

000		054																																																																																										
001		055																																																																																										
002		056																																																																																										
003		057																																																																																										
004	Context Aware Online Adaptation of Activity Recognition Models	058																																																																																										
005		059																																																																																										
006		060																																																																																										
007		061																																																																																										
008		062																																																																																										
009	Anonymous ICCV submission	063																																																																																										
010		064																																																																																										
011	Paper ID 1873	065																																																																																										
012		066																																																																																										
013		067																																																																																										
014		068																																																																																										
015		069																																																																																										
016		070																																																																																										
017		071																																																																																										
018	Table of Contents																																																																																											
019	<table border="1"><thead><tr><th>Number</th><th>Contents</th><th>Section Link</th></tr></thead><tbody><tr><td colspan="3">High resolution plots - Figure 5 of the main paper 1</td></tr><tr><td>1</td><td>VIRAT</td><td>1.1</td></tr><tr><td>2</td><td>UCLA-Office</td><td>1.2</td></tr><tr><td>3</td><td>MPII-Cooking</td><td>1.3</td></tr><tr><td>4</td><td>UCF50</td><td>1.4</td></tr><tr><td colspan="3">Dataset description 2</td></tr><tr><td>5</td><td>VIRAT</td><td>2.1</td></tr><tr><td>6</td><td>UCLA-Office</td><td>2.2</td></tr><tr><td>7</td><td>MPII-Cooking</td><td>2.3</td></tr><tr><td>8</td><td>UCF50</td><td>2.4</td></tr><tr><td colspan="3">Context features 3</td></tr><tr><td>9</td><td>VIRAT</td><td>3.1</td></tr><tr><td>10</td><td>UCLA-Office</td><td>3.2</td></tr><tr><td>11</td><td>MPII-Cooking</td><td>3.3</td></tr><tr><td>12</td><td>UCF50</td><td>3.4</td></tr><tr><td colspan="3">Evaluation of continuous learning on individual activities 4</td></tr><tr><td>13</td><td>VIRAT</td><td>4.1</td></tr><tr><td>14</td><td>UCLA-Office</td><td>4.2</td></tr><tr><td>15</td><td>MPII-Cooking</td><td>4.3</td></tr><tr><td>16</td><td>UCF50</td><td>4.4</td></tr><tr><td colspan="3">Activity class-wise performance evaluation 5</td></tr><tr><td>17</td><td>VIRAT</td><td>5.1</td></tr><tr><td>18</td><td>UCLA-Office</td><td>5.2</td></tr><tr><td>19</td><td>MPII-Cooking</td><td>5.3</td></tr><tr><td>20</td><td>UCF50</td><td>5.4</td></tr><tr><td colspan="3">Parameter Values 6</td></tr><tr><td>21</td><td>All Datasets</td><td>6</td></tr><tr><td colspan="3">Parameter Sensitivity 6.1</td></tr><tr><td>22</td><td>VIRAT</td><td>6.1.1</td></tr></tbody></table>	Number	Contents	Section Link	High resolution plots - Figure 5 of the main paper 1			1	VIRAT	1.1	2	UCLA-Office	1.2	3	MPII-Cooking	1.3	4	UCF50	1.4	Dataset description 2			5	VIRAT	2.1	6	UCLA-Office	2.2	7	MPII-Cooking	2.3	8	UCF50	2.4	Context features 3			9	VIRAT	3.1	10	UCLA-Office	3.2	11	MPII-Cooking	3.3	12	UCF50	3.4	Evaluation of continuous learning on individual activities 4			13	VIRAT	4.1	14	UCLA-Office	4.2	15	MPII-Cooking	4.3	16	UCF50	4.4	Activity class-wise performance evaluation 5			17	VIRAT	5.1	18	UCLA-Office	5.2	19	MPII-Cooking	5.3	20	UCF50	5.4	Parameter Values 6			21	All Datasets	6	Parameter Sensitivity 6.1			22	VIRAT	6.1.1	072
Number	Contents	Section Link																																																																																										
High resolution plots - Figure 5 of the main paper 1																																																																																												
1	VIRAT	1.1																																																																																										
2	UCLA-Office	1.2																																																																																										
3	MPII-Cooking	1.3																																																																																										
4	UCF50	1.4																																																																																										
Dataset description 2																																																																																												
5	VIRAT	2.1																																																																																										
6	UCLA-Office	2.2																																																																																										
7	MPII-Cooking	2.3																																																																																										
8	UCF50	2.4																																																																																										
Context features 3																																																																																												
9	VIRAT	3.1																																																																																										
10	UCLA-Office	3.2																																																																																										
11	MPII-Cooking	3.3																																																																																										
12	UCF50	3.4																																																																																										
Evaluation of continuous learning on individual activities 4																																																																																												
13	VIRAT	4.1																																																																																										
14	UCLA-Office	4.2																																																																																										
15	MPII-Cooking	4.3																																																																																										
16	UCF50	4.4																																																																																										
Activity class-wise performance evaluation 5																																																																																												
17	VIRAT	5.1																																																																																										
18	UCLA-Office	5.2																																																																																										
19	MPII-Cooking	5.3																																																																																										
20	UCF50	5.4																																																																																										
Parameter Values 6																																																																																												
21	All Datasets	6																																																																																										
Parameter Sensitivity 6.1																																																																																												
22	VIRAT	6.1.1																																																																																										
020		073																																																																																										
021		074																																																																																										
022		075																																																																																										
023		076																																																																																										
024		077																																																																																										
025		078																																																																																										
026		079																																																																																										
027		080																																																																																										
028		081																																																																																										
029		082																																																																																										
030		083																																																																																										
031		084																																																																																										
032		085																																																																																										
033		086																																																																																										
034		087																																																																																										
035		088																																																																																										
036		089																																																																																										
037		090																																																																																										
038		091																																																																																										
039		092																																																																																										
040		093																																																																																										
041		094																																																																																										
042		095																																																																																										
043		096																																																																																										
044		097																																																																																										
045		098																																																																																										
046		099																																																																																										
047		100																																																																																										
048		101																																																																																										
049		102																																																																																										
050		103																																																																																										
051		104																																																																																										
052		105																																																																																										
053		106																																																																																										
		107																																																																																										

108

1. High Resolution Plots

109

1.1. VIRAT

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

216 1.2. UCLA-Office
217

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

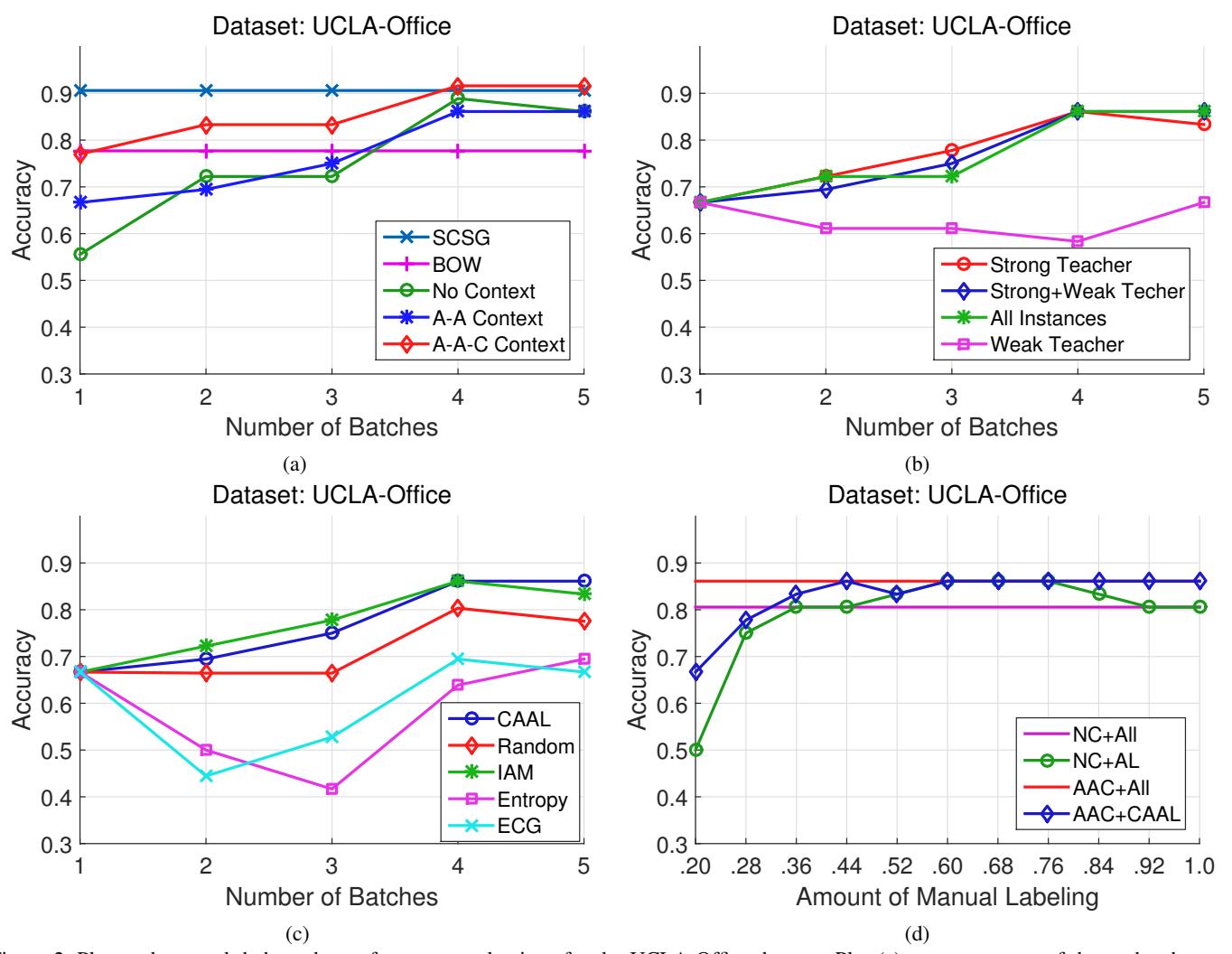
319

320

321

322

323



251 Figure 2. Plots a, b, c, and d show the performance evaluations for the UCLA-Office dataset. Plot (a) compares state-of-the-art batch
252 and incremental methods against the three variants of the proposed approach. These variants are No Context (it only uses the appearance
253 features with a softmax classifier), A-A context (it uses the spatial-temporal inter-relationship contexts of the activities represented by a
254 CRF), and A-A-C context (it uses both of the associated object contexts and spatial-temporal inter-relationship context represented by
255 a CRF). A-A-C test case performs better than all other methods with very less amount of manually labeled data. Plot (b) analyzes the
256 performances of the four variants of our proposed active learning system. These variants are based on the use of two types of teachers.
257 These variants are Strong Teacher (it manually labels a fraction of the incoming instances), Weak Teacher (it does not use any manually
258 labeled data, only uses the highly confident labels provided by the classifier), Strong+Weak Teacher (it uses the labels provided by both of
259 the above teachers.), and All Instances (it is the most expensive case, where all of the instances are manually labeled). Performance of the
260 Strong+Weak Teacher is almost similar to All Instances but with very less amount of manually labeled data. It proves the robustness of
261 our proposed framework. Plot (c) compares our proposed context aware active learning system (CAAL) against the recent active learning
262 methods and random sampling. CAAL uses the labels provided by both of the strong and the weak teachers. Recent active learning methods
263 are - incremental activity modeling (IAM), uncertainty of the nodes (Entropy), and expected change of gradients (ECG). Our proposed
264 active learning system performs better than other active learning approaches. Plot (d) compares the accuracy against the amount of manual
265 labeling. NC does not use any context information, whereas AAC use both of the spatial-temporal and the object contexts. NC+All and
266 AAC+All test cases use all manually labeled data. It is evident that AAC+CAAL uses approximately forty percent manually labeled data
267 in order to achieve similar performance comparing to batch method that use all manually labeled data.

324

1.3. MPII-Cooking

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

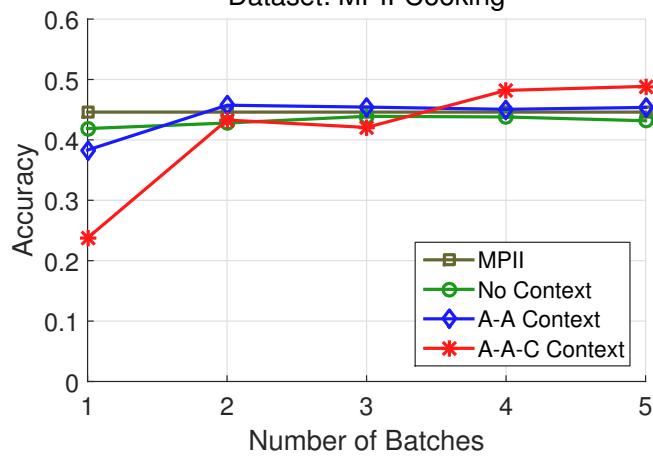
428

429

430

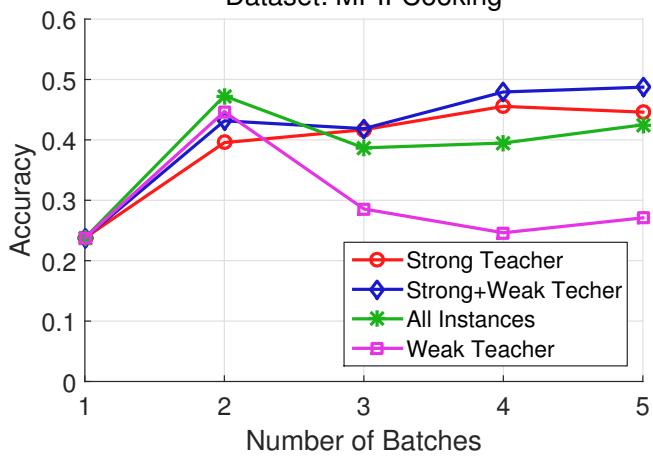
431

Dataset: MPII-Cooking



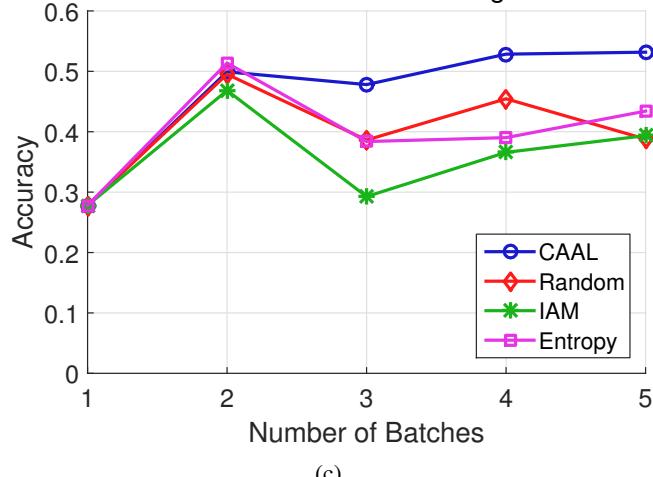
(a)

Dataset: MPII-Cooking



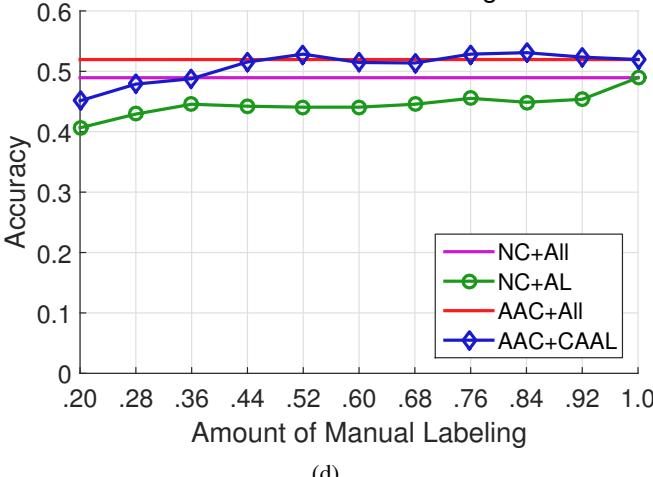
(b)

Dataset: MPII-Cooking



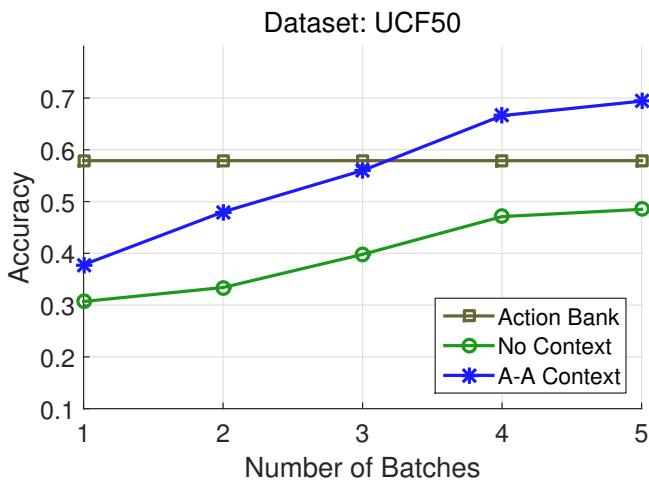
(c)

Dataset: MPII-Cooking

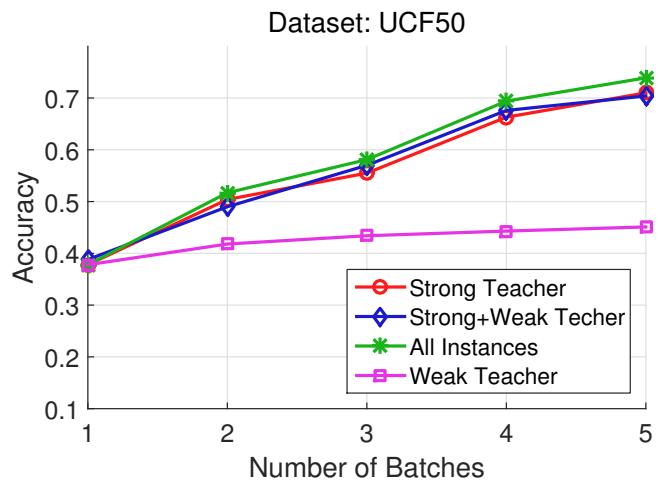


(d)

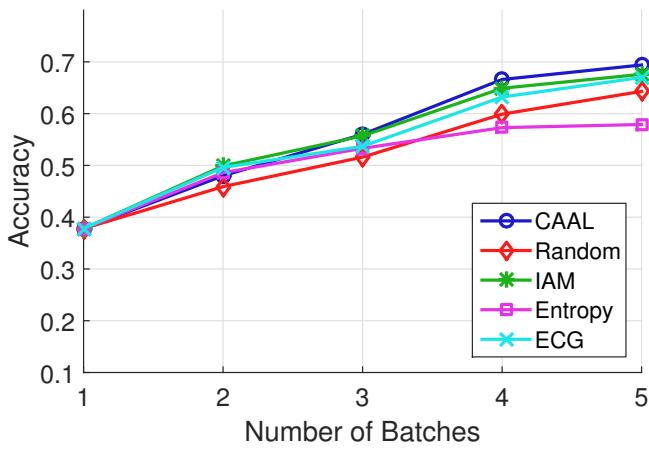
Figure 3. Plots a, b, c, and d show the performance evaluations for the MPII-Cooking dataset. Plot (a) compares state-of-the-art batch and incremental methods against the three variants of the proposed approach. These variants are No Context (it only uses the appearance features with a svm classifier), A-A context (it uses the spatial-temporal inter-relationship contexts of the activities represented by a CRF), and A-A-C context (it uses both of the associated object contexts and spatial-temporal inter-relationship context represented by a CRF). A-A-C test case performs better than all other methods with very less amount of manually labeled data. Plot (b) analyzes the performances of the four variants of our proposed active learning system. These variants are based on the use of two types of teachers. These variants are Strong Teacher (it manually labels a fraction of the incoming instances), Weak Teacher (it does not use any manually labeled data, only uses the highly confident labels provided by the classifier), Strong+Weak Teacher (it uses the labels provided by both of the above teachers.), and All Instances (it is the most expensive case, where all of the instances are manually labeled). Performance of the Strong+Weak Teacher is almost similar to All Instances but with very less amount of manually labeled data. It proves the robustness of our proposed framework. Plot (c) compares our proposed context aware active learning system (CAAL) against the recent active learning methods and random sampling. CAAL uses the labels provided by both of the strong and the weak teachers. Recent active learning methods are - incremental activity modeling (IAM), uncertainty of the nodes (Entropy), and expected change of gradients (ECG). Our proposed active learning system performs better than other active learning approaches. Plot (d) compares the accuracy against the amount of manual labeling. NC does not use any context information, whereas AAC use both of the spatial-temporal and the object contexts. NC+All and AAC+All test cases use all manually labeled data. It is evident that AAC+CAAL uses approximately forty percent manually labeled data in order to achieve similar performance comparing to batch method that use all manually labeled data.

432 1.4. UCF50
433
434
435
436

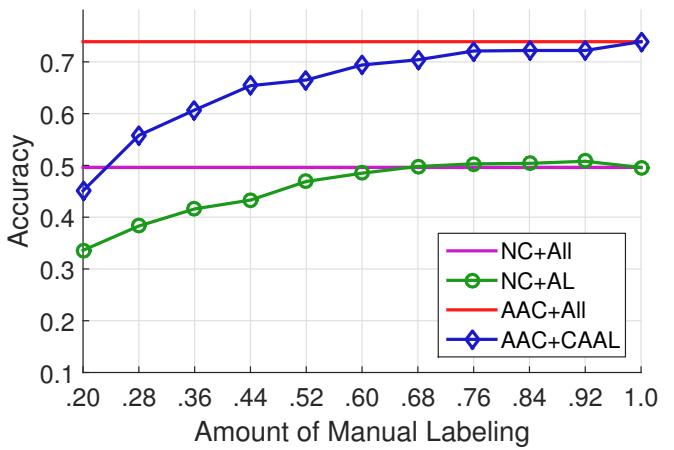
(a)



(b)



(c)



(d)

467 Figure 4. Plots a, b, c, and d show the performance evaluations for the UCF50 dataset. Plot (a) compares state-of-the-art batch and
468 incremental methods against the three variants of the proposed approach. These variants are No Context (it only uses the appearance
469 features with a softmax classifier) and A-A context (it uses the spatial-temporal inter-relationship contexts of the activities represented
470 by a CRF). A-A test case performs better than all other methods with very less amount of manually labeled data. Plot (b) analyzes the
471 performances of the four variants of our proposed active learning system. These variants are based on the use of two types of teachers.
472 These variants are Strong Teacher (it manually labels a fraction of the incoming instances), Weak Teacher (it does not use any manually
473 labeled data, only uses the highly confident labels provided by the classifier), Strong+Weak Teacher (it uses the labels provided by both of
474 the above teachers.), and All Instances (it is the most expensive case, where all of the instances are manually labeled). Performance of the
475 Strong+Weak Teacher is almost similar to All Instances but with very less amount of manually labeled data. It proves the robustness of
476 our proposed framework. Plot (c) compares our proposed context aware active learning system (CAAL) against the recent active learning
477 methods and random sampling. CAAL uses the labels provided by both of the strong and the weak teachers. Recent active learning methods
478 are - incremental activity modeling (IAM), uncertainty of the nodes (Entropy), and expected change of gradients (ECG). Our proposed
479 active learning system performs better than other active learning approaches. Plot (d) compares the accuracy against the amount of manual
480 labeling. NC does not use any context information, whereas AAC use both of the spatial-temporal and the object contexts. NC+All and
481 AAC+All test cases use all manually labeled data.

482 Go to Table of Contents
483
484
485

540

2. Dataset Description

594

541

2.1. VIRAT

595

542

Description	The VIRAT dataset is a state-of-the-art human action dataset with many challenging characteristics such as wide variation in the activities and a high amount of occlusion and clutter. It consists of surveillance videos such as parking lot videos involving single vehicle activities, person and vehicle interactions, and people interactions. There are also some group activities. This dataset consists of scenes captured on a single camera although the viewpoint can differ from one scene to the next. In any scene, the activities can occur at different orientations depending on the location. However, since these are wide-area videos, persons of interest are usually far away from the camera.
Number of Scenes	11
Number of Sequences	329
Number of Activities	1555
Number of Activity Types	11
Activity Types	Loading an object to a vehicle, Unloading an object from a vehicle, Opening a vehicle trunk, Closing a vehicle trunk, Getting into a vehicle, Getting out of a vehicle, Gesturing, Person carrying an object, Person running, Person entering a facility, and Person exiting a facility.
Associated Object Types	Person, Car, Vehicle, Carrying objects, and Bike.
Video Resolution	1920 × 1080
Total Video Duration	About 5 hours
Wild?	Yes
Segmented?	No
Background	Fixed for a sequence

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

Go to Table of Contents

2.2. UCLA-Office

625

626

627

Description	The UCLA office dataset consists of indoor and outdoor videos of single and two-person activities. Here, we perform experiments on the lab scene containing close to 35 minutes of video captured with a single fixed camera in a room. There is very little variation in viewpoint, occlusion and scale here. Each activity occurs 6 to 15 times in the dataset.
Number of Scenes	1
Number of Sequences	3
Number of Activities	157
Number of Activity Types	10
Activity Types	EnterRoom, ExitRoom, SitDown, StandUp, WorkLaptop, WorkPaper, ThrowTrash, PourDrink, PickPhone, PlacePhone.
Associated Object Types	Laptop, Phone, Paper, Trash, etc.
Video Resolution	1280 × 720
Total Video Duration	35 minutes
Wild?	No
Segmented?	No
Background	Fixed

Go to Table of Contents

648

2.3. MPII-Cooking

702

649

703

650

704

651

705

652

706

653

707

654

708

655

709

656

710

657

711

658

712

659

713

660

714

661

715

662

716

663

717

664

718

665

719

666

720

667

721

668

722

669

723

670

724

671

725

672

726

673

727

674

728

Description	This dataset contains fine-grained cooking activities in an indoor settings. 12 participants performed 65 different cooking activities. Participants were asked to prepare one to six of a total of 14 dishes such as fruit salad or cake containing several cooking activities. Activities have low inter-class variability and high intra-class variability due to diverse subjects and cooking ingredients. However, activities have very little occlusion, clutter, or change of viewpoints.
Number of Subjects	12
Number of Sequences	44
Number of Activities	5609
Number of Activity Types	65
Activity Types	background, changeTemp, cutApart, cutDice, cutIn, cutOffEnds, cutOutInside, cutSlices, cutStripes, dry, fillWaterFromTap, grate, lidPutOn, lidRemove, mix, move, openEgg, openTin, openCloseCupboard, openCloseDrawer, openCloseFridge, openCloseOven, package, peel, plugInOut, pour, pullOut, puree, putInBowl, putInPanPot, putOnBreadDough, putOnCuttingBoard, putOnPlate, read, removeFromPackage, ripOpen, scratchOff, screwClose, screwOpen, shake, smell, spice, spread, squeeze, stamp, stir, strew, takePutInCupboard, takePutInDrawer, takePutInFridge, takePutInOven, TakePutInSpiceHolder, takeIngredientApart, takeOutFromCupboard, takeOutFromDrawer, takeOutFromFridge, takeOutFromOven, takeOutFromSpiceHolder, taste, throwInGarbage, unrollDough, washHands, washObjects, whisk, wipeClean.
Associated Object Types	hand, bottle, cap opener, mug, beer, tomato, etc.
Video Resolution	1624 × 1224
Total Video Duration	8 hours 20 minutes
Wild?	No
Segmented?	No
Background	Fixed

681

735

682

736

683

737

684

738

685

739

686

740

687

741

688

742

689

743

690

744

691

745

692

746

693

747

694

748

695

749

696

750

697

751

698

752

699

753

700

754

701

755

756

2.4. UCF50

810

757

811

758

812

759

813

760

814

761

815

762

816

763

817

764

818

765

819

Description	This dataset contains various types of sports activities in the wild. It has very low inter-class and very high intra-class variability. Activities have very high amount of occlusion, clutter and change of viewpoints.
Number of Subjects	25
Number of Sequences	6676
Number of Activities	6676
Number of Activity Types	50
Activity Types	BaseballPitch, Basketball, BenchPress, Biking, Billiards, BreastStroke, CleanAndJerk, Diving, Drumming, Fencing, GolfSwing, HighJump, HorseRace, HorseRiding, HulaHoop, JavelinThrow, JugglingBalls, JumpRope, JumpingJack, Kayaking, Lunges, MilitaryParade, Mixing, Nunchucks, PizzaTossing, PlayingGuitar, PlayingPiano, PlayingTabla, PlayingViolin, PoleVault, PommelHorse, PullUps, Punch, PushUps, RockClimbingIndoor, RopeClimbing, Rowing, SalsaSpin, SkateBoarding, Skiing, Skijet, SoccerJuggling, Swing, TaiChi, TennisSwing, ThrowDiscus, TrampolineJumping, VolleyballSpiking, WalkingWithDog, YoYo.
Associated Object Types	—
Video Resolution	320 × 240
Total Video Duration	About 8 hours
Wild?	Yes
Segmented?	Yes
Background	Dynamic

782

829

783

830

784

831

785

832

786

833

787

834

788

835

789

836

790

837

791

838

792

839

793

840

794

841

795

842

796

843

797

844

798

845

799

846

800

847

801

848

802

849

803

850

804

851

805

852

806

853

807

854

808

855

809

856

[Go to Table of Contents](#)

864	3. Context Features	918
865		919
866	Contexts may vary based on the application domain. Our proposed generalized framework can handle any number and	920
867	types of contexts. In this paper, we use different contexts for different datasets. We describe them below.	921
868		922
869	3.1. VIRAT	923
870	We use both of the spatial-temporal relationships among the activities context and object contexts for this dataset. These	924
871	contextual features have been described in the main paper in Equations 2-9.	925
872		926
873	3.2. UCLA-Office	927
874	We only use the spatial-temporal relationships among the activities context for this dataset. This context feature has been	928
875	described in the main paper in Equations 2 and 6.	929
876		930
877	3.3. MPII-Cooking	931
878	We use both of the spatial-temporal relationships among the activities context and object contexts for this dataset. Spatial-	932
879	temporal relationship context remains the same as in Equations 2 and 6 in the main paper. Activities in this dataset involve	933
880	three types of objects - tools (c_i^1), ingredients (c_i^2), and containers (c_i^3). We use each of them as a separate context and	934
881	formulate them as like in Equations 3, 4, 7, and 8. So the Equations 3 and 7 will become,	935
882		936
883	$\phi(c_i, z_i) = \phi(c_i^1, z_i) \odot \phi(c_i^2, z_i) \odot \phi(c_i^3, z_i)$	937
884	$\psi(a_i, c_i) = \psi(a_i, c_i^1) \otimes \psi(a_i, c_i^2) \otimes \psi(a_i, c_i^3)$	938
885		939
886	3.4. UCF50	940
887		941
888	Since the activities in UCF50 dataset are segmented, there are no natural spatial-temporal relationships exist among the	942
889	activities. Also, each activity involves a person and a particular tool. Use of object context might overfit the model. So, we	943
890	improvise a relationship among the activities. We roughly categorize fifty activity classes into eight groups and assume that in	944
891	each group activities are inter-related. These groups are - Outdoor Group Sports (BaseballPitch, Basketball, VollyballSpiking,	945
892	TennisSwing, HorseRace, and Rowing), Outdoor Individual Sports (GolfSwing, HighJump, JavelinThrow, Kayaking, Skiing,	946
893	SoccerJuggling, ThrowDiscuss, and PoleVolt), Indoor Sports (Billiards, CleanAndJerk, Fencing, PommelHorse, Punch, and	947
894	RockClimbing), Outdoor Activity (Biking, Diving, MilitaryParade, NunChucks, HorseRiding, RopeClimbing, SkateBoard-	948
895	ing, SkiJet, Swing, and TampolineJumping), Indoor Activity (SalsaSpin, BreastStroke, HulaHoop, JugglingBalls, and YoYo),	949
896	Physical Exercise (BenchPress, JumpingJack, JumpRope, TaiChi, Walking, PullUps, PushUps, and Lunges), Kitchen (Mix-	950
897	ing and PizzaTossing), and Instrumental (Drumming, PlayingGuitar, PlayingPiano, PlayingTabla, and PlayingViolin). The	951
898	corresponding mathematical formulations remain same as like Equations 2 and 6 of the main paper.	952
899		953
900	Go to Table of Contents	954
901		955
902		956
903		957
904		958
905		959
906		960
907		961
908		962
909		963
910		964
911		965
912		966
913		967
914		968
915		969
916		970
917		971

ICCV
#1873

4. Evaluation of Continuous Learning on Individual Activities

4.1. VIRAT

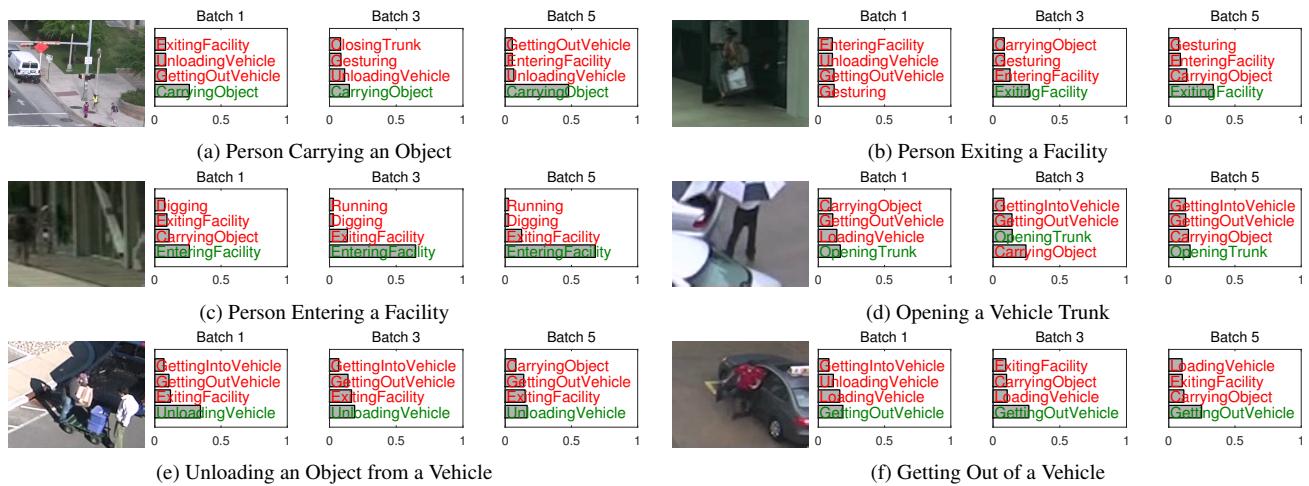


Figure 5. Evaluation of continuous learning on individual activities. Activity with green color means the ground truth class, whereas activities with red color means false predictions. Grey bars represent probability scores. Here, we show the results obtained after the arrival of batch 1, 3, and 5 data. The plots in this figure show some of the successful examples, where continuous learning helps to obtain the correct label with a higher probability even though some of them were miss-classified initially. Best viewable in color.



Figure 6. Evaluation of continuous learning on individual activities. Activity with green color means the ground truth class, whereas activities with red color means false predictions. Grey bars represent probability scores. Here, we show the results obtained after the arrival of batch 1, 3, and 5 data. The plots in this figure show few failure cases. Best viewable in color.

[Go to Table of Contents](#)

1080

4.2. UCLA-Office

1134

1081

1135

1082

1136

1083

1137

1084

1138

1085

1139

1086

1140

1087

1141

1088

1142

1089

1143

1090

1144

1091

1145

1092

1146

1093

1147

1094

1148

1095

1149

1096

1150

1097

1151

1098

1152

1099

1153

1100

1154

1101

1155

1102

1156

1103

1157

1104

1158

1105

1159

1106

1160

1107

1161

1108

1162

1109

1163

1110

1164

1111

1165

1112

1166

1113

1167

1114

1168

1115

1169

1116

1170

1117

1171

1118

1172

1119

1173

1120

1174

1121

1175

1122

1176

1123

1177

1124

1178

1125

1179

1126

1180

1127

1181

1128

1182

1129

1183

1130

1184

1131

1185

1132

1186

1133

1187

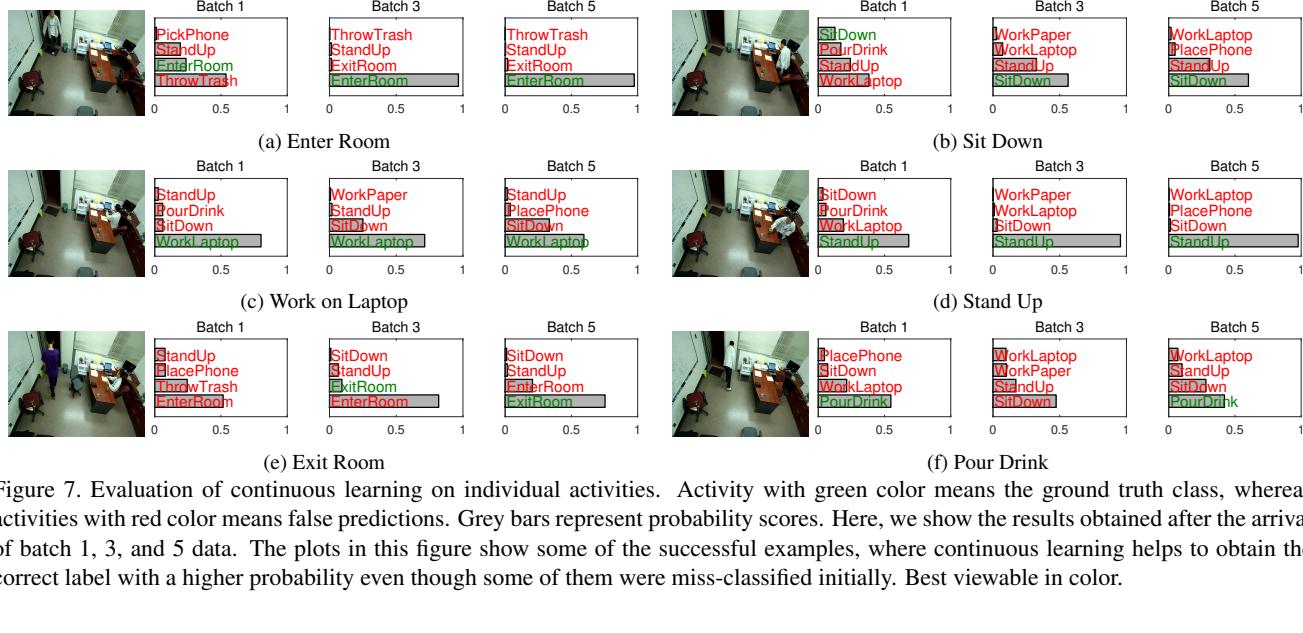


Figure 7. Evaluation of continuous learning on individual activities. Activity with green color means the ground truth class, whereas activities with red color means false predictions. Grey bars represent probability scores. Here, we show the results obtained after the arrival of batch 1, 3, and 5 data. The plots in this figure show some of the successful examples, where continuous learning helps to obtain the correct label with a higher probability even though some of them were miss-classified initially. Best viewable in color.



Figure 8. Evaluation of continuous learning on individual activities. Activity with green color means the ground truth class, whereas activities with red color means false predictions. Grey bars represent probability scores. Here, we show the results obtained after the arrival of batch 1, 3, and 5 data. The plots in this figure show few failure cases. Best viewable in color.

[Go to Table of Contents](#)

1188 **4.3. MPII-Cooking** 1242

1189 1243

1190 1244

1191 1245

1192 1246

1193 1247

1194 1248

1195 1249

1196 1250

1197 1251

1198 1252

1199 1253

1200 1254

1201 1255

1202 1256

1203 1257

1204 1258

1205 1259

1206 1260

1207 1261

1208 1262

1209 1263

1210 1264

1211 1265

1212 1266

1213 1267

1214 1268

1215 1269

1216 1270

1217 1271

1218 1272

1219 1273

1220 1274

1221 1275

1222 1276

1223 1277

1224 1278

1225 1279

1226 1280

1227 1281

1228 1282

1229 1283

1230 1284

1231 1285

1232 1286

1233 1287

1234 1288

1235 1289

1236 1290

1237 1291

1238 1292

1239 1293

1240 1294

1241 1295

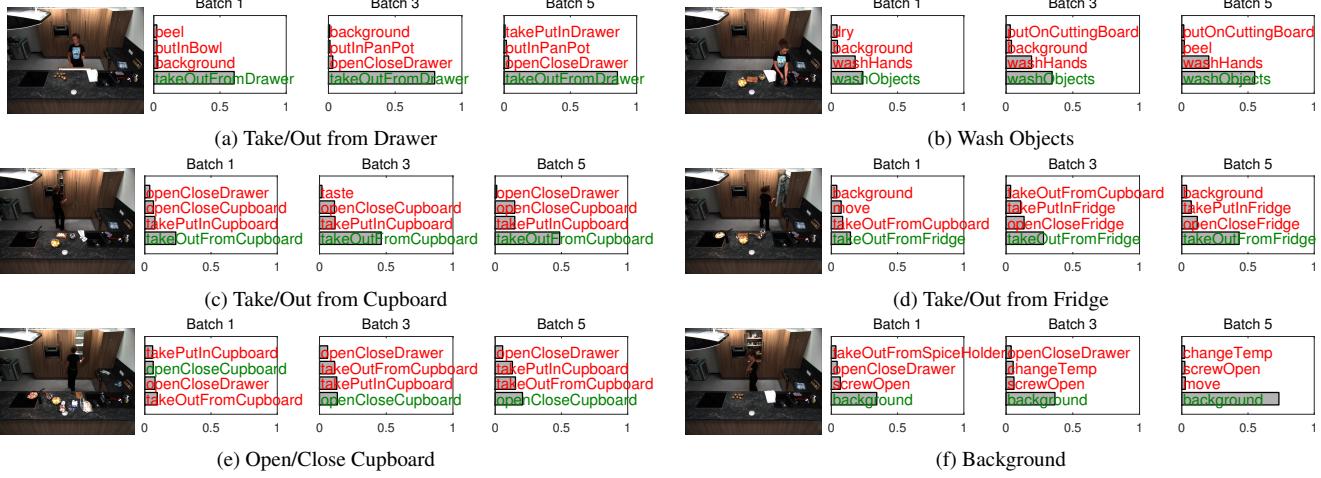


Figure 9. Evaluation of continuous learning on individual activities. Activity with green color means the ground truth class, whereas activities with red color means false predictions. Grey bars represent probability scores. Here, we show the results obtained after the arrival of batch 1, 3, and 5 data. The plots in this figure show some of the successful examples, where continuous learning helps to obtain the correct label with a higher probability even though some of them were miss-classified initially. Best viewable in color.

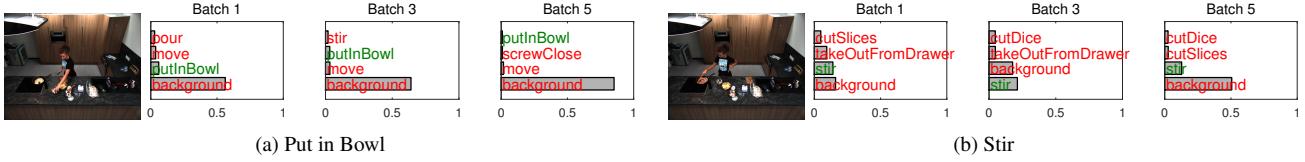


Figure 10. Evaluation of continuous learning on individual activities. Activity with green color means the ground truth class, whereas activities with red color means false predictions. Grey bars represent probability scores. Here, we show the results obtained after the arrival of batch 1, 3, and 5 data. The plots in this figure show few failure cases. Best viewable in color.

[Go to Table of Contents](#)

1296

4.4. UCF50

1350

1297

1351

1298

1352

1299

1353

1300

1354

1301

1355

1302

1356

1303

1357

1304

1358

1305

1359

1306

1360

1307

1361

1308

1362

1309

1363

1310

1364

1311

1365

1312

1366

1313

1367

1314

1368

1315

1369

1316

1370

1317

1371

1318

1372

1319

1373

1320

1374

1321

1375

1322

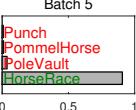
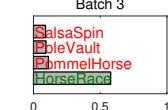
1376

1323

1377



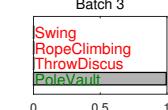
Batch 1



(a) Horse Race



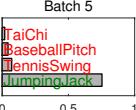
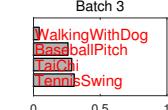
Batch 1



(c) Pole Vault



Batch 1



(b) Golf Swing

(d) Breast Stroke

(e) Jumping Jack

(f) Walking With Dog

Figure 11. Evaluation of continuous learning on individual activities. Activity with green color means the ground truth class, whereas activities with red color means false predictions. Grey bars represent probability scores. Here, we show the results obtained after the arrival of batch 1, 3, and 5 data. The plots in this figure show some of the successful examples, where continuous learning helps to obtain the correct label with a higher probability even though some of them were miss-classified initially. Best viewable in color.



Batch 1



(a) Jumping Jack

(b) Walking With Dog

Figure 12. Evaluation of continuous learning on individual activities. Activity with green color means the ground truth class, whereas activities with red color means false predictions. Grey bars represent probability scores. Here, we show the results obtained after the arrival of batch 1, 3, and 5 data. The plots in this figure show few failure cases. Best viewable in color.

Go to Table of Contents

1324

1378

1325

1379

1326

1380

1327

1381

1328

1382

1329

1383

1330

1384

1331

1385

1332

1386

1333

1387

1334

1388

1335

1389

1336

1390

1337

1391

1338

1392

1339

1393

1340

1394

1341

1395

1342

1396

1343

1397

1344

1398

1345

1399

1346

1400

1347

1401

1348

1402

1349

1403

1404

5. Activity Class-wise Performance Evaluation

1405

5.1. VIRAT

1406

Dataset: VIRAT, Activity: GettingIntoVehicle

1407

1409

1410

1411

1412

1413

1414

1415

1416

1417

1418

1419

1420

1421

1422

1423

1424

1425

1426

1427

1428

1429

1430

1431

1432

1433

1434

1435

1436

1437

1438

1439

1440

1441

1442

1443

1444

1445

1446

1447

1448

1449

1450

1451

1452

1453

1454

1455

1456

1457

1458

1459

1460

1461

1462

1463

1464

1465

1466

1467

1468

1469

1470

1471

1472

1473

1474

1475

1476

1477

1478

1479

1480

1481

1482

1483

1484

1485

1486

1487

1488

1489

1490

1491

1492

1493

1494

1495

1496

1497

1498

1499

1500

1501

1502

1503

1504

1505

1506

1507

1508

1509

1510

1511

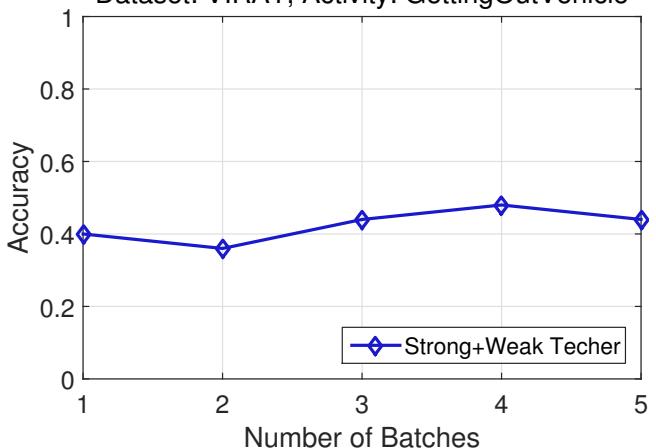
5. Activity Class-wise Performance Evaluation

5.1. VIRAT

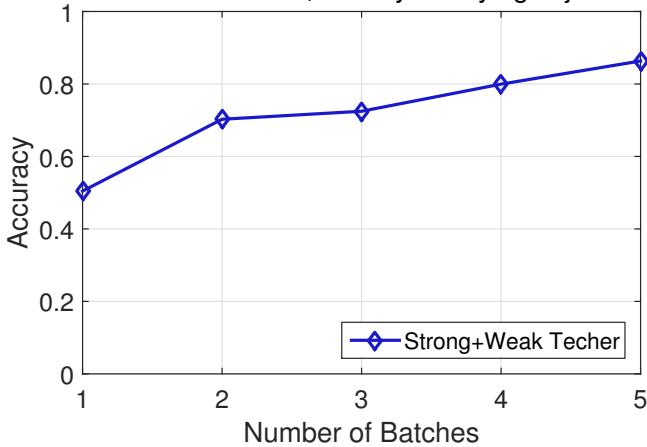
Dataset: VIRAT, Activity: GettingIntoVehicle



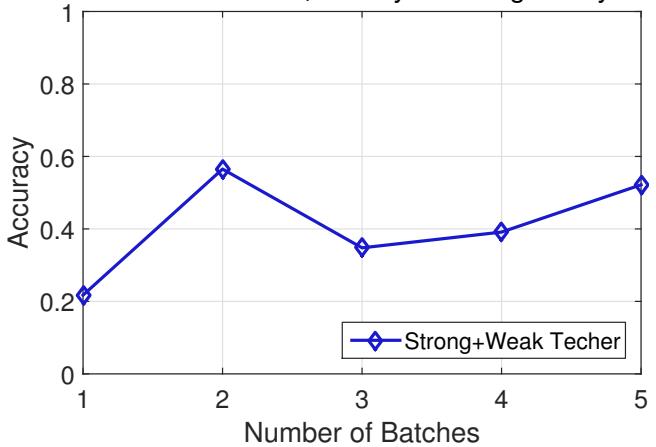
Dataset: VIRAT, Activity: GettingOutVehicle



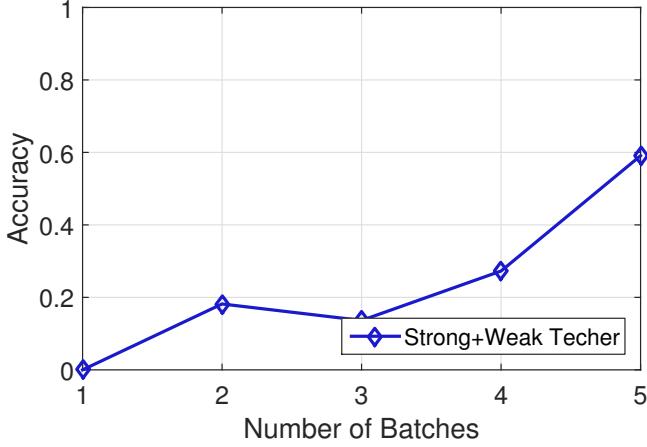
Dataset: VIRAT, Activity: CarryingObject



Dataset: VIRAT, Activity: EnteringFacility



Dataset: VIRAT, Activity: ExitingFacility



Dataset: VIRAT, Activity: OpeningTrunk

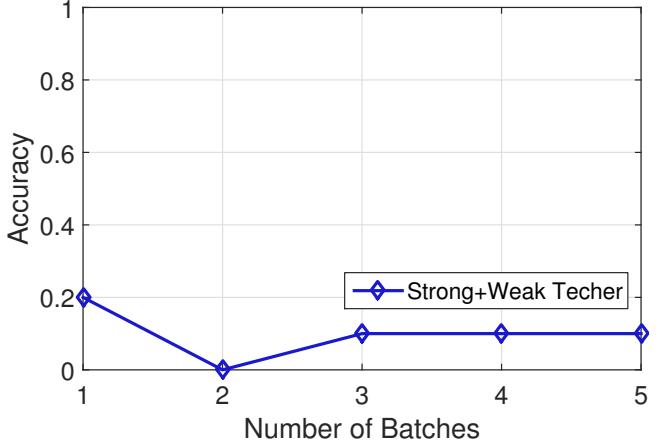


Figure 13. Activity class-wise performance evaluation of our proposed framework. This figure shows some selected activity classes of VIRAT dataset. In most of the cases, recognition performance of our framework for a particular class improves with the availability of new training data.

[Go to Table of Contents](#)

1512

5.2. UCLA-Office

1513

1514

1515

1516

1517

1518

1519

1520

1521

1522

1523

1524

1525

1526

1527

1528

1529

1530

1531

1532

1533

1534

1535

1536

1537

1538

1539

1540

1541

1542

1543

1544

1545

1546

1547

1548

1549

1550

1551

1552

1553

1554

1555

1556

1557

1558

1559

1560

1561

1562

1563

1564

1565

1566

1567

1568

1569

1570

1571

1572

1573

1574

1575

1576

1577

1578

1579

1580

1581

1582

1583

1584

1585

1586

1587

1588

1589

1590

1591

1592

1593

1594

1595

1596

1597

1598

1599

1600

1601

1602

1603

1604

1605

1606

1607

1608

1609

1610

1611

1612

1613

1614

1615

1616

1617

1618

1619

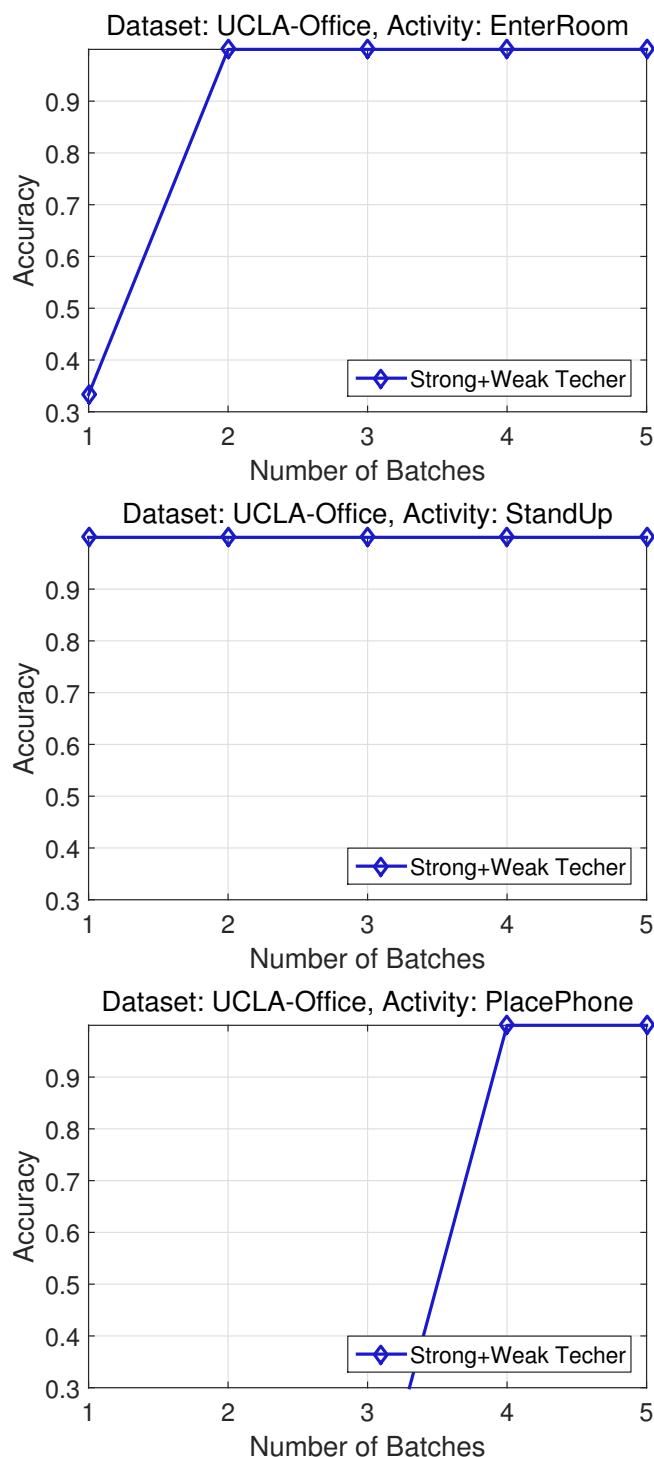


Figure 14. Activity class-wise performance evaluation of our proposed framework. This figure shows some selected activity classes of UCLA-Office dataset. In most of the cases, recognition performance of our framework for a particular class improves with the availability of new training data.

[Go to Table of Contents](#)

1620

5.3. MPII-Cooking

1621

1622

1623

1624

1625

1626

1627

1628

1629

1630

1631

1632

1633

1634

1635

1636

1637

1638

1639

1640

1641

1642

1643

1644

1645

1646

1647

1648

1649

1650

1651

1652

1653

1654

1655

1656

1657

1658

1659

1660

1661

1662

1663

1664

1665

1666

1667

1668

1669

1670

1671

1672

1673

1674

1675

1676

1677

1678

1679

1680

1681

1682

1683

1684

1685

1686

1687

1688

1689

1690

1691

1692

1693

1694

1695

1696

1697

1698

1699

1700

1701

1702

1703

1704

1705

1706

1707

1708

1709

1710

1711

1712

1713

1714

1715

1716

1717

1718

1719

1720

1721

1722

1723

1724

1725

1726

1727

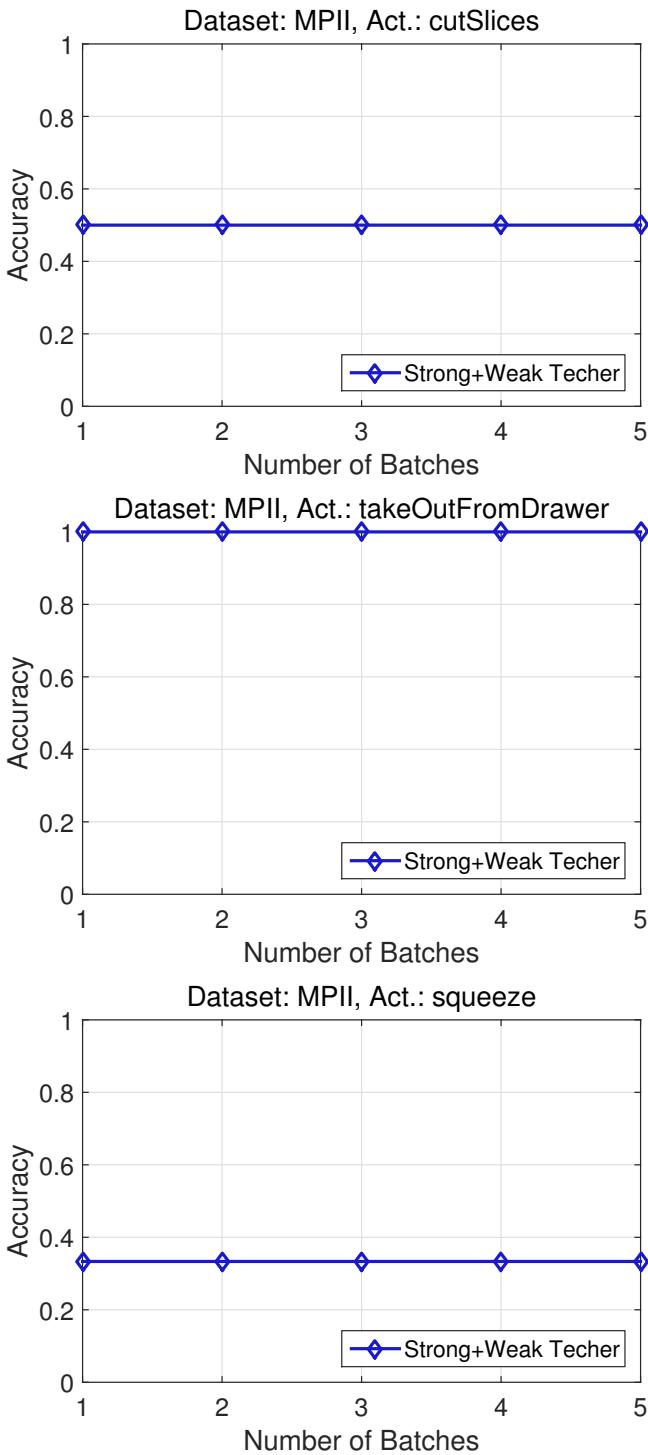


Figure 15. Activity class-wise performance evaluation of our proposed framework. This figure shows some selected activity classes of MPII-Cooking dataset. In most of the cases, recognition performance of our framework for a particular class improves with the availability of new training data.

[Go to Table of Contents](#)

1728

5.4. UCF50

1729

1730

1731

1732

1733

1734

1735

1736

1737

1738

1739

1740

1741

1742

1743

1744

1745

1746

1747

1748

1749

1750

1751

1752

1753

1754

1755

1756

1757

1758

1759

1760

1761

1762

1763

1764

1765

1766

1767

1768

1769

1770

1771

1772

1773

1774

1775

1776

1777

1778

1779

1780

1781

1782

1783

1784

1785

1786

1787

1788

1789

1790

1791

1792

1793

1794

1795

1796

1797

1798

1799

1800

1801

1802

1803

1804

1805

1806

1807

1808

1809

1810

1811

1812

1813

1814

1815

1816

1817

1818

1819

1820

1821

1822

1823

1824

1825

1826

1827

1828

1829

1830

1831

1832

1833

1834

1835

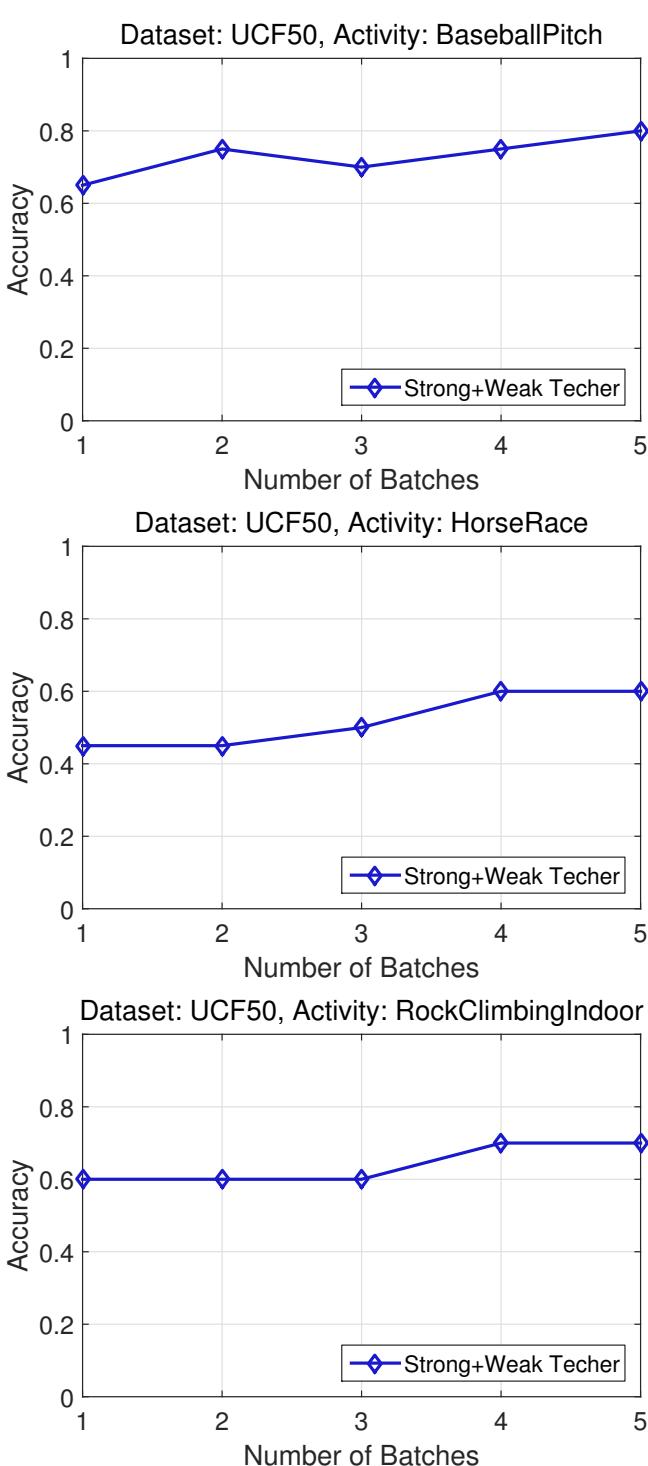


Figure 16. Activity class-wise performance evaluation of our proposed framework. This figure shows some selected activity classes of UCF50 dataset. In most of the cases, recognition performance of our framework for a particular class improves with the availability of new training data.

[Go to Table of Contents](#)

1836

6. Parameter Values

1837

We learn most of the parameters from training data. We manually set only three parameters - manual labeling percentage (K), weight decay parameter (λ) of our baseline softmax classifier (Section 4.1 of the main paper), and the weak teacher threshold parameter (δ). We provided the sensitivity analysis of K as the accuracy vs. manual labeling plot in the Figure 5(d, h, l, and p) of the main paper for all datasets. Here, we present the values of K , λ , and δ that we used during our experiments for all datasets followed by the sensitivity analysis of λ and δ for VIRAT dataset.

1843

Parameters	Dataset			
	VIRAT	UCLA-Office	MPII-Cooking	UCF50
K	0.5	0.5	0.5	0.5
λ	10^{-1}	10^{-2}	(used lin. svm)	10^{-4}
δ	0.9	0.9	0.9	0.9

Table 1. Parameter Values

1849

1850

6.1. Parameter Sensitivity

1851

6.1.1 VIRAT

1852

1853

1854

1855

1856

1857

1858

1859

1860

1861

1862

1863

1864

1865

1866

1867

1868

1869

1870

1871

1872

1873

1874

1875

1876

1877

1878

1879

1880

1881

1882

1883

1884

1885

1886

1887

1888

1889

1890

1891

1892

1893

1894

1895

1896

1897

1898

1899

1900

1901

1902

1903

1904

1905

1906

1907

1908

1909

1910

1911

1912

1913

1914

1915

1916

1917

1918

1919

1920

1921

1922

1923

1924

1925

1926

1927

1928

1929

1930

1931

1932

1933

1934

1935

1936

1937

1938

1939

1940

1941

1942

1943

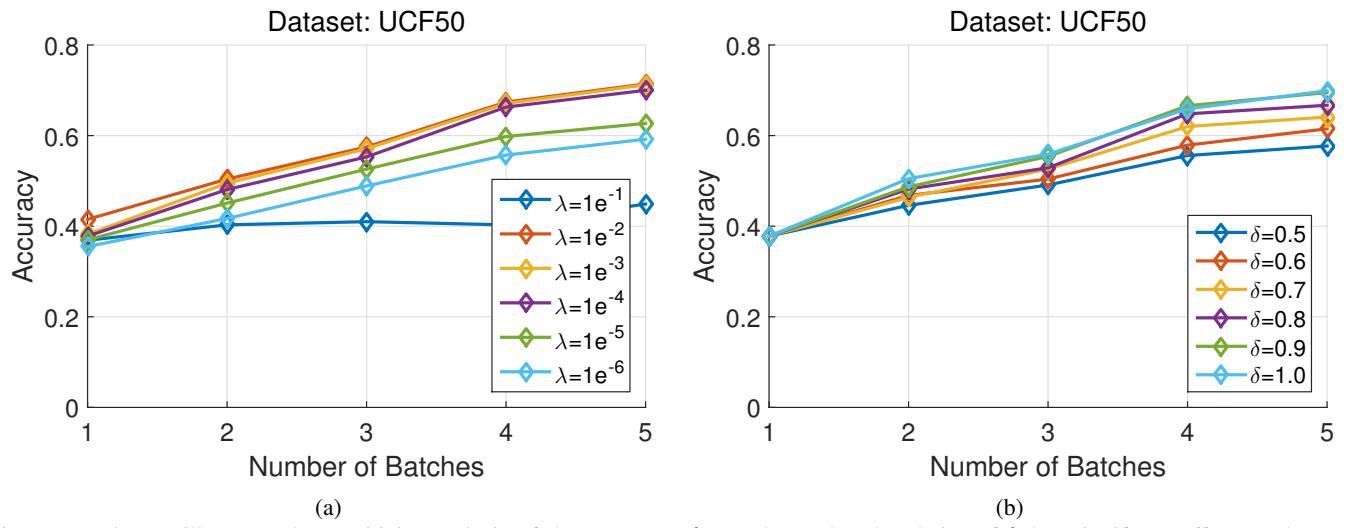


Figure 17. Plot (a) illustrates the sensitivity analysis of the parameter λ . It shows that the choice of λ has significant effect on the performance, however performance of the framework is quite similar in range of $10^{-2} - 10^{-4}$. Plot (b) illustrates the sensitivity analysis of the parameter δ . Our framework performs better for the higher values of δ . It means that for a higher value of δ the framework will use very high confident labels from the classifier to retrain it. For a lower value of δ , it may be possible that some of the misclassified instances are used for retraining, which is the reason for inferior performance. Above two experiments use Strong+Weak Teacher active learning system.

[Go to Table of Contents](#)