Method	F1@0.05	F1@0.1	F1@0.15	F1@0.2	F1@0.25	F1@0.3	F1@0.35	F1@0.4	F1@0.45	F1@0.5	Average
Supervised Offline											
ISBA [34]	0.106	0.170	0.227	0.265	0.298	0.326	0.348	0.348	0.348	0.348	0.330
TCN [34]	0.237	0.312	0.331	0.339	0.342	0.344	0.347	0.348	0.348	0.348	0.330
CTM [34]	0.244	0.312	0.336	0.351	0.361	0.369	0.374	0.381	0.383	0.385	0.350
TransParser [22]	0.289	0.381	0.435	0.475	0.500	0.514	0.527	0.534	0.540	0.545	0.474
PC [34]	0.522	0.595	0.628	0.646	0.659	0.665	0.671	0.676	0.679	0.683	0.642
DDM-Net [39]	0.604	0.681	0.715	0.735	0.747	0.753	0.757	0.760	0.763	0.767	0.728
Temporal Perceiver [38]	0.552	0.663	0.713	0.738	0.757	0.765	0.774	0.779	0.784	0.788	0.732
SC-Transformer [23]	0.618	0.694	0.728	0.749	0.761	0.767	0.771	0.774	0.777	0.780	0.742
EfficientGEBD [56]	**0.631**	**0.705**	–	–	–	**0.774**	–	–	–	0.786	**0.748**
LCVSL [52]	0.618	0.694	0.728	0.749	0.761	0.767	0.771	0.774	0.777	0.780	0.742
DyBDet [55]	0.625	0.701	**0.734**	**0.756**	**0.767**	0.772	**0.775**	**0.779**	**0.781**	**0.784**	0.747
Offline Unsupervised											
SceneDetect [34]	0.035	0.045	0.047	0.051	0.053	0.054	0.055	0.056	0.057	0.058	0.051
PA-Random [34]	0.158	0.233	0.273	0.310	0.331	0.347	0.357	0.369	0.376	0.384	0.314
PA [34]	0.360	0.459	0.507	0.543	0.567	0.579	0.592	0.601	0.609	0.615	0.543
FlowGEBD [12]	**0.375**	**0.502**	**0.569**	**0.624**	**0.658**	**0.677**	**0.695**	**0.703**	**0.711**	**0.717**	**0.623**
Online Supervised											
TeSTra-BC [54]	0.364	0.417	0.452	0.478	0.496	0.511	0.523	0.533	0.542	0.550	0.487
Sim-On-BC [41]	0.225	0.269	0.303	0.329	0.350	0.367	0.381	0.394	0.405	0.415	0.344
Oad-TR-BC [44]	0.263	0.319	0.361	0.394	0.422	0.445	0.465	0.483	0.497	0.510	0.416
MiniROAD-BC [2]	**0.422**	**0.472**	**0.502**	0.522	0.537	0.549	0.558	0.566	0.572	0.578	0.528
ESTimator	0.394	0.455	0.499	**0.532**	**0.558**	**0.578**	**0.594**	**0.608**	**0.619**	**0.629**	**0.547**
Our Baseline ambiguity fixed											
Our Baseline (1, 4)	0.223	0.336	0.385	0.425	0.453	0.474	0.474	0.490	0.498	0.498	0.426
Our Baseline (1, 8)	0.262	0.375	0.417	0.464	0.492	0.511	0.525	0.539	0.548	0.548	0.468
Our Baseline (1, 12)	0.227	0.310	0.408	0.439	0.455	0.475	0.481	0.486	0.486	0.501	0.427
Our Baseline (1, 16)	0.194	0.291	0.349	0.406	0.440	0.457	0.474	0.474	0.474	0.486	0.405
Our Baseline (3, 4)	0.177	0.210	0.234	0.246	0.256	0.260	0.262	0.263	0.265	0.265	0.244
Our Baseline (3, 8)	0.231	0.290	0.334	0.349	0.367	0.378	0.382	0.387	0.393	0.393	0.351
Our Baseline (3, 12)	0.264	0.353	0.405	0.424	0.435	0.441	0.457	0.466	0.466	0.466	0.418
Our Baseline (3, 16)	0.265	0.378	0.439	0.472	0.498	0.517	0.523	0.540	0.543	0.546	0.472
Our Baseline (5, 4)	0.190	0.262	0.322	0.342	0.355	0.366	0.372	0.388	0.388	0.391	0.338
Our Baseline (5, 8)	0.234	0.330	0.371	0.399	0.416	0.440	0.450	0.457	0.467	0.471	0.403
Our Baseline (5, 12)	0.226	0.327	0.397	0.440	0.471	0.490	0.525	0.537	0.549	0.549	0.451
Our Baseline (5, 16)	**0.268**	**0.386**	**0.443**	**0.496**	**0.526**	**0.553**	**0.561**	**0.575**	**0.583**	**0.588**	**0.498**
Our Baseline (1, 4, threshold 0.7)	0.189	0.270	0.323	0.361	0.394	0.431	0.442	0.480	0.491	0.491	0.387
Our Baseline (1, 8, threshold 0.7)	0.189	0.277	0.333	0.377	0.396	0.428	0.459	0.465	0.478	0.491	0.389
Our Baseline (1, 12, threshold 0.7)	0.162	0.242	0.316	0.343	0.370	0.397	0.424	0.444	0.451	0.458	0.361
Our Baseline (1, 16, threshold 0.7)	0.129	0.245	0.320	0.367	0.388	0.401	0.442	0.456	0.476	0.476	0.370
Our Baseline (2, 4, threshold 0.7)	0.200	0.330	0.400	0.425	0.450	0.475	0.480	0.490	0.495	0.500	0.424
Our Baseline (2, 8, threshold 0.7)	0.248	0.336	0.403	0.434	0.481	0.506	0.512	0.527	0.532	0.537	0.452
Our Baseline (2, 12, threshold 0.7)	0.179	0.295	0.370	0.376	0.410	0.445	0.457	0.468	0.474	0.474	0.395
Our Baseline (2, 16, threshold 0.7)	0.167	0.256	0.308	0.359	0.404	0.429	0.449	0.468	0.481	0.487	0.381
Our Baseline (3, 4, threshold 0.7)	0.202	0.271	0.330	0.351	0.376	0.393	0.410	0.424	0.431	0.431	0.362
Our Baseline (3, 8, threshold 0.7)	0.247	0.340	0.369	0.394	0.428	0.453	0.465	0.474	0.495	0.499	0.416
Our Baseline (3, 12, threshold 0.7)	0.219	0.330	0.409	0.451	0.479	0.507	0.544	0.553	0.567	0.572	0.463
Our Baseline (3, 16, threshold 0.7)	**0.247**	**0.375**	**0.434**	**0.461**	**0.483**	**0.504**	**0.531**	**0.552**	**0.568**	**0.584**	**0.474**
Our Baseline (1, 4, threshold 0.5)	0.185	0.271	0.352	0.399	0.428	0.461	0.466	0.485	0.499	0.499	0.404
Our Baseline (1, 8, threshold 0.5)	0.188	0.288	0.329	0.382	0.418	0.447	0.482	0.488	0.500	0.500	0.402
Our Baseline (1, 12, threshold 0.5)	0.170	0.252	0.340	0.371	0.403	0.428	0.440	0.465	0.478	0.484	0.383
Our Baseline (1, 16, threshold 0.5)	0.118	0.236	0.315	0.361	0.407	0.420	0.446	0.466	0.492	0.498	0.376
Our Baseline (2, 4, threshold 0.5)	0.225	0.340	0.395	0.437	0.459	0.484	0.488	0.497	0.501	0.501	0.433
Our Baseline (2, 8, threshold 0.5)	0.264	0.361	0.413	0.452	0.486	0.500	0.519	0.534	0.543	0.543	0.462
Our Baseline (2, 12, threshold 0.5)	0.201	0.286	0.397	0.418	0.439	0.466	0.466	0.471	0.471	0.487	0.410
Our Baseline (2, 16, threshold 0.5)	0.187	0.286	0.344	0.402	0.443	0.461	0.478	0.484	0.490	0.501	0.408
Our Baseline (3, 4, threshold 0.5)	0.208	0.282	0.340	0.349	0.370	0.390	0.402	0.419	0.419	0.419	0.360
Our Baseline (3, 8, threshold 0.5)	0.236	0.326	0.364	0.385	0.409	0.430	0.440	0.447	0.454	0.458	0.395
Our Baseline (3, 12, threshold 0.5)	0.218	0.319	0.387	0.433	0.466	0.487	0.525	0.534	0.550	0.555	0.447
Our Baseline (3, 16, threshold 0.5)	**0.265**	**0.392**	**0.454**	**0.492**	**0.520**	**0.548**	**0.567**	**0.577**	**0.582**	**0.586**	**0.498**
Our Baseline (1, 4, threshold 0.9)	0.169	0.254	0.302	0.326	0.363	0.405	0.429	0.465	0.483	0.489	0.369
Our Baseline (1, 8, threshold 0.9)	0.140	0.246	0.306	0.339	0.372	0.399	0.419	0.452	0.472	0.478	0.362
Our Baseline (1, 12, threshold 0.9)	0.153	0.211	0.276	0.305	0.327	0.371	0.378	0.400	0.407	0.415	0.324
Our Baseline (1, 16, threshold 0.9)	0.119	0.186	0.253	0.268	0.305	0.327	0.357	0.364	0.387	0.394	0.296
Our Baseline (2, 4, threshold 0.9)	0.158	0.282	0.339	0.373	0.418	0.452	0.458	0.463	0.469	0.469	0.388
Our Baseline (2, 8, threshold 0.9)	0.182	0.257	0.332	0.364	0.401	0.414	0.414	0.439	0.445	0.445	0.369
Our Baseline (2, 12, threshold 0.9)	0.154	0.244	0.328	0.367	0.392	0.431	0.450	0.450	0.463	0.463	0.374
Our Baseline (2, 16, threshold 0.9)	0.155	0.269	0.323	0.357	0.377	0.397	0.418	0.424	0.451	0.451	0.362
Our Baseline (3, 4, threshold 0.9)	0.231	0.332	0.378	0.399	0.437	0.458	0.475	0.487	0.496	0.496	0.419
Our Baseline (3, 8, threshold 0.9)	0.253	0.347	0.382	0.412	0.452	0.476	0.511	0.526	0.536	0.541	0.444
Our Baseline (3, 12, threshold 0.9)	0.207	0.330	0.413	0.458	0.480	0.514	0.536	0.542	0.547	0.547	0.458
Our Baseline (3, 16, threshold 0.9)	**0.231**	**0.314**	**0.373**	**0.391**	**0.444**	**0.456**	**0.485**	**0.527**	**0.538**	**0.550**	**0.431**
B.surprise (0.5s, 16)	0.248	0.350	0.447	0.505	0.544	0.558	0.573	0.587	0.587	0.602	0.500
B.surprise (0.5s, 16) logits	0.234	0.335	0.421	0.447	0.467	0.508	0.533	0.533	0.538	0.538	0.455
Cross datasets											
EfficientGEBD R50L4 $Kine \rightarrow Tapos$	0.3594	0.4571	0.4986	0.5183	0.5319	0.5401	0.5459	0.5493	0.5525	0.5549	0.5108

Table 1: Comparison of F1 scores at different temporal thresholds. Column-wise maxima are bolded per group.