Online supplementary material: "Livestock Detection and Counting in Kenyan Rangelands Using Aerial Imagery and Deep Learning Techniques"

Table S1. Distribution of test images and annotations across the three scenarios.

	Test images	Test annotations
Scenario 1		
Combined dataset	277	2642
Scenario 2 and Scenario 3		
Lumo	537	6211
THWS	356	3106
Choke	423	2966

Genetic evolution algorithm

We used a genetic evolution algorithm to tune hyperparameters for YOLOv5 on the training dataset. The best values were attained after 56 generation of ten epochs and with a batch size of eight. Table S2 shows the new tuned hyperparameters and the definition of these parameters can be found at the URL[1]

Table S1. Fine-tuned YOLOv5 hyperparameters values attained from generic evolution.

Parameter	Initial	Evolved	Parameter	Initial	Evolve
Learning rate (lr)	0.01	0.01131	HSV-H Augment	0.015	001598
Final lr	0.01	0.01139	HSV-S Augment	0.7	0.75931
Momentum	0.937	0.94372	HSV-V Augment	0.4	0.58615
Weight Decay	0.0005	0.00036	Translation Augment	0.1	0.0726
Warmup Epochs	3.0	2.5532	Scale Augment	0.5	0.34732
Warmup Momentum	0.8	0.92029	Mosaic Augment	1.0	0.97171
Warmup Bias lr	0.1	0.0957	Object Loss Gain	1.0	0.83949
Box Loss Gain	0.05	0.0413	Object BCE Loss	1.0	0.82926
Class Loss Gain	0.5	0.42629	Anchor-Multiple Threshold	4.0	3.743
Class BCE Loss	1.0	1.1091			

Table S3. Comparison of accuracy metrics for YOLOv5 on scenario 2 on test dataset when the models are trained on raw dataset, raw and augmented dataset, and combination of raw, augmented and background images. Models trained with Choke dataset had the highest gains in AP ($AP_{0.5-0.95}$). Inclusion of background data (negative images) had a slight improvement with exception to model trained on Choke and tested on Lumo, where the AP decreased.

Training	Test		Raw dataset				Raw + Augmented dataset				Raw + Augmented + Background data			
			P	R	$AP_{0.5}$	AP	P	R	$AP_{0.5}$	AP	P	R	$AP_{0.5}$	AP
Lumo	THWS	v5s	88.6	79.8	84.7	33.4	86.6	80.5	84.4	32.2	90.1	81.8	88.2	35.8
		v5m	86.8	80.8	83.9	32.8	85.1	81.9	85.9	33.0	90.1	82.4	88.4	35.7
		v5l	84.0	79.1	81.9	31.4	88.3	77.1	84.5	32.7	90.4	82.5	88.6	36.2
		v5x	86.7	79.4	82.6	32.1	86.1	82.3	86.4	33.7	90.6	82.6	89.1	36.4
	Choke	v5s	83.8	66.1	72.9	26.3	85.3	72.7	80.4	29.6	84.9	72.5	81.2	32.2
		v5m	84.4	68.3	74.0	27.1	83.0	79.5	83.6	31.3	85.6	73.9	81.9	32.0
		v5l	81.6	68.7	72.6	25.7	86.1	76.3	82.5	30.3	88.4	73.2	83.1	33.0
		v5x	81.8	70.4	73.9	27.5	86.3	77.5	82.5	30.6	84.6	72.9	80.4	31.7
THWS	Choke	v5s	82.8	71.9	77.2	27.2	85.9	77.4	82.0	29.7	87.4	73.5	80.8	31.1
		v5m	82.5	71.4	76.9	28.1	84.8	79.1	82.7	30.0	84.7	72.7	79.8	31.2
		v5l	82.2	69.7	73.8	26.4	84.4	78.1	82.5	30.2	85.9	71.7	80.6	32.1
		v5x	83.2	73.4	76.8	28.1	83.2	79.0	82.0	30.0	85.5	75.1	81.8	32.1
	Lumo	v5s	79.8	73.5	77.9	28.8	81.8	78.7	82.6	32.0	85.3	76.3	82.2	32.4
		v5m	79.7	74.9	77.3	28.3	81.4	79.8	83.8	32.5	84.4	77.9	83.8	33.2
		v5l	79.2	74.1	76.8	28.4	82.9	79.2	83.9	33.4	85.7	76.2	83.6	33.0
		v5x	80.4	74.4	78.7	29.8	82.5	78.8	83.8	32.6	86.0	77.0	85.0	33.8
Choke	Lumo	v5s	64.5	52.0	53.2	16.2	78.2	66.2	73.0	25.9	76.5	65.6	71.6	25.5
		v5m	71.7	61.9	67.2	23.2	79.4	72.0	77.6	28.2	77.5	67.2	73.3	25.7
		v5l	69.0	59.2	62.0	21.1	76.7	69.0	73.8	27.1	76.5	65.7	71.0	25.3
		v5x	62.4	52.4	53.3	17.6	78.3	67.8	73.7	26.9	79.5	63.9	71.8	25.8
	THWS	v5s	73.2	59.4	62.4	19.7	82.4	73.9	77.6	27.5	85.6	74.0	80.7	29.8
		v5m	78.2	68.4	73.1	25.5	84.3	75.0	80.2	28.6	84.6	76.1	81.3	29.6
		v5l	73.9	63.9	68.5	23.3	83.0	75.9	79.6	27.9	83.1	74.0	79.7	29.6
		v5x	71.2	58.5	63.7	21.9	83.8	74.7	79.9	28.7	85.6	75.6	81.9	30.7

P-Precision, R-Recall, $AP_{0.5}$ - Average precision with IoU @ 0.5; AP- Average precision with IoU @0.5-0.95

Table S4. Accuracy metrics for YOLOv8 on scenario 2 test datasets when the models are trained on raw dataset, raw and augmented dataset, and combination of raw, augmented and background images.

Training	Test		Raw dataset				Raw + Augmented dataset				Raw + Augmented + Background data			
			P	R	$AP_{0.5}$	AP	P	R	$AP_{0.5}$	AP	P	R	$AP_{0.5}$	AP
Lumo	THWS	v8s	86.0	75.8	82.7	36.9	88.6	78.3	84.1	36.2	90.6	82.6	87.6	38.7
		v8m	87.6	76.9	83.3	36.4	89.3	78.6	84.8	36.4	90.3	81.9	87.6	38.7
		v8l	85.7	78.2	84.5	36.6	88.4	78.0	83.7	36.4	90.4	82.6	87.9	39.6
		v8x	84.9	77.7	83.1	36.3	87.8	78.6	84.1	36.6	90.0	82.1	87.4	39.2
	Choke	v8s	87.4	64.5	77.2	33.7	86.8	76.1	82.0	34.3	90.9	69.7	80.9	35.9
		v8m	86.8	59.8	73.9	32.1	86.4	72.5	80.3	33.9	90.2	70.6	81.0	35.1
		v8l	82.6	68.2	76.9	32.6	87.9	75.6	82.8	34.8	89.9	69.3	80.5	35.8
		v8x	87.1	59.9	74.2	31.1	87.8	77.3	83.6	35.7	90.6	71.2	81.6	36.0
THWS	Choke	v8s	85.4	67.2	77.4	32.7	82.7	79.2	83.1	34.1	87.7	72.9	81.7	35.7
		v8m	83.2	64.4	74.5	30.7	82.4	76.6	81.5	32.6	88.8	69.8	80.7	36.5
		v8l	84.0	64.2	75.4	31.0	85.1	75.9	81.2	32.8	87.9	73.4	82.2	36.6
		v8x	87.9	63.7	76.4	32.6	84.7	76.3	82.3	34.0	87.9	74.8	82.8	36.3
	Lumo	v8s	78.9	72.1	77.8	32.1	80.4	79.8	83.4	35.2	84.1	77.7	83.6	36.4
		v8m	76.4	72.0	77.2	31.7	80.3	76.9	82.4	34.7	86.5	74.4	82.8	37.1
		v8l	78.8	69.1	76.8	31.9	79.9	79.5	83.4	35.9	84.4	76.4	83.1	36.8
		v8x	78.6	71.8	78.6	33.5	80.5	77.5	82.6	34.9	84.7	74.7	82.0	36.1
Choke	Lumo	v8s	61.5	49.9	54.5	19.4	77.1	66.1	74.9	29.8	76.8	60.4	71.5	28.9
		v8m	60.5	46.5	52.4	18.7	76.7	67.3	75.3	30.7	78.5	62.8	73.4	29.3
		v8l	53.8	42.9	46.4	16.0	76.0	65.9	74.5	30.1	75.0	65.7	73.9	29.6
		v8x	64.3	56.1	61.0	22.6	72.6	65.2	72.6	28.8	73.3	63.6	71.2	28.3
	THWS	v8s	69.5	58.5	65.9	25.1	83.8	79.8	79.1	31.2	79.3	73.9	80.0	32.7
		v8m	69.6	51.7	62.0	23.7	81.2	70.4	78.4	31.4	81.9	75.6	81.5	33.2
		v8l	63.7	50.3	57.0	21.5	81.5	70.2	78.3	31.5	85.1	75.1	82.4	34.3
		v8x	74.3	65.4	71.8	27.1	81.2	70.1	77.7	30.7	80.5	75.5	81.5	33.0

Table S5. Model performance across YOLOv5 and YOLOv8 in scenario 3, when models are trained using data from two sites and tested in the third site. The models were compared on the impact of background data on the accuracy metrics and count error. Inclusion of background data significance varied across the different model with slight changes in the $AP_{0.5-0.95}$. YOLOv5 architectures had the least count error across the three sites despite lower $AP_{0.5-0.95}$ compared to YOLOv8 architectures.

Test site	Model	Raw + Augmented data					Raw + Augmented + Background data				
		P	R	$AP_{0.5}$	$AP_{0.5-0.95}$	CE	P	R	$AP_{0.5}$	$AP_{0.5-0.95}$	CE
THWS	YOLOv5s	91.2	83.4	88.9	39.3	-1.3	90.7	84.0	89.0	39.5	-1.2
	YOLOv8s	89.8	85.3	89.0	39.5	-4.2	91.0	84.0	89.0	39.9	-6.8
	YOLOv5m	89.7	84.0	89.2	39.6	-0.7	90.5	84.2	89.2	39.1	-0.7
	YOLOv8m	90.0	85.4	89.5	39.9	-3.2	90.6	83.6	88.1	39.9	-6.4
	YOLOv5l	89.0	84.4	89.1	39.5	-2.9	90.7	84.0	88.9	39.7	-2.9
	YOLOv8l	90.8	84.8	88.9	40.3	-5.5	90.3	84.3	88.7	40.3	-5.0
	YOLOv5x	90.6	85.1	89.9	40.4	-0.7	91.0	83.6	89.1	40.4	**-0.1**
	YOLOv8x	90.1	85.0	89.2	40.0	-5.2	90.3	83.8	88.5	40.1	-6.2
Choke	YOLOv5s	87.0	75.7	83.0	36.9	-9.8	87.9	77.4	84.0	37.0	**-7.9**
	YOLOv8s	89.4	75.6	83.5	37.2	-13.0	89.7	76.8	84.4	37.4	-12.9
	YOLOv5m	87.8	72.8	81.8	36.7	-11.6	90.1	86.5	84.7	37.3	-10.4
	YOLOV8m	90.2	76.8	84.4	37.4	-11.7	91.3	73.6	82.9	37.4	-16.8
	YOLOv5l	88.1	76.4	84.0	37.2	-9.5	89.0	75.6	83.6	36.6	-12.4
	YOLOv8l	90.5	73.8	82.9	37.0	-16.1	89.9	76.2	83.9	37.0	-13.6
	YOLOv5x	88.7	76.8	84.4	37.5	-11.3	89.4	74.6	82.5	36.1	-14.0
	YOLOv8x	90.8	74.7	83.7	37.5	-16.2	90.5	76.9	84.5	37.7	-11.4
Lumo	YOLOv5s	85.9	76.8	84.2	36.3	-1.9	85.9	77.1	84.3	36.0	-1.0
	YOLOV8s	85.4	78.8	84.7	36.7	-7.4	85.4	79.6	84.5	37.1	-5.9
	YOLOv5m	85.8	76.1	84.1	36.2	-1.4	85.5	77.0	84.2	36.7	**-0.7**
	YOLOv8m	86.7	79.1	85.3	37.3	-7.2	86.3	79.6	85.3	37.6	-6.7
	YOLOv5l	85.3	76.3	83.8	36.6	-0.7	86.4	77.2	84.9	36.7	-2.2
	YOLOv8l	84.4	79.3	84.8	37.2	-6.3	86.3	77.9	84.5	37.3	-9.3
	YOLOv5x	85.3	77.7	84.4	36.7	-0.8	84.5	78.8	85.1	36.6	2.6
	YOLOv8x	84.8	78.2	84.2	36.9	-7.0	86.0	77.3	84.0	37.1	-9.5

CE: Count error