Supplementary Materials
EmotioNet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild

C. Fabian Benitez-Quiroz*, Ramprakash Srinivasan*, Aleix M. Martinez
Dept. Electrical and Computer Engineering
The Ohio State University
*These authors contributed equally to this paper.

1. Extended Experimental Results

The within-database classification results on the shoulder database were compared to methods described in the literature which report results using ROC (Receiver Operating Characteristic) curves. ROC curves are used to visually and analytically evaluate the performance of binary classifiers. Recall that our classifiers are binary, i.e., AU present (active) in the image or not. ROC plots display the *true positive rate* against the *false positive rate*. The true positive rate is the sensitivity of the classifier, which we have preciously defined as Recall in the main paper. The false positive rate is the number of negative test samples classified as positive (i.e., the image does not include AU i but is classified as having AU i present) over the total number of false positives plus true negatives. Note that the derived algorithm only provides a result, but this can be plotted in ROC space and compared to state-of-the-art methods. Furthermore, since we run a five-fold cross validation, we actually have five results plus the mean reported in the main document. Thus, we can plot six results in ROC space. These results are in Figure S1. Figure S2 provides the same ROC plots for the DISFA database.

As mentioned above, our proposed approach does not yield an ROC curve but rather a set of points in ROC space. We can nevertheless estimate an ROC curve by changing the value of the prior of each AU i. In the results reported in the main paper, we assumed equal priors for AU i active and not active. Reducing the prior of AU i active will decrease the false detection rate, i.e., it is less likely to misclassify a face that does not have AU i active as such. Increasing the prior of AU i active will increase the true positive detection rate. This is *not* what our algorithm does, but it is a simple extension of what can be obtained in applications where the use of priors is needed. Figures S3 and S4 provide the ROC curves thus computed on two of the databases used in the main paper, shoulder pain and DISFA.

The plots in Figures S3 allow us to compute the area under the curve for the results of our algorithm on the shoulder pain database. These and comparative results against the algorithms of [5] and [11] are in Table S1. Once again, we see that the results obtained with the proposed algorithm are superior than those reported in the literature.

We also computed the results on a recent database of spontaneous facial expressions, AM-FED [7]. Our F1 scores where as follows: .93 (AU 2), .89 (AU 4), .94 (AU 5), .82 (AU 9), .92 (AU 12), .75 (AU 14), .82 (AU 15), .92 (AU 17), .90 (AU 18), .72 (AU 26).

2. EmotioNet: Facial Expressions of Emotion in the Wild

We collected one million images of facial expressions of emotion in the wild. Images were downloaded from several popular web search engines by using the emotive keywords defined as nodes of the word "feeling" in WordNet [8] and with the requirement that a face be present in the image. The number of concepts (i.e., words with the same meaning) given by WordNet was 421. These words are listed in Tables S2-S5.

This search yielded a large number of images. These images were further evaluated to guarantee they included a face. This was done in two stages. First, we used the face detector of [10] to detect faces in these images. Images where a face was not detected by this algorithm were discarded. Second, the resulting images were visually inspected by the authors. Images that did not have a face, had a drawing of a face or pornography were eliminated. The end result was a dataset of one million images. This set of images in the wild was the one used in the present work. The number of images in these categories varies from a low of 47 to a maximum of $6,300$, and more than $1,000$ categories have $> 1,000$ images. The average number of sample images/category is 600 (805 stdv).

As described in the main paper, images were automatically annotated by our algorithm. First, our algorithm anno-

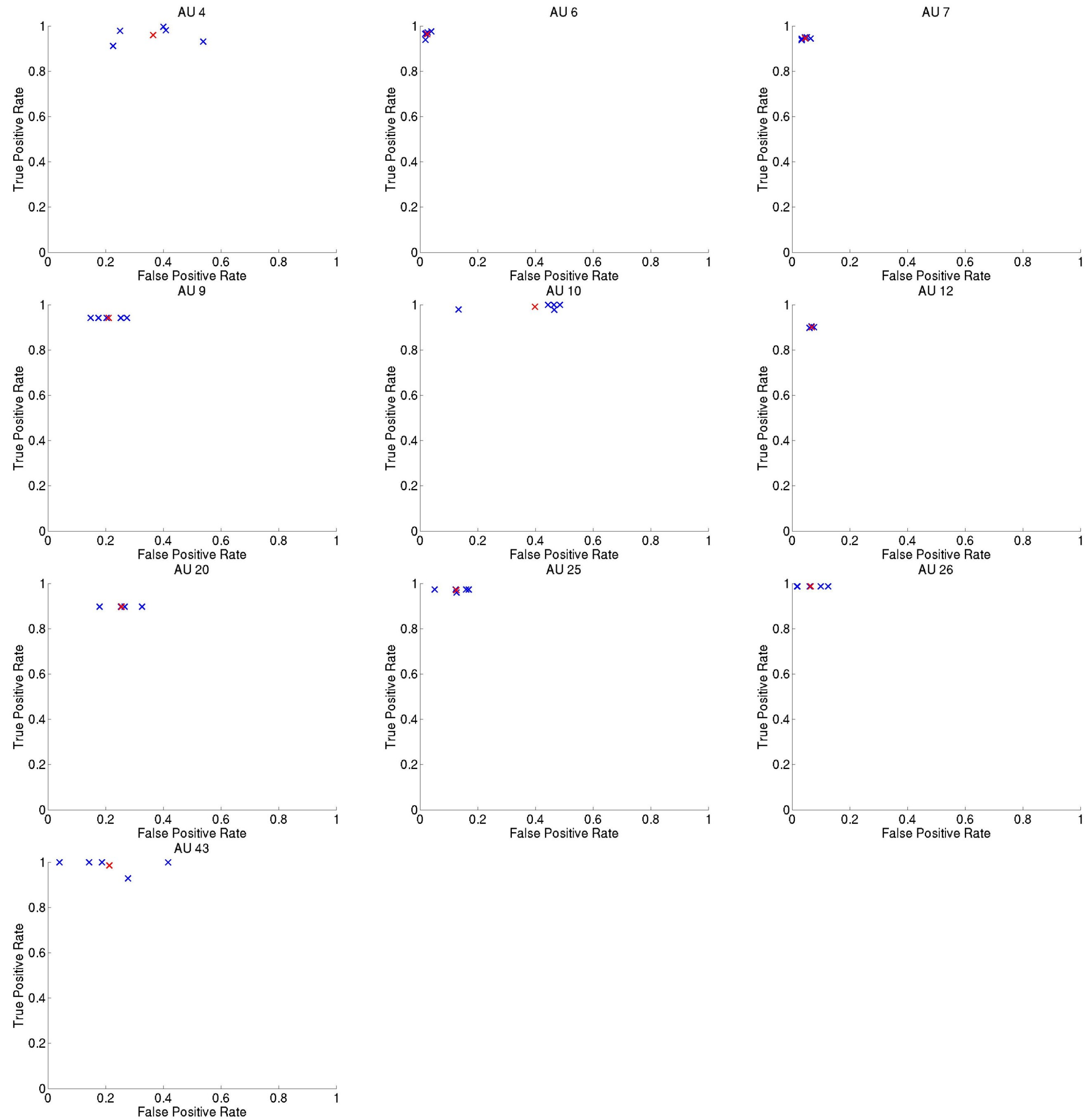


Figure S1: True positive rate against false positive rate of the proposed algorithm for each of the AUs automatically recognized in the images of the shoulder pain database. Shown in the figure are the five results of the five-fold cross-validation test (shown in blue) and the mean (shown in red).

tated AUs and AU intensities. The AUs we annotated were 1, 2, 4, 5, 6, 9, 12, 15, 17, 20, 25 and 26, since these were the well represented ones in the databases used for training the system. Note that we need a set of accurately annotated AUs and AU intensities to be included during training.

Figure S5a shows the percentages of images in our database of facial expressions in the wild that where automatically annotated with AU i. For example, AU 1 was automatically annotated in over $200,000$ images.

Importantly, we manually FACS-coded 10% of this database. That is, a total of $100,000$ images were manually annotated with AUs by experienced coders in our lab-

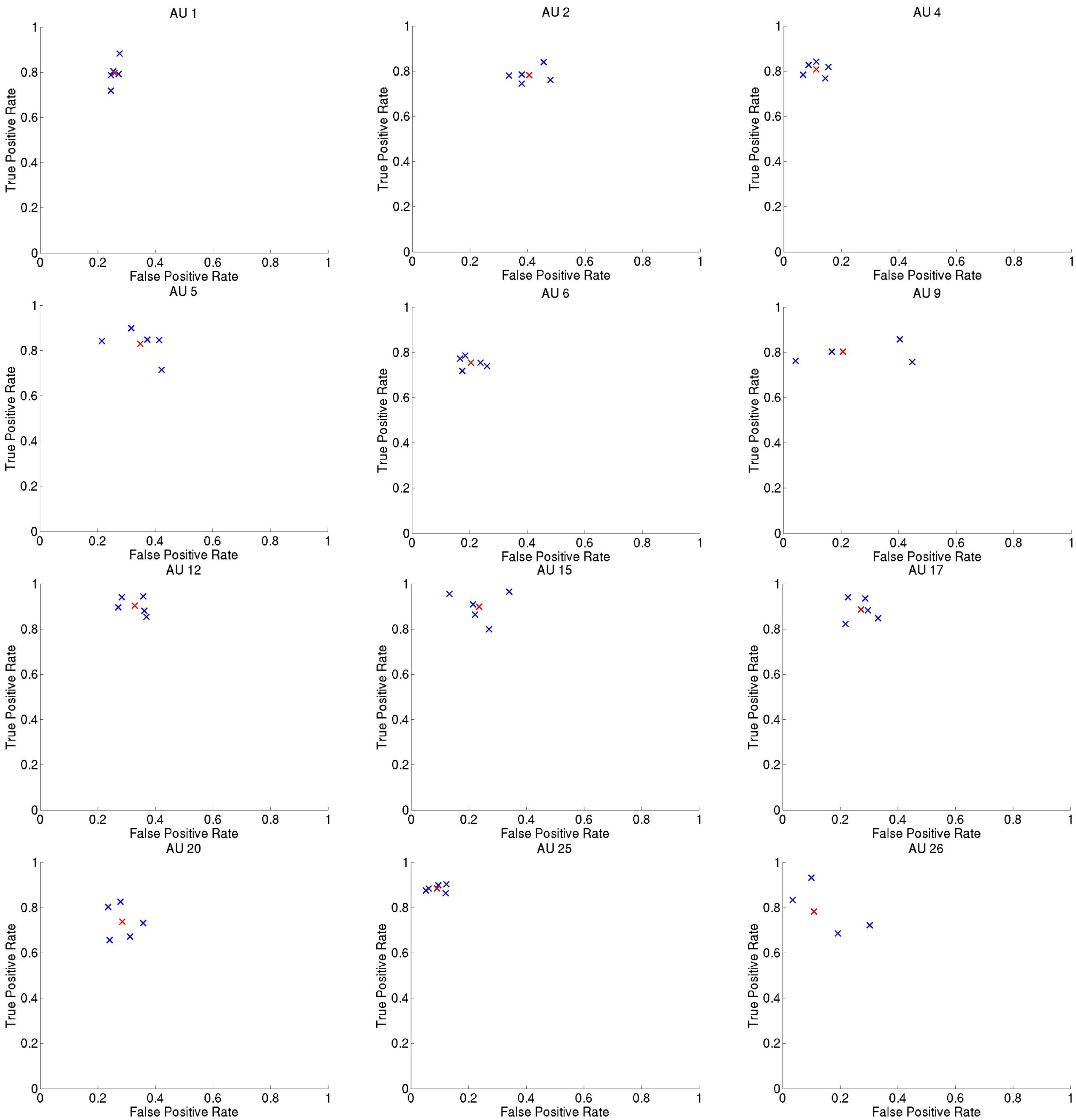


Figure S2: True positive rate against false positive rate of the proposed algorithm for each of the AUs automatically recognized in the images of the DISFA dataset.

AU	4	6	7	9	10	12	20	25	26	43
This paper	**82.45**	**93.48**	**88.57**	**92.56**	**86.15**	**98.54**	**91.13**	**81.46**	**87.19**	**95.47**
Lucey et al. [5]	53.7	86.2	70	79.8	75.4	85.6	66.8	73.3	52.3	90.9
Zafar et al. [11]	78.77	**91.2**		**92.1**						**96.53**

Table S1: Area under the curve for the results shown in Figure S3.

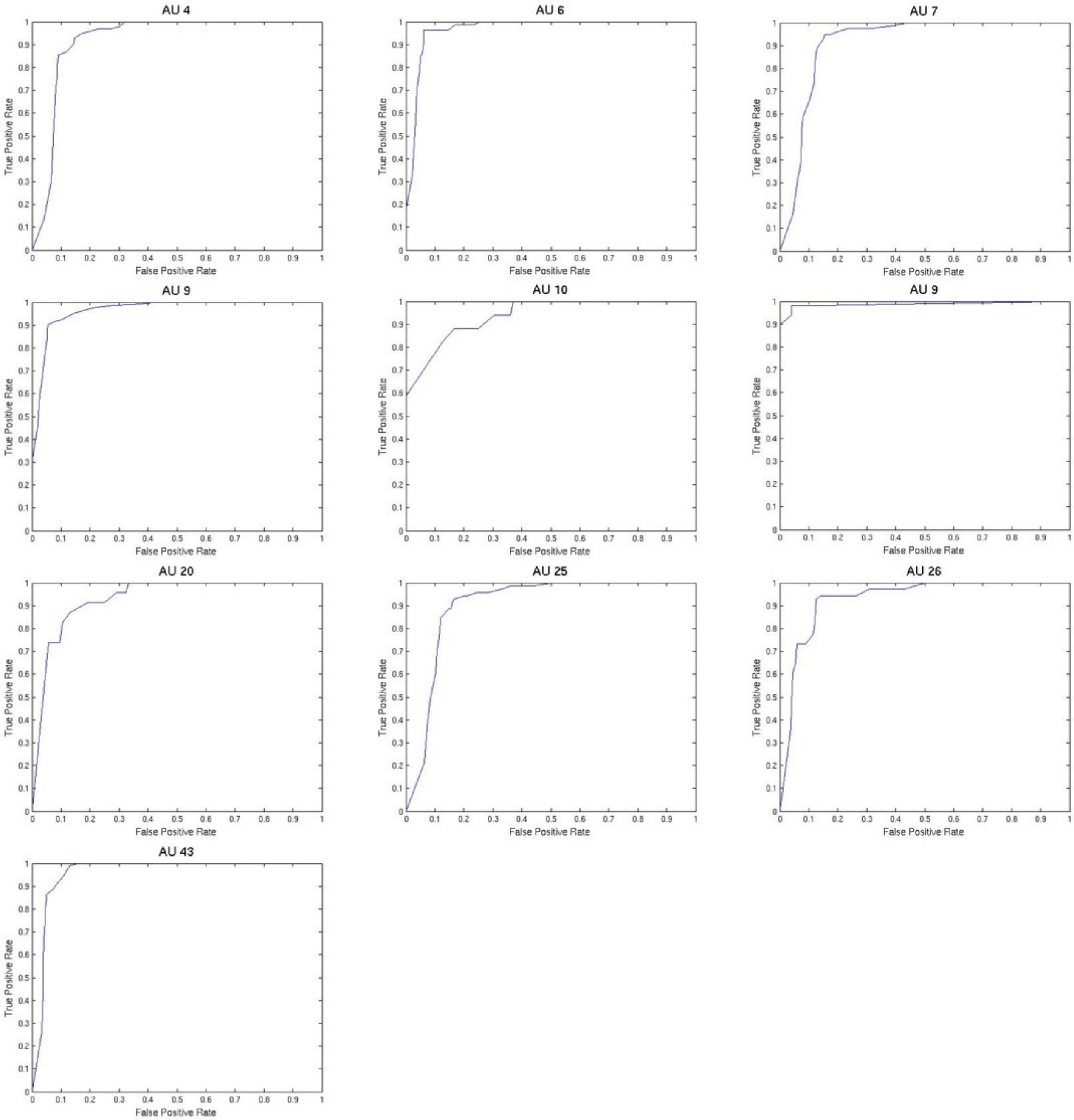


Figure S3: ROC curves for each AU on the shoulder pain database. ROC curves were computed by varying the value of the priors for AU i present and AU i not present.

oratory. This allowed us to estimate the AU detection accuracy of our algorithm, which was about 80%. Note this is extremely accurate given the heterogeneity of the images in the EmotioNet dataset. However, this number only considers correct true positive and true negatives, but does *not* include false negative. Additional work is needed to provide a full analysis of our proposed method on millions of images.

Once an image had been annotated with AUs and AU intensities, we used Table 1 to determine if the face in the image expressed one of the 23 basic or compound emotion categories described in [2, 3]. Note that a facial expression needs not belong to one of these categories. Only when the unique pattern of AU activation described in Ta-

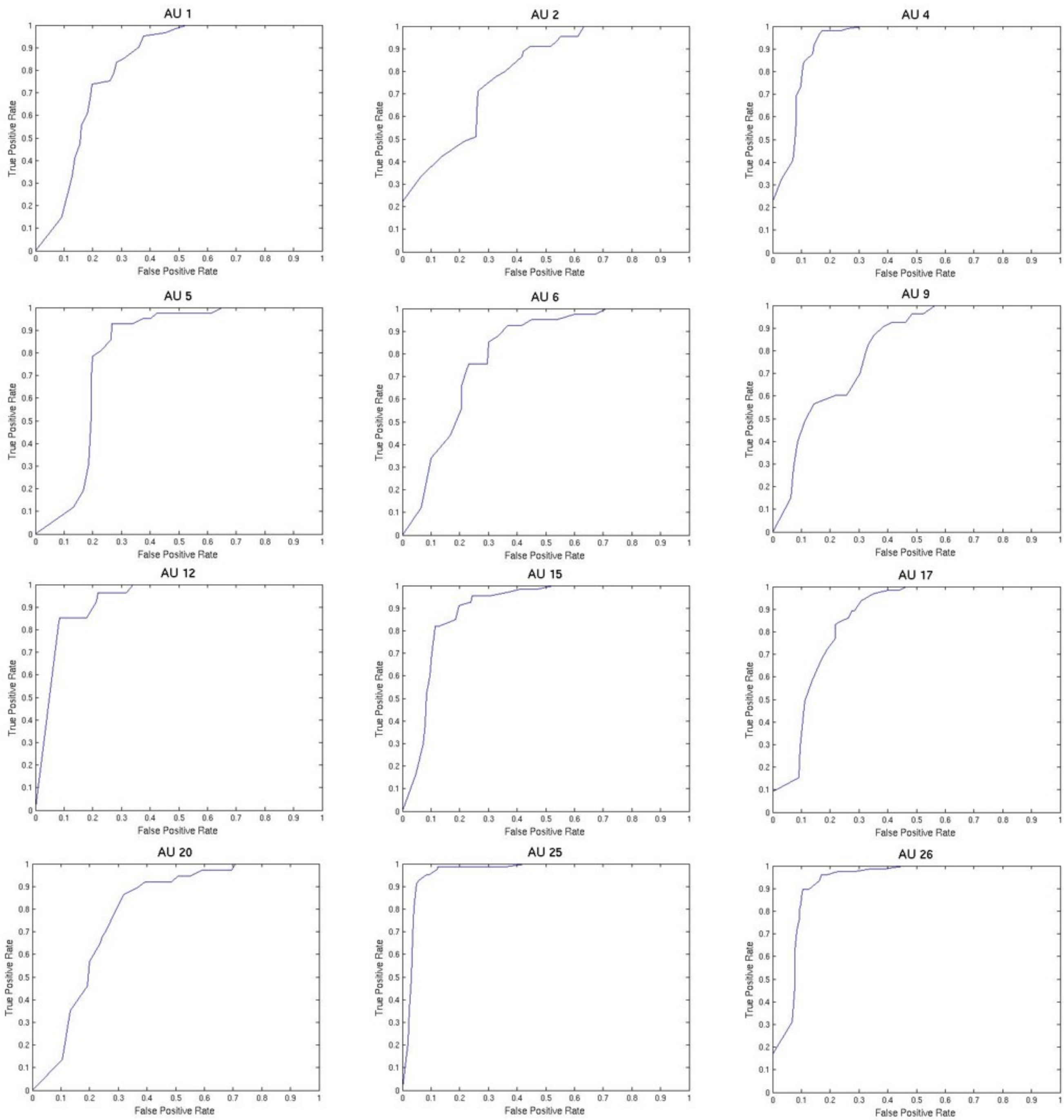


Figure S4: ROC curves for each AU on the DISFA database. ROC curves were computed by varying the value of the priors for AU i present and AU i not present.

ble 1 was present was the face classified as expressing one of these emotions. Figure S5b shows the percentage of images of each emotion category in our database. For example, over $78,000$ images include a facial expression of anger and about $76,000$ have an expression of sadly disgusted. Our algorithm has also been successfully used to detect the "not face" in images in the wild [1]. The "not face" is a grammmatical marker of negation and a facial expression of negation and disapproval.

The above two sections have shown additional quantitative results and analyses of the approach and database of facial expressions of emotion in the wild defined in the main paper. Figure S6 now shows qualitative examples of the images automatically annotated with AU 12 active (present).

3. Rank ordering AU classification

To retrieve images with AU i active, we rank-ordered images according to the posterior probability given by the logistic regression function in the face space of AU i. More formally, let $\mathbf{z}^{\varphi}$ be the sample feature vector of image $\mathbf{I}$ in the kernel space of AU i, then the posterior probability is given by,

$$\log\left(\frac{P(\text{AU } i \text{ active}|\mathbf{Z}=\mathbf{z}_{\varphi})}{P(\text{AU } i \text{ inactive}|\mathbf{Z}=\mathbf{z}^{\varphi})}\right)=\mathbf{b}_i+\mathbf{n}_i^T\mathbf{z}^{\varphi}, \quad \text{(S1)}$$

where $\mathbf{b}_i$ and $\mathbf{n}_i$ are the bias and normal of the hyperplane defining the classifier of AU i in kernel space. It is easy to show that (S1) above is equivalent to,

$$P(\text{AU } i \text{ active}|\mathbf{Z}=\mathbf{z}^{\varphi})=\frac{1}{1+e^{-\left(\mathbf{b}_i+\mathbf{n}_i^T\mathbf{z}^{\varphi}\right)}}. \quad \text{(S2)}$$

The parameters $\mathbf{b}_i$ and $\mathbf{n}_i$ are estimated with iterative re-weighted least squares on the training data $\mathcal{D}_i$ and by optimizing the following function,

$$(\mathbf{b}^*,\mathbf{n}_i^*)=\arg\min_{\mathbf{b},\mathbf{n}_i}\sum_{j=1}^{n_i+m_i}\left\{y_{ij}(\mathbf{b}_i+\mathbf{n}_i^T\mathbf{z}_i^{\varphi})-\log\left(1+e^{-\left(\mathbf{b}_i+\mathbf{n}_i^T\mathbf{z}^{\varphi}\right)}\right)\right\}. \quad \text{(S3)}$$

The images we previously shown in Figure S6 are rank-ordered by (S3) such that the images in the first row have a greater posterior than those in the second row and, in general, the images of a top row have a larger posterior than those in its bottom row. Images in the same row have a similar posterior.

4. Ordinal Regression Metrics

The evaluation of AU intensities is a bit trickier than that of AU active/inactive because these are defined by ordinal variables. Unfortunately, evaluation of ordinal variables is a difficult problem. One popular solution is to use the Mean Zero-one Error (MZE), given by $n^{-1}\sum L(f(\mathbf{z}_i)\neq y_i)$, where n is the number of samples, $L(.)$ is an indicator function, $\mathbf{z}_i$ are the samples, y_i are the ordinal labels, and $f(.)$ is the function that estimates the ordinal variable y. Note that this metric does not take the ordinal nature of the labels y_i into account and thus misclassifying a sample $\mathbf{z}_i$ with ordinal value k by any other value but k is considered equally bad. This is not applicable to our case because misclassifying AU intensity by one ordinal step is better than misclassifying it by two which, in turn, is better than misclassifying it by three and so on.

Two other popular methods for evaluating one's estimates of ordinal variables are the Mean Absolute Error (MAE) and the Mean Square Error (MSE). Here, a function $g(.)$ is employed to assign real values to the ordinal categories, e.g., AU intensity $a=1$, $b=2$, $c=3$, $d=4$ and $e=5$. The error is then measured as $n^{-1}\sum|y_i-f(\mathbf{z}_i)|^b$, where y_i and $f(.)$ are now real numbers, and $b=1$ for MAE and $b=2$ for MSE. This is a popular option and was the one chosen to analyze the results in the main paper (with $b=1$).

The main problem with the aforementioned approach is that it assumes that the distance between any two ordinal values is the same, i.e., the distance between AU intensity a and b is the same as the distance between c and d. This is of course not necessarily true.

While the distance between any pair of AU intensities is difficult to define generally, its definition can be readily obtained in most applications. For example, in some applications, misclassifying intensity a as c is twice as bad as misclassifying a as b, and misclassifying intensity a as e is twice as bad as misclassifying a as c. This corresponds to a linear function and thus MSE or MAE are the most appropriate measurements. However, when misclassifying intensity a as c is only a little worse than misclassifying a as b, MAE and MSE need to be modified. This can be easily done by defining

$$\frac{1}{n}\sum_{i=1}^{n}|M(y_i,f(\mathbf{z}_i))|^b, \quad \text{(S4)}$$

where y_i and $\mathbf{z}_i$ now take values from the ordinal set $\{a,b,c,d,e\}$, $M(.,.)$ is a 5×5 matrix with each (p,q) entry specifying how bad our estimation of AU intensity is in our application. For example, we can define $M(.,.)$ as

	a	b	c	d	e
a	0	1	1.2	1.3	1.4
b	1	0	1	1.2	1.3
c	1.2	1	0	1	1.2
d	1.3	1.2	1	0	1
e	1.4	1.3	1.2	1	0

Using the above defined metric (and $b=1$) to calculate the AU intensity estimation errors of our derived algorithm across databases yields the following errors: .73 for AU 4, .62 for AU 6, .58 for AU 9, .46 for AU 12, .7 for AU 20, .43 for AU 25, and .49 for AU 26. These results would substitute those previously reported in Figure 5b and are based on the idea that misclassifying by one ordinal value is almost as bad as any other misclassification.

1	Feeling	62	Anxiousness, disquiet
2	Affect	63	Insecurity
3	Emotion	64	Disquietude, edginess, inquietude, uneasiness
4	Conditioned emotional response	65	Care,concern, fear
5	Anger, choler, ire	66	Willies
6	Fury, madness, rage	67	Sinking
7	Wrath	68	Misgiving, qualm, scruple
8	Lividity	69	Jitteriness, jumpiness, nervousness, restiveness
9	Enragement, infuriation	70	Angst
10	Offence, offense, umbrage	71	Joy, joyfulness, joyousness
11	Indignation, outrage	72	Elation, lightness
12	Dudgeon	73	Euphoria, euphory
13	Huffiness	74	Exultation, jubilance
14	Dander, hackles	75	Triumph
15	Irascibility, spleen	76	Excitement, exhilaration
16	Conniption, fit, scene, tantrum	77	Bang, boot, charge, flush, kick, rush, thrill
17	Annoyance,chafe, vexation	78	Intoxication
18	Irritation,pique, temper	79	Titillation
19	Frustration	80	Exuberance
20	Aggravation, exasperation	81	Love
21	Harassment, torment	82	Adoration, worship
22	Displeasure	83	Agape
23	Fear, fearfulness, fright	84	Crush, infatuation
24	Alarm, consternation, dismay	85	Amorousness, enamoredness
25	Creeps	86	Ardor, ardour
26	Chill, frisson, quiver, shiver, shudder, thrill, tingle	87	Devotedness, devotion
27	Horror	88	Benevolence
28	Hysteria	89	Beneficence
29	Affright, panic, terror	90	Heartstrings
30	Swivet	91	Caring, lovingness
31	Scare	92	Warmheartedness, warmth
32	Apprehension, apprehensiveness, dread	93	Hate, hatred
33	Trepidation	94	Loyalty
34	Boding, foreboding, premonition, presentiment	95	Abhorrence, abomination, detestation, execration, loathing, odium
35	Shadow	96	Misanthropy
36	Presage	97	Misogamy
37	Suspense	98	Misogynism, misogyny
38	Gloom, gloominess, somberness, somberness	99	Misology
39	Chill, pall	100	Misoneism
40	Timidity, timidness, timorousness	101	Murderousness
41	Shyness	102	Despising
42	Diffidence, self-distrust, self-doubt	103	Enmity, hostility
43	Hesitance, hesitancy	104	Animosity, animus
44	Unassertiveness	105	Antagonism
45	Intimidation	106	Aggression, aggressiveness
46	Awe, fear, reverence, veneration	107	Belligerence, belligerency
47	Anxiety	108	Warpath
48	Discomfiture, discomposure, disconcertion, disconcertment	109	Bitterness, gall, rancor, rancor, resentment
49	Trouble, worry	110	Huffishness, sulkiness
50	Grievance, grudge, score	111	Comfort
51	Enviousness, envy	112	Felicity, happiness
52	Covetousness	113	Beatification, beatitude, blessedness
53	Jealousy	114	Enlightenment, nirvana
54	Malevolence, malignity	115	Radiance
55	Maleficence	116	State
56	Malice, maliciousness, spite, spitefulness, venom	117	Unhappiness
57	Vengefulness, vindictiveness	118	Embitterment
58	Spirit	119	Sadness, sorrow, sorrowfulness
59	Embarrassment	120	Huffishness, sulkiness
60	Ecstasy, exaltation, rapture, raptus, transport	121	Bereavement, mourning
61	Gratification, satisfaction	122	Poignance, poignancy

Table S2: List of the WordNet concepts used as keywords to search images of faces in a variety of web search enginees.

123	Glow	184	Sex
124	Faintness	185	Pleasance, pleasure
125	Soul, soulfulness	186	Afterglow
126	Passion	187	Delectation, delight
127	Infatuation	188	Entrancement, ravishment
128	Abandon, wildness	189	Amusement
129	Ardor, ardor, fervency, fervor, fervor, fire	190	Schadenfreude
130	Zeal	191	Enjoyment
131	Storminess	192	Gusto, relish, zest, zestfulness
132	Sentiment	193	Pleasantness
133	Sentimentality	194	Comfort
134	Bathos, mawkishness	195	Consolation, solace, solacement
135	Complex	196	Alleviation, assuagement, relief
136	Ambivalence, ambivalency	197	Algolagnia, algophilia
137	Conflict	198	Sadism
138	Apathy	199	Sadomasochism
139	Emotionlessness, impassiveness, impassivity, indifference, phlegm, stolidity	200	Masochism
140	Languor,lassitude, listlessness	201	Pain, painfulness
141	Desire	202	Unpleasantness
142	Ambition,aspiration, dream	203	Hurt, suffering
143	Emulation	204	Agony, torment, torture
144	Nationalism	205	Throes
145	Bloodlust	206	Discomfort, irritation, soreness
146	Temptation	207	Distress, hurt, suffering
147	Craving	208	Anguish, torment, torture
148	Appetence, appetency, appetite	209	Self-torment, self-torture
149	Stomach	210	Tsoris
150	Addiction	211	Wound
151	Want, wish, wishing	212	Pang, stab, twinge
152	Velleity	213	Liking
153	Hungriness, longing, yearning	214	Leaning,propensity, tendency
154	Hankering, yen	215	Fancy, fondness, partiality
155	Pining	216	Captivation, enchantment, enthrallment, fascination
156	Lovesickness	217	Penchant,predilection, preference, taste
157	Wistfulness	218	Weakness
158	Nostalgia	219	Mysophilia
159	Homesickness	220	Inclination
160	Discontent,discontentment	221	Stomach
161	Disgruntlement	222	Undertow
162	Dysphoria	223	Friendliness
163	Dissatisfaction	224	Amicability, amicableness
164	Boredom ,ennui, tedium	225	Goodwill
165	Blahs	226	Brotherhood
166	Fatigue	227	Approval
167	Displeasure	228	Favor, favour
168	Disappointment, letdown	229	Approbation
169	Defeat, frustration	230	Admiration, esteem
170	Concupiscence, eros	231	Anglophilia
171	Love	232	Philhellenism
172	Aphrodisia	233	Philogyny
173	Passion	234	Dislike
174	Sensualism, sensuality, sensualness	235	Disinclination
175	Amativeness, amorousness, eroticism, erotism, sexiness	236	Anglophobia
176	Carnality, lasciviousness, lubricity, prurience, pruriency	237	Unfriendliness
177	Fetish	238	Alienation, disaffection, estrangement
178	Libido	239	Isolation
179	Lecherousness, lust, lustfulness	240	Antipathy, aversion, distaste
180	Nymphomania	241	Disapproval
181	Satyriasis	242	Contempt, despite, disdain, scorn
182	Itch, urge	243	Disgust
183	Caprice, impulse, whim	244	Abhorrence, abomination, detestation, execration, loathing, odium

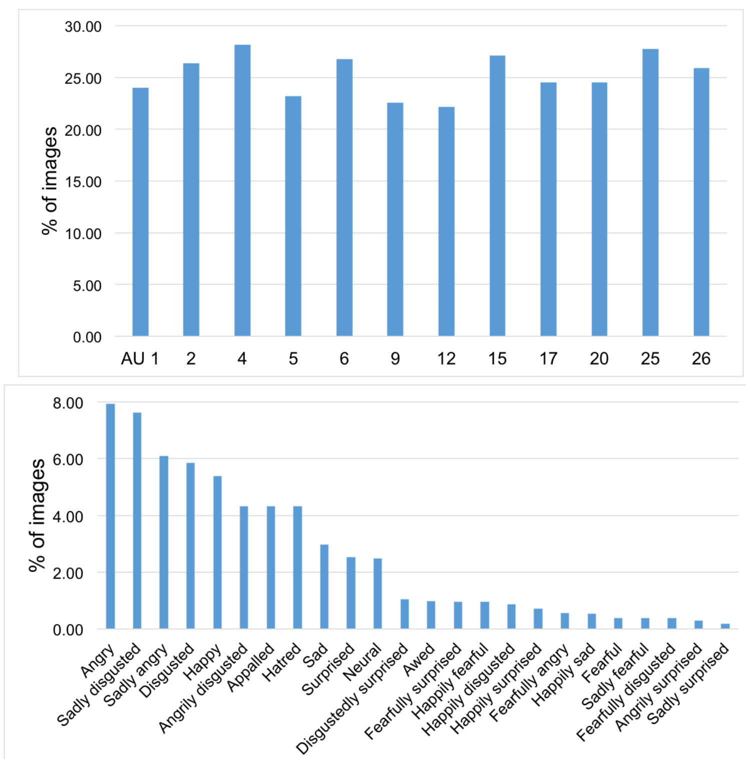
Table S3: Continues from Table S2.

245	Horror, repugnance, repulsion, revulsion	306	Sensation
246	Nausea	307	Tumult, turmoil
247	Creepy-crawlies	308	Calmness
248	Scunner	309	Placidity, placidness
249	Technophobia	310	Coolness, imperturbability
250	Antagonism	311	Dreaminess, languor
251	Gratitude	312	Bravery, fearlessness
252	Appreciativeness, gratefulness, thankfulness	313	Security
253	Ingratitude, ungratefulness	314	Confidence
254	Unconcern	315	Quietness, quietude, tranquility, tranquillity
255	Indifference	316	Ataraxis, heartsease, peace, peacefulness, repose, serenity
256	Aloofness, distance	317	Easiness,relaxation
257	Detachment, withdrawal	318	Happiness
258	Coldheartedness, hardheartedness, heartlessness	319	Bonheur
259	Cruelty, mercilessness, pitilessness, ruthlessness	320	Gladfulness, gladness, gladsomeness
260	Shame	321	Gaiety, merriment
261	Conscience	322	Glee, gleefulness, hilarity, mirth, mirthfulness
262	Self-disgust, self-hatred	323	Jocularity, jocundity
263	Embarrassment	324	Jolliness, jollity, joviality
264	Self-consciousness, uncomfortableness, uneasiness	325	Rejoicing
265	Shamefacedness,sheepishness	326	Belonging
266	Chagrin, humiliation, mortification	327	Comfortableness
267	Confusion, discombobulation	328	Closeness, intimacy
268	Abashment, bashfulness	329	Togetherness
269	Discomfiture, discomposure, disconcertion, disconcertment	330	Blitheness, cheerfulness
270	Pride, pridefulness	331	Buoyancy, perkiness
271	Dignity, self-regard, self-respect, self-worth	332	Carefreeness, insouciance, lightheartedness, lightsomeness
272	Self-esteem, self-pride	333	Contentment
273	Ego, egotism, self-importance	334	Satisfaction
274	Conceit, self-love, vanity	335	Pride
275	Humbleness, humility	336	Complacence, complacency, self-complacency, self-satisfaction
276	Meekness, submission	337	Smugness
277	Self-depreciation	338	Fulfillment, fulfilment
278	Amazement, astonishment	339	Gloat, gloating
279	Admiration,wonder, wonderment	340	Sadness, unhappiness
280	Awe	341	Dolefulness
281	Surprise	342	Heaviness
282	Stupefaction	243	Melancholy
283	Daze, shock, stupor	344	Gloom, gloominess, somberness, somberness
284	Devastation	345	Heavyheartedness
285	Expectation	346	Brooding, pensiveness
286	Anticipation, expectancy	247	Weltschmerz, world-weariness
287	Suspense	248	Misery
288	Fever	349	Desolation, forlornness, loneliness
289	Hope	350	Tearfulness, weepiness
290	Levity	351	Sorrow
291	Gaiety, playfulness	352	Brokenheartedness, grief, heartache, heartbreak
292	Gravity, solemnity	353	Dolor, dolour
293	Earnestness, seriousness, sincerity	354	Mournfulness, ruthfulness, sorrowfulness
294	Sensitiveness, sensitivity	355	Woe, woefulness
295	Sensibility	356	Plaintiveness
296	Insight, perceptiveness, perceptivity	357	Self-pity
297	Sensuousness	358	Regret, rue, ruefulness, sorrow
298	Feelings	359	Attrition, contriteness, contrition
299	Agitation	360	Compunction, remorse, self-reproach
300	Unrest	361	Guilt
301	Fidget, fidgetiness, restlessness	362	Penance, penitence, repentance
302	Impatience	363	Cheerlessness, uncheerfulness
303	Stewing	364	Joylessness
304	Stir	365	Depression
305	Electricity	366	Demoralization

Table S4: Continues from Tables S2-S3.

367	Helplessness	395	Jolliness, jollity, joviality
368	Despondence, despondency, disconsolateness, heartsickness	396	Distemper
369	Oppression, oppressiveness	397	Moodiness
370	Weight	398	Glumness, moroseness, sullenness
371	Dysphoria	399	Testiness, tetchiness, touchiness
372	Dejectedness, dispiritedness, downheartedness, low-spiritedness, lowness	400	Technophilia
373	Hope	401	Pet
374	Hopefulness	402	Sympathy
375	Encouragement	403	Concern
376	Optimis	404	Solicitousness, solicitude
377	Sanguineness, sanguinity	405	Softheartedness, tenderness
378	Despair	406	Kind-heartedness, kindheartedness
379	Hopelessness	407	Mellowness
380	Resignation, surrender	408	Exuberance
381	Defeatism	409	Compassion, compassionateness
382	Discouragement, disheartenment, dismay	410	Heartstrings
383	Intimidation	411	Tenderheartedness, tenderness
384	Pessimism	412	Ardor, ardour, elan, zeal
385	Cynicism	413	Mercifulness, mercy
386	Affection, affectionateness, fondness, heart, philia, tenderness warmheartedness, warmness	414	Choler, crossness, fretfulness, fussiness, irritability, peevishness, petulance
387	Attachment	415	Forgiveness
389	Protectiveness	416	Commiseration, pathos, pity, ruth
390	Regard, respect	417	Compatibility
391	Humor, mood, temper	418	Empathy
392	Peeve	419	Enthusiasm
393	Sulk, sulkiness	420	Gusto, relish, zest, zestfulness
394	Amiability	421	Avidity, avidness, eagerness, keenness

Table S5: Continues from Tables S2-S4.



(a)

(b)

Figure S5: (a) Percentage of images (y-axis) automatically annotated with AU i (x-axis). (b) Percentage of images (y-axis) automatically annotated with one of the 23 basic or compound emotion categories (x-axis) listed in Table 1.

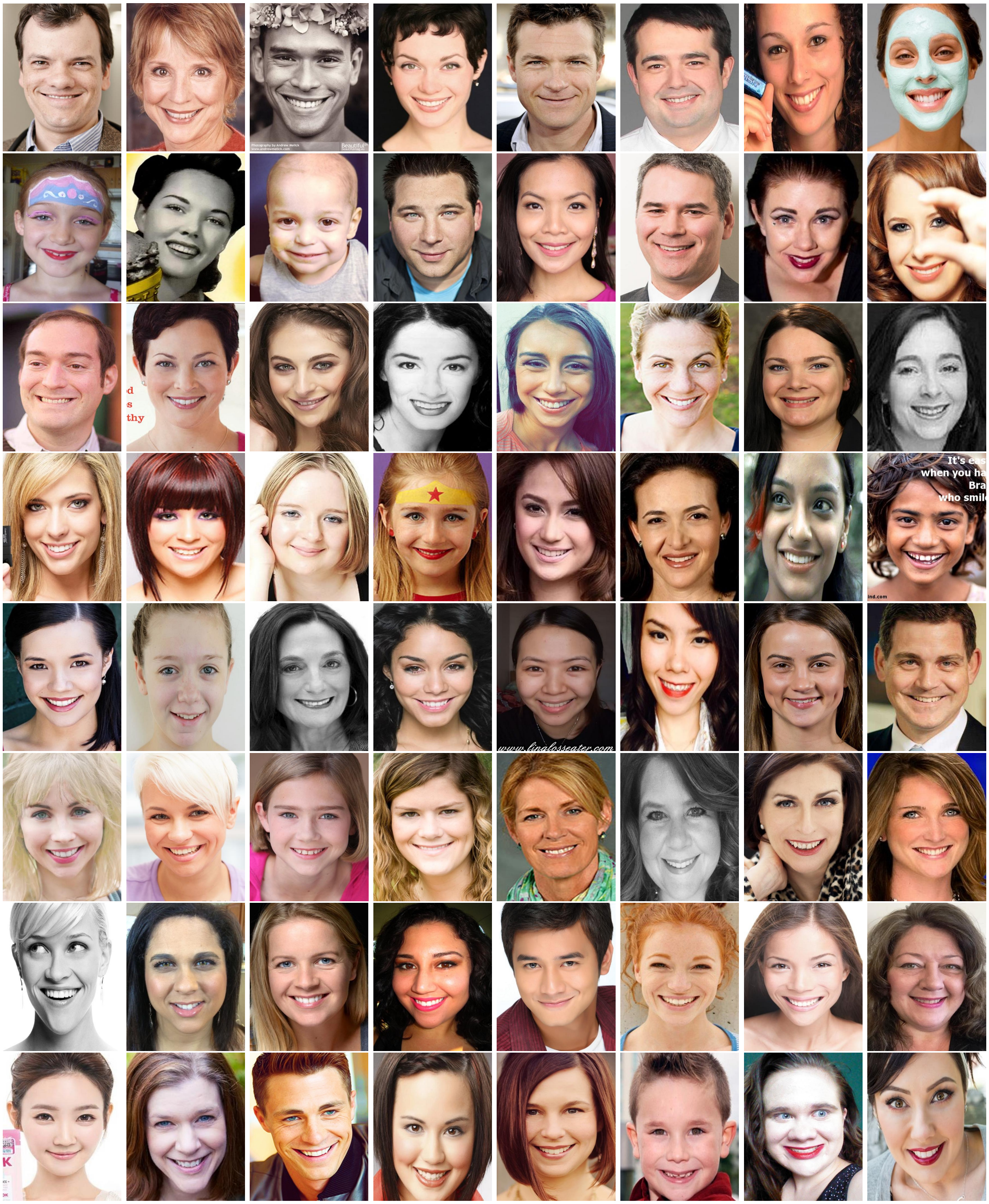


Figure S6: Sample images with AU 12 automatically annotated by our algorithm. The images are ranked according to the probability of AU 12 being active in the image.

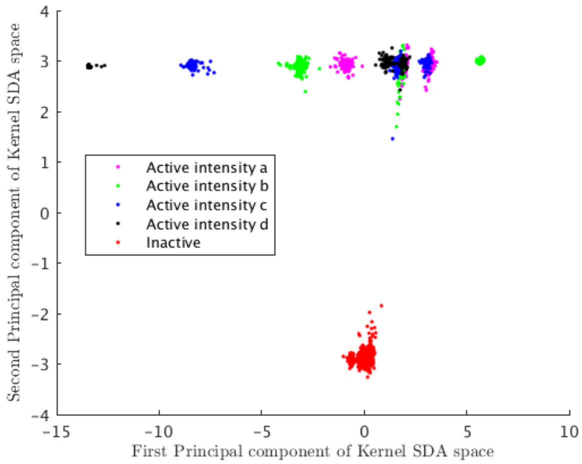


Figure S7: The first two KSDA components of the face space of an AU. Different colors correspond to distinct intensities of the AU. Note how some intensities are divided into subclasses, whereas others are not.

5. Subclass-based Representation of AUs

A key component of our algorithm is to assign the images with AU i active to distinct subclasses as a function of their intensity of activation. That is, images that show AU i active at intensity a are assigned to a subclass of class i, images showing AU i active at intensity b are assigned to a second subclass of class i, images showing AU i active at intensity c are assigned to a third subclass of class i, and images showing AU i active at intensity d are assigned to a fourth subclass of class i. This innovative approach is what allows us to simultaneously identify AUs and their intensities quickly and accurately in images.

This approach is related to the work of Subclass Discriminant Analysis (SDA) [12], which is a mathematical formulation specifically derived to identify the optimal number of subclasses to maximize spreadability of samples in different classes even when these are not defined by a Normal distribution. This is achieved by minimizing a criterion defined in [6], which guarantees Bayes optimality in this classification process under mild conditions.

The approach derived in the present paper is different in that we specify the initial subclass division, rather than using the Bayes criterion defined in [6]. Specifically, we derive a Kernel (SDA-inspired) algorithm to learn to simultaneously identify AUs and their intensities in images. This is done by first dividing the training data of each AU into five sets – one for each of the four intensities, $\mathcal{D}_i(a)$ to $\mathcal{D}_i(d)$, and another set to include the images that do not have that AU active $\mathcal{D}_i(\texttt{not active}) = \mathcal{D}_i - \cup_{j=a,b,c,d}\mathcal{D}_i(j)$. Thus, the initial number of subclasses for class AU i active is 4, i.e., $h_{i1} = 4$, and, the initial number of subclasses for AU i not active (i.e., not present) in the images is 1, i.e., $h_{i2} = 1$. This was illustrated in Figure 3 in the main paper. A 2D plot, for one of the AUs, with real data is now shown in Figure S7. Also, the sample images in each of these five sets are sorted using the nearest-neighbor algorithm of [12].

Next, we use the criterion derived in the main paper, $Q_i(\varphi_i, h_{i1}, h_{i2})$, to further optimize the number of classes and the parameters of the kernel mapping function. This criterion maximizes spherical-homoscedasticity in the RBF kernel space, which is known to minimize the Bayes classification error [4]. Note that, in the paper, we used the RBF kernel, but other options are possible, with each one yielding a different optimization function $Q(.)$.

6. Extended Discussion

The ability to automatically annotate facial action units in the wild in real time is likely to revolutionize research in the study or non-verbal communication and emotion theory. To date, most studies have focused on the analysis of data collected in the laboratory, even when this data corresponds to spontaneous facial expressions. Extending these studies to facial expressions in the wild is a necessary step.

The algorithm described in the present work achieves this goal, allowing researchers to analyze their data quickly and reliably. As a plus, our system is consistent with what is known about the visual perception of facial expressions of emotion by humans [3]. In fact, a recent result from our laboratory has identified a small region of interest in the human brain dedicated to the visual interpretation of facial actions [9]. The computer vision system defined in this paper could thus also help us advance our understanding of human perception.

References

[1] C. F. Benitez-Quiroz, R. B. Wilbur, and A. M. Martinez. The not face: A grammaticalization of facial expressions of emotion. *Cognition*, 150:77–84, 2016. 5

[2] S. Du and A. M. Martinez. Compound facial expressions of emotion: from basic research to clinical applications. *Dialogues in clinical neuroscience*, 17(4):443, 2015. 4

[3] S. Du, Y. Tao, and A. M. Martinez. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, 111(15):E1454–E1462, 2014. 4, 12

[4] O. C. Hamsici and A. M. Martinez. Spherical-homoscedastic distributions: The equivalency of spherical and normal distributions in classification. *Journal of Machine Learning Research*, 8(1583-1623):1–3, 2007. 12

[5] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin. Automatically detecting pain in video through facial action units. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 41(3):664–674, 2011. 1, 3

[6] A. M. Martinez and M. Zhu. Where are linear feature extraction methods applicable? *IEEE Transactions on Pat-*

tern Analysis and Machine Intelligence, 27(12):1934–1944, 2005. 12

[7] D. McDuff, R. Kaliouby, T. Senechal, M. Amr, J. Cohn, and R. Picard. Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Workshops*, pages 881–888, 2013. 1

[8] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995. 1

[9] R. Srinivasan, J. D. Golomb, and A. M. Martine. A neural basis of facial action recognition in humans. *The Journal of Neuroscience*, 2016. 12

[10] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004. 1

[11] Z. Zafar and N. A. Khan. Pain intensity evaluation through facial action units. In *Pattern Recognition (ICPR), 22nd International Conference on*, pages 4696–4701. IEEE, 2014. 1, 3

[12] M. Zhu and A. M. Martinez. Subclass discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1274–1286, 2006. 12