TABLE I

DATA AUGMENTATION CONFIGURATIONS

Augmentation	**Magnitude**	**Probability**
Mosica	-	0.5
Mixup	-	0.5
Random Translation	±0.1	1.0
Random Scale	±0.1	1.0
Horizontal Flip	-	0.5

TABLE II

PERFORMANCE COMPARISON OF EVENT REPRESENTATIONS

Method	**Representation**	**mAP@50**	**mAP@50:95**
YOLOv5s	Event Histogram	0.822	0.690
	Time Surface	0.819	0.682
	Event Volume	**0.850**	**0.711**
YOLOv5m	Event Histogram	0.856	0.727
	Time Surface	0.848	0.715
	Event Volume	**0.861**	**0.728**
YOLOv8s	Event Histogram	**0.833**	**0.701**
	Time Surface	0.813	0.680
	Event Volume	0.831	0.699
YOLOv8m	Event Histogram	0.868	0.737
	Time Surface	0.843	0.714
	Event Volume	**0.871**	**0.743**
YOLO11s	Event Histogram	**0.863**	**0.726**
	Time Surface	0.839	0.708
	Event Volume	0.859	0.725
YOLO11m	Event Histogram	**0.873**	**0.742**
	Time Surface	0.856	0.725
	Event Volume	0.865	0.739
RT-DETR-R18	Event Histogram	**0.826**	**0.680**
	Time Surface	0.801	0.663
	Event Volume	0.809	0.659

TABLE III

DETECTION PERFORMANCE ON EVENT HISTOGRAM INPUTS

Method	**AP@50**	**AP@50:95**	**APs**
YOLOv5s	0.822	0.690	0.458
YOLOv5m	0.856	0.727	0.508
YOLOv8s	0.833	0.701	0.453
YOLOv8m	0.868	0.737	<u>0.512</u>
YOLO11s	0.863	0.726	0.490
YOLO11m	<u>0.873</u>	<u>0.742</u>	<u>0.512</u>
RT-DETR-R18	<u>0.826</u>	<u>0.680</u>	<u>0.499</u>
PoolFormer-S12	0.768	0.610	0.420
RepViT-M0.9	0.783	0.624	0.447
MobileNetV2	0.768	0.623	0.431
M2Former	0.822	0.663	0.466
RT-DETR-R18 + AAL	0.833	0.694	0.516
RT-DETR-R18 + AAL + Aug	0.869	0.710	0.543
M2Former + AAL	0.826	0.677	0.511
M2Former + AAL + Aug	**<u>0.903</u>**	**<u>0.743</u>**	**<u>0.580</u>**

TABLE IV

MODEL EFFICIENCY COMPARISON

Method	**Params (M)**	**GFLOPs**
YOLOv5s	9.1	23.6
YOLOv5m	20.9	63.7
YOLOv8s	11.1	28.6
YOLOv8m	25.9	78.9
YOLO11s	9.4	21.5
YOLO11m	20.1	68.0
RT-DETR-R18	20.1	58.2
PoolFormer-S12	20.3	53.7
RepViT-M0.9	13.6	37.9
MobileNetV2	10.6	28.8
M2Former	**9.7**	**27.5**

TABLE V
ABLATION STUDY ON M^{2}FORMER COMPONENTS

Method	**AP@50**	**AP@50:95**
M2Former (baseline)	0.822	0.663
w/o Res2Net	0.795 (-0.027)	0.644 (-0.019)
w/o Spatial Attention	0.809 (-0.013)	0.658 (-0.005)
w/o Channel Attention	0.805 (-0.017)	0.650 (-0.013)
w/o SPD-Conv	0.814 (-0.008)	0.651 (-0.012)

TABLE VI
ABLATION STUDY ON DATA AUGMENTATION

Method	**AP@50**	**AP@50:95**
M2Former (baseline)	0.822	0.663
with Mosica	0.851 (+0.029)	0.698 (+0.026)
with Mixup	0.847 (+0.025)	0.688 (+0.025)
with Transformations	0.837 (+0.015)	0.680 (+0.017)

TABLE VII
DETECTION PERFORMANCE ON EVENT HISTOGRAM
UNDER LOWER RESOLUTION INPUT

Method	**AP@50**	**AP@50:95**	**APs**
YOLOv5s	0.719	0.556	0.414
YOLOv5m	0.796	0.617	0.506
YOLOv8s	0.771	0.598	0.482
YOLOv8m	0.790	0.617	0.503
YOLO11s	0.726	0.564	0.448
YOLO11m	0.728	0.572	0.423
RT-DETR-R18	0.733	0.562	0.492
PoolFormer-S12	0.621	0.448	0.392
RepViT-M0.9	0.659	0.487	0.421
MobileNetV2	0.589	0.434	0.376
M2Former	0.699	0.524	0.472
RT-DETR-R18 + AAL	0.752	0.580	0.524
RT-DETR-R18 + AAL + Aug	0.760	0.593	0.526
M2Former + AAL	0.721	0.546	0.488
M2Former + AAL + Aug	**0.809**	**0.622**	**0.546**

TABLE VIII
ZERO-SHOT DETECTION PERFORMANCE FROM
SYNTHETIC DOMAIN TO REAL DOMAIN

Method	Modality	Light Condition	AP50	AP50:95
YOLOv8s	RGB	Normal Exposure	0.025	0.006
		Overexposure	0.003	0.001
		Underexposure	0.000	0.000
		Average	0.009	0.002
YOLOv8s	Event	Normal Exposure	0.257	0.154
		Overexposure	0.135	0.090
		Underexposure	0.274	0.142
		Average	0.222	0.129
RT-DETR-R18	Event	Normal Exposure	0.230	0.102
		Overexposure	0.053	0.012
		Underexposure	0.748	0.372
		Average	0.344	0.162
M2Former	Event	Normal Exposure	0.442	0.211
		Overexposure	0.142	0.072
		Underexposure	0.753	0.309
		Average	0.446	0.197