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

Model	Dataset(s)	Species to classification	Accuracy (%)	F-measure (%)	Reference
AlexNet	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	94.91	91.48	[39]
	LHI-Animal-Faces dataset. Total 2,200 images	20	91.48		[66]
	COCO: 1,000 images each class, total 10,000	10	56.93		[53]
	SS. 757,000 animal images + 757,000 random non-animal images (1.4 million for training, 105,000 for testing)	2	95.8		[37]
	Head-140, body-124, scale-174, total 438 images (348 for testing, 90 for testing)	6	Head-85.79, body-89.63, scale-78.9		[47]
ResNet-18	North American Camera Trap Images (NACTI) dataset. Total 3,741,656 images (90% for training, 10% for testing)	27	81.8		[64]
	Snapshot Australia (SA), SS, Snapshot Wisconsin (SW), 8 classes each dataset, 750 images each class	8	10 epochs-88, 20 epochs-90, 50 epochs-93, 150 epochs-95, 500 epochs-97, 1000 epochs-97		[58]
	SS. 757,000 animal images + 757,000 random non-animal images (1.4 million for training, 105,000 for testing)	2	96.3		[37]
	SS-7.3 million, Camera CATalogue (CC)-0.52 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, CC-97.7, EE-96.6, SW-92.4		[77]
Fast R-CNN	6,300 (4,800 for training, 1,500 for testing)	8	83.89	77.95	[32]
	17,000 images (12,000 training, 5,000 Testing) (all, side, frontal, back)	1	92.1		[52]
	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]
	Virunga and Garamba dataset. Total 10,239 images (70% for training, 10% for validation, 20% for testing	6	74		[42]
ResNet-50	561 images, 4,305 labeled animals (70% for validation, 20% for testing)	3	86.3		[24]
	Head-140, body-124, scale-174, total 438 images (348 for testing, 90 for testing)	6	Head-86.85, body-92.52, scale-84.03		[47]
	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	96.11	93.36	[39]
	SS. 757,000 animal images + 757,000 random non-animal images (1.4 million for training, 105,000 for testing)	2	96.3		[37]
	Snapshot Australia (SA), SS, Snapshot Wisconsin (SW), 8 classes each dataset, 750 images each class	8	10 epochs-88, 20 epochs-91, 50 epochs-93, 150 epochs-95, 500 epochs-97, 1000 epochs-98		[58]
Mask R-CNN	1,462 images (1,187 for training, 275 for validation)	5	80.43		[68]
	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 training, 10% for testing)	2	92.4		[55]

YOLO V3	96 aerial images (89 for training, 7 for testing)	1	83		[63]
	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]