SCAN: Socio-Cultural Attitudinal Networks MURI Project

Dr. Purush Iyer

US Army Research Office

Summer 2020 Webinar Series

Date	Time (EST)	Speaker	Title
June 4	3 – 4:30 pm	**Dr. Purush Iyer** US Army Research Office **V.S. Subrahmanian** Dartmouth College	Introduction to the SCAN Project and Deception Detection from Online Videos
June 10	4 – 5 pm	**Judee Burgoon** University of Arizona	A Novel Approach to Investigating Deception during Group Interaction
June 15	12-1 pm	**Norah Dunbar** UCSB	Persuasive Deception and Dyadic Power Theory
June 22	3 – 4 pm	**Jure Leskovec** Stanford University	Dynamic Embeddings of Temporal Interaction Networks
June 29	3 – 3:40 pm	**Pan Li** Stanford University	An Interpretable Representation Learning Framework for Dynamic Social Networks
July 13	3 – 4 pm	**Chongyang Bai** Dartmouth College	Dominance Detection in Group Interaction Videos
July 20	4 --5 pm	**Dimitris Metaxas** Rutgers University	Detection and Tracking of Humans and Faces using Machine Learning
July 27	4 – 5 pm	**Dimitris Metaxas** Rutgers University	Video-based Deception Detection and Corresponding Feature Discovery
August 11	2 – 3 pm	**Jay Nunamaker** University of Arizona	Going the Last Mile for SCAN Transition
August 18	1 – 2 pm	**Miriam Metzger** UCSB	International Data Collection using Human Subjects: Logistics and Challenges

SCAN: Socio-Cultural Attitudinal Networks

Major Accomplishments to Date

Deception Detection Research

V.S. Subrahmanian

Dartmouth College

vs@dartmouth.edu

Research funded by the US Army Research Office

Joint work with many collaborators from the SCAN Team and Beyond.

Ground Rules for SCAN Project Webinar Series

- **Please use the Q&A button at the bottom of your Zoom screen to ask questions. Questions will be moderated.**
- **The Q&A session will follow the panel presentation.**
- **The entire session is being recorded and will be made available on the ISTS YouTube channel.**

SCAN Team

SCAN Project Goals

Identify non-verbal behaviors and develop predictive models that enable us to better understand and predict

- **Dominance/deference relationships**
- **Trust/distrust relationships**
- **Like/dislike relationships**
- **Deception**

in group settings where multiple people interact with each other.

Potential SCAN Project Applications

Negotiations

Meetings

Sales Events

Interviews

Security Interviews

Talk Outline

Overview of the SCAN Project

- **How Humans Detect Deception and Dominance**
- How AI Algorithms Detect Deception and Dominance
- Other Major Contributions

Deception Detection

- Deception in Real-world Courtroom Videos
- Deception in Multi-Player Face to Face Games

Accomplishment I: The SCAN Dataset

The world's **most extensive dataset** on human-human communications in a setting that is

- Multinational
- Multicultural
- Designed to elicit behaviors such as
  - Like/dislike
  - Trust/distrust
  - Dominance/deference
  - Deception
- 6 countries, 8 sites, almost 700 participants in all.
- Developed training manual and game software to support replicating our Resistance-style game

Accomplishment I: The SCAN Dataset

How Humans Detect Deception and Dominance

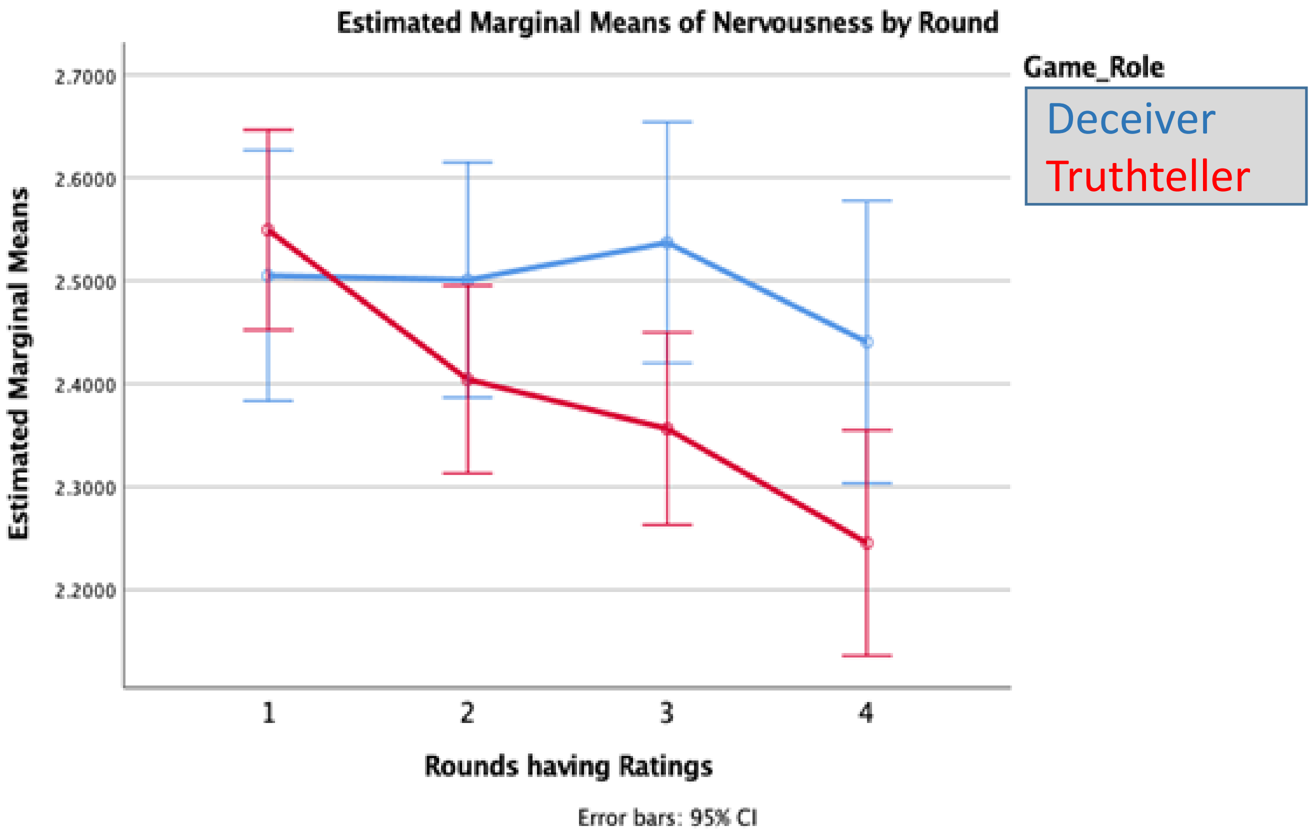
These results either study how humans use communication cues to detect deception and dominance, or how human-provided inputs in conjunction with statistical models can do so using the SCAN dataset.

Accomplishment II: Discovering the Cues used by Humans to Detect Deception

- Eye blinks
- Stretched lips, lips up
- Eyebrows – frown, raised
- Deceivers are more nervous over time
- Deceivers are less trusted over time
- Deceivers are less dominant
- Interaction with other Deceivers

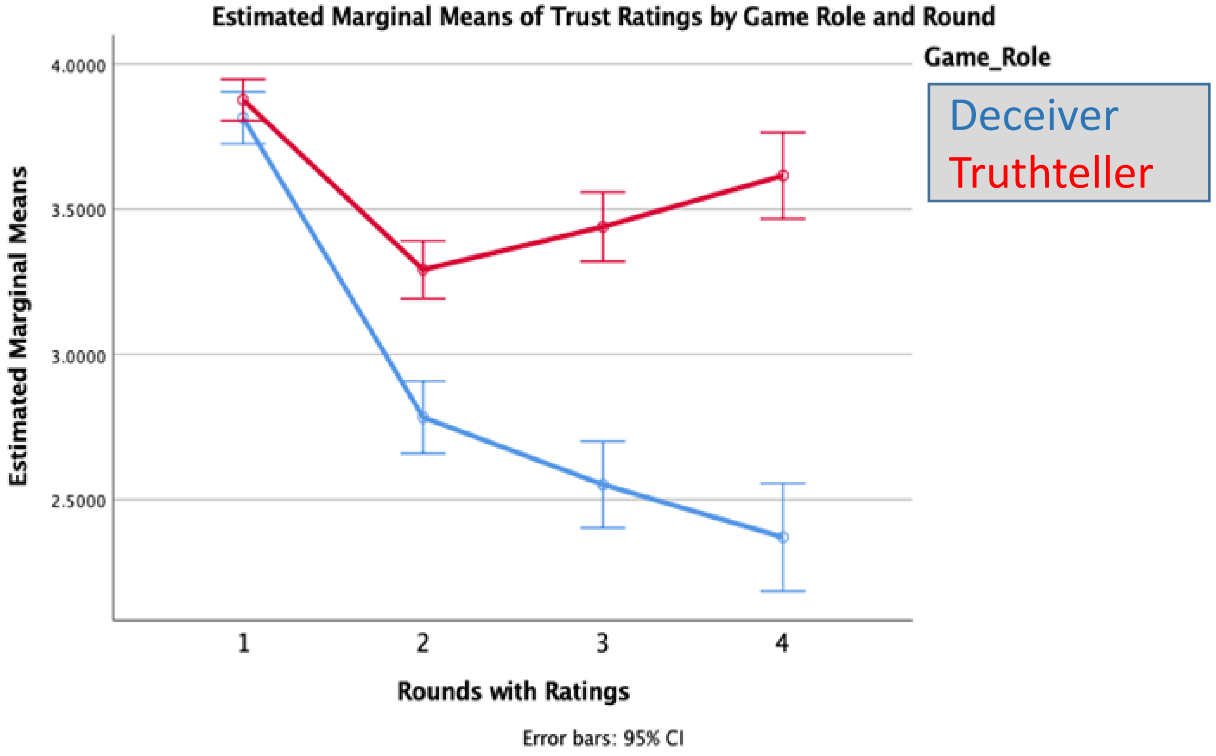
Accomplishment II: Discovering the Cues used by Humans to Detect Deception

- Eye blinks
- Stretched lips, lips up
- Eyebrows – frown, raised
- **Deceivers are more nervous over time**
- Deceivers are less trusted over time
- Deceivers are less dominant
- Interaction with other Deceivers

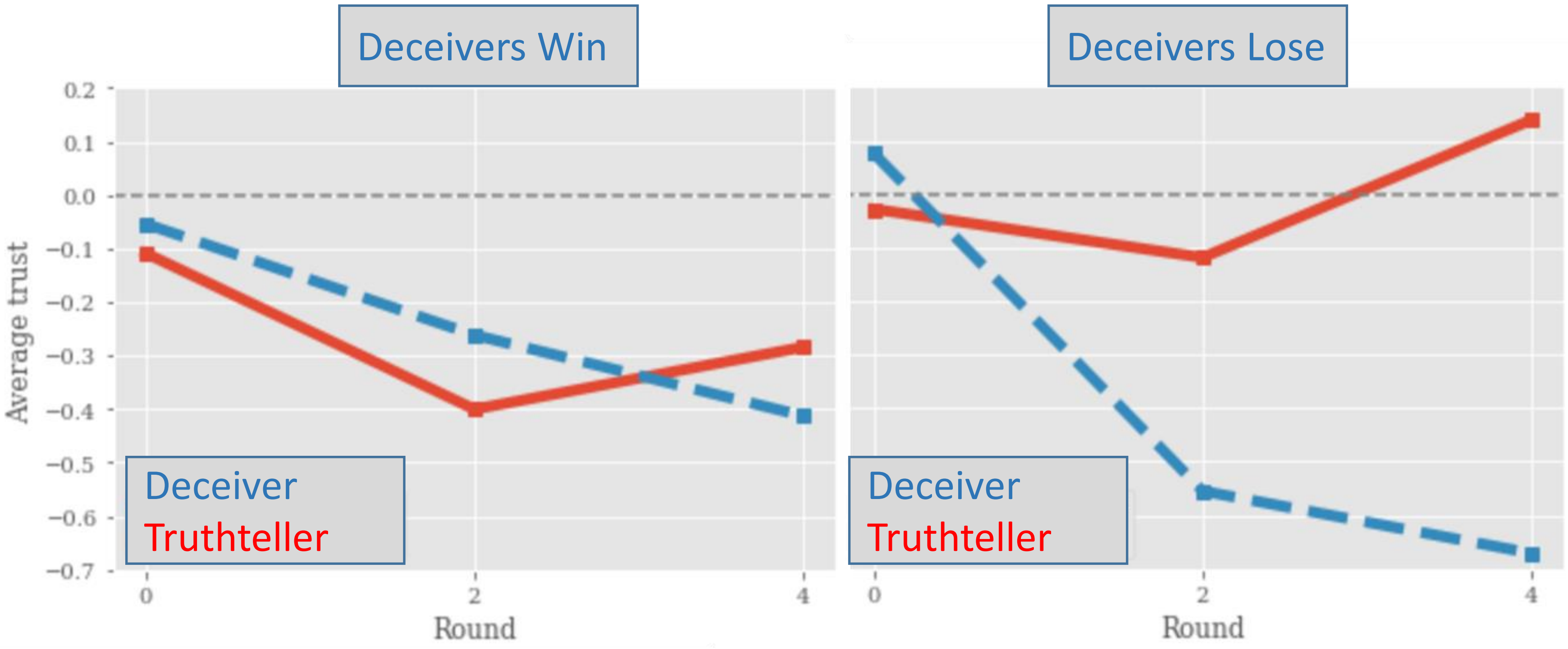


Accomplishment II: Discovering the Cues used by Humans to Detect Deception

- Eye blinks
- Stretched lips
- Deceivers are more nervous over time
- **Deceivers are less trusted over time**
- Deceivers are less dominant
- Interaction with other Deceivers

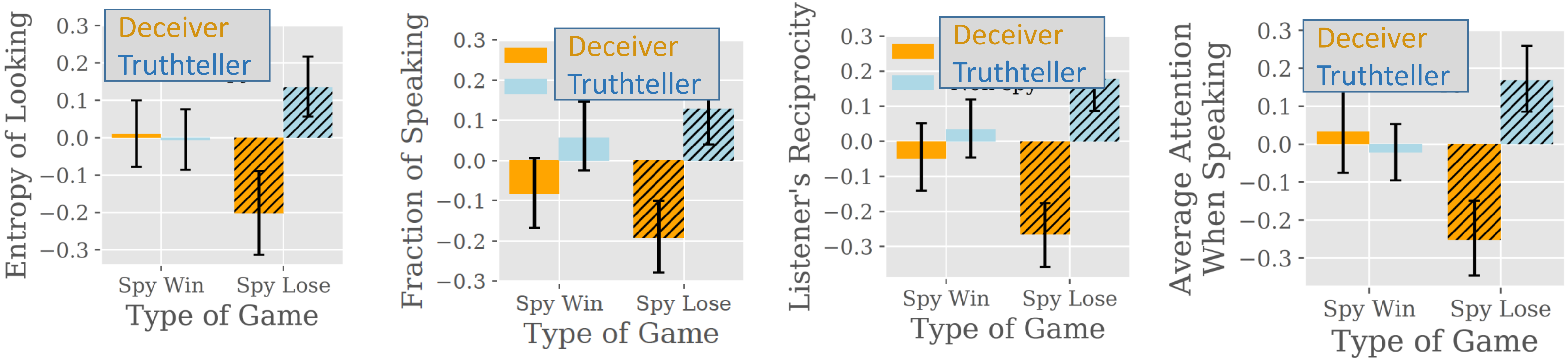


Accomplishment III: Deceivers are less trusted over time



Accomplishment IV: Signals of Deception from Looking and Speaking Networks

- Deceivers are indistinguishable from non-deceivers in games where deceivers win.
- Deceivers speak less, are not listened to, and get less attention in games where they lose.



Accomplishment V: Linear Regression Model for Deception Detection

- A combination of last round trust, second round trust, last round dominance and baseline dominance yield the best predictive results.
- Can identify truthtellers at 81% accuracy, truth tellers at 65% accuracy.

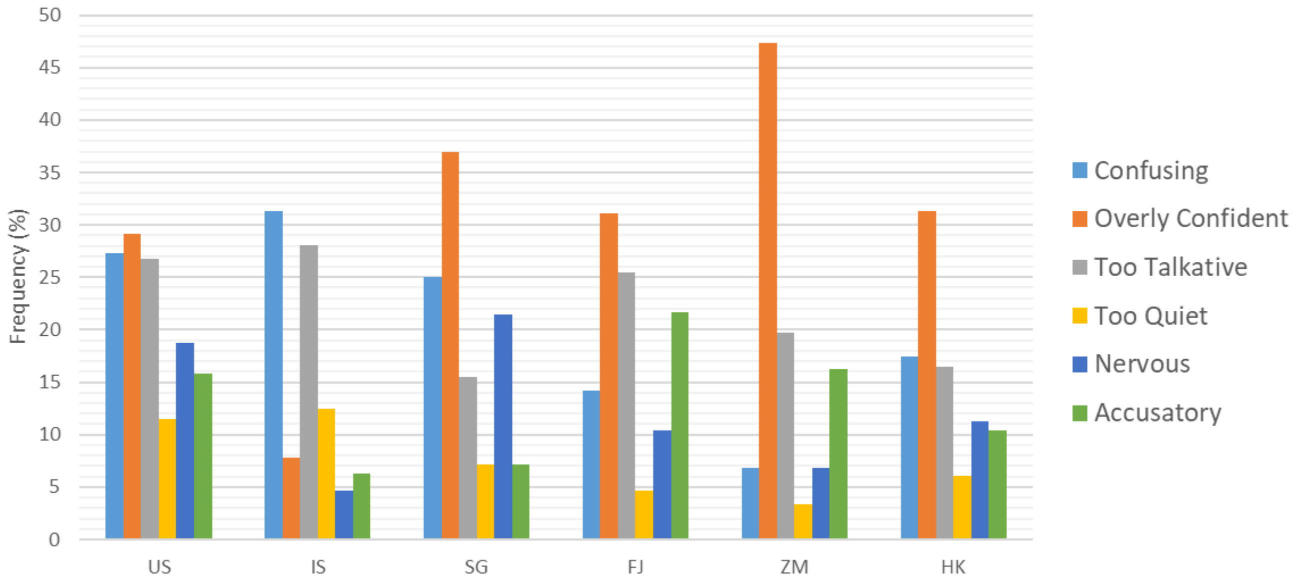
Discriminant Analysis of Relational Communication Dimensions as Discriminating between Deceivers and Truthteller, Test of Equality Between Means

	Wilks' Lambda	F	df1	df2	Sig.
Trust	.848	123.197	1	687	.000
Dominance	.964	25.505	1	687	.000
Arousal	.993	5.173	1	687	.023

Accomplishment VI: Discovering the Cues used by Humans to Detect Deception

- Near **universal distrust for overly confident** statements
    - Less in Israel (horizontal individualist)
    - Most strong in Zambia (horizontal collectivist)
- More problematic to be **too talkative vs. too quiet** universally
- **Nervousness is more problematic in vertical societies** (US, SG, FJ, HK) than in horizontal ones
- **Confusion was perceived as less problematic** in the most highly collectivist societies (ZM, FJ, HK)

Key takeaway – Culture seems to matter



Accomplishment VII: Culture and Accuracy in Deception Detection

Villagers' Deception Detection Accuracy Rates (Proportions) by Country

	Prop. of villagers winning game	Accuracy in detecting spies and villagers	True Positive rate (accurately detecting spies)	True Negative rate (accurately detecting villagers)
Singapore	.827	.730	.546	.840
Fiji	.424	.682	.413	.849
U.S.	.496	.676	.422	.846
Hong Kong	.803	.647	.295	.862
Zambia	.203	.636	.236	.866
Israel	.205	.620	.226	.867

Pearson correlations between cultural dimensions and deception detection accuracy

	Villagers winning game	Accuracy of detecting spies and villagers	True Positive rate (accurately detecting spies)	True Negative rate (accurately detecting villagers)
Horizontal Collectivism	.010	.064	.058	.041
Horizontal Individualism	.030	-.031	-.021	-.044
Vertical Collectivism	.038	**.100***	**.100***	.015
Vertical Individualism	**.136****	.092+	**.101***	.050
Negative Face	.046	.076	.093+	.009
Positive Face	-.017	.020	.020	-.015

Analyses controlled for prior game experience

Villagers win most in SG and HK:

- SG has highest overall accuracy and high true positive rate (best at detecting spies) but HK does not follow this same pattern
- True positive rate seems to be the most important factor in accuracy

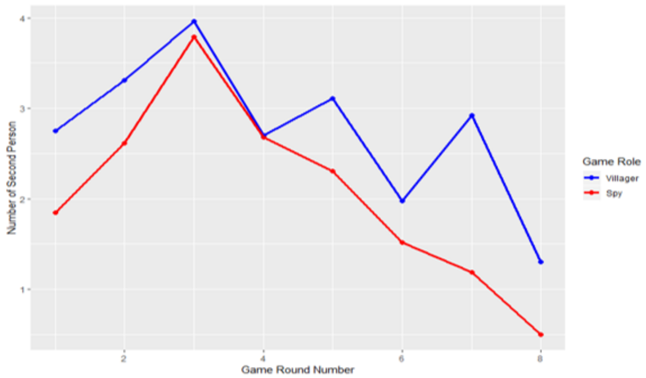
Culture matters in deception detection success:

- Cultural verticalism (competition and sacrifice for group) is associated with highest success in villagers' ability to accurately detect deception

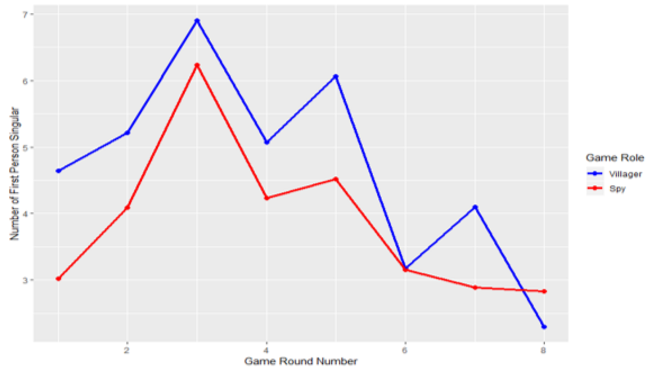
Culture seems to matter less than other factors => Need a deeper dive into this

Accomplishment VIII: Linguistic Analysis of Deceivers vs. Truth Tellers

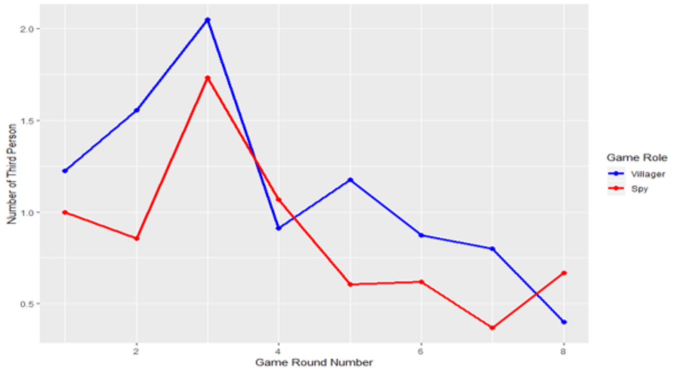
Number of Words



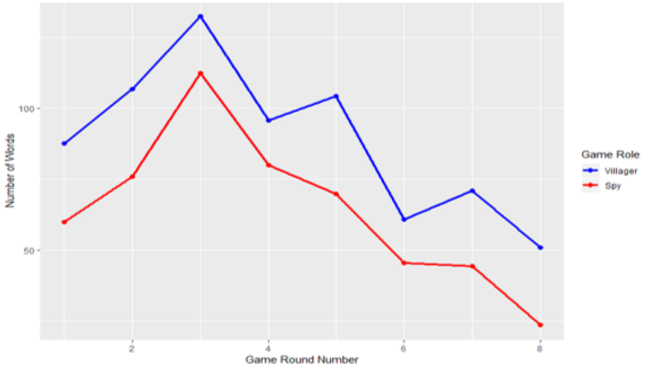
Number of First Person Singular Pronouns



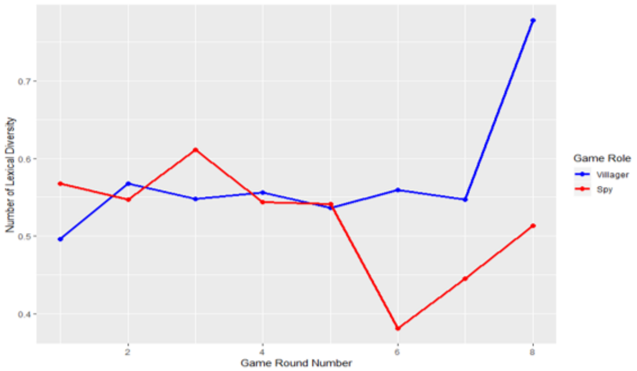
Number of Third Person Pronouns



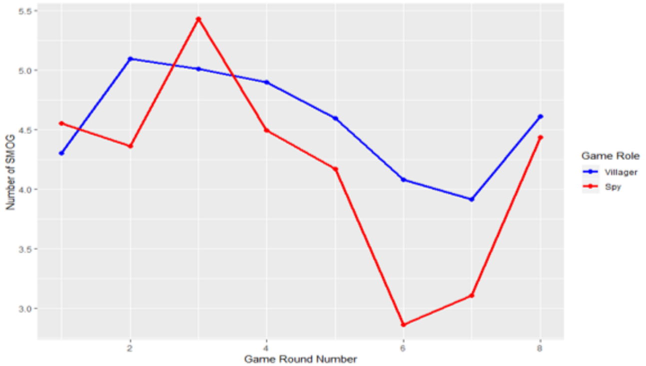
Number of Second Person Pronouns



Avg. Lexical Diversity

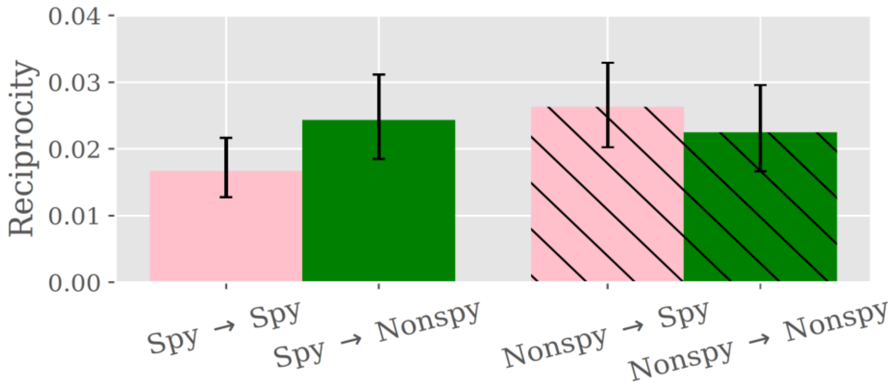


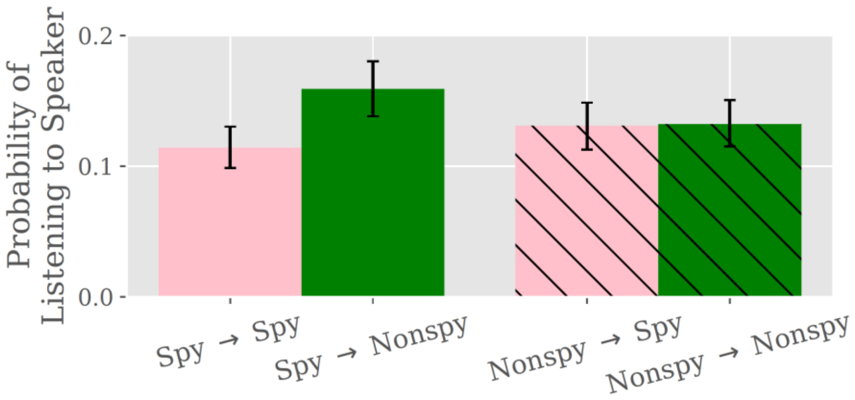
Avg. Comprehensibility (SMOG)



Accomplishment IX: Interactions between Deceivers and Non-Deceivers

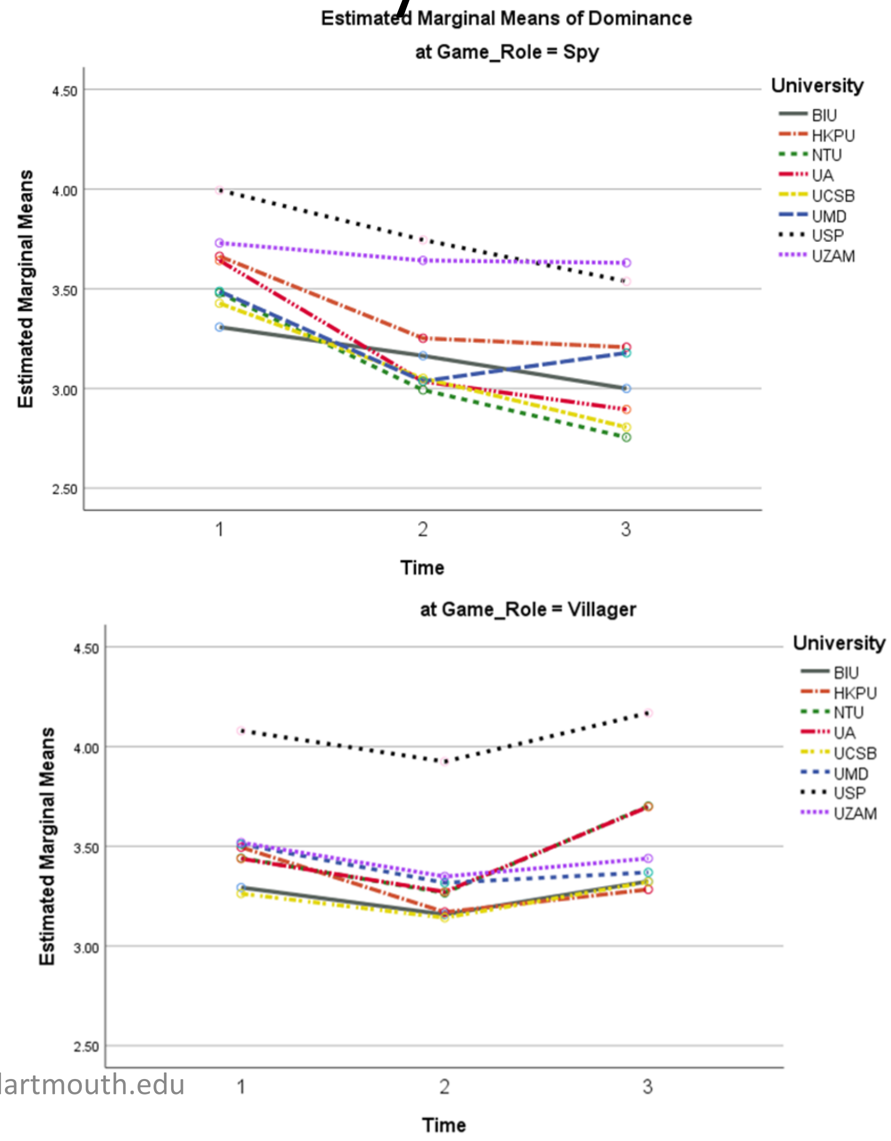
- Truth-tellers interact equally with everyone while deceivers interact more with truth-tellers.





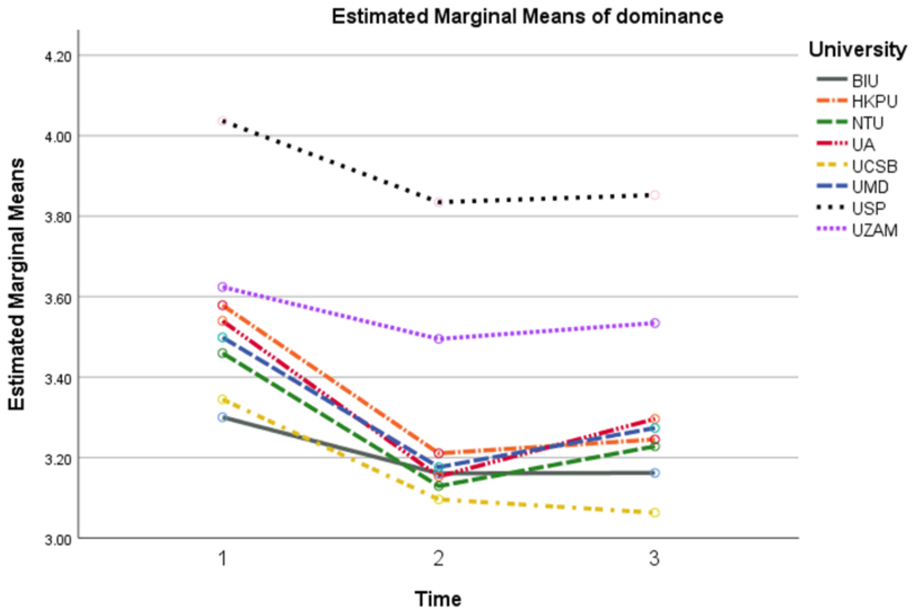
Accomplishment X:Features Used by Humans to Detect Dominance

- Significant features
  - Mean pitch in final round
  - Variance of loudness
  - Mean voice quality
  - Variance in voice quality (harmonics to noise ratio)
  - Utterance length in words
- Deceivers diminish in dominance over time.
- But dominance and deception appear to be more culture sensitive – deceivers in Fiji and Zambia are more dominant. *Needs further investigation.*



Accomplishment XI: Effect of Culture on Dominance

- Overall, dominance did not differ by location
- But, two locations stand out as different from the rest:
  - Fiji
  - Zambia
- Seen as more dominant than other locations



Talk Outline

Overview of the SCAN Project

- How Humans Detect Deception and Dominance
- **How AI Algorithms Detect Deception and Dominance**
- Other Major Contributions

Deception Detection

- Deception in Real-world Courtroom Videos
- Deception in Multi-Player Face to Face Games

How AI Algorithms Detect Deception and Dominance

These results show how the design of novel, state of the art AI algorithms to predict a host of factors linked to deception and dominance on the SCAN dataset.

Accomplishment XII: Predicting Deception in Groups, 1st Attempt

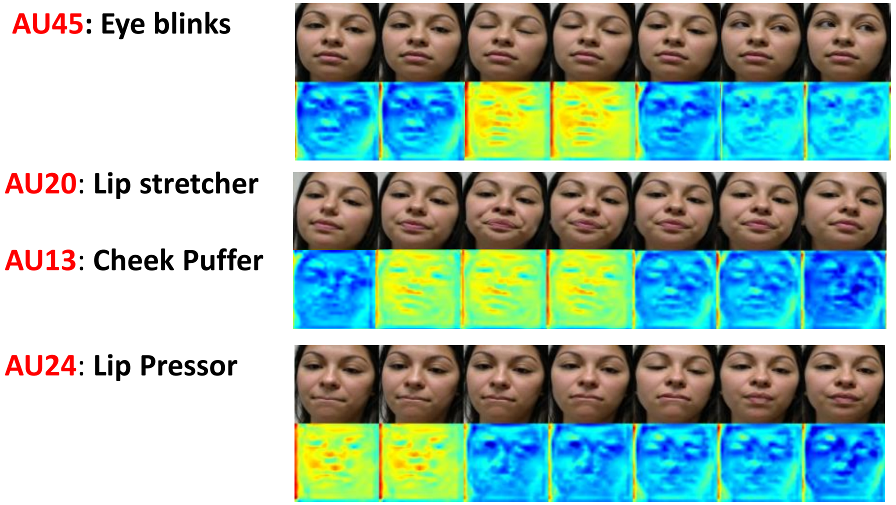
- Prior work by us (AAAI 2018) was able to predict deception in court-room settings with AUC of 0.877.
- But long term deception in a much more free environment is harder to detect
- A fully automated system (LiarOrNot) for predicting long- term deception in videos
- A new class of histogram-based features
- A novel “meta-feature” called LiarRank that builds on the basic features
- An ensemble based prediction model
- Achieves an **AUC of 0.705** in predicting the role of a player in the game
- AUC for prediction by humans is 0.583

Accomplishment XIII: Predicting Deception in Groups, Attention-Based Facial Behavior Analytics

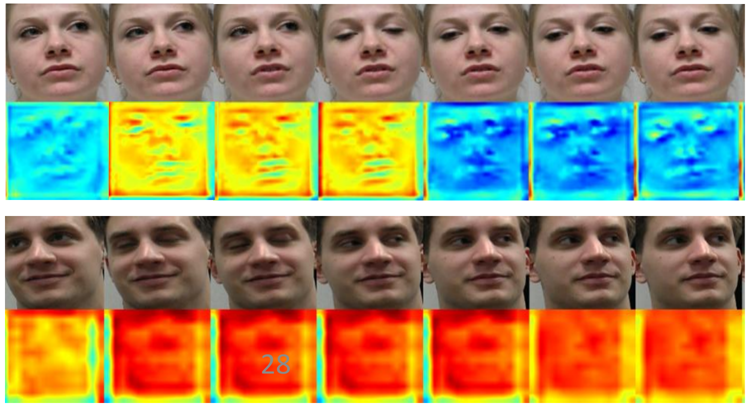
- Attention technique discovers the important **spatial and temporal** information on the face for deceiver/truth-teller detection
- Quantitative results liar vs. truth-teller: model trained with attention-based sampling (giving more weight to the video data with higher attention probabilities) **achieves ~4% higher accuracy** than conventional training
- Qualitative results on the fact that our attention NN is capable of discovering cues for deceivers, which are related to what is known from communication theory for deception.

Accomplishment XIV: Predicting Deception in Groups, Attention-Based Facial Behavior Analytics

- We show that players exhibiting some Facial Action Units (AUs:13,20,24,45 are more likely to be classified as deceivers.
- According to the communication theory:
  - AUs 20 and 45 are related to deception, which is consistent to our expectation that deceivers are more willing to lie, but not always.
  - AU 20 = stretched lips
  - AU 45 = eye blinks
- Our approach can detect small facial movements related to deception like **eye blinking** in the top row, and detect the **fake smile** (bottom) so as to correctly classify the type of player's role.

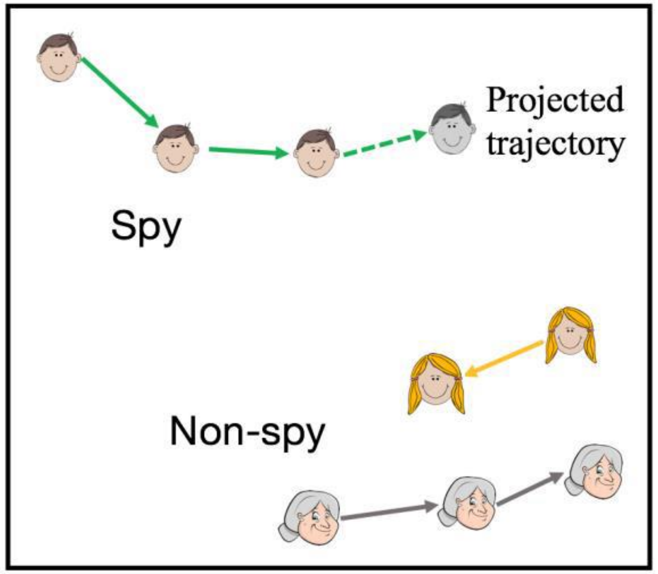


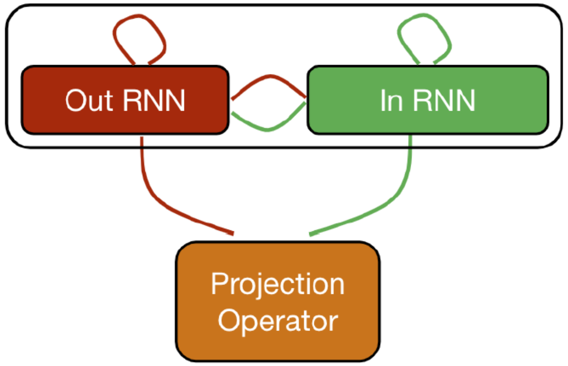
(a)


(b)

Accomplishment XV: Predicting Deception with Graph Convolution Models

- To predict deception, we used interaction networks to train
  - A Temporal Graph Convolutional Network model,
  - a Belief Propagation Model (on the negative network),
  - A Deep Temporal Model that uses Dynamic Embeddings
- •Tested and evaluated all models on deception prediction in the context of the SCAN game.
- •Current AUC is 0.73 using one minute of video



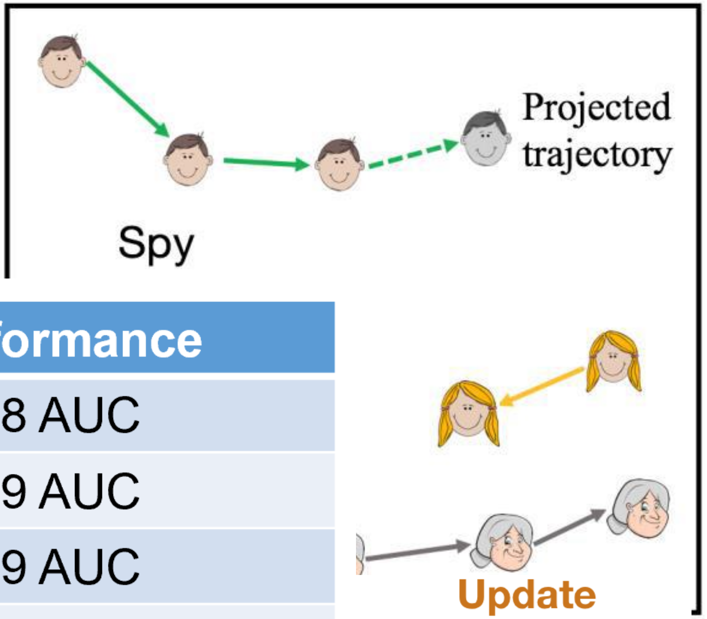


Update Component: updates the embeddings after each interaction

Project Component: generates future embeddings to make future predictions

Accomplishment XV: Predicting Deception with Graph Convolution Model

- To predict deception, we used interaction networks to train
  - A Tei Netv
  - a Bel nega
  - A De Dyna
- •Testec on dece context
- •Currer minute



Baselines

Our new models

Method	Performance
Emotion	0.538 AUC
Head and eye movement	0.549 AUC
Facial action unit	0.569 AUC
Late fusion	0.587 AUC
Graph convolution network model	0.596 AUC
Belief propagation on negative network	**0.73 AUC**

Update Component: updates the embeddings after each interaction

Operator

Project Component: generates future embeddings to make future predictions

Accomplishment XVI: Predicting the Most Dominant Person in a Group

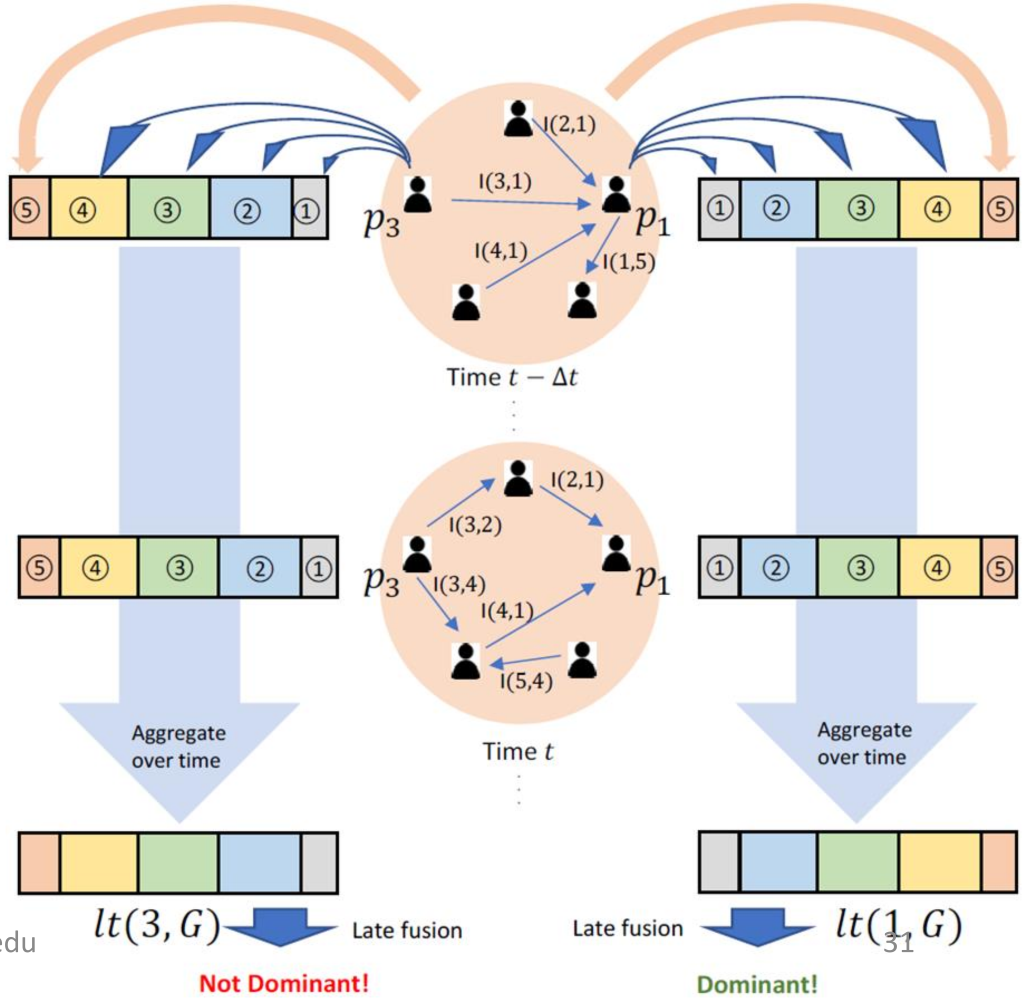
Features:
1. Speaking probability
2. Facial Action Units
3. Emotions
4. Audio features (MFCC)
5. Dominance Rank feature (new!)

Aggregation:
1. Fisher Vector
2. Histograms

Ensemble:

$$S = \sum_{i=1}^{5} \alpha_i S_i ,$$

where S_i are scores for individual feature types.



Accomplishment XVI: Predicting the Most Dominant Person in a Group: The DELF algorithms

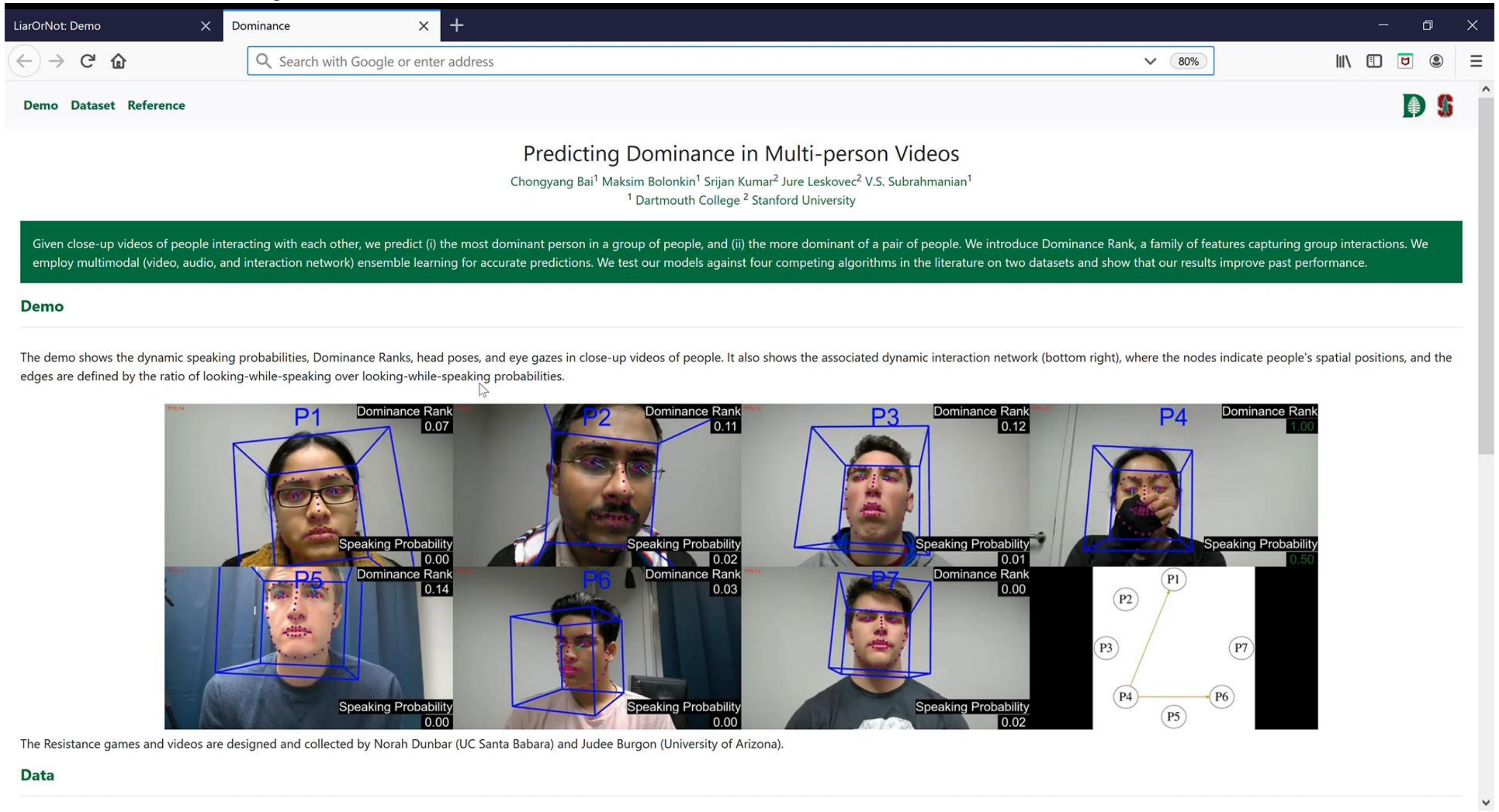
	MPD-All			MDP-Distinct			PDP-All			PDP-Distinct		
Features	**AUC**	**FPR**	**Acc.**	**AUC**	**FPR**	**Acc.**	**AUC**	**FPR**	**Acc.**	**AUC**	**FPR**	**Acc.**
DELF	**0.791**	0.027	0.769	**0.894**	0.021	0.889	**0.874**	0.281	0.792	**0.949**	0.189	0.876
DR (LS/LL, 1 sec) + FV	0.754	0.056	0.761	0.855	0.017	0.89	0.77	0.281	0.694	0.832	0.235	0.741
DR (LS/LL, 1 sec) + Hist.	0.754	0.252	0.711	0.836	0.209	0.868	0.788	0.314	0.724	0.861	0.392	0.768
DR (LS/LL, 5 sec) + FV	**0.773**	0.064	0.761	**0.861**	0.167	0.868	0.771	0.328	0.695	0.835	0.28	0.74
DR (LS/LL, 5 sec) + Hist.	0.770	0.252	0.720	0.844	0.179	0.879	0.793	0.441	0.709	0.861	0.347	0.788
Speaking + FV	0.741	0.279	0.689	0.838	0.030	0.875	**0.853**	0.261	0.762	**0.92**	0.179	0.825
Speaking + Hist.	0.756	0.066	0.770	0.821	0.150	0.879	0.847	0.258	0.778	0.91	0.164	0.860
Baseline (speak.)	0.738	0.103	0.730	0.769	0.200	0.879	0.800	0.274	0.738	0.893	0.198	0.845
Baseline (comb.)	0.767	0.252	0.716	0.764	0.214	0.879	0.828	0.290	0.759	0.906	0.168	0.863

Also predicting the more dominant person in a group of two people

Accomplishment XVIII: Predicting the Most Dominant Person in a Group: The GDP algorithms

Feature	Classif.	AUC	FPR	Acc.
MDP-All				
Speaking + FV	MLP	0.809	0.219	0.745
Speaking + FV	RF	**0.817**	0.133	0.770
DR (LS/LL, 5sec) + FV	MLP	0.783	0.222	0.733
DR (LS/LL, 5sec) + Hist.	MLP	0.772	0.157	0.746
MDP-Distinct				
Speaking + FV	MLP	**0.936**	0.048	0.917
Speaking + FV	RF	0.902	0.088	0.849
DR (LS/LL, 5sec) + FV	RF	0.878	0.071	0.878
DR (LS/LL, 5sec) + FV	MLP	0.850	0.065	0.889

Accomplishment XVII: DELF/GDP Dominance Prediction System



Accomplishment XVIII: Predicting Dominance on Related Datasets, Cooperative Environment

Key question: Does Dominance Rank work for datasets that already exist with a similar goal of predicting dominance?

• Swiss group developed the ELEA dataset in which participants were assigned a winter survival task and were asked to elect a leader.

• Difference with SCAN dataset: task is cooperative, everyone wants to survive.

• **Dominance Rank based Features yielded the best results.**

[Okada *et al.*, 2018]	58.82	64.71
[Aran and Gatica-Perez, 2013]	65.69	59.80
[Okada *et al.*, 2015]	67.65	68.63
DR (LS/LL) + FV (ours)	**76.47**	67.65
DR (LS/LL) + Hist. (ours)	74.51	**71.57**
Human scores	68.63	—

Predicting Pairwise Dominance

Dominance Rank based Features Outperform Humans

Accomplishment XIX: Key Factors Linked to Dominance Prediction

- Used ablation testing to identify which features' exclusion led to the greatest drop in AUC.
- Dominance Rank Features dominate for Most Dominant Player Prediction
- Audio Features dominate for Pairwise Dominance Prediction
- FAU features AU15, AU20, AU25 all significant
  - AU 15 = lip corner depressor
  - AU20 = lip stretcher
  - AU 25 = lips parted

Excluded Feature	AUC
MDP-All	
All features present	0.790
FAU (AU15, AU20, AU25)	0.790
MFCC	0.775
DR (LS/LL, 5sec) + FV	**0.757**
Emotions (Angry, Surprised, Calm)	0.772
Speaking+Hist.	0.775
MDP-Distinct	
All features present	0.894
FAU (AU05, AU14, AU20)	0.888
MFCC	0.890
DR (LS/LL, 5sec) + FV	**0.849**
Emotions (Angry, Confused)	0.891
Speaking+FV	0.884
PDP-All	
All features present	0.874
FAU (AU15, AU20, AU25)	0.824
MFCC	0.867
DR (LS/LL, 5sec) + Hist.	0.866
Emotions (Smile, Angry, Surprised)	0.866
Speaking+ FV	**0.816**
PDP-Distinct	
All features present	0.949
FAU (AU14, AU15, AU25)	0.948
MFCC	**0.921**
DR (LS/LL, 1sec) + Hist.	0.934
Emotions (Happy, Angry, Calm)	0.945
Speaking + FV	0.949

Accomplishment XIX: Key Factors Linked to Dominance Prediction

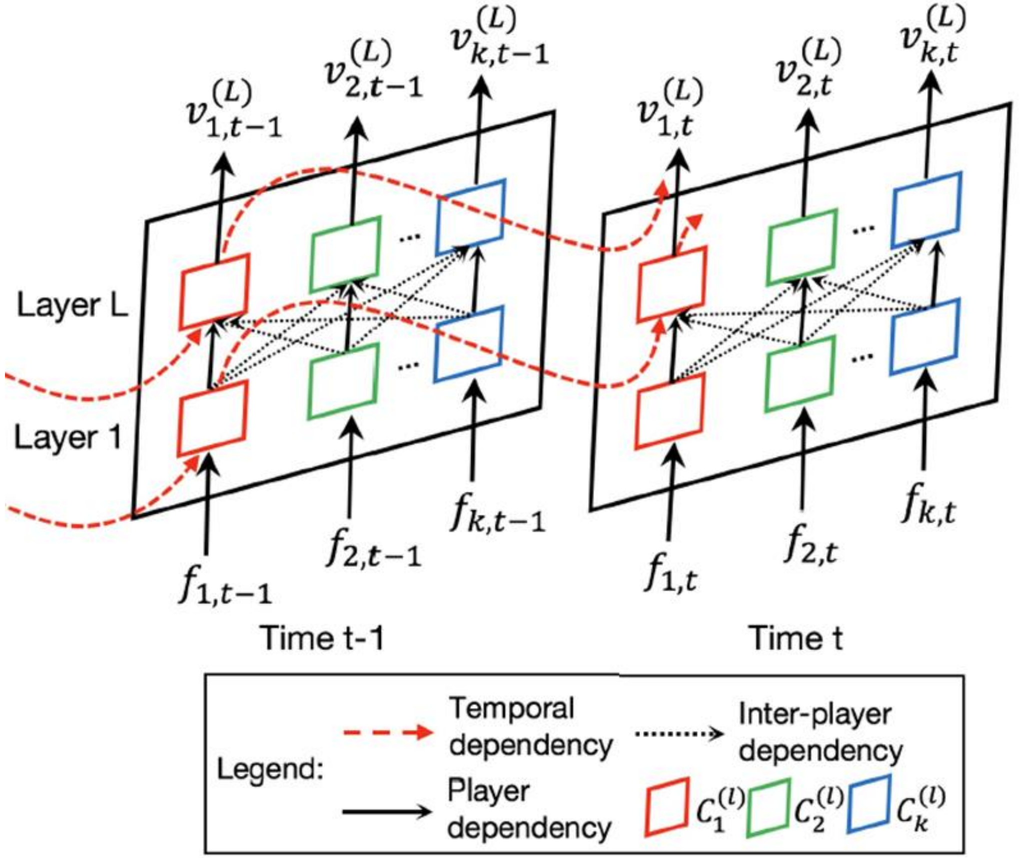
FAU features AU15, AU20, AU25 all significant

- Lip corner depressor
- Lip stretcher
- Lips part

Sample FAU images from: https://www.cs.cmu.edu/~face/facs.htm

Accomplishment XX: Predicting Who is Looking at Who

- Raw features at time *t-1* or *t* are at the bottom
- Novel collective classification algorithm used at each time point to capture player-player dependencies.
- Novel temporal dependency metric used to capture dependency on solution at time *t-1* to predict solution at time *t*



Talk Outline

Overview of the SCAN Project

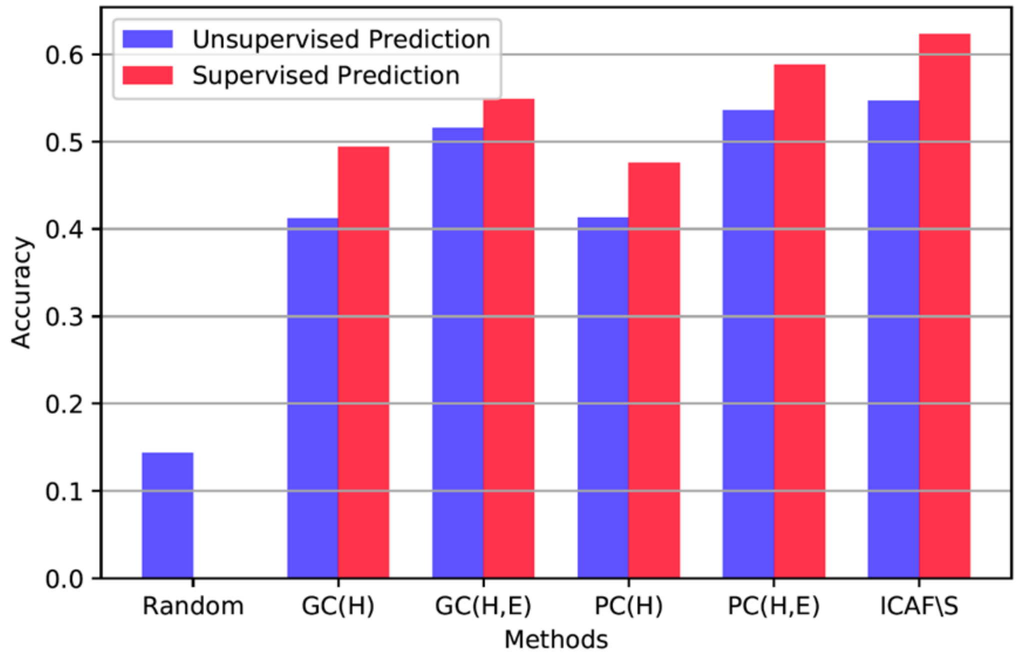
- How Humans Detect Deception and Dominance
- How AI Algorithms Detect Deception and Dominance
- **Other Major Contributions**

Deception Detection

- Deception in Real-world Courtroom Videos
- Deception in Multi-Player Face to Face Games

Accomplishment XX: Predicting Who is Looking at Who

- Developed ICAF (Iterative Collective Attention Focus) algorithm and system
- Predictive accuracy is over 60% for the best algorithm compared to a baseline of 11-16% for random guessing.



Accomplishment XXI:
Building Out Who is Looking at Who Network

- Developed ICAF (Iterative Collective Attention Focus) algorithm and system
- Predictive accuracy is over 60% for the best algorithm compared to a baseline of 11-16% for random guessing.
- ICAF automatically generates networks! For each game
  - Weighted network measures the probability score of looking at another player
  - Binary network has edges with the highest probability of looking at another player

Dataset statistics	
Number of networks	62
Number of nodes	451
Number of edges	3,126,993
Average number of edges per network	50,435
Total temporal length	142,005 seconds
Average temporal length per network	2,290 seconds

Accomplishment XXI: System to Construct SCAN Network

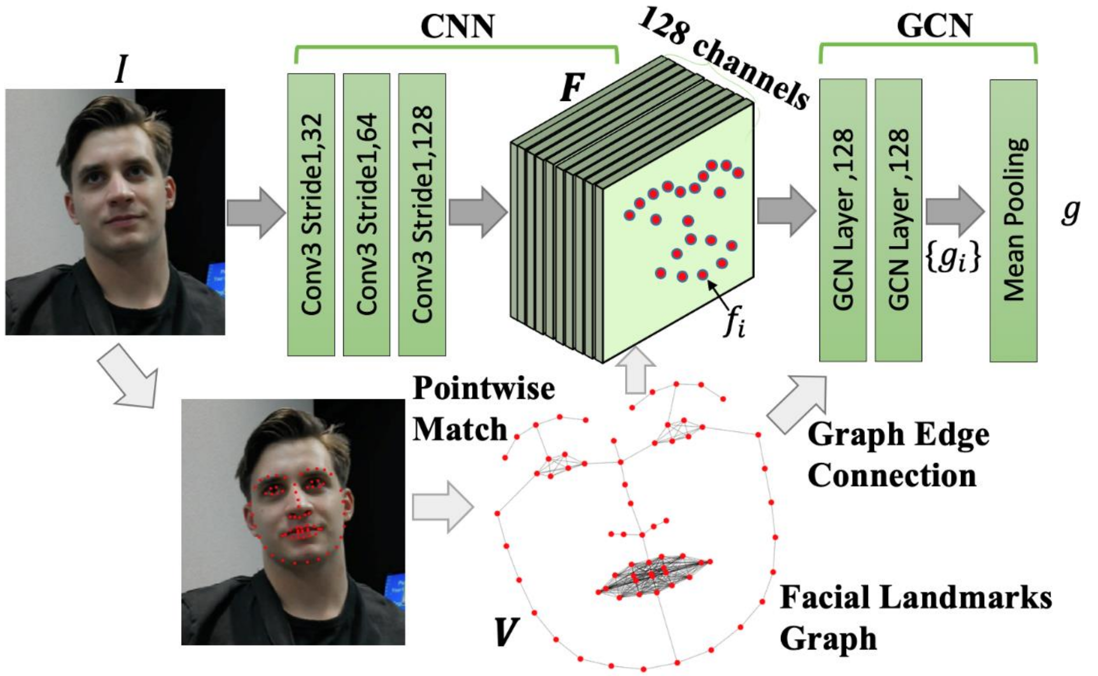
PLAYED VIA A SEPARATE FILE

Accomplishment XXII: Relative Nervousness Prediction

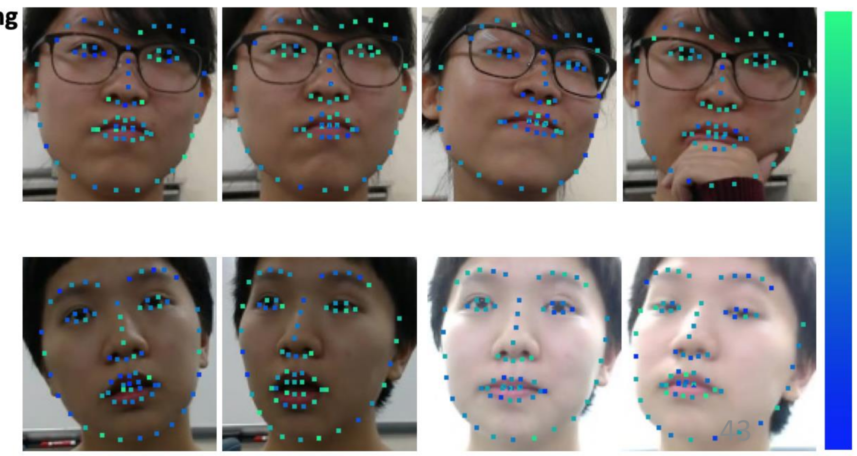
- Tasks considered:
  - Pairwise Nervousness Prediction (PNP)
  - PNP-Distinct
  - Nervousness Change Prediction (NCP)
- Combine positive/negative emotions toward speaker and relative dominance of speaker with listeners to generate nervousness scores.
- Audio and Visual Nervousness Scores

$$NS_t\,(v) = \alpha NS_{pos,t}(v) + (1-\alpha) NS_{neg\ t}(v)$$

- Facial Emotion-oriented Graph Convolutional Network (FE-GCN)



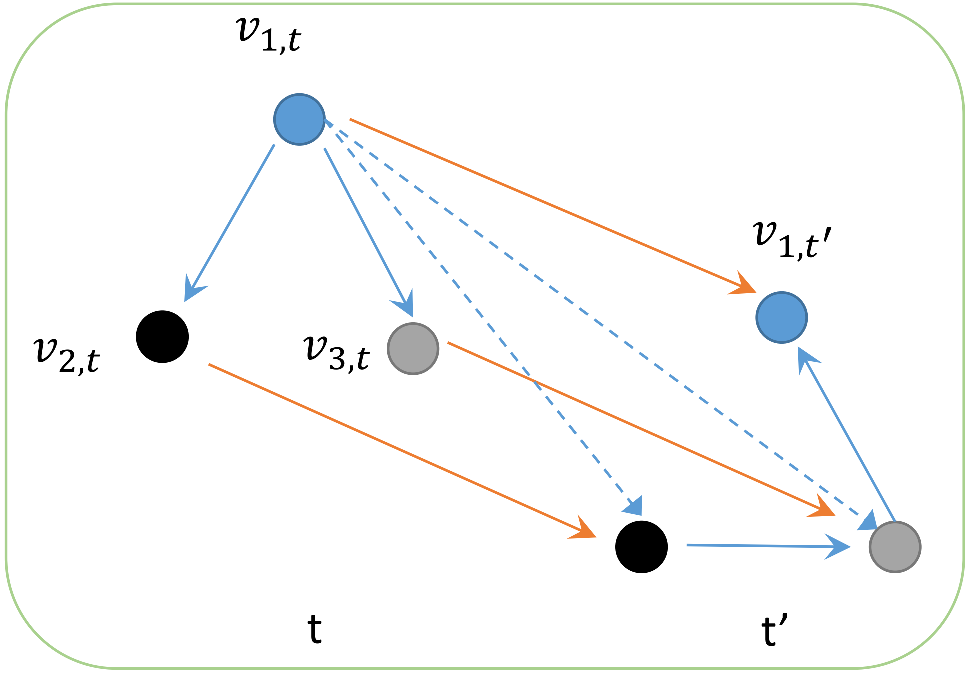
	The Resistance PNP	The Resistance PNP-Distinct	The Resistance NCP	ELEA PNP
ANS	0.635	0.723	**0.724**	0.623
VNS	0.668	**0.765**	0.667	0.760
FE-GCN	**0.681**	0.744	0.634	**0.802**



Accomplishment XXIII: Like/Dislike Prediction via Graph Convolutional Networks (GCNs)

- Vertices: Players in each frame $v_{i,t}$
- Edges
  - (1) Interaction (e.g. look-at)
    - $(v_{i,t}, v_{j,t}) \in E_1$
    - $c(v_{i,t}, v_{j,t})$ is the interaction probability
  - (2) Identification
    - $(v_{i,t}, v_{i,t'}) \in E_2,\ t' - t \le \Gamma$
    - $c(v_{i,t}, v_{i,t'}) = \gamma^{t'-t},\ 0 < \gamma < 1.$
  - (3) Delayed Influence
    - $(v_{i,t}, v_{j,t'}) \in E_3,$ $iff.\ (v_{i,t}, v_{j,t}) \in E_1, (v_{i,t}, v_{i,t'}) \in E_2, i \ne j$
    - $c(v_{i,t}, v_{j,t'}) = c(v_{i,t}, v_{j,t})\, c(v_{i,t}, v_{i,t'})$
- Train the node embeddings for Like/Dislike prediction:

Task	Interaction	AUC	F1
Cold or Warm	Talk To	0.649	0.397
	Listen To	0.615	0.362
	Look At	0.635	0.358
Negative or Positive	Talk To	0.611	0.354
	Listen To	0.638	0.367
	Look At	0.574	0.320



GCN block for node embedding update:

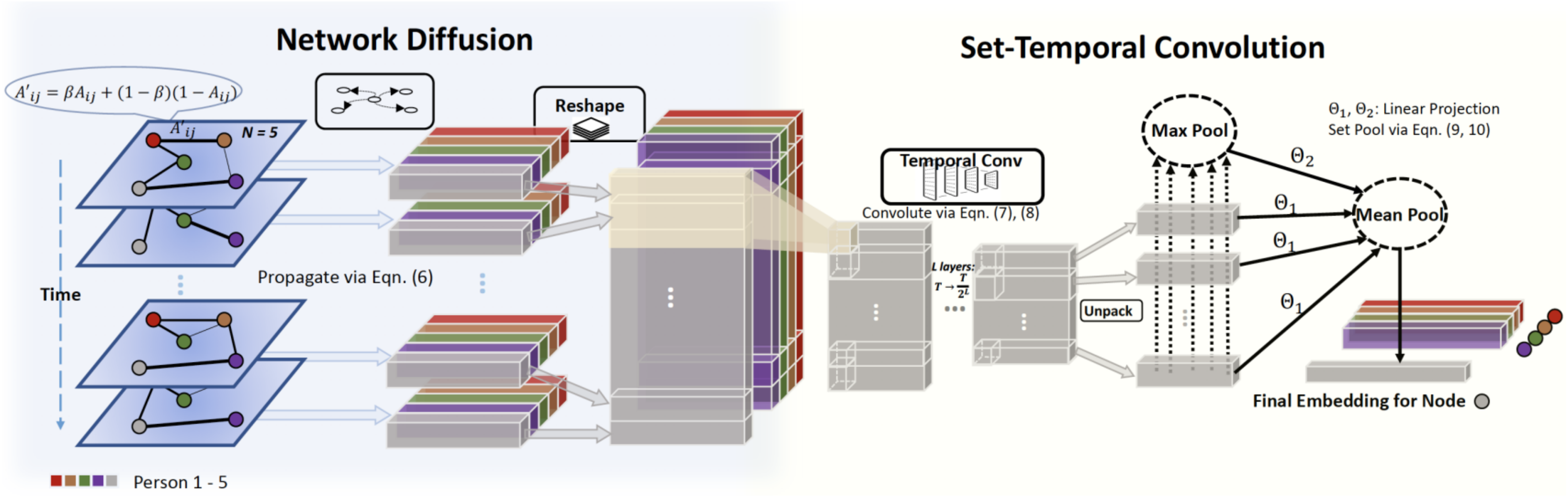
$$\tilde{v} = \sum_{k=1}^{3}\Big(\sum_{u \in IN_k(v)} c(u,v)\boldsymbol{w_k}(u,v)\boldsymbol{f_k}(v) + \sum_{u \in OUT_k(v)} c(v,u)\boldsymbol{w_k}(v,u)\boldsymbol{f_k}(v)\Big)$$

$$\boldsymbol{w_k}(u,v) = \boldsymbol{attn}(\boldsymbol{f_k}(u), \boldsymbol{f_k}(v))$$

k goes over edge types, f_k is a learnable layer,
w_k is the learned edge weight from the attention from v to u

Accomplishment XXIV: Representation Learning Framework for Dynamic Social Interaction Networks

- Temporal Network-Diffusion Convolution Networks (TN-DCN)
  - Network Diffusion
    - Weighted combination of both network (interaction) and complement network (avoid interaction)
    - Multi-hop diffusion for node features
  - Set-Temporal Convolution
    - Aggregate the node features over time via 1D convolutions
    - Max-pooling and mean-pooling over time to get the final embedding for each node.
    - The node embeddings can be used to learn various tasks

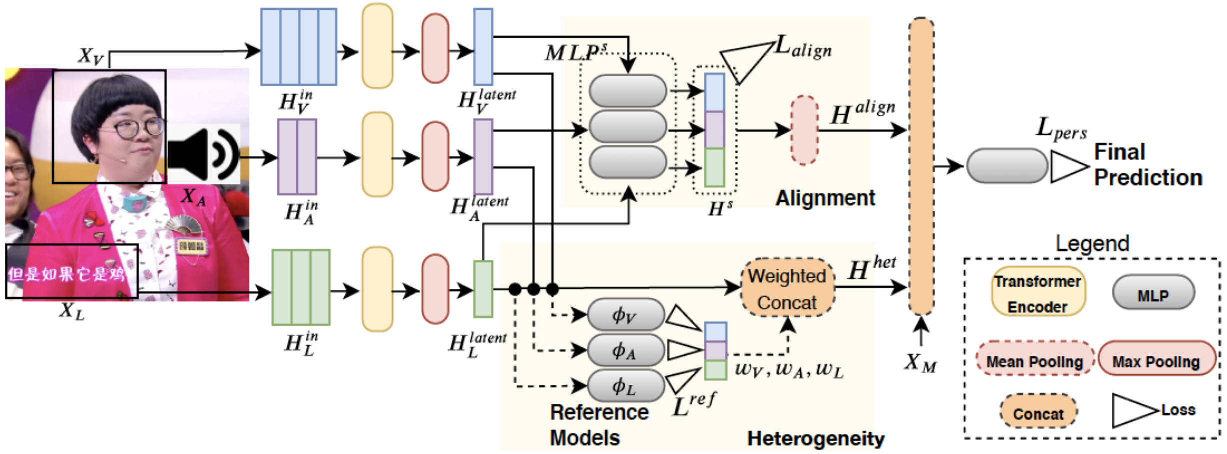


Comparison of performance on on RESISTANCE (first three) and CIAW(last one)

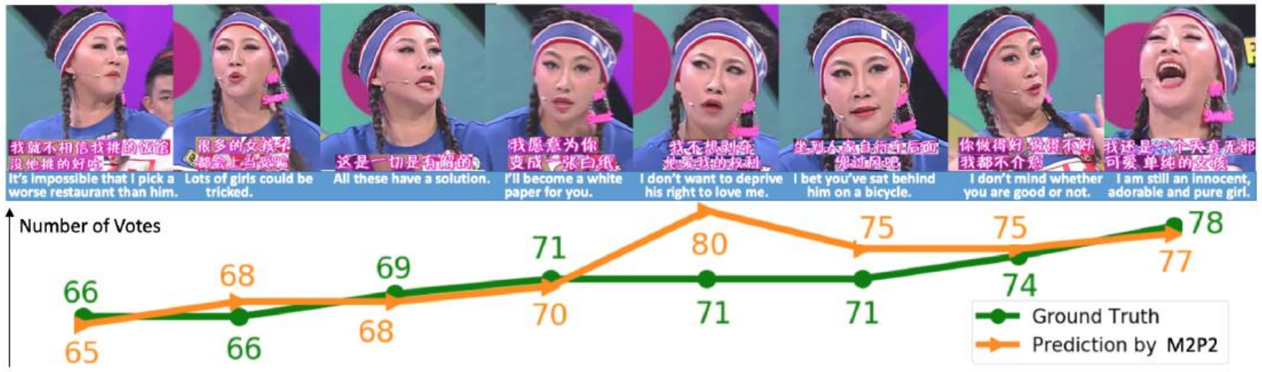
	Dominance Identification		Deception Detection		Nervousness Detection		Community Detection	
	Method	Performance	Method	Performance*	Method	Performance	Method	Perform.
Baselines	MKL [6]	0.879	FAU [12]	0.608	LR.	0.493	WD-GCN [27]	0.813
	DELF [4]	0.889	TGCN-L†[26]	0.550	RF.	0.678	CD-GCN [27]	0.819
	GDP-MLP [4]	0.917	LiarOrNot [3]	0.665	GCN-LSTM [39]	0.702	GCN-LSTM [39]	0.601
	GDP-RF [4]	0.878	ADD [46]	0.632	Facial Cues [16]	0.733	EvolveGCN[32]	0.893
Ours	-	**0.923** (±0.009)	-	**0.689** (±0.021)	-	**0.769** (±0.023)	-	**0.929** (±0.011)

Accomplishment XXV: Multimodal Persuasion Prediction

- **M2P2 architecture.**
  - Audio, face and language sequences are extracted from a video clip and fed to extract primary input embeddings X.
  - Each of embeddings is fed to a Transformer encoder and max pooling to the latent embeddings H^{latent}.
  - The latent embeddings are fed to the **alignment** and **heterogeneity** modules to generate the embeddings H^{align} and H^{het}.
  - Concatenate H^{align} and H^{het} and the debate meta-data X_M, and feed to an MLP for persuasiveness prediction.
  - H^{latent} interact with two procedures alternately:
    - Optimize the alignment loss L_{align} and persuasiveness loss L_{pers}
    - Learn weights through 3 reference models ϕ.



- **Real-time prediction of debate persuasiveness** using M2P2. The debate is from a Chinese debate TV show, Qipashuo. M2P2 closely predicts the ground truth number of votes.



- **Experiments on two tasks**
  - Debate outcome prediction (DOP)
  - Intensity of persuasion prediction (IPP)

Methods	Acc. on DOP	MSE on IPP
Brilman et al.[1]	0.614	0.016
Nojavanasghari et al.[2]	0.615	0.016
Santos et al. [3]	0.598	0.02
M2P2	**0.635**	**0.012**

Talk Outline

Overview of the SCAN Project

- How Humans Detect Deception and Dominance
- How AI Algorithms Detect Deception and Dominance
- Other Major Contributions

Deception Detection

- Deception in Real-world Courtroom Videos
- Deception in Multi-Player Face to Face Games

Goal

- **Deception Detection**: Detect whether the subject in a video is deceptive or truthful.

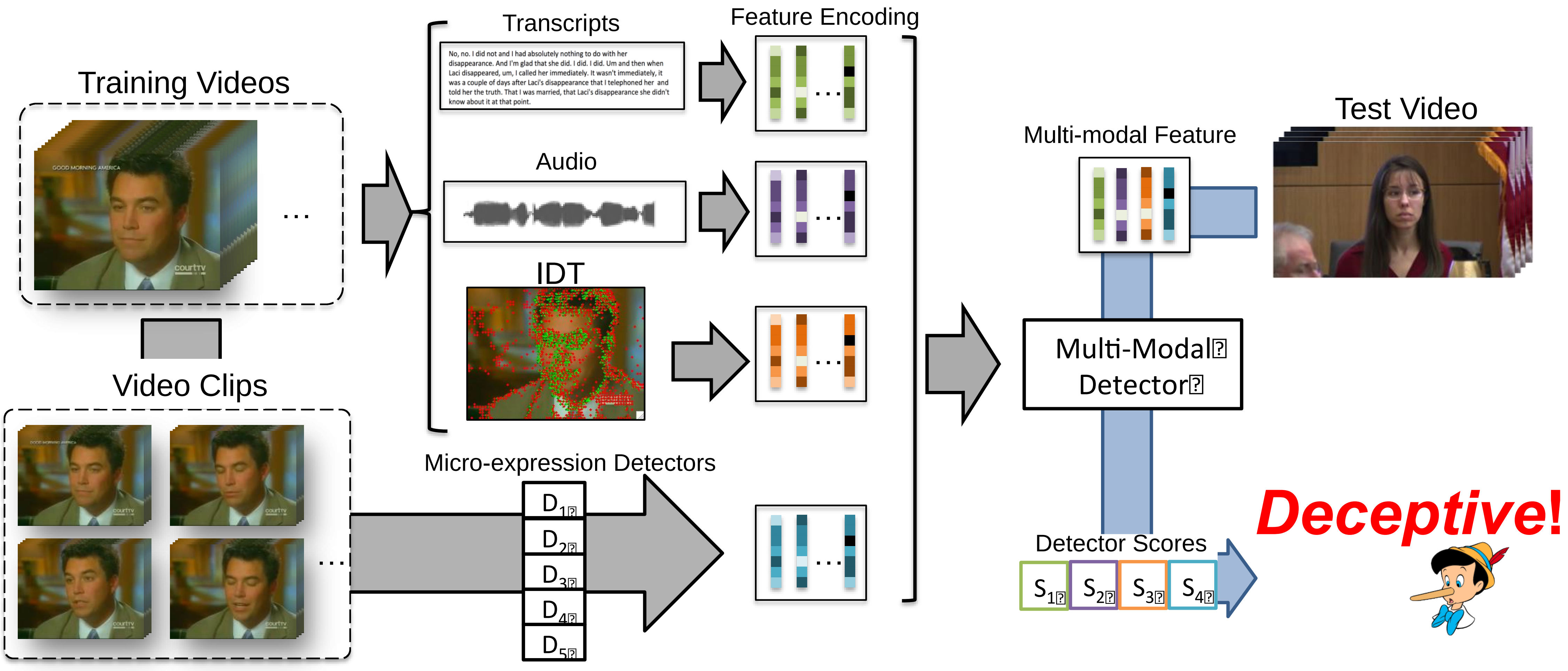

- We evaluate our method on a dataset of real-life court room videos [1] in a high-stakes situation.
- Our automated multi-modal system considers visual, audio and verbal modalities.

[1] Pe´rez-Rosas, V.; Abouelenien, M.; Mihalcea, R.; and Burzo, M. 2015. Deception detection using real-life trial data. In ICMI 2015.

Main Contributions

- We study different modalities, e.g. visual, audio and verbal, for detecting deceptive behavior.
- We show the effectiveness of visual features, including low-level motion features and high-level feature prediction scores of micro-expressions, and audio features, e.g. MFCC.
- Our system outperforms the start-of-the-art method that used human annotation as features by 5%. When combined with human annotations of micro-expressions, our AUC improves to 0.922.
- A user-study to analyze how well do average humans perform on this task, what modalities they use for deception detection and how they perform if only one modality is accessible. We show that our automated system is better than average humans.

Framework


Training Videos
Video Clips
Transcripts
No, no. I did not and I had absolutely nothing to do with her disappearance. And I'm glad that she did. I did. I did. Um and then when Laci disappeared, um, I called her immediately. It wasn't immediately, it was a couple of days after Laci's disappearance that I telephoned her and told her the truth. That I was married, that Laci's disappearance she didn't know about it at that point.
Audio
IDT
Feature Encoding
Micro-expression Detectors
D_1
D_2
D_3
D_4
D_5
Multi-modal Feature
Test Video
Multi-Modal Detector
Detector Scores
S_1 S_2 S_3 S_4
Deceptive!

Approach

- **Visual Features** include low-level motion features and high-level micro-expression features.
    - Low-level motion feature
    - High-level motion feature
- **Audio features:** Mel-frequency Cepstral Coefficients (MFCC) which ARE popular in speech recognition.
- **Verbal features:** Glove, Global Vectors for Word Representation.
- **Feature Encoding:** Fisher Vector Encoding, which aggregates a variable number of features to a fixed-length vector.
- **Deception Detection** is performed via late fusion of deception scores from all different modalities.

Micro-Expressions

- We investigate 5 micro-expressions that are reported to be most effective among all micro-expressions in prior art


1) Frown


2) Eyebrows Raise


3) Lips Corner Up


4) Lips Protruded


5) Head side turn

Experiments

- We evaluate four individual features, as well as their different combinations, using several classifiers to test the robustness

Features	L-SVM	K-SVM	NB	DT	RF	LR	Adaboost
IDT	0.7731	0.6374	0.5984	0.5895	0.5567	0.6425	0.6591
MicroExpression	0.7502	0.7540	0.7629	0.7269	0.8064	0.7398	0.7507
Transcript	0.6457	0.4667	0.6625	0.5251	0.6172	0.5643	0.6416
MFCC	0.7694	0.8171	0.6726	0.4369	0.7393	0.6683	0.6900
IDT+MicroExpression	0.8347	0.7540	0.7629	0.7687	0.8184	0.7419	0.7507
IDT+MicroExpression+Transcripts	0.8347	0.7540	0.7776	0.7777	0.8184	0.7419	0.7507
IDT+MicroExpression+MFCC	0.8596	0.8233	0.7629	0.7687	0.8477	0.7894	0.7899
All Modalities	**0.8773**	0.8233	0.7776	0.7777	0.8477	0.7894	0.7899

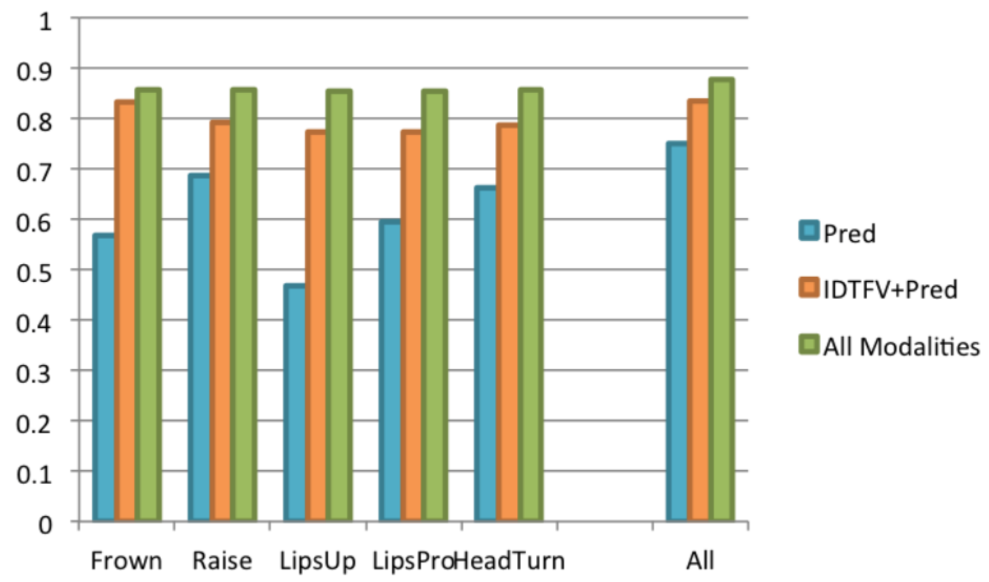
Table 1: Deception Detection results using different feature and classifier combinations. First 4 rows are results of independent features. Last 4 rows are late fusion results of multi-modal features.

Features	L-SVM	K-SVM	NB	DT	RF	LR	Adaboost
GTMicroExpression	0.7964	0.8102	0.8325	0.7731	0.8151	0.8275	0.8270
GTMicroExpression+IDT	0.8456	0.8137	0.8468	0.7834	0.8205	0.8988	0.8270
GTMicroExpression+IDT+Transcript	0.8594	0.8137	0.8923	0.8074	0.8205	0.8988	0.8270
GTMicroExpression+IDT+MFCC	0.8969	0.9002	0.8668	0.7834	0.8319	0.9221	0.8320
GTMicroExpression+All Modalities	0.9065	0.9002	0.8905	0.8074	0.8731	**0.9221**	0.8321

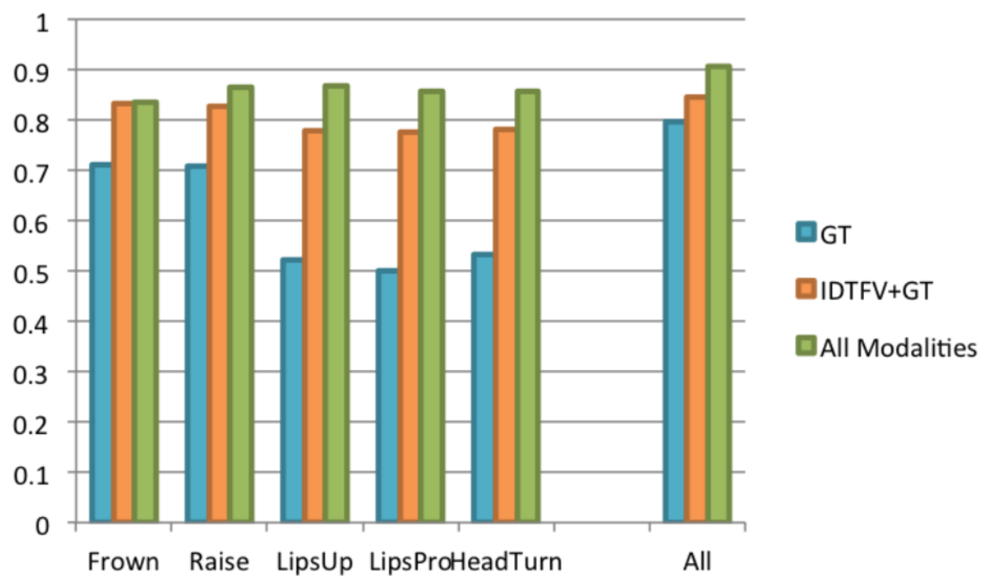
Table 2: Deception Detection results with Ground Truth micro-expression features and other feature modalities.

Micro-Expression Analysis

- We investigate the effectiveness of each micro-expression



(a) Predicted Micro-Expressions

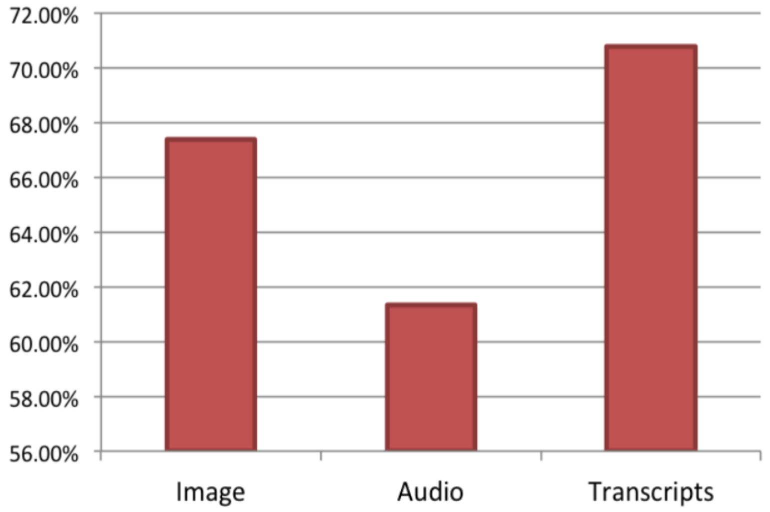


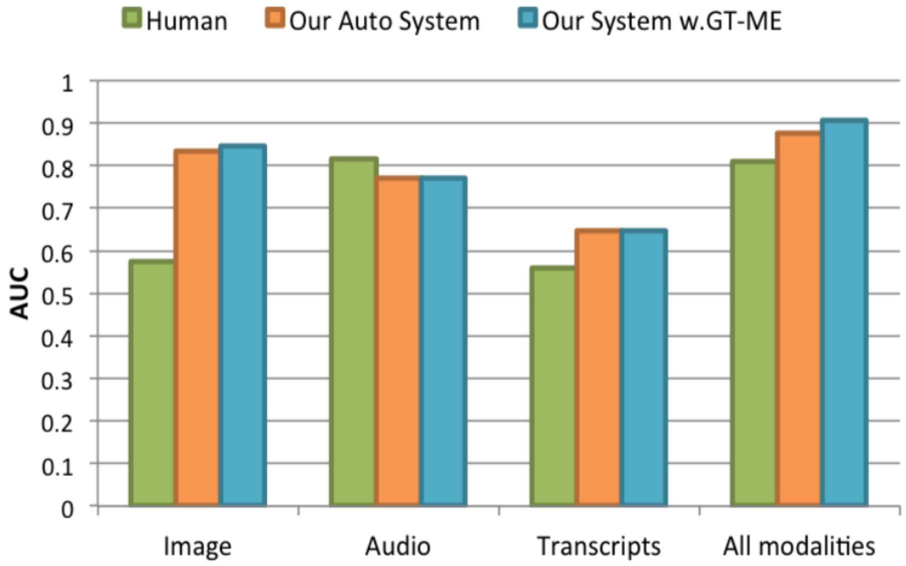
(b) Ground Truth Micro-Expressions

Most important features from past literature: Frowning, Raised Eyebrows, Lip corners up, Lips protruded, Head Side Turn
Eyebrows Raised, Head Side Turn seemed most predictive in our analysis

User Study

- We perform a user-study to analyze how well average humans perform on this task, what modalities they use for deception detection and how they perform if only one modality is accessible.

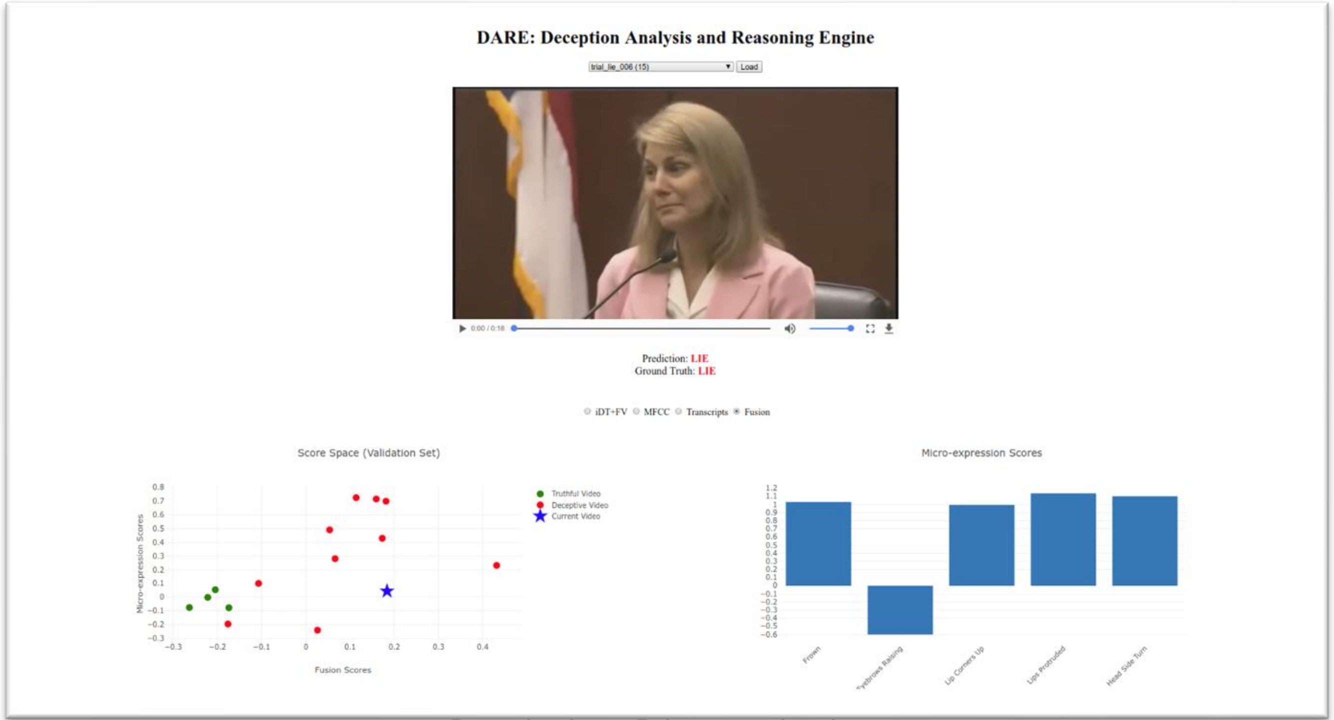




Audio features (MFCC) were the most important features for DARE's deception prediction.
Human's thought visual features were most important for them, but audio was actually more important.

DARE System Demo

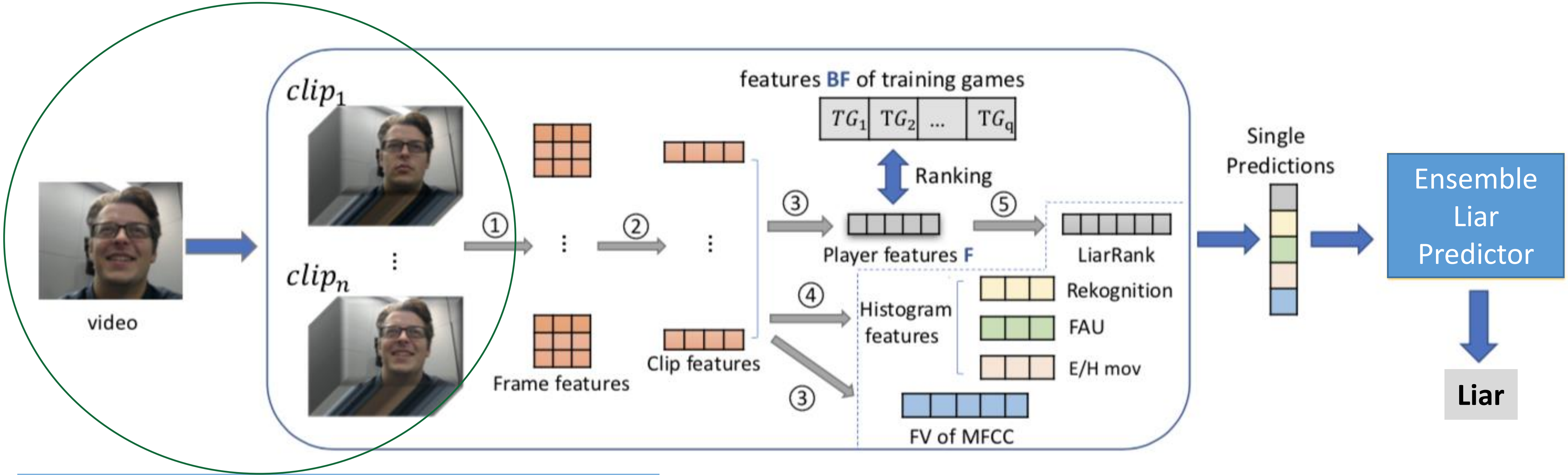
- Try our demo at our project page:
  - http://www.cs.dartmouth.edu/~mbolonkin/dare/demo/



LiarOrNot: Detecting Deception in a Group Setting

- A fully automated system (LiarOrNot) for predicting long-term deception in videos
- A class of histogram-based features
- A novel “meta-feature” called LiarRank that builds on the basic features
- An ensemble based prediction model
- Achieves an **AUC of 0.705** in predicting the role of a player in the game
- AUC for human prediction is 0.583

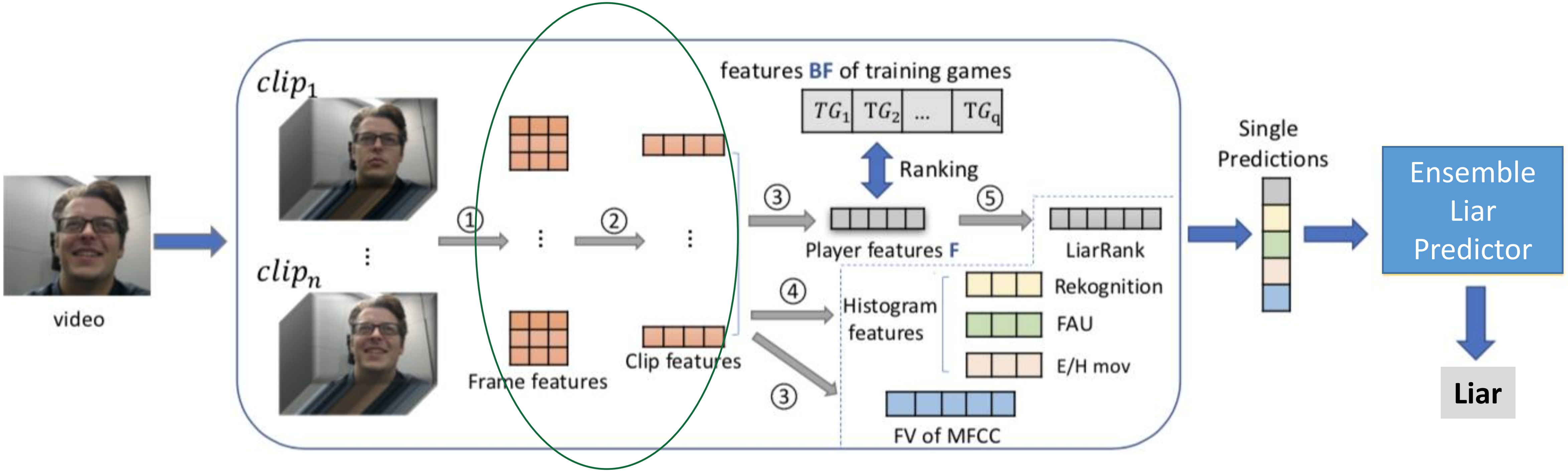
LiarOrNot Architecture



Step 1: uniformly sample 10-second clips in every 30 seconds.

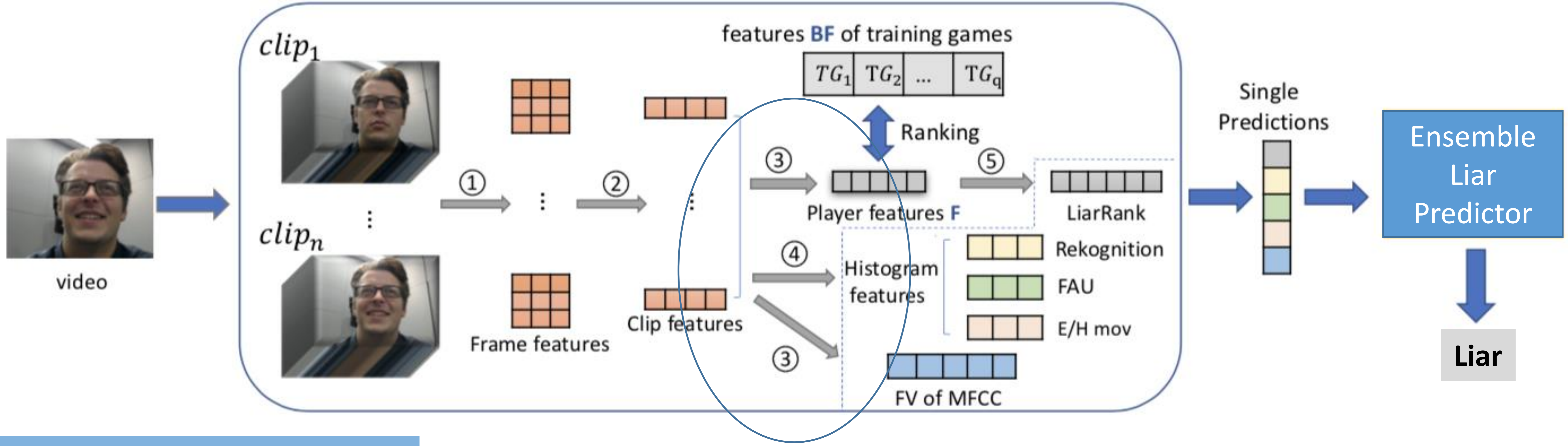
To resolve the challenge of long videos

LiarOrNot Architecture



Step 2: Extract visual and audio features for frame and clips
1. VGG Face
2. Facial Action Units
3. Emotions (from Amazon Rekognition)
4. Eye/Head Movements
5. Mel-Frequency Cepstral Coefficients (MFCC)

LiarOrNot Architecture

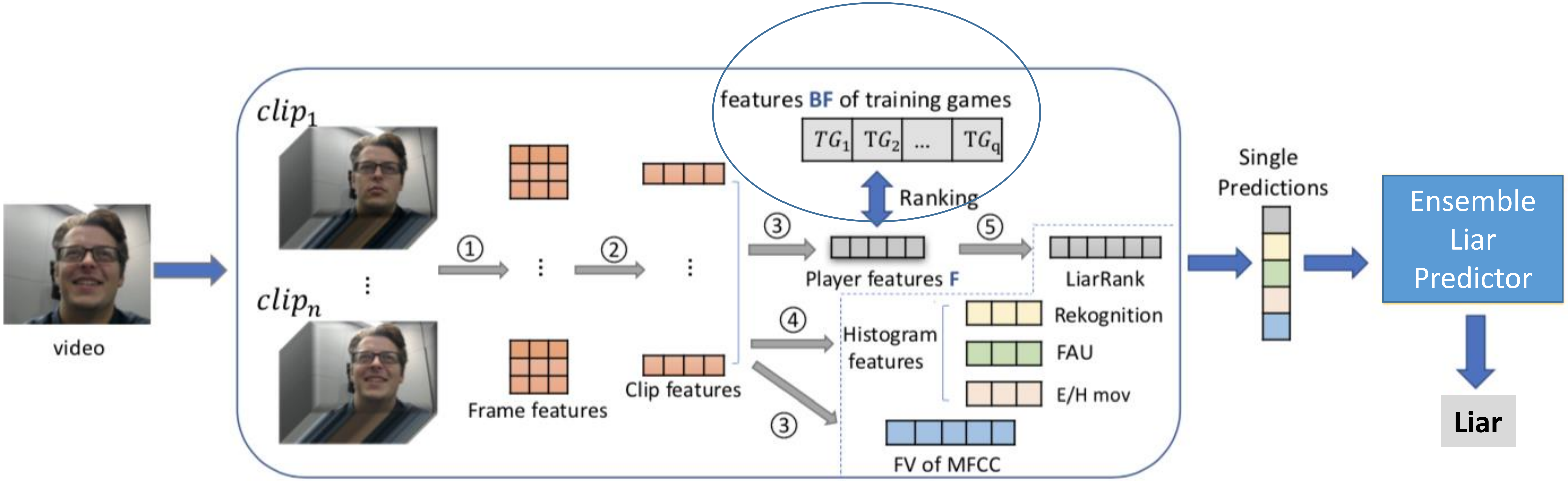


Step 3: Feature aggregation
1. Fisher Vector
2. Histogram

Different games have different number of clips and frames, so their feature vectors may be of different lengths.

We use these 2 aggregation methods to normalize these to a single length feature for each player.

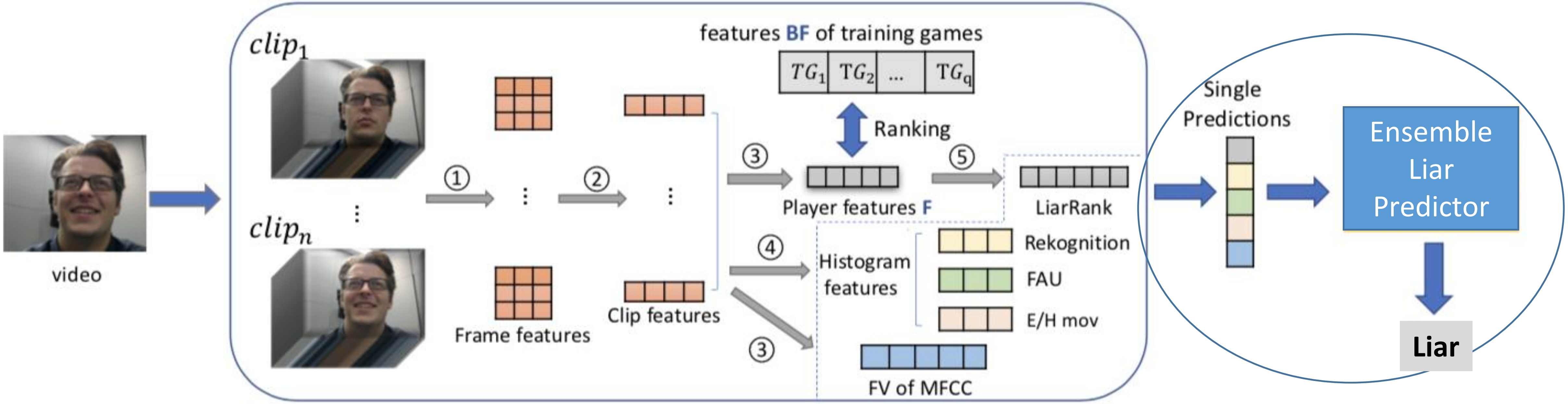
LiarOrNot Architecture



Step 4: LiarRank meta features

Capture the game-level information for a player comparing to all games in the training set

LiarOrNot Architecture

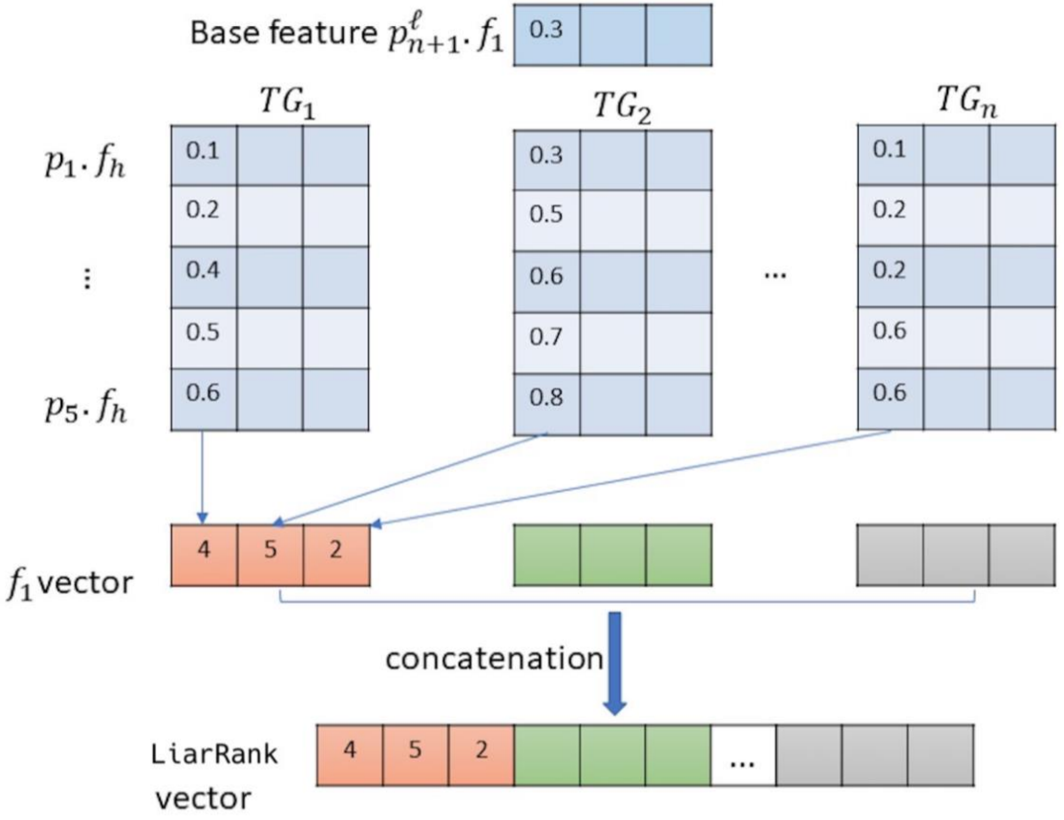


Step 5: Ensemble prediction

Optimize weights of 5 predictors (each from a kind of features)
Final prediction is the weighted sum of the 5 predictors.

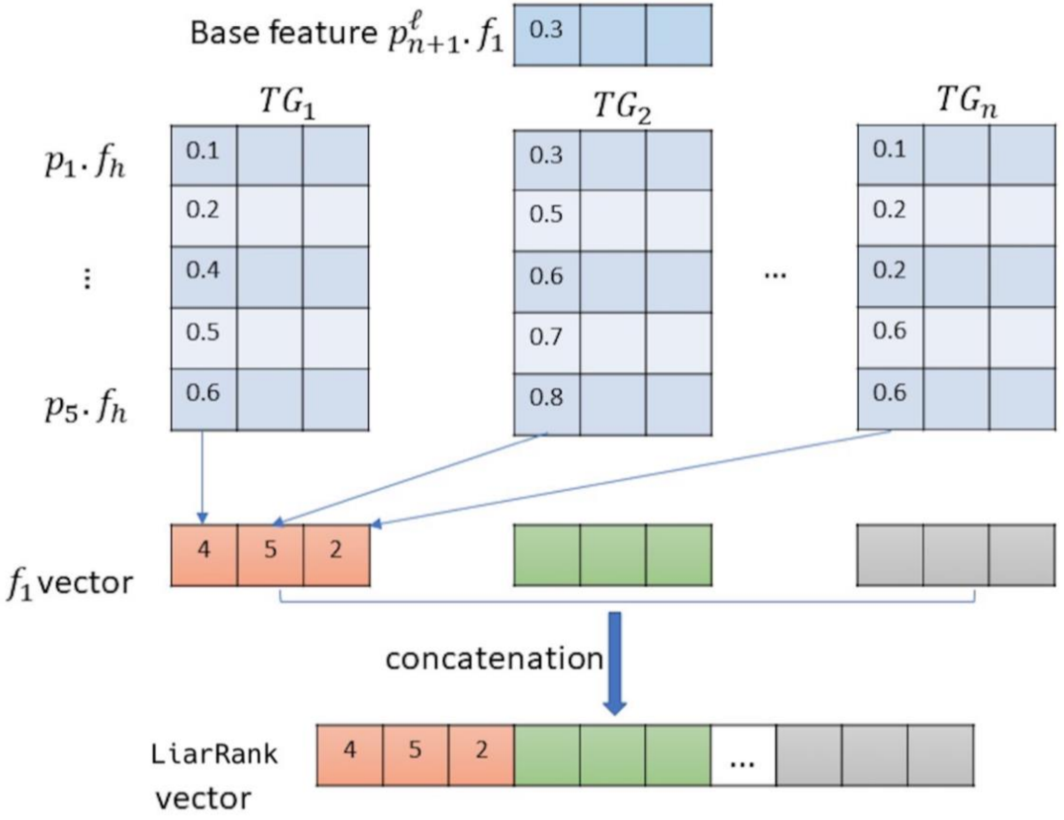
LiarRank meta-feature

- In the training data, we know who the liars are
- *n* games in training set denoted by TG
- Given a player p_{n+1} in a clip c_{n+1} not in the training set, pretend he was in all training games and compare each of his features with those of the truth-tellers and liars in the training data.
- LiarRank of p_{n+1} is the rank of a base feature f's value in a game
- Resulting in |F| * |G| features for each player
  - |F| is the dimension of basic features
  - |G| is the number of games in training set



LiarRank meta-feature

- Build upon any base feature
- Example on the right:
  - 3 games TG_1, TG_2, TG_3 in training set
  - 3 base features f_1, f_2, f_3
  - For f_1 of a given player, his/her rank is 4, 5, 2 in the three games in the training set.
  - Also generate ranks for f_2 (green) and f_3 (gray)



Result: Single feature models (Fisher vectors)

Features	RF	L-SVM	NB	LR	KNN
Average VGG Face (baseline)	0.516	0.533	0.549	0.546	0.50
VGG Face clip-level voting	0.503	0.520	0.550	0.527	0.479
FV of VGG Face	0.468	0.573	0.502	0.584	0.502
FV of VGG Face + FS	0.506	0.470	0.491	0.467	0.522
LiarRank of FV of VGG Face + FS	0.639	0.647	**0.663**	0.652	0.603
FV of MFCC frame-level	0.606	0.395	0.56	0.608	0.579
FV of MFCC clip-level	0.586	0.441	0.533	0.579	0.595

The LiarRank meta feature boosts the performance of VGG Face + Fisher Vector
LiarRank is robust across all classifiers.

Result: Single feature models (Histogram vectors)

Amazon Rekognition					
Frame hist.		Clip hist.		Combined	
Disgusted, Surprised	0.630	Smile, Angry, Disgusted	0.634	Smile, Angry, Disgusted	**0.676**
Surprised	0.622	Smile , Angry	0.623	Smile, Disgusted	0.647
Calm	0.622	Smile, Disgusted, Calm	0.618	Angry	0.638
All features	0.557	All features	0.544	All features	0.563
Facial Action Units					
Frame hist.		Clip hist.		Combined	
AU07+AU10+AU12	**0.621**	AU06+AU14	0.609	AU07+AU09+AU10	**0.621**
AU12+AU23+AU25	0.614	AU07+AU09+AU10	0.606	AU07+AU10+AU23	0.617
AU09+AU10+AU12	0.612	AU07+AU14+AU45	0.603	AU12+AU25	0.611
All features	0.592	All features	0.577	All features	0.608
Eye/Head movement					
Frame hist.		Clip hist.		Combined	
3+8	0.632	1+6+8	**0.671**	1+3+4+5+6+8	0.643
3	0.624	1+6	0.642	1+3+5+8	0.627
3+7	0.615	1+3+6+8	0.636	1+3+5+6+8	0.625
All features	0.591	All features	0.560	All features	0.618

For **expression** features, the combination of Smile, Angry and Disgusted gave the highest AUC: 0.676

For **Facial Action Units**, the combination of AU07(Lid tightener), AU09(Nose wrinkler) and AU10(Upper lip raiser) gave the highest AUC: 0.621

For **Eye/Head movements**, the combination of horizontal eye movements, and x, z head movements gave the highest AUC: 0.671

Result: Ensemble model

- **Ensemble: 0.705 AUC**
- **Ablation test**

Removed feature	AUC
MFCC	0.703
E/H Movement	0.703
FAUs	0.702
Amazon Rek.	**0.688**
LiarRank	**0.688**

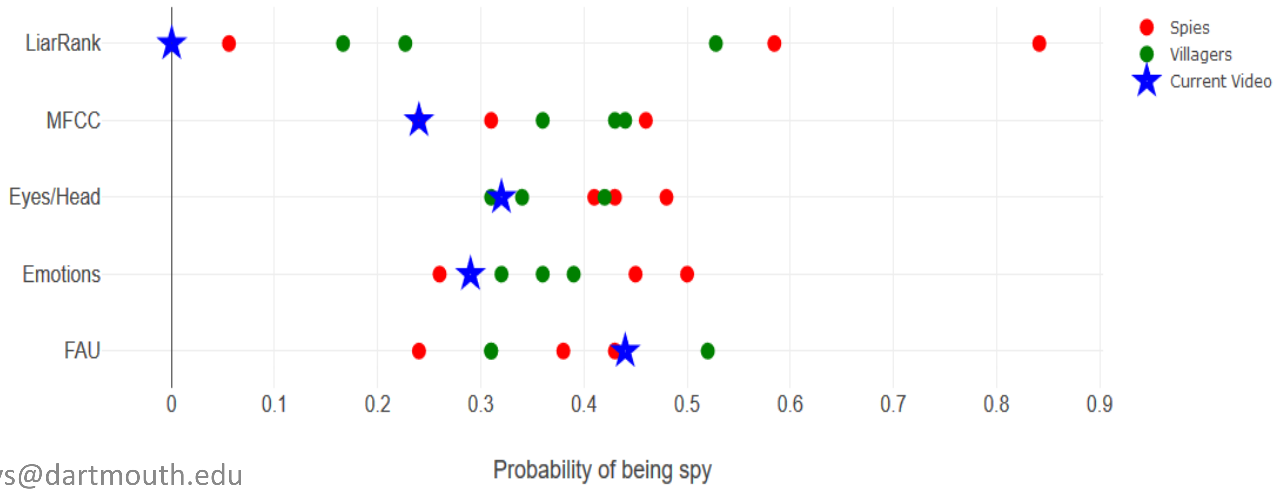
Emotion features and LiarRank are the most important features for predicting deception in a group setting

Demo

Human annotators answers:
Worker 1: SPY
Worker 2: SPY
Worker 3: SPY

LiarOrNot answer:
VILLAGER

Ground truth:
VILLAGER



Demo available at
https://cs.dartmouth.edu/dsail/demos/liar-or-not

@vssubrah, vs@dartmouth.edu

Summer 2020 Webinar Series

Date	Time (EST)	Speaker	Title
June 4	3 – 4:30 pm	**Dr. Purush Iyer** US Army Research Office **V.S. Subrahmanian** Dartmouth College	Introduction to the SCAN Project and Deception Detection from Online Videos
June 10	4 – 5 pm	**Judee Burgoon** University of Arizona	A Novel Approach to Investigating Deception during Group Interaction
June 15	12-1 pm	**Norah Dunbar** UCSB	Persuasive Deception and Dyadic Power Theory
June 22	3 – 4 pm	**Jure Leskovec** Stanford University	Dynamic Embeddings of Temporal Interaction Networks
June 29	3 – 3:40 pm	**Pan Li** Stanford University	An Interpretable Representation Learning Framework for Dynamic Social Networks
July 13	3 – 4 pm	**Chongyang Bai** Dartmouth College	Dominance Detection in Group Interaction Videos
July 20	4 --5 pm	**Dimitris Metaxas** Rutgers University	Detection and Tracking of Humans and Faces using Machine Learning
July 27	4 – 5 pm	**Dimitris Metaxas** Rutgers University	Video-based Deception Detection and Corresponding Feature Discovery
August 11	2 – 3 pm	**Jay Nunamaker** University of Arizona	Going the Last Mile for SCAN Transition
August 18	1 – 2 pm	**Miriam Metzger** UCSB	International Data Collection using Human Subjects: Logistics and Challenges

Contact Info

V.S. Subrahmanian
Dept. of Computer Science
&
Institute for Security, Tech., & Society
Dartmouth College
Hanover,NH 03755.

vs@dartmouth.edu

https://home.cs.dartmouth.edu/~vs/

Research funded by the US Army Research Office