TABLE: Top 7 network models in selected publications and their performance

ImageNet dataset. Training: 55,000 animal images and 25,000 non-aimal images. Validation: 15,000 animal images and 8,500 non-aimal images. Validation: 15,000 animal images and 8,500 non-aimal images. Clarifilition images each class. Intel 10,000 10 56,93 [53] [53] [55] [16] [17] [17] [18] [18] [18] [18] [18] [18] [19] [19] [19] [19] [19] [19] [19] [19	Model	Dataset(s)	Species to classification	Accuracy (%)	F- measure (%)	Reference
Lill-Animar-Paces dataset. Iotal 2,200 ms 20 91.48 [66] CCCO 1.000 images each class, total 10,000 10 56.93 [53] CCCO 1.000 images each class, total 10,000 10 56.93 [53] S. 757,000 animal images (1.4 million for training, 10,000 for testing) 10 100,000 for testing, 90 for testi	AlexNet	images and 25,000 non-animal images. Validation: 18,500 animal images and 8,500 non-animal images	6	94.91	91.48	[39]
SS. 757,000 animal images + 1.757,000 random non-animal images (1.4 million for training, 105,000 for testing) Head-140, body-124, scale-174, total 438 images (348 for testing, 90 for testing) Head-140, body-124, scale-174, total 438 images (348 for testing, 90 for testing) The state of the stat		,	20	91.48		[66]
		COCO: 1,000 images each class, total 10,000	10	56.93		[53]
ResNet-19 Sease-174, total 438 lm ages (348 for testing, 90 for testing, 91 for testing, 91 for testing, 92 for testing, 92 for testing, 92 for testing, 96 for testing, 97 for testing, 96 for testing, 96 for testing, 97 for testing, 96 for testing, 97 for testing, 96 for testing, 97 for testing, 96 for testing, 96 for testing, 97 for testing, 96 for testing, 97 for testing, 97 for testing, 96 for testing, 97 for testing, 97 for testing, 97 for testing, 96 for testing, 97 for testing, 9		non-animal images (1.4 million for training,	2	95.8		[37]
(NACTI) dataset. Total 3,741,656 images 27 81.8 [64] (90% for training, 10% for testing) 10 epochs-88, 20 epochs-90, 50 epochs-90,		ages (348 for testing, 90 for testing)	6	body-89.63,		[47]
Snapshot Australia (SA), SS, Snapshot Wisconsin (SW), 8 classes each dataset, 750 images each class So openches 93, 150 epochs 95, 500 epochs 97, 1000 epo	ResNet-18	(NACTI) dataset. Total 3,741,656 images	27			[64]
105,000 for testing) 2 96.3 [37] 105,000 for testing) SS-7.3 million, Camera CATalogue (CC)-0.52 million, Elephant Expendition (EE)-0.42 mills on, Snapshot Wisconsin (SW)-0.5 million, 90% for training, 10% for testing) SS-51, CC-42, EE-9, SW-31 SW-92.4 [77] SW-92.4 SW-92.4 [77] SW-92.4 SW-92.4 [77] SW-92.4 SW-92.4 [77] SW-92.4 SW-9		consin (SW), 8 classes each dataset, 750 images each class	8	20 epochs-90, 50 epochs-93, 150 epochs- 95, 500 epochs-97, 1000 epochs-		[58]
million, Elephant Expendition (EE)-0.42 million, Snapshot Wisconsin (SW)-0.5 million, 90% for training, 10% for testing) SS-51, CC-42, EE-9, SW-31 SS-96, SW-92.4 SW-92.		non-animal images (1.4 million for training, 105,000 for testing)	2	96.3		[37]
Fast R-CNN 17,000 images (12,000 training, 5,000 Testing) (all, side, frontal, back) 96 aerial images (89 for training, 7 for testing) 1 89.1 [63] 151,106 (113466 for training, 20,178 for validation, 17,462 for testing) 4 95.44 [67]		million, Elephant Expendition (EE)-0.42 million, Snapshot Wisconsin (SW)-0.5 million,		97.7, EE-96.6,		[77]
Fast R-CNN 17,000 images (12,000 training, 5,000 Testing) (all, side, frontal, back) 96 aerial images (89 for training, 7 for testing) 1 89.1 [63] 151,106 (113466 for training, 20,178 for validation, 17,462 for testing) 4 95.44 [67]		6.300 (4.800 for training, 1.500 for testing)	8	83.89	77.95	[32]
96 aerial images (89 for training, 7 for testing) 1 1 1 1 1 1 1 1 1	ResNet-50	17,000 images (12,000 training, 5,000 Test-				
dation, 17,462 for testing)		96 aerial images (89 for training, 7 for testing)	1	89.1		[63]
images (70% for training, 10% for validation, 20% for testing) 561 images, 4,305 labeled animals (70% for validation, 20% for testing) 3 86.3 [24]		dation, 17,462 for testing)	4	95.44		[67]
Validation, 20% for testing 3 80.3 [24]		images (70% for training, 10% for validation, 20% for testing	6	74		[42]
Head-140, body-124, scale-1/4, total 438 images (348 for testing, 90 for testing) 6 body-92.52, scale-84.03 [47]			3	86.3		[24]
images and 25,000 non-animal images. Validation: 18,500 animal images and 8,500 nonanimal images SS. 757,000 animal images + 757,000 random non-animal images (1.4 million for training, 105,000 for testing) Snapshot Australia (SA), SS, Snapshot Wisconsin (SW), 8 classes each dataset, 750 images each class 10 epochs-88, 20 epochs-91, 50 epochs-93, 150 epochs-93, 150 epochs-93, 150 epochs-95, 500 epochs-97, 1000 epochs-97, 1000 epochs-97, 1000 epochs-98 1,462 images (1,187 for training, 275 for validation) Mask R-CNN 910 images (720 for training, 190 for validation) Mask R-CNN 910 images (80% for training, 20% for testing) 1 Image-95.4, video-86.8 10 epochs-88, 20 epochs-93, 150 epochs-95, 500 epochs-97, 1000 epochs-95, 500 epochs-97, 1000 epochs-98 Image-95.4, video-86.8			6	body-92.52,		[47]
non-animal images (1.4 million for training, 105,000 for testing) 2 96.3 [37]		images and 25,000 non-animal images. Validation: 18,500 animal images and 8,500 non-animal images	6	96.11	93.36	[39]
Snapshot Australia (SA), SS, Snapshot Wisconsin (SW), 8 classes each dataset, 750 images each class 1,462 images (1,187 for training, 275 for validation) Mask R-CNN Mask R-CNN 6,080 images (80% for training, 20% for testing) 1000 images each bird species (90% for training) 20 epochs-91, 50 epochs-93, 150 epochs-95, 500 epochs-97, 1000 epochs-98 8 0.43 [68] Image-95.4, video-86.8 [66]		non-animal images (1.4 million for training,	2	96.3		[37]
Mask R-CNN 910 images (720 for training, 190 for validation) 7 93.8 [56] 6,080 images (80% for training, 20% for testing) 1 Image-95.4, video-86.8 [46] 1000 images each bird species (90% for train-12) 100.24 [55]		consin (SW), 8 classes each dataset, 750	8	20 epochs-91, 50 epochs-93, 150 epochs- 95, 500 epochs-97, 1000 epochs-		[58]
Mask R-CNN 910 images (720 for training, 190 for validation) 7 93.8 [56] 6,080 images (80% for training, 20% for testing) 1 Image-95.4, video-86.8 [46] 1000 images each bird species (90% for train-1000 images each bird species) 2 02.4 [55]		, 9 ()	5	80.43		[68]
6,080 images (80% for training, 20% for testing) 1 Image-95.4, video-86.8 [46] 1000 images each bird species (90% for train-		910 images (720 for training, 190 for valida-	7	93.8		[56]
1000 images each bird species (90% for train-		6,080 images (80% for training, 20% for	1			[46]
			2			[55]

	96 aerial images (89 for training, 7 for testing)	1	83		[63]
YOLO V3	151,106 (113466 for training, 20178 for validation, 17462 for testing)	4	91.8		[67]
	Rabbit 1,246, Kangaroo 4,211, pig 6,000, total 11,457 images (85% for training, 15% for testing)	3	10 epochs- 92.31, 20 epochs-95.84, 30 epochs- 96.86, 40 epochs-97.39, 50 epochs- 98.38		[72]
	SS. 757,000 animal images + 757,000 random non-animal images (1.4 million for training, 105,000 for testing)	2	96.3		[37]
VGG-16	ImageNet dataset. Training: 55,000 animal images and 25,000 non-animal images. Validation: 18,500 animal images and 8,500 non-animal images	6	96.6	93.72	[39]
	LHI-Animal-Faces dataset, total 2200 images	20	94.39		[66]
	Head-140, body-124, scale-174, total 438 images (348 for testing, 90 for testing)	6	Head-70.54, body-81.52, scale-84.03		[47]