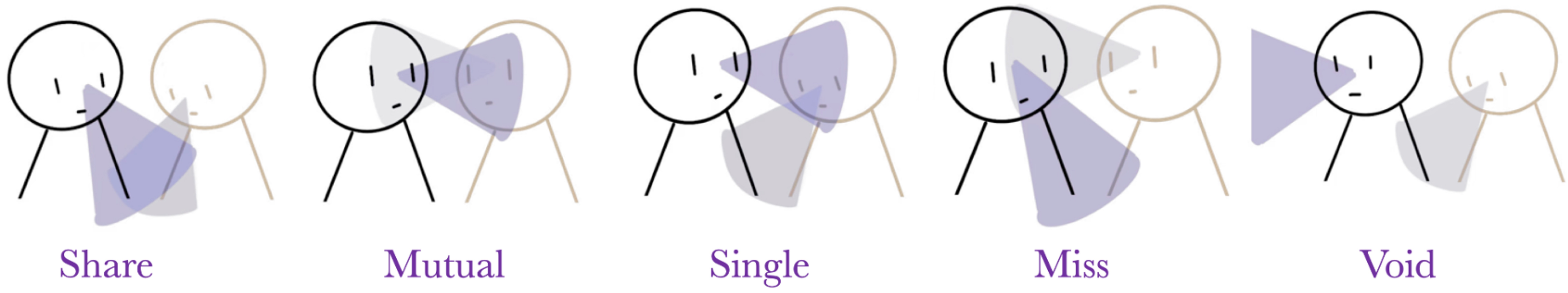
Gaze Pattern Recognition in Dyadic Communication

Fei Chang[1,2], Jiabei Zeng[1], Qiaoyun Liu[4], Shiguang Shan[1,3]
Institute of Computing Technology, Chinese Academy of Sciences(CAS), Beijing, China[1]
Peng Cheng Laboratory, Shenzhen, China[2]
University of CAS, Beijing, China[3]
East China Normal University, Shanghai, China[4]

Motivation:

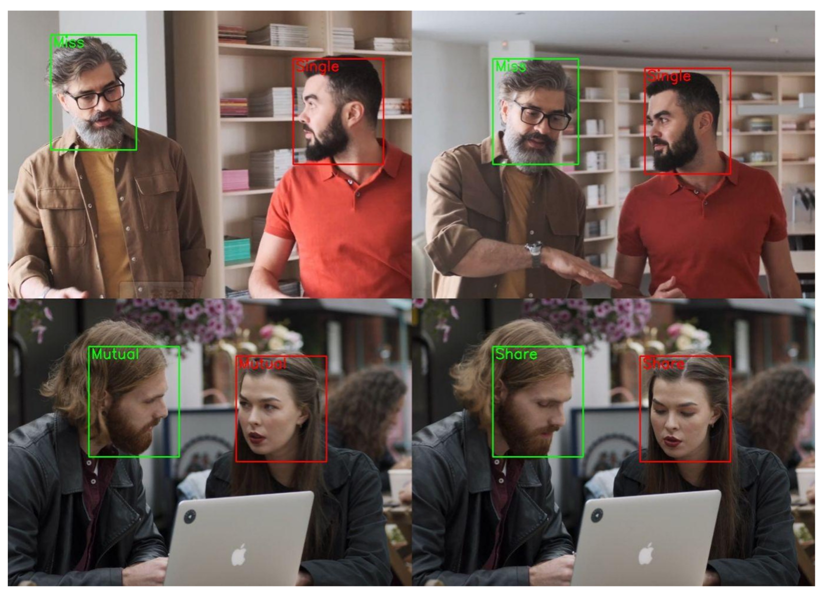
Gaze behavior is a primitive yet a significant mechanism to express interests and reveal emotions during communication. Most previous works in computer science focus on detecting a single gaze pattern. To investigate gaze exhaustively, we propose to group the atomic-level gaze status of two individuals in a dyadic communication into five exclusive patterns: **Share**, **Mutual**, **Single**, **Miss** and **Void**.



Contributions:

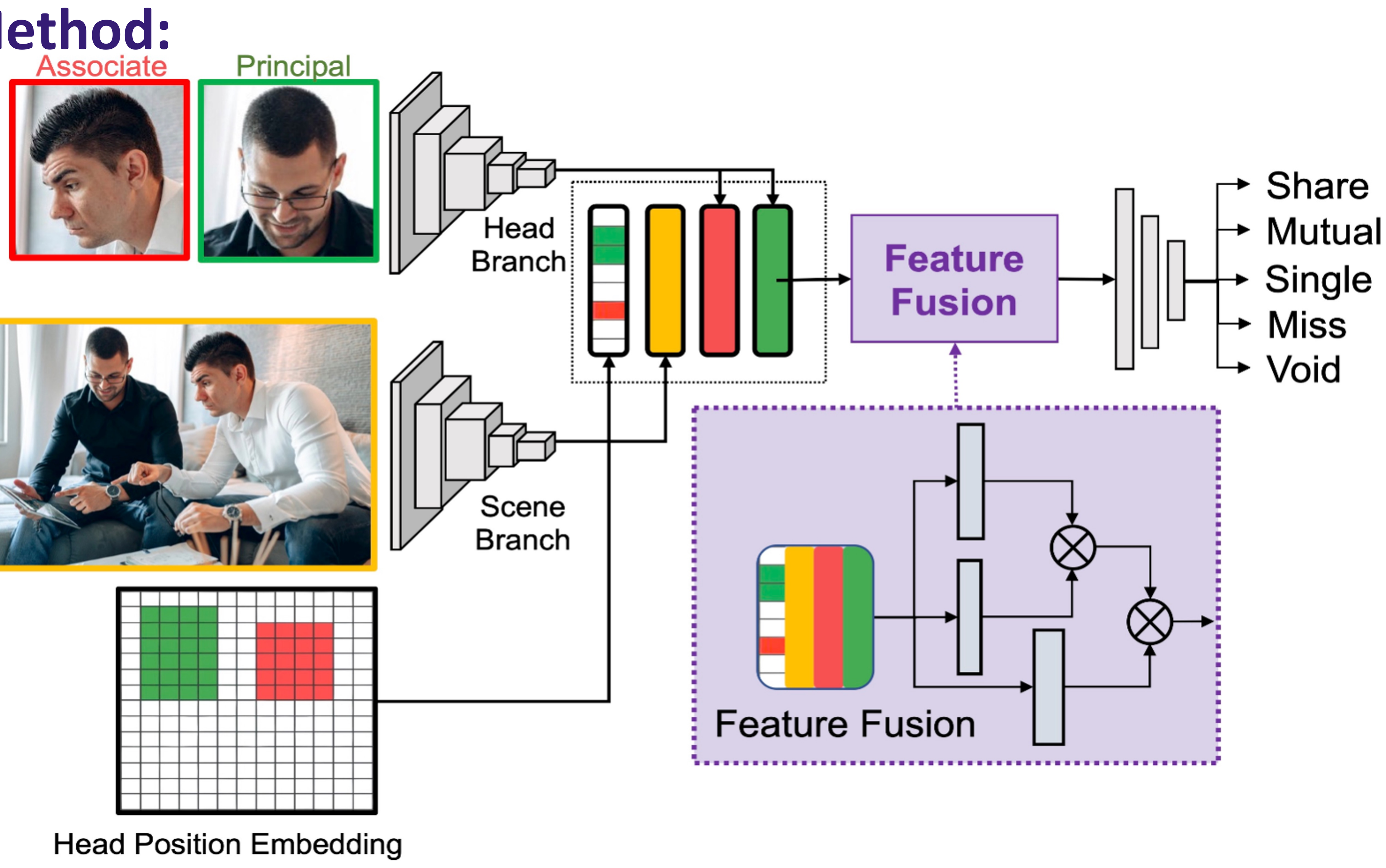
- ➢A taxonomy of five gaze patterns that comprehensively describe the possible stationary gaze status of an individual in dyadic communications
- ➢A benchmark dataset, GP-Static, containing 370 videos of dyadic interactions with frame-level gaze pattern annotations.
- ➢A framework to automatically classify gaze patterns given an image.

GP-Static Benchmark Dataset:



	Train	Test
Share	23,244	3,794
Mutual	41,376	8,482
Single	26,858	5,573
Miss	26,858	5,573
Void	21,124	6,482
Total	139,460	29,904

Method:



- ➢The **head branch** and the **scene branch** are convolution pathways to encode information from heads of two individuals and the surrounding environment.
- ➢The **head position embedding** is derived from two binary images in which pixels inside the head bounding box of each individual are designated with value one and the rest with zero.
- ➢The **feature fusion** consists of three linear layers, which combines the features into a combined representation:

$$\mathbf{x}'_i = \sum_j \alpha_{i,j}\mathbf{x}_j\mathbf{W}_3. \qquad \alpha_{i,j} = \frac{e^{\mathbf{x}_i\mathbf{W}_1(\mathbf{x}_j\mathbf{W}_2)^\top}}{\sum_k e^{\mathbf{x}_i\mathbf{W}_1(\mathbf{x}_k\mathbf{W}_2)^\top}}$$

where $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3$ are weights of the three layers, and $\mathbf{x}_i$, $\mathbf{x}'_i$ are features before and after feature fusion.

Experiments and Results:

➢Quantitative evaluation results on Static Gaze Pattern Classification Task.
(f1): f1-score; Avg. Acc.:Average Accuracy. The best scores are marked in bold.

Method	Share (f1)	Mutual (f1)	Single (f1)	Miss (f1)	Void (f1)	Avg. Acc.
GF-Fixed	0.18	0.46	0.31	0.31	0.36	0.35
GF-Modified	0.34	0.61	0.26	0.26	0.42	0.43
Ours	**0.73**	**0.79**	**0.59**	**0.59**	**0.60**	**0.67**

➢Quantitative evaluation results on Single Gaze Pattern Detection Task.(AP.): Average Precision; (Acc.) :Prediction Accuracy.
The best scores are marked in bold.

Method	Looking-At-Each-Other(AP.)			Share(Acc.)
	UCO-LAEO	AVA-LAEO	OI-MG	VideoCoAtt
LAEO-Net	79.5	50.6	-	-
AAAI'21	65.1	72.2	70.1	-
CVPR'18	-	-	-	71.4
Ours	**80.3**	**82.5**	**72.1**	**73.9**

➢T-Test results on the gaze pattern statistics between children with and without autism.
Gaze patterns are obtained from videos on 20 pre-school children during their interaction with a teacher, among which 10 are diagnosed with autism.

Null hypothesis	t-statistics	p-value
The duration of 'Share' pattern is the same between children with and without autism	-0.46	0.66
The duration of 'Mutual' pattern is the same between children with and without autism	-2.12	0.048(**)
The duration of 'Single' pattern is the same between children with and without autism	-19.00	0.000(***)
The duration of 'Miss' pattern is the same between children with and without autism	4.54	0.000(***)
The duration of 'Void' pattern is the same between children with and without autism	-3.07	0.006(**)