeling deception during the relevant phase. However, due the small size of the dataset (N=220), we used a simple two layer feed forward network with regularization in Weighted Fusion Network. It may be surprising that deception detection accuracy (70%) is greater than distinguishing imagining vs. remembering accuracy (60%). But we constructed an *Imagination indices* by evaluating at least ten questions independently; therefore the deception detection model has the knowledge of whether each question is representing imagining or remembering. With such knowledge, the model has the capability of detecting deception more successfully, therefore such boost in accuracy is anticipated.

We also conducted an unpaired t-test analysis on the facial expressions between remembering and imagining answers to see if there is any difference. The most notable results (p-value< 0.05) are demonstrated in Table IV. One of the most notable differences is in AU12 (Lip Corner Puller), which is higher for the imagining group. This suggests that when people are creating something, they have greater expressions of smile than when they are simply directly remembering. This finding of greater smile expressions when imagining a thought has also been found in deceptive communication [47]. Furthermore, when people answer remembering questions, they show more expressions of AU45 (Blink) and Eye Closure than when they are answering imagining questions. As shown in previous research, it is typical for a person to close their eyes in order to disengage themselves from their environment when trying to remember an event and thus facilitate memory [48]. However, after conducting Bonferroni correction, with a correction factor of 74, none of the findings remain significant. Nonetheless, some of these results are interesting and aligned with the previous research findings.

IX. Conclusion

In this paper, we introduce a neural network architecture called Baseline Knowledge Model that can distinguish when people are imagining vs. remembering based on their facial expressions. We achieve an accuracy of 60% on classifying imagining vs. remembering whereas human performance for this task is 51%. We use this pre-trained model to develop a novel facial expression based *Imagination Index* to quantify the level of imagination. Using a transfer learning approach, *Imagination Index* feature is applied to detect deception in an interrogation based dyadic communication game. We achieve the performance of 70% accuracy in detecting Truth vs. Bluff, significantly outperforming the human performance of 52%.

Acknowledgment

This research was supported in part by grant W911NF-15-1-0542 and W911NF-19-1-0029 with the US Defense Advanced Research Projects Agency (DARPA) and the Army Research Office (ARO). Fig.1. Sequences 1 and 4 depict the facial expressions over time while an individual is remembering an experience, while sequences 2 and 3 depict when an individual is imagining an experience.

References

[1] S. Tomkins, *Affect imagery consciousness: Volume I: The positive affects*. Springer publishing company, 1962.

[2] M. R. Ali, T. Sen, D. Crasta, V.-D. Nguyen, R. Rogge, and M. E. Hoque, "The what, when, and why of facial expressions: An objective analysis of conversational skills in speed-dating videos," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 203–209.

[3] I. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque, "Automated analysis and prediction of job interview performance," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 191–204, 2018.

[4] T. Sen, M. R. Ali, M. E. Hoque, R. Epstein, and P. Duberstein, "Modeling doctor-patient communication with affective text analysis," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, Oct 2017, pp. 170–177.

[5] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual categorization: A survey," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 5, pp. 1019–1034, 2015.

[6] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[7] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, p. 9, 2016.

[8] B. Romera-Paredes, M. S. Aung, M. Pontil, N. Bianchi-Berthouze, A. C. d. C. Williams, and P. Watson, "Transfer learning to account for idiosyncrasy in face and body expressions," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, 2013, pp. 1–6.

[9] T. Almaev, B. Martinez, and M. Valstar, "Learning to transfer: transferring latent task structures and its application to person-specific facial action unit detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3774–3782.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] T. Sen, M. K. Hasan, Z. Teicher, and M. E. Hoque, "Automated dyadic data recorder (addr) framework and analysis of facial cues in deceptive communication," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, p. 163, 2018.

[12] R. Ekman, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA, 1997.

[13] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[14] P. Ekman, W. V. Freisen, and S. Ancoli, "Facial signs of emotional experience." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1125, 1980.

[15] A. Zadeh, M. Chen, S. Poria, E. Cambria, and L.-P. Morency, "Tensor fusion network for multimodal sentiment analysis," *arXiv preprint arXiv:1707.07250*, 2017.

[16] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969.

[17] G. Warren, E. Schertler, and P. Bull, "Detecting deception from emotional and unemotional cues," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 59–69, 2009.

[18] T. Sen, M. K. Hasan, M. Tran, M. Levin, Y. Yang, and M. E. Hoque, "Say cheese: Common human emotional expression set encoder and its application to analyze deceptive communication," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 357–364.

[19] S. Song, L. Shen, and M. Valstar, "Human behaviour-based automatic depression analysis using hand-crafted statistics and deep learned spectral features," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 158–165.

[20] D. Hazarika, S. Poria, A. Zadeh, E. Cambria, L.-P. Morency, and R. Zimmermann, "Conversational memory network for emotion recognition in dyadic dialogue videos," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018, pp. 2122–2132.

[21] C. Navarretta, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, and B. Maegaard, "Mirroring facial expressions and emotions in dyadic conversations." in *LREC*, 2016.

[22] S. Samrose, R. Zhao, J. White, V. Li, L. Nova, Y. Lu, M. R. Ali, and M. E. Hoque, "Coco: Collaboration coach for understanding team dynamics during video conferencing," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 4, pp. 160:1–160:24, Jan. 2018. [Online]. Available: http://doi.acm.org/10.1145/3161186

[23] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102–107, 2016.

[24] A. Zadeh, P. P. Liang, N. Mazumder, S. Poria, E. Cambria, and L.-P. Morency, "Memory fusion network for multi-view sequential learning," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[25] L. Yang, D. Jiang, X. Xia, E. Pei, M. C. Oveneke, and H. Sahli, "Multimodal measurement of depression using deep learning models," in *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*. ACM, 2017, pp. 53–59.

[26] T. Al Hanai, M. Ghassemi, and J. Glass, "Detecting depression with audio/text sequence modeling of interviews," in *Proc. Interspeech*, 2018, pp. 1716–1720.

[27] D. Han, Q. Liu, and W. Fan, "A new image classification method using cnn transfer learning and web data augmentation," *Expert Systems with Applications*, vol. 95, pp. 43–56, 2018.

[28] N. Zhuang, Y. Yan, S. Chen, H. Wang, and C. Shen, "Multi-label learning based deep transfer neural network for facial attribute classification," *Pattern Recognition*, vol. 80, pp. 225–240, 2018.

[29] M. Harel and S. Mannor, "Learning from multiple outlooks," *arXiv preprint arXiv:1005.0027*, 2010.

[30] C. B. Do and A. Y. Ng, "Transfer learning for text classification," in *Advances in Neural Information Processing Systems*, 2006, pp. 299–306.

[31] S. Lu, G. Tsechpenakis, D. N. Metaxas, M. L. Jensen, and J. Kruse, "Blob analysis of the head and hands: A method for deception detection," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. IEEE, 2005, pp. 20c–20c.

[32] A. K. Webb, C. R. Honts, J. C. Kircher, P. Bernhardt, and A. E. Cook, "Effectiveness of pupil diameter in a probable-lie comparison question test for deception," *Legal and Criminological Psychology*, vol. 14, no. 2, pp. 279–292, 2009.

[33] K. Fukuda, "Eye blinks: new indices for the detection of deception," *International Journal of Psychophysiology*, vol. 40, no. 3, pp. 239–245, 2001.

[34] L. Caso, F. Maricchiolo, M. Bonaiuto, A. Vrij, and S. Mann, "The impact of deception and suspicion on different hand movements," *Journal of Nonverbal behavior*, vol. 30, no. 1, pp. 1–19, 2006.

[35] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, "Deception detection using real-life trial data," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 2015, pp. 59–66.

[36] Z. Wu, B. Singh, L. S. Davis, and V. Subrahmanian, "Deception detection in videos," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[37] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2017, pp. 797–806.

[38] W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648*, 2017.

[39] Y. Ren and D. Ji, "Neural networks for deceptive opinion spam detection: An empirical study," *Information Sciences*, vol. 385, pp. 213–224, 2017.

[40] P. Ekman, E. R. Sorenson, and W. V. Friesen, "Pan-cultural elements in facial displays of emotion," *Science*, vol. 164, no. 3875, pp. 86–88, 1969.

[41] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–10.

[42] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby, "Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit," in *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. ACM, 2016, pp. 3723–3726.

[43] Y. Li and Y. Yuan, "Convergence analysis of two-layer neural networks with relu activation," in *Advances in Neural Information Processing Systems*, 2017, pp. 597–607.

[44] K.-R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," in *International Conference on Artificial Neural Networks*. Springer, 1997, pp. 999–1004.

[45] G. F. Smits and E. M. Jordaan, "Improved svm regression using mixtures of kernels," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, vol. 3. IEEE, 2002, pp. 2785–2790.

[46] A. Zadeh, R. Zellers, E. Pincus, and L.-P. Morency, "Mosi: multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos," *arXiv preprint arXiv:1606.06259*, 2016.

[47] L. Ten Brinke, S. Porter, and A. Baker, "Darwin the detective: Observable facial muscle contractions reveal emotional high-stakes lies," *Evolution and Human Behavior*, vol. 33, no. 4, pp. 411–416, 2012.

[48] A. M. Glenberg, J. L. Schroeder, and D. A. Robertson, "Averting the gaze disengages the environment and facilitates remembering," *Memory & cognition*, vol. 26, no. 4, pp. 651–658, 1998.