LUD-YOLO: A novel lightweight object detection network for unmanned aerial vehicle

Universidad del País Vasco / Euskal Herriko Unibertsitatea
Konputazio Ingeniaritza eta Sistema Adimentsuak Unibertsitate Masterra
Máster Universitario en Ingeniería Computacional y Sistemas Inteligentes

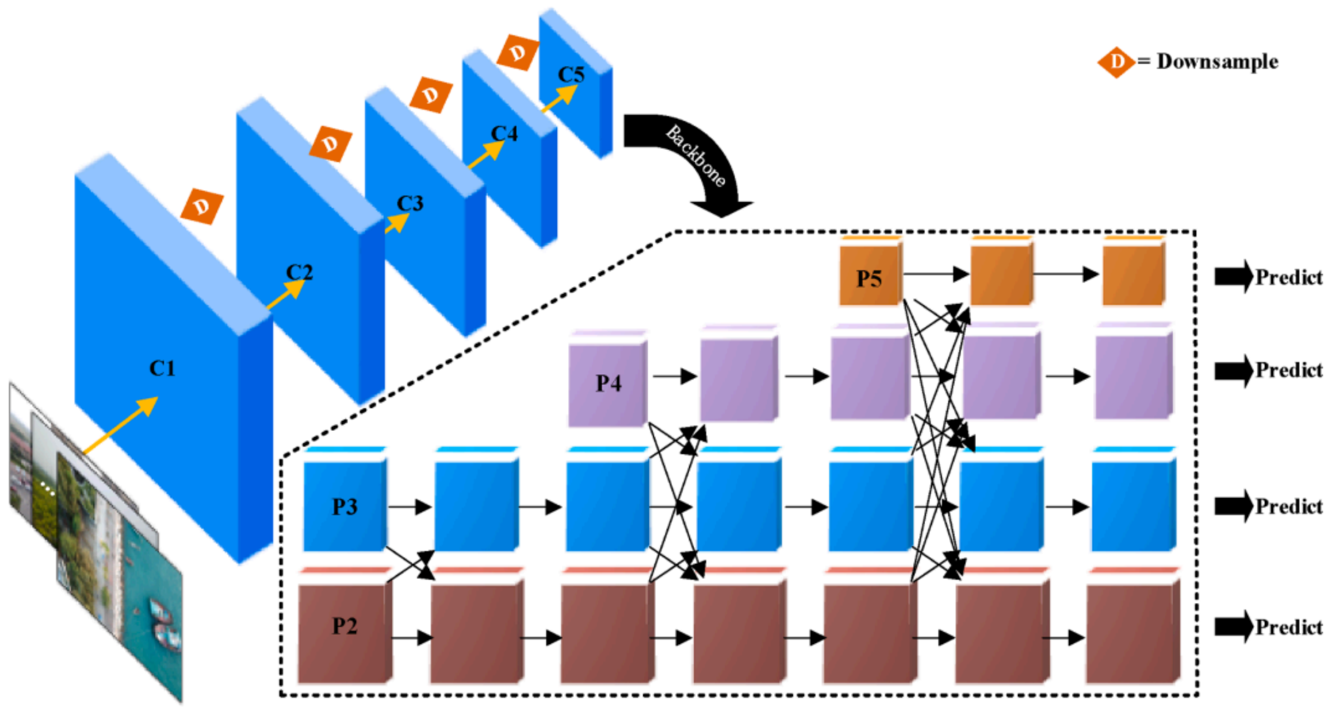
AUTHORS

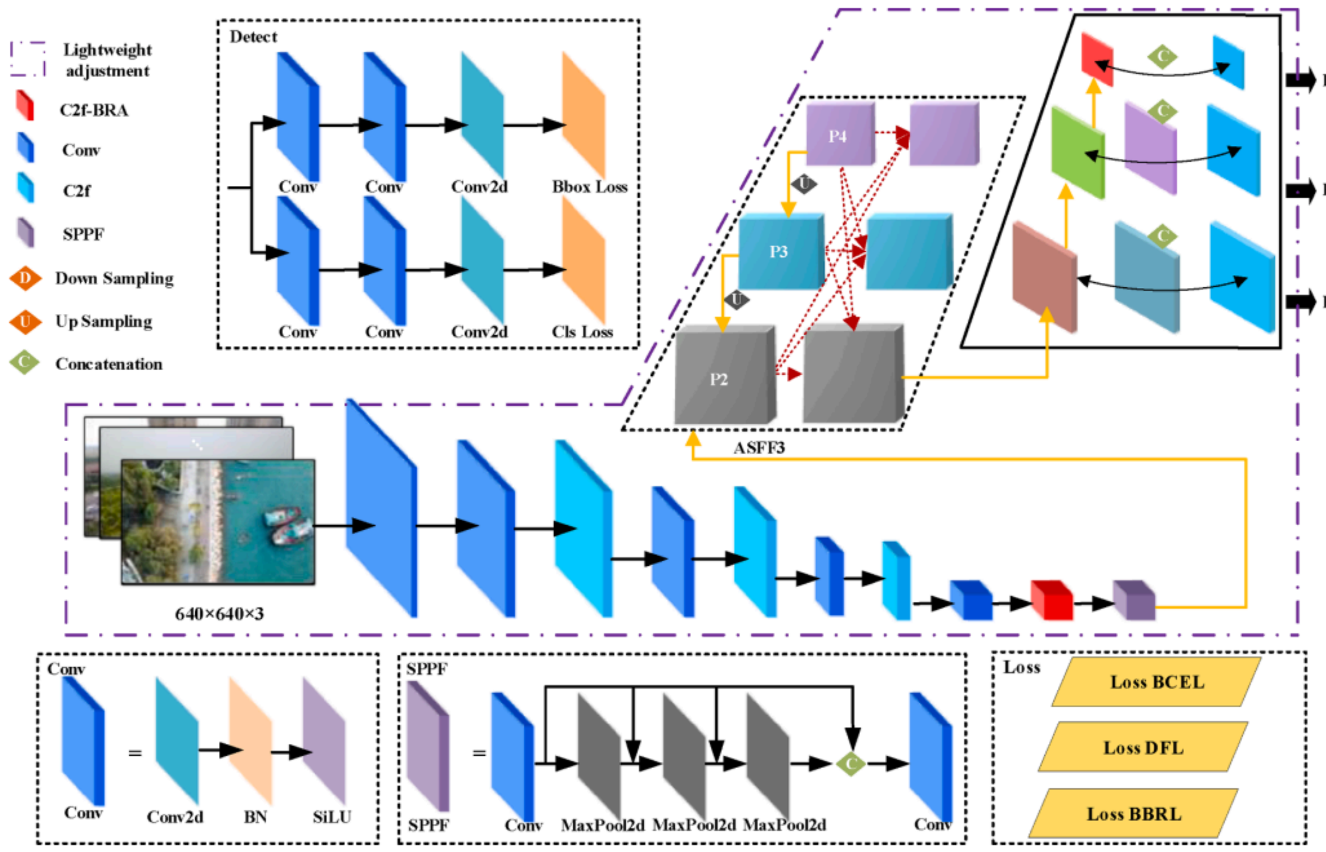
Qingsong Fan, Seifedine Kadry , Yiting Li , Muhammet Deveci, Kaiyang Zhong g ,

LUD-YOLO is a lightweight object detection algorithm for UAVs, based on YOLOv8. It incorporates dynamic sparse attention for efficient computing and is optimized for UAV small devices..

1. Motivation

- Models with better trade-off between computational cost, inference time and accuracy are necessary for UAVs
- Two-Staged Object Detection models have very accurate results but with low inference speed
- One-Staged Object Detection models have show good detection results with better performance but are still large models for the UAVs design domain





3. Contributions

3.1. Improvements in feature fusion

YOLOv8 feature fusion: combines multi-scale features, but...

- Feature degradation (low-level feature dilution)
- Combines information only from adjacent feature levels

Solution proposal:

- Use of **Asymptotic Feature Pyramid Network (AFPN)** in YOLOv8 for progressive distinct-level feature interaction, improving FPN-PAN limitations
- Faster inference by one-step C2-C4 adjustment yields **ASFF3 module**

3.2. Improvements in feature extraction

Convolution only operates on local information

Proposal:

- **BiLevel Routing Attention** combines local information by convolutions and global information by self-attention, adapting the C2f of YOLOv8 → **C2f-BRA**
  - Image is divided into regions and attention computed only with important regions.
  - Extract better features.

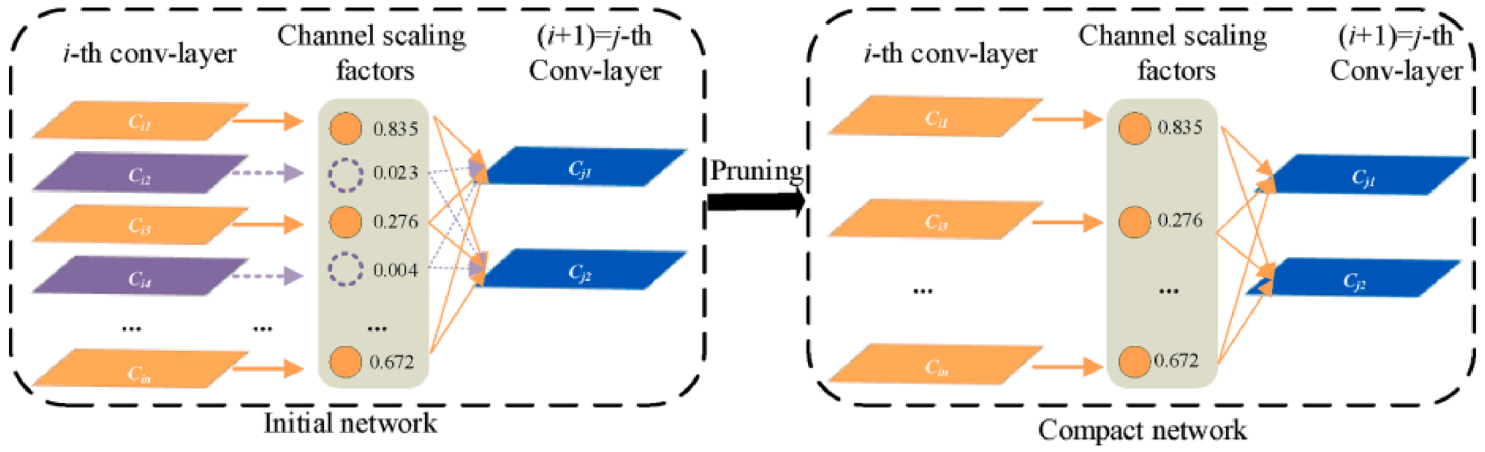
3.3. Lightweight adjustment of the LUD-YOLO

Crucial factors on models executed on UAVs:

- Size of the model (disk memory)
- Memory occupied on runtime
  - Convolutions are memory-expensive operations!
- Amount of calculations

Solution proposal: Network Slimming adaptation:

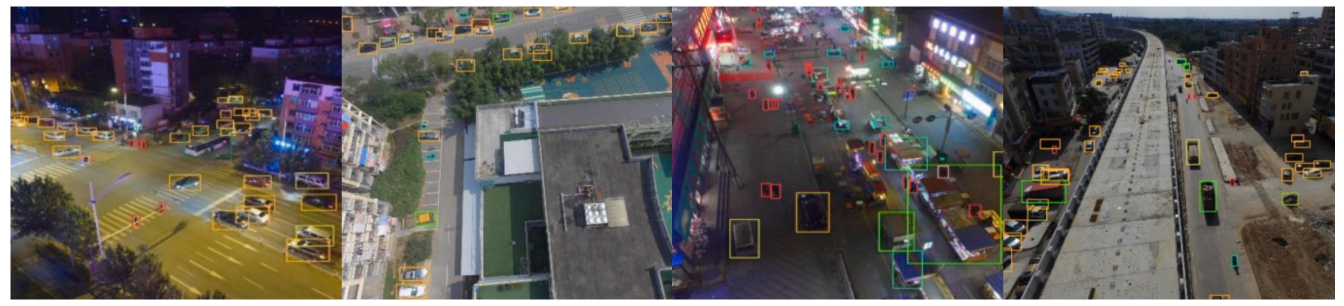
- Adding L1 regularization constraints
- Acquiring sensitivity
- Pruning convolution channels that do not meet the threshold
- Fine-tuning the model to compensate for accuracy loss



4. Experiments-VisDrone2019 Dataset

To evaluate the models VisDrone2019 and UAVDT datasets are used:

- 6000 videos and 25000 images.
  - Target categories: pedestrians, bicycles, motorcycles...
  - Labels: object bounding boxes, object categories...
  - Class unbalance Data augmentation



4.1 Comparisons

Results of the LUDY-N ablation experiment.

Data	AFPN	Biform	Pruning	P	R	mAP	FPS	Parameters/million	Model size/MB
Val	×	×	×	0.412	0.325	0.315	205	3.008	6.082
	√	×	×	0.458	0.344	0.348	171	3.491	7.090
	×	√	×	0.427	0.322	0.319	218	2.552	5.183
	√	√	×	0.471	0.349	0.353	180	3.035	6.191
	√	√	√	0.470	0.347	0.352	218	2.812	5.560
Test	×	×	×	0.365	0.286	0.257	218	3.008	6.082
	√	×	×	0.396	0.302	0.281	197	3.491	7.090
	×	√	×	0.382	0.285	0.256	257	2.552	5.183
	√	√	×	0.401	0.305	0.281	199	3.035	6.191
	√	√	√	0.401	0.303	0.279	263	2.812	5.560

Results of the LUDY-S ablation experiment.

Data	AFPN	Biform	Pruning	P	R	mAP	FPS	Parameters/million	Model size/MB
Val	×	×	×	0.504	0.373	0.381	186	11.129	22.203
	√	×	×	**0.535**	0.411	0.422	173	13.048	25.789
	×	√	×	0.507	0.373	0.384	179	9.301	18.389
	√	√	×	0.531	**0.413**	**0.424**	164	11.209	22.204
	√	√	√	0.525	0.408	0.417	**194**	10.339	20.492
Test	×	×	×	0.430	0.331	0.309	209	11.129	22.203
	√	×	×	**0.471**	0.352	0.339	193	13.048	25.789
	×	√	×	0.447	0.326	0.311	209	9.301	18.389
	√	√	×	0.468	**0.347**	**0.336**	197	11.209	22.204
	√	√	√	0.464	0.343	0.333	**213**	10.339	20.492

Model	P	R	mAP(50:95)	FPS	Parameters/million	Model size/MB
YOLOv8-s	0.978	0.975	0.831	238.00	11.127	21.968
YOLOv8-n	0.971	0.963	0.773	278.00	3.011	6.079
YOLOv7tiny-silu	0.961	0.960	0.709	263.00	6.012	11.974
YOLOv7tiny-mobilevit-odconv	0.945	0.886	0.639	208.00	8.572	17.072
YOLOv5-MobileNetv3	0.964	0.963	0.728	257.00	3.553	7.431
YOLOv5n-Bifpn	0.969	0.964	0.768	268.00	**2.513**	**5.203**
YOLOXs	0.960	0.952	0.710	55.00	8.928	35.172
MobileNetv3-SSD	0.933	0.875	0.613	156.00	3.926	15.680
LUDY-N	0.971	0.974	0.809	**277.00**	3.262	5.313
LUDY-S	**0.981**	**0.982**	**0.862**	251.00	9.716	19.210

5. Conclusions

- High detection accuracy under the premise of real-time detection
- Number of parameters and model size greatly reduced compared with benchmark model

6. Future work

- Only usable on high-quiality labeled data scenarios

7. Ackowledgments

This paper was published by Elsevier Inc. This is an open access article under the CC BY license, 16 August 2024 , State Key Laboratory of Intelligent Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China.

Alex Amenabar, Iosu Abal, Alan Garcia