Supplementary: Adversarial Invariant Feature Learning with Accuracy Constraint for Domain Generalization

Kei Akuzawa[1], Yusuke Iwasawa[1], and Yutaka Matsuo[1]

The University of Tokyo, Tokyo, Japan
{akuzawa-kei,iwasawa,matsuo}@weblab.t.u-tokyo.ac.jp

Sample Sizes for PACS, IEMOCAP, and WISDM

Tables 1-3 show sample sizes for each class and domain for PACS, IEMOCAP, and WISDM, respectively.

Table 1. Sample sizes for each <domain, class> pair in PACS dataset. The column shows category name while and index shows style.

	Guitar	House	Giraffe	Person	Horse	Dog	Elephant
Art Painting	184	295	285	449	201	379	255
Cartoon	135	288	346	405	324	389	457
Photo	186	280	182	432	199	189	202
Sketch	608	80	753	160	816	772	740

Table 2. Sample sizes for each <domain, class> pair in IEMOCAP dataset. The column shows emotion name while and index shows actor id.

	hap	ang	neu	sad
Ses01F	45	67	222	191
Ses02F	54	31	456	124
Ses03F	40	45	127	198
Ses04F	16	208	190	194
Ses05F	62	79	371	186
Ses01M	25	69	161	129
Ses02M	67	44	263	218
Ses03M	89	128	267	174
Ses04M	98	39	163	164
Ses05M	178	13	292	189

Table 3. Sample sizes for each <domain, class> pair in WISDM dataset. The column shows activity name and the index shows user id.

	Jogging	Walking	Upstairs	Downstairs	Sitting	Standing
User 1	49	248	36	75	54	26
User 2	48	161	94	62	0	0
User 3	215	218	80	77	260	89
User 4	213	207	79	72	38	26
User 5	205	217	77	70	19	27
User 6	213	192	34	29	0	0
User 7	196	206	27	23	27	11
User 8	200	207	54	57	34	27
User 9	200	103	90	69	41	32
User 10	199	209	40	40	24	32
User 11	204	206	63	39	50	27
User 12	209	119	0	0	26	17
User 13	207	202	73	44	0	0
User 14	0	208	23	26	49	32
User 15	106	204	56	54	27	25
User 16	201	217	71	63	0	27
User 17	0	236	48	49	0	21
User 18	198	220	60	63	0	0
User 19	221	230	136	47	0	0
User 20	204	104	50	48	11	9
User 21	206	179	44	47	38	27
User 22	205	109	80	32	0	0
User 23	14	101	22	29	20	0
User 24	0	209	70	64	25	51
User 25	214	222	65	47	26	22
User 26	171	285	74	55	44	54
User 27	234	281	77	64	35	43
User 28	159	208	80	67	26	47
User 29	183	216	56	55	26	47
User 30	103	117	90	60	0	0
User 31	184	214	52	49	0	0
User 32	0	215	0	0	0	0
User 33	108	116	0	0	0	0
User 34	196	195	0	0	0	0
User 35	153	183	60	37	42	39
User 36	270	293	71	43	42	35

Network architectures for BMNISTR, IEMOCAP, and WISDM

Tables 4-6 show details of the DNN architectures. In these table, "Conv$n$D(C_{in}, C_{out}, k, s)" denotes a n-dimensional convolution layer. Here, C_{in} and C_{out} mean the number of channels of input and output, and k and s denote the convolution window size and stride width, respectively. "Linear(in, out)" denotes a linear layer with in-dimensional input and out-dimenstional output features. "MaxPool$n$D(k)" denotes a n-dimensional max pooling layer with window size k. "$DO(p)$" means dropout whose ratio was set to p.

Table 4. DNN architectures used on BMNISTR.

Encoder	
Conv2D(1, 32, 5, 1)	ReLU
Conv2D(32, 48, 5, 1)	ReLU MaxPool2D(2)
Linear(768, 100)	ReLU
Linear(100, 100)	ReLU
Classifier	
Linear(100, 100)	ReLU
Linear(100, 100)	ReLU
Linear(100, 10)	
Discriminator	
Linear(100, 100)	ReLU
Linear(100, 5)	

Table 5. DNN architectures used on WISDM.

Encoder	
Conv1D(3, 50, 5, 1)	ReLU MaxPool1D(2)
Conv1D(50, 40, 5, 1)	ReLU MaxPool1D(2)
Conv1D(40, 20, 3, 1)	ReLU DP(0.5)
Linear(200, 400)	ReLU DP(0.5)
Classifier	
Linear(400, 6)	
Discriminator	
Linear(400, 400)	ReLU DP(0.5)
Linear(400, 20)	

Table 6. DNN architectures used on IEMOCAP.

Encoder	
Conv2D(40, 64, (10, 3), (4, 2))	ReLU DP(0.5)
Conv2D(64, 128, (10, 3), (4, 2))	ReLU DP(0.5)
Conv2D(128, 128, (5, 3), (4, 2))	ReLU DP(0.5)
Linear(3200, 128)	ReLU
Classifier	
Linear(128, 4)	
Discriminator	
Linear(128, 100)	ReLU
Linear(100, 8)	