2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) | 978-1-7281-3079-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/FG47880.2020.00025

Recognizing Perceived Emotions from Facial Expressions

Saurabh Hinduja[1], Shaun Canavan[1] and Lijun Yin[2]
[1] Department of Computer Science and Engineering, University of South Florida, Tampa, Florida
[2] Department of Computer Science, Binghamton University, Binghamton, New York

***Abstract*— Expression recognition has seen an increase in research in past years, however, little work has been on recognizing perceived emotion (i.e. subject self-reporting of emotion). Considering this, we investigate the perceived emotion of subjects that perform tasks meant to elicit emotion. To facilitate this investigation, we use the BP4D+ multimodal spontaneous emotion corpus. We first statistically analyze the subject's perceived emotions across 10 tasks available in BP4D+. We show the percentage of subjects that felt specific emotions for each of the tasks. This is done across all tested subjects, as well as male and female subjects independently. Along with our statistical analysis, we also propose a 3D convolutional neural network (CNN) architecture to recognize multiple emotions felt for each task sequence. We report accuracy, F1-binary and AUC for all subjects, as well as male and female subjects.**

I. Introduction

In recent years, there has been a great deal of research into inferring emotion through the use of facial expressions, using both 2D and 3D information. Yang et al. [12] proposed the use of FACS3D-Net to detect action units (AU). Their method integrates 2D and 3D convolutional neural networks for this task. They showed that combining spatial and temporal information yields an increase in AU detection results. Fabiano et al. [3] created synthetic 3D facial data used to train deep neural networks. They showed the use of this synthetic training data allowed for generalizing facial expressions across multiple state-of-the-art datasets. Yang et al. [11] proposed de-expression residue learning which recognizes facial expressions by extracting the expressive component through a generative model.

Although there has been encouraging results on recognizing emotion from facial expressions, it has been shown that expressions vary across cultures and people [8]. Barrett et al. [1] found that while there is evidence to suggest people follow a common view [1] of emotion vs expression (e.g. smile when happy and frown when sad), how this is communicated varies widely across people even within the same situation. It has also been shown that multiple emotions can occur within facial expressions [5], [10]. Considering this, it is important to investigate what emotions (possibly multiple) the subjects themselves felt (i.e. perceived emotion) to further learn how expressions are related to the emotion of the subject. Many of the works on recognition of emotion focus on recognizing the emotion that was meant to be elicited, however, Girard et al. [4] have begun to investigate subject-self-reporting as related to facial expressions, with a specific focus on the felt emotion when a smile occurs. They found that while a smile occurs in emotions such as amusement, embarrassed, fear, and pain, the smiles also looked different as evidenced by their measurement of action units. This work motivated our current investigation into the perceived emotion of subjects. Bias and the impact of gender are also important when analyzing perceived emotions. Recently, Zou et al [14], call for AI to become more fair. They highlight the problem is due to the fact that most of the AI training data is collected in the United States. AI not only learns identification features from the data given to it, but it also learns the bias and distribution of data. This effect has been also highlighted by Boulamwini et al. [2], where they analyzed the accuracy of gender classifiers for people with different demographic traits and found that gender classifiers performed best on light skinned males and worst on dark skinned females.

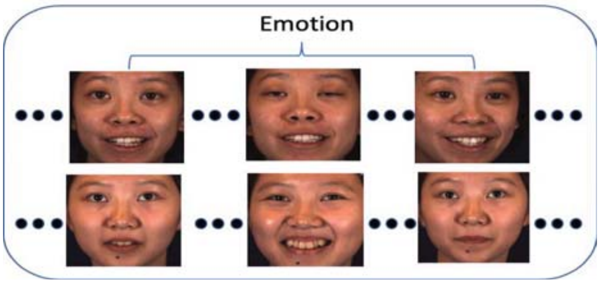

Fig. 1: Different subjects from same task meant to elicit happiness/amusement. Perceived emotion (top) - relaxed:5; amused:4; sympathetic:2; startled:0; surprise:1. (bottom) - relaxed:1; amused:1; sympathetic:0; startled:2; surprise:1. All other perceived emotions are 0 for both subjects.

As it has been shown that expressions vary among people, multiple emotions can occur with facial expressions, and is difficult to infer emotion directly from them (See Fig. 1), we are motivated to investigate the perceived emotion of subjects (self-reporting of emotion), as they relate to facial expressions. The contribution of this work is 3-fold:

1) We analyze subject self-reporting (perceived emotion) from the BP4D+ multimodal spontaneous emotion corpus [13], across gender in relation to the emotion that was meant to be elicited from the task.
2) We propose a 3D CNN architecture for recognizing multiple perceived emotions from facial expressions.
3) To the best of our knowledge, this is the first work to report results for recognizing multiple perceived emotions, across all tasks in BP4D+.

II. Analysis of Perceived Emotions

A. Dataset

BP4D+ [13] consists of 140 subjects (58 male/82 female) with an age range of 18-66. It contains multiple ethnicities in-

978-1-7281-3079-8/20/$31.00 ©2020 IEEE
DOI 10.1109/FG47880.2020.00025

TABLE I: Percent of subjects which felt emotion in each task. Rows are subject self-reporting, columns are tasks. *NOTE: Darker color corresponds to higher percent.*

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Relaxed	0.63	0.25	0.01	0.00	0.09	0.09	0.04	0.07	0.06	0.03
Amused	0.96	0.50	0.00	0.12	0.14	0.43	0.25	0.02	0.14	0.01
Disgusted	0.00	0.04	0.04	0.00	0.00	0.02	0.01	0.00	0.01	0.97
Afraid	0.01	0.00	0.53	0.35	0.01	0.02	0.71	0.12	0.15	0.13
Angry	0.00	0.00	0.42	0.12	0.04	0.01	0.04	0.09	0.47	0.09
Frustrated	0.01	0.00	0.06	0.04	0.06	0.12	0.02	0.08	0.41	0.04
Sad	0.00	0.01	0.65	0.00	0.00	0.01	0.00	0.00	0.09	0.00
Sympathetic	0.05	0.00	0.36	0.01	0.01	0.01	0.01	0.00	0.07	0.01
Nervous	0.22	0.05	0.27	0.10	0.09	0.40	0.55	0.12	0.33	0.16
Pained	0.00	0.01	0.02	0.04	0.00	0.00	0.02	0.99	0.00	0.04
Embarrassed	0.05	0.32	0.00	0.08	0.03	0.94	0.01	0.00	0.27	0.02
Startled	0.03	0.17	0.37	0.99	0.09	0.02	0.33	0.17	0.20	0.17
Surprised	0.23	0.81	0.09	0.64	0.46	0.17	0.31	0.12	0.19	0.12
Skeptical	0.04	0.03	0.01	0.00	0.98	0.04	0.13	0.03	0.28	0.04

TABLE II: Percentage of male subjects that felt emotion in each task. *NOTE: Rows/columns same as Table I.*

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Relaxed	0.66	0.29	0.00	0.00	0.10	0.16	0.03	0.14	0.07	0.02
Amused	0.93	0.53	0.00	0.21	0.17	0.38	0.29	0.03	0.16	0.00
Disgusted	0.00	0.02	0.10	0.00	0.00	0.03	0.00	0.00	0.02	0.98
Afraid	0.00	0.00	0.50	0.28	0.00	0.03	0.66	0.10	0.12	0.05
Angry	0.00	0.00	0.33	0.12	0.07	0.03	0.03	0.10	0.52	0.07
Frustrated	0.02	0.00	0.09	0.03	0.03	0.14	0.02	0.07	0.36	0.09
Sad	0.00	0.02	0.55	0.00	0.00	0.02	0.00	0.00	0.12	0.00
Sympathetic	0.07	0.00	0.38	0.02	0.00	0.02	0.02	0.00	0.03	0.02
Nervous	0.21	0.03	0.34	0.09	0.07	0.33	0.57	0.14	0.34	0.17
Pained	0.00	0.02	0.02	0.03	0.00	0.00	0.02	0.98	0.00	0.05
Embarrassed	0.07	0.22	0.00	0.12	0.00	0.93	0.00	0.00	0.19	0.00
Startled	0.03	0.19	0.28	1.00	0.14	0.03	0.26	0.09	0.14	0.12
Surprised	0.28	0.78	0.07	0.69	0.45	0.17	0.29	0.07	0.21	0.14
Skeptical	0.02	0.05	0.00	0.00	0.98	0.02	0.21	0.03	0.29	0.05

cluding Caucasian, African American, Asian, and Hispanic. It contains 2D and thermal images, 3D models, physiological data, action units, and facial landmarks (2D and 3D). Ten tasks were performed to elicit the following emotions: (T1) Happy, (T2) Surprise, (T3) Sad, (T4) Startled, (T5) Skeptical, (T6) Embarrassed, (T7) Fear (T8) Pain, (T9) Anger, and (T10) Disgust. Along with this multimodal data, subject self-reporting on the emotions they felt during each task was collected, for 138 of the subjects. The subjects were allowed to choose multiple emotions for each task, such as relaxed, surprised, sad, happy, etc. The intensity of the self-reported emotions was also collected using a 5-point-Likert scale. Our analysis (Section II-B) is done on all 138 subjects.

B. Subject self-reporting

Motivated by work that has shown multiple emotions can be felt during facial expressions [5], [10], we have investigated which emotions were felt across all of the tasks in BP4D+, for all subjects with self-reports. As can be seen in Table I, many of the subjects reported multiple emotions for each task. For example, in task 1, where happy was meant to be elicited, a large percentage of the subjects felt amused, relaxed, surprised, and nervous (96%, 63%, 23% and 22%, respectively). Along with these emotions a small percentage (<5%) also felt afraid, frustrated, sympathetic, embarrassed, startled, and surprised. For each task, the emotion that was meant to be elicited (or similar emotion) is felt by the largest percentage of all subjects (e.g. for happy task 96% of the subjects felt amused, for pain task, 99% of the subjects felt pain). What is interesting, is that there are many instances of complementary emotions being felt for the tasks. For example, in task 3 (sad) 65% of the subjects felt sad, however, fear, anger, sympathy, nervous, and startled were all felt with relatively high percentages (53%, 42%, 36%, 27% and 37%, respectively). This analysis agrees with the literature [1] that people react differently even within the same situation (i.e. task). Another interesting task is 9 (anger), where frustration is felt with similar frequency as anger (41% and 47%, respectively). Multiple emotions being felt within the same task, can partially explain the difficulty in inferring emotion from facial expressions alone. It is possible similar expressions are shown in a facial image, however, the subject is feeling a different emotion.

TABLE III: Percentage of female subjects that felt emotion in each task. *NOTE: Rows/columns same as Table I.*

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Relaxed	0.60	0.22	0.01	0.00	0.07	0.04	0.05	0.02	0.05	0.04
Amused	0.98	0.48	0.00	0.06	0.12	0.47	0.22	0.01	0.12	0.01
Disgusted	0.00	0.06	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.96
Afraid	0.01	0.00	0.54	0.40	0.01	0.01	0.74	0.12	0.17	0.19
Angry	0.00	0.00	0.49	0.12	0.02	0.00	0.04	0.09	0.43	0.10
Frustrated	0.01	0.00	0.05	0.04	0.07	0.11	0.02	0.09	0.44	0.01
Sad	0.00	0.01	0.72	0.00	0.00	0.00	0.00	0.00	0.07	0.00
Sympathetic	0.04	0.00	0.35	0.00	0.01	0.00	0.00	0.00	0.10	0.01
Nervous	0.22	0.06	0.22	0.11	0.10	0.46	0.54	0.11	0.32	0.15
Pained	0.00	0.00	0.02	0.05	0.00	0.00	0.02	1.00	0.00	0.02
Embarrassed	0.04	0.38	0.00	0.05	0.05	0.95	0.01	0.00	0.32	0.04
Startled	0.02	0.15	0.44	0.99	0.06	0.01	0.38	0.23	0.25	0.21
Surprised	0.20	0.84	0.11	0.60	0.47	0.16	0.32	0.15	0.17	0.11
Skeptical	0.05	0.01	0.01	0.00	0.98	0.05	0.07	0.02	0.27	0.02

Along with analyzing all subjects with self-report, we also independently analyzed male and female subjects. As can be seen in Tables II and III, many of the self-reported emotions are similar across the tasks for both male and female subjects. For example with task T5 (skeptical), 98% of the male and female subjects reported that they felt skeptical, with 45% of the male and 47% of the female subjects reported they felt surprised. Although, the majority of female and male subjects reported similar feelings of emotion across the majority of the tasks, there are some instances where they differ. For example, in task 3 (sad), 72% of female subjects reported feeling sad, while only 55% of the male subjects reported this. To further investigate this, we evaluated the statistical significance between Tables II and III, by conducting paired t-tests (Table V). As mentioned in Section II-A, the self-reports from BP4D+ also contains the intensity of emotion felt, therefore, we also calculated the statistical significance between males and females and the intensities of their perceived emotions (Table IV).

As can be seen in Table V, while most of the differences between male and female, occurrences of self-reported emotion, are not significant, there are some notable exceptions to this, especially tasks 3 (Sad), 4 (Startled), and 8 (Pain). This suggests that there are some differences in how gender emotionally responded to some of the tasks, although more

TABLE IV: Significance of differences between male and female intensity of self-reported emotion. *Note: 'n.s' stands for not significant, * p< 0.05, ** p< 0.01, *** p<0.005*

Task	Relaxed	Amused	Disgusted	Afraid	Angry	Frustrated	Sad	Sympathetic	Nervous	Pained	Embarrassed	Startled	Surprised	Skeptical
T1	*n.s*	*n.s*	-	*n.s*	-	*n.s*	-	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T2	*n.s*	*n.s*	*n.s*	-	-	-	*n.s*	-	*n.s*	*n.s*	*	*n.s*	*n.s*	*n.s*
T3	*n.s*	-	**	*n.s*	*n.s*	*n.s*	***	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*
T4	-	***	-	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-
T5	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T6	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T7	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*
T8	***	*n.s*	-	*n.s*	*n.s*	*n.s*	-	-	*n.s*	*	-	*	*	*n.s*
T9	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T10	*n.s*	*n.s*	*n.s*	*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*	*	*n.s*	*n.s*

TABLE V: Significance of differences between male and female occurrence of self-reported emotion. *Note: 'n.s' stands for not significant, * p< 0.05, ** p< 0.01, *** p<0.005*

Task	Relaxed	Amused	Disgusted	Afraid	Angry	Frustrated	Sad	Sympathetic	Nervous	Pained	Embarrassed	Startled	Surprised	Skeptical
T1	*n.s*	*n.s*	-	-	-	-	-	-	*n.s*	-	*n.s*	-	*n.s*	-
T2	*n.s*	*	*n.s*	-	-	-	-	-	*n.s*	-	**	*n.s*	*n.s*	*n.s*
T3	-	-	*n.s*	*n.s*	*n.s*	*n.s*	***	*n.s*	*n.s*	*n.s*	-	*n.s*	*	-
T4	-	*	-	*n.s*	*n.s*	*n.s*	-	-	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-
T5	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T6	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T7	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	-	*n.s*	-	*n.s*	*	*n.s*	*n.s*
T8	***	*n.s*	-	*n.s*	*n.s*	*n.s*	-	-	*n.s*	*n.s*	-	*	*n.s*	*n.s*
T9	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	*n.s*	-	*n.s*	*n.s*	*n.s*	*n.s*
T10	*n.s*	-	*n.s*	*	*n.s*	*n.s*	-	-	*n.s*	*n.s*	*n.s*	**	*n.s*	*n.s*

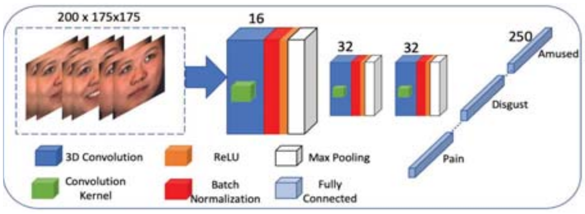

Fig. 2: Proposed 3D convolutional neural network for recognizing multiple perceived emotions for one task.

statistical analysis is needed to determine this. It is also interesting to note, that the differences change when intensity is accounted for (Table IV). For example, in task 2 (surprise), there is no significance between males and females in how often they reported feeling amusement, however, in the subjects that reported this emotion, there is significance in the intensity of the emotion that they reported. To further investigate this, we calculated the average intensity of each reported emotion, for each task, for male and female subjects. We calculated this value for two cases. First, the range of [0, 5], which also includes when an emotion was not felt. Secondly, we calculated it for the range [1,5], to compare the intensities only when an emotion was felt. For the first case, the average intensity across all tasks and emotions is 0.49 and 0.42 for females and males, respectively. For the second case, the average intensity across all tasks and emotions is 3.67 and 3.43 for females and males, respectively. Table VI, shows the average intensities for all perceived emotions across all tasks (range [0,5]), which shows that for the majority of tasks, female subjects had stronger intensities of emotion.

III. Recognizing Perceived Emotion

A. Experimental Design

1) Data Preprocessing: First, we tracked and normalized faces in terms of rotation, scaling and centered using Dlib [6]. This resulted in face images of size 256 x 256 which we scaled down to 175 x 175 pixels. Next, 3D CNN architectures require each input sequence to have the same number of frames; to satisfy this requirement one could take the N number of consecutive frames, however, this would not be representative of the whole sequence [12]. Therefore, we sampled the entire sequence to obtain an equal number of frames from each sequence. As adjacent frames are highly correlated [12] we sampled 200 equidistant frames from each sequence preserving maximum temporal information.

2) Proposed 3D CNN: Inspired by FACS3D-Net [12], we extend this work by proposing a multi-tail architecture for multi-emotion recognition that utilizes shared 3D Convolution layers. The 3D convolution captures the temporal information of emotion from the onset to offset and the multi-tail performs regression for each independent emotion. The 3D convolution layers capture the temporal information of the emotion. The multiple tails of the network take the deep features of the shared 3D CNN [9] to predict the presence of the perceived emotion, treating each emotion independently while preserving the co-occurrence of emotions.

The proposed multi-tail 3D CNN (Fig. 2) has 3 3D CNN layers, each followed by batch normalization and max pooling. The first layer has 16 filters with a kernel size of (3,3,3), the second and third layers were identical with 32 filters and a kernel size of (3,3,3), all three 3D CNN were followed Max pooling layers with kernel size of (3,3,3) and

TABLE VI: Average intensity (range [0,5]) of perceived emotions across all tasks (BP4D+) for male (M) and female (F). White cells intensity $[0, < 1)$; yellow cells intensity [1, 2); green cells intensity [2, 3); red cells intensity [3, 5].

Emotion	T1		T2		T3		T4		T5		T6		T7		T8		T9		T10	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Relaxed	1.97	2.00	0.90	0.60	0.00	0.01	0.00	0.00	0.28	0.24	0.38	0.12	0.09	0.15	0.41	0.04	0.16	0.11	0.05	0.12
Amused	2.91	3.11	1.71	1.22	0.00	0.00	0.60	0.11	0.41	0.28	0.95	1.23	0.74	0.59	0.10	0.01	0.36	0.30	0.00	0.01
Disgusted	0.00	0.00	0.02	0.10	0.22	0.00	0.00	0.00	0.00	0.00	0.05	0.04	0.00	0.05	0.00	0.00	0.02	0.00	3.83	4.22
Afraid	0.00	0.01	0.00	0.00	1.48	1.93	0.97	1.54	0.00	0.04	0.07	0.02	2.07	2.01	0.30	0.31	0.28	0.47	0.12	0.59
Angry	0.00	0.00	0.00	0.00	1.26	1.82	0.28	0.46	0.21	0.06	0.09	0.00	0.09	0.12	0.20	0.26	1.55	1.30	0.16	0.25
Frustrated	0.03	0.01	0.00	0.00	0.22	0.16	0.14	0.11	0.07	0.14	0.41	0.21	0.07	0.09	0.16	0.19	1.72	1.40	0.24	0.06
Sad	0.00	0.00	0.02	0.02	1.81	2.85	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.35	0.21	0.00	0.00
Sympathetic	0.12	0.05	0.00	0.00	1.33	1.30	0.04	0.00	0.00	0.04	0.02	0.00	0.04	0.00	0.00	0.00	0.05	0.25	0.02	0.01
Nervous	0.24	0.33	0.05	0.14	1.07	0.71	0.24	0.38	0.16	0.31	0.86	1.35	1.70	1.50	0.50	0.27	0.90	1.00	0.43	0.37
Pained	0.00	0.00	0.02	0.00	0.02	0.10	0.16	0.17	0.00	0.00	0.00	0.00	0.02	0.05	3.60	4.10	0.00	0.00	0.17	0.05
Embarrassed	0.10	0.09	0.40	0.96	0.00	0.00	0.26	0.14	0.00	0.19	3.20	3.31	0.00	0.04	0.00	0.00	0.60	0.90	0.00	0.15
Startled	0.05	0.04	0.40	0.40	1.00	1.54	4.50	4.61	0.36	0.12	0.05	0.05	0.75	1.15	0.20	0.70	0.36	0.70	0.28	0.70
Surprised	0.62	0.42	2.10	2.20	0.10	0.30	2.74	2.60	1.20	1.40	0.50	0.40	0.90	0.83	0.10	0.41	0.50	0.47	0.47	0.26
Skeptical	0.03	0.08	0.14	0.03	0.00	0.01	0.00	0.00	3.86	4.14	0.09	0.12	0.52	0.17	0.03	0.06	0.90	0.90	0.12	0.09

TABLE VII: Perceived emotion evaluation metrics.

	All			Female			Males		
	Acc	F1 Bin	AUC	Acc	F1 Bin	AUC	Acc	F1 Bin	AUC
Relaxed	0.89	0.45	0.70	0.89	0.48	0.72	0.89	0.42	0.68
Amused	0.80	0.42	0.65	0.81	0.45	0.68	0.78	0.38	0.62
Nervous	0.81	0.42	0.68	0.79	0.44	0.68	0.84	0.38	0.66
Pained	0.86	0.35	0.66	0.84	0.31	0.64	0.89	0.40	0.68
Embarrassed	0.86	0.44	0.70	0.84	0.39	0.66	0.87	0.49	0.75
Surprised	0.75	0.39	0.62	0.75	0.39	0.61	0.74	0.39	0.63
Average	**0.83**	**0.41**	**0.67**	**0.82**	**0.41**	**0.67**	**0.84**	**0.41**	**0.67**

TABLE VIII: Variance and standard deviation of accuracy for recognizing perceived emotion.

	Standard Deviation	**Variance**
Tasks	0.101	0.01
Subjects	0.16	0.026

ReLu activation. For the multi-tailed part of the network; each emotion is predicted by a tail of the network; the deep features from the shared 3D CNN are connected to a fully connected layer with 250 neurons, followed by a single neuron output layer for regression. We used adam [7] optimizer and as we had only 9 training sequences (i.e. 9 tasks) per subject, batch size = 2.

3) Data Validation: Since it has been shown that emotions can vary widely across people even within the same situation [1], for our experiments, we conduct subject specific task out validation (e.g. train on 9 tasks from same subject, test on 1). This leads to 10 experiments per subject on 70 subjects giving a total of 700 experiments. The average performance of all subjects is reported. For training, the self-reported emotions (multiple) are used as the ground-truth class labels for each task sequence. In BP4D+ not all emotions were reported equally across all tasks, to recognize multiple emotions across tasks only the 6 emotions which were reported more than 100 times were used (Table VII).

B. Results

In Table VII we report the average accuracy, F1-binary, and AUC scores for all subjects, male, and female subjects. Our network achieved an average accuracy of 83%, F1-binary of 0.41, and AUC of 0.67. As can be seen in Table VIII, the standard deviation and variance for recognizing perceived emotion is low for both tasks and subjects, validating the proposed approach. As shown in Section II, there are many similarities between the occurrence of self-reported emotions across male and female subjects. This partially explain the results that the evaluation metrics are similar for the tested emotions across gender (Table VII). To the best of our knowledge, this is the first work to report these evaluation metrics on perceived emotions in BP4D+, therefore we did not have any works to compare against.

IV. Conclusion

We have analyzed perceived emotion, showing that while male and female subjects generally respond to the tasks in a similar manner, there are instances where the two classes diverge. We have also shown which emotions are statistically significant across the different tasks for male and female subjects, for occurrence and intensity of emotion. Results suggest that on average female subjects have a higher intensity when the emotion is elicited, in BP4D+. Our analysis agrees with literature [1], that emotions can vary across subjects within the same situation.

We have also proposed a 3D CNN architecture for recognizing perceived emotions from task sequences. We report multiple evaluation metrics, and to the best of our knowledge, this is the first work to do this on perceived emotions. While these results are encouraging, there are some limitations of this work. First, only a subset of subjects are used from BP4D+. All subjects need analyzed, as well as different datasets containing subject self-report. Secondly, while perceived emotions vary between subjects it will be interesting to see results on leave-one-subject-out validation. Finally, the proposed approach needs compared to other approaches (e.g. random forest).

Acknowledgement

We would like to thank the support from Amazon (Machine Learning Research Award), as well as from NSF (grant CNS-1629898 and CNS-1205664), and the Center of Imaging, Acoustics, and Perception Science (CIAPS) of the Research Foundation of Binghamton University.

REFERENCES

[1] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, 20(1):1–68, 2019.

[2] J. Buolamwini and T. Gebru. Gender shades: Intersec. accuracy disparities in commercial gender class. In *CFAT*, pages 77–91, 2018.

[3] D. Fabiano and S. Canavan. Deformable synthesis model for emotion recognition. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, pages 1–5, 2019.

[4] J. M. Girard, G. Shandar, Z. Liu, J. F. Cohn, L. Yin, and L.-P. Morency. The shape of a smile. *8th International Conference on Affective Computing and Intelligent Interaction (ACII 2019)*, 2019.

[5] J. A. Hall and D. Matsumoto. Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, 4(2):201, 2004.

[6] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10(Jul):1755–1758, 2009.

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

[8] R. Srinivasan and A. M. Martinez. Cross-cultural and cultural-specific production and perception of facial expressions of emotion in the wild. *IEEE Transactions on Affective Computing*, 2018.

[9] Y. Wei, W. Xia, M. Lin, J. Huang, B. Ni, J. Dong, Y. Zhao, and S. Yan. Hcp: A flexible cnn framework for multi-label image classification. *IEEE transactions on pattern analysis and machine intelligence*, 38(9):1901–1907, 2015.

[10] J. Winston, J. O'doherty, and R. J. Dolan. Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *Neuroimage*, 20(1):84–97, 2003.

[11] H. Yang, U. Ciftci, and L. Yin. Facial expression recognition by de-expression residue learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2168–2177, 2018.

[12] L. Yang, I. O. Ertugrul, J. F. Cohn, Z. Hammal, D. Jiang, and H. Sahli. Facs3d-net: 3d convolution based spatiotemporal representation for action unit detection. *International Conference on Affective Computing and Intelligent Interaction*, 2019.

[13] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, J. Cohn, Q. Ji, and L. Yin. Multimodal spontaneous emotion corpus for human behavior analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3438–3446, 2016.

[14] J. Zou and L. Schiebinger. Ai can be sexist and racist-its time to make it fair, 2018.